

CHOIX PUBLICS STRATEGIQUES ET SYSTEMES SOCIAUX

Etat de l'art sur les théories de la décision et méthodologies de l'approche système

étude coordonnée par

Jacky MONTMAIN¹

Tel. : 0466387058

jacky.montmain@ema.fr

Jean michel PENALVA¹

Tel. : 0466387036

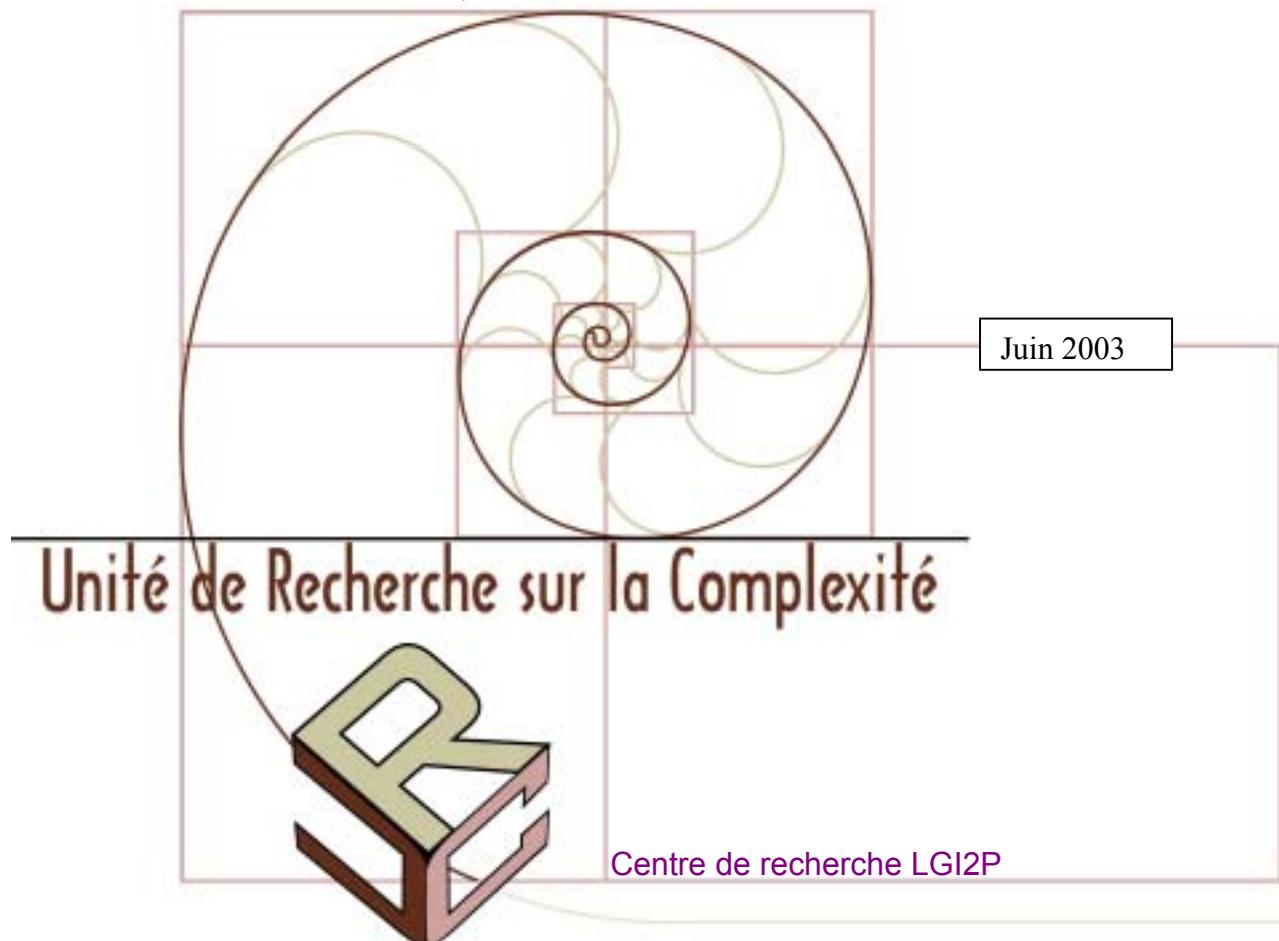
jean-michel.penalva@ema.fr

avec la collaboration de

Abdellah AKHARRAZ², Vincent CHAPURLAT², Pierre COUTURIER², Michel CRAMPES²,
Gérard DRAY², Stefan JANAQI², Isabelle MARC², Pascal PONCELET², Pierre-Michel RICCIO², Pierre
RUNTZ², Béatrice VACHER² et Michel VASQUEZ²

¹ COMMISSARIAT A L'ENERGIE ATOMIQUE,

² ECOLE DES MINES D'ALES







PARTIE I : PREPARER UNE DECISION COLLECTIVE

1. LA DÉCISION COMME ENSEMBLE D'ACTIVITÉS.....14

1.1.	RECHERCHE DE L'INFORMATION	14
1.2.	ELABORATION DES HYPOTHÈSES	14
1.3.	RECHERCHE DE SOLUTIONS	15
1.4.	SÉLECTION DE SOLUTIONS	15

2. L'APPROCHE SYSTÈME DES SITUATIONS COMPLEXES.....16

2.1.	APPROCHE SYSTÉMIQUE DE LA SITUATION	16
2.1.1.	LA FORME CANONIQUE D'UNE SITUATION	18
2.1.2.	COMPRENDRE ET ANTICIPER UNE SITUATION COMPLEXE	19
2.1.3.	PRISE DE DÉCISION ET MANAGEMENT DU RISQUE	20
2.1.4.	COGNITION ET SITUATION.....	22
2.1.5.	LA QUALIFICATION D'UNE SITUATION	23
2.1.6.	LE RECADRAGE DES SITUATIONS COMPLEXES	23
2.1.7.	SYSTÈME DE PRÉSENTATION ET SYSTÈME DE RÉFÉRENCE	25
2.1.8.	DE LA SITUATION AU PROJET.....	27
2.2.	APPROCHE COMMUNICATIONNELLE DE LA SITUATION	29
2.2.1.	APPROCHE SÉMIO-CONTEXTUELLE DE LA COMMUNICATION.....	29
2.2.2.	APPROCHE SYSTÉMIQUE DES COMMUNICATIONS.....	31
2.2.3.	CONCLUSION	34
2.3.	APPROCHE ORGANISATIONNELLE DE LA SITUATION	35
2.3.1.	RAPPELS SUR LA NOTION D'INFORMATION	35
2.3.2.	THÉORIE DE LA DÉCISION ET SCIENCES DE L'INFORMATION	36
2.3.3.	L'ACTION EST SITUÉE ET LA COGNITION DISTRIBUÉE	38
3.1.4 -	DONNER DU SENS À L'ACTION.....	40

3. APPROCHE SYSTÈME DES COMPORTEMENTS COLLECTIFS INTELLIGENTS....42

3.1.	SYSTÈMES MULTI-AGENTS	42
3.1.1.	PROCESSUS NATURELS D'AUTO-ORGANISATION	43
3.1.2.	APPRENTISSAGE ET SYSTÈMES DISTRIBUÉS	49
3.1.3.	MODÉLISATION DES PHÉNOMÈNES DE COMPORTEMENT COLLECTIF INTELLIGENT	54
3.1.4.	L'APPROCHE DES PROBLÈMES DE TRANSPORT PAR DES SIMULATIONS MULTI-AGENTS ? ..	58
3.2.	INTELLIGENCE HUMAINE COLLECTIVE	60
3.2.1.	MÉMOIRE COLLECTIVE.....	60
3.2.2.	CONNAISSANCES ACTIONNABLES	62
3.2.3.	EXPERTISE INDIVIDUELLE	63
3.2.4.	INTELLIGENCE COLLECTIVE	65
3.2.5.	LES SYSTÈMES DE GESTION DYNAMIQUE DES CONNAISSANCES (SGDC)	68
3.2.6.	RETOUR D'EXPÉRIENCES : UNE TYPOLOGIE	70
3.2.7.	CONCLUSION	73



4. LA CRÉATIVITÉ.....	73
4.1. QUELQUES DÉFINITIONS	73
4.2. LA CRÉATIVITÉ : UNE APPROCHE PHYSIOLOGIQUE	74
4.3. L'ATTITUDE CRÉATIVE	75
4.3.1. VAINCRE L'INERTIE PSYCHOLOGIQUE DES ACTEURS.....	75
4.3.2. UTILISER DES RÉFÉRENTIELS.....	76
4.3.3. IMPLIQUER, INCITER ET PILOTER LES ACTEURS DE LA CRÉATIVITÉ	76
4.4. CLASSIFICATION DES MÉTHODES D'AIDE À LA CRÉATIVITÉ	77
4.4.1. MÉTHODES COMBINATOIRES.....	78
4.4.2. MÉTHODES ANTITHÉTIQUES.....	78
4.4.3. MÉTHODES ASSOCIATIVES	78
4.4.4. MÉTHODES ANALOGIQUES	78
4.4.5. MÉTHODES EMPATHIQUES.....	79
4.4.6. MÉTHODES ALÉATOIRES	79
4.4.7. SYNTHÈSE.....	79
4.5. MISE EN ŒUVRE DES MÉTHODES	79
4.5.1. APPROCHE DIRECTE OU INDIRECTE	79
4.5.2. LES PHASES DE DIVERGENCE.....	80
4.5.3. LES PHASES DE CONVERGENCE	90
4.6. LA RÉSOLUTION DE PROBLÈMES INVENTIFS : LA MÉTHODE TRIZ	91
4.6.1. PRINCIPES	93
4.6.2. LEVÉE D'INERTIE PSYCHOLOGIQUE.....	95
4.6.3. MODÉLISATION DU PROBLÈME.....	95
4.6.4. RECHERCHER DES PISTES DE SOLUTION	96
4.6.5. CONCLUSION	100
4.7. L'INNOVATION	101
4.7.1. MODÉLISATION DES PROCESSUS D'INNOVATION	102
4.7.2. L'INNOVATION DANS LES SERVICES	102
4.7.3. LE MANAGEMENT DE L'INNOVATION	103
4.7.4. LES MÉTHODES D'AIDE À L'INNOVATION.....	105
4.8. CONCLUSION	106



PARTIE II : LES THEORIES DE LA DECISION

1. LES APPROCHES MATHÉMATIQUES DE LA DÉCISION 107

1.1.	EXTRACTION DE CONNAISSANCES ET FOUILLE DE DONNÉES	107
1.1.1.	INTRODUCTION	107
1.1.2.	LE PROCESSUS D'EXTRACTION DE CONNAISSANCES À PARTIR DE DONNÉES	110
1.1.3.	CLASSIFICATION	119
1.1.4.	LES RÈGLES D'ASSOCIATIONS	145
1.1.5.	LES MOTIFS SÉQUENTIELS	149
1.1.6.	EXEMPLES D'APPLICATIONS	153
1.2.	LA RECHERCHE OPÉRATIONNELLE	157
1.2.1.	PROGRAMMATION LINÉAIRE	157
1.2.2.	PROGRAMMATION NON LINÉAIRE.	170
1.2.3.	LA THÉORIE DES GRAPHES ET RÉSEAUX	180
1.3.	LES CHOIX ALÉATOIRES	186
1.3.1.	CARACTÈRE DES PROBLÈMES ALÉATOIRES	186
1.3.2.	LES PHÉNOMÈNES D'ATTENTE	187
1.3.3.	ENTRETIEN PRÉVENTIF ET RENOUVELLEMENT DES ÉQUIPEMENTS	187
1.3.4.	GESTION SCIENTIFIQUE DES STOCKS	188
1.4.	STRATÉGIES EN SITUATION DE CONCURRENCE, THÉORIE DES JEUX	188
1.4.1.	LA CARACTÉRISTIQUE DES PROBLÈMES DE CONCURRENCE	188
1.4.2.	LES JEUX CONTRE LA NATURE	188
1.4.3.	LES JEUX À DEUX PERSONNES	189
1.4.4.	RESPONSABLE ET ANALYSTE DANS LA PRISE DE DÉCISION	191
1.5.	LES MÉTHODES MULTICRITÈRES	192
1.5.1.	JUSTIFICATION DES MÉTHODES MULTICRITÈRES	192
1.5.2.	CONCEPTS FONDAMENTAUX	193
1.5.3.	APPROCHES D'AGRÉGATION MULTICRITÈRE	194

2. LES APPROCHES COGNITIVES DE LA DÉCISION 202

2.1.	LE PROCESSUS DE LA DÉCISION	202
2.2.	LE KALÉIDOSCOPE DE LA DÉCISION ET L'APPORT DES SCIENCES HUMAINES	203
2.3.	LES PIÈGES DE LA RAISON	211
2.3.1.	LA LOGIQUE DES ERREURS	211
2.3.2.	DÉCIDER DANS L'URGENCE	212
2.3.3.	ENTRE STRATÉGIE CONSCIENTE ET FORCE AVEUGLE	213
2.3.4.	DÉCISION RAISONNÉE OU RAISONNABLE	214
2.4.	LE MODÈLE S.T.I DE H.A. SIMON	215
2.4.1.	RATIONALITÉ SUBSTANTIVE VERSUS RATIONALITÉ PROCÉDURALE	215
2.4.2.	ACQUISITION ET TRAITEMENT DE L'INFORMATION	217
2.5.	LE COUPLAGE D'UN SIADG AVEC UN SYSTÈME DE GESTION DES CONNAISSANCES	219
2.5.1.	CONNAISSANCES ACTIONNABLES	220
2.5.2.	EVALUATION MULTICRITÈRE	221



2.5.3.	ACCEPTABILITÉ D'UNE DÉCISION COLLECTIVE : ARGUMENTATION ET GESTION DU RISQUE	
	222	

2.6.	LES MODÈLES GRAPHIQUES PROBABILISTIQUES POUR LA REPRÉSENTATION DES CROYANCES ET POUR LA DÉCISION STRATÉGIQUE	224
2.6.1.	INTRODUCTION : VISUALISATION, INTUITION ET RAISONNEMENT	224
2.6.2.	LES MODÈLES GRAPHIQUES PROBABILISTES DE DIAGNOSTIC ET DÉCISION	225
2.6.3.	APPLICATIONS ET BIBLIOGRAPHIE DES RB ET DES DI	227
2.6.4.	ANALYSE CRITIQUE DES REPRÉSENTATIONS GRAPHIQUES PROBABILISTES	230
2.6.5.	FACILITATIONS DE L'USAGE DES MODÈLES GRAPHIQUES PROBABILISTIQUES PAR INTERFACES ADAPTATIVES	231
2.6.6.	CONCLUSION	231

<u>3. DISCUSSION</u>	<u>232</u>
-----------------------------	-------------------

<u>BIBLIOGRAPHIE</u>	<u>235</u>
-----------------------------	-------------------



LA DECISION, ENTRE CALCUL ET COGNITION

Aborder une situation complexe

Toute décision raisonnée repose d'abord sur une analyse de la situation qui permet d'expliquer et de caractériser les phases de perception, de représentation et de sélection inhérentes à un processus décisionnel.

La confrontation d'un opérateur humain à la gestion d'une situation complexe définit une « situation cognitive » où il doit s'adapter à la situation de décision ou d'action, c'est-à-dire accorder ses structures cognitives aux circonstances du réel perçu afin de prendre des dispositions pertinentes pour obtenir le but visé. Il doit ainsi combiner : un savoir observer sur l'objet (il perçoit l'objet au travers de caractères descriptifs qui lui donnent une forme particulière), un savoir faire (il fabrique des représentations de l'objet adéquate au projet), un savoir pratique (il développe un savoir se comporter dans l'action en fonction de connaissances acquises).

Dans un premier temps, l'objet prend forme dans le projet : ce processus *d'information* produit les données de base, les processus d'évaluation de l'objet dans son environnement et les concepts qui se forment au cours de l'histoire de l'objet. Puis le projet mobilise le sujet. L'objet est évoqué dans la relation projet-sujet, c'est à dire qu'il est à la fois désigné et interprété par le sujet, processus que l'on qualifie de *représentation*, par les savoirs, articulés autour de systèmes de symboles et intégrés en compétences. Enfin, l'objet prend sens, la relation objet-sujet prend sens par un processus de *signification* qui explicite les connaissances élémentaires actionnables développées par le sujet, les raisonnements stratégiques, fondés sur des calculs au sens large, qui s'appliquent à l'objet et font référence à des principes directeurs.

Si chacune de ces étapes apparaît comme naturelle dans le cadre de l'apprentissage et de l'appropriation de la situation par l'individu, la combinaison de ces savoirs (savoir-observer, savoir-faire et savoir-pratique) dans le cadre d'une décision collective peut nécessiter le recours à des processus artificiels d'ingénierie des données, de la connaissance et de la décision.

Ce document offre donc un panorama de ces ingénieries sous l'angle particulier de l'aide à la réflexion stratégique.

Notons encore que les savoirs évoqués ci-avant peuvent être exprimés selon plusieurs niveaux de formalisation allant des modèles logico-mathématiques aux corpus de connaissances en langage naturel. Parce qu'ils sont la résultante de processus de valorisation des données, d'interprétation d'informations et de délibération sur les connaissances acquises, les savoirs sont inéluctablement l'objet d'imprécision, d'incertitude et d'incomplétude. Ces niveaux de formalisation et de fiabilité déterminent le choix des outils pour les « manipuler » dans le processus décisionnel et expliquent la diversité des approches de la décision dont nous proposons un descriptif dans ce document.

(inter)Agir collectivement

La question posée par la mission Transport sur la problématique des Choix publics stratégiques et systèmes sociaux est le problème du responsable politique qui, à son niveau,



doit faire des choix de politique publique, supposés cohérents avec le positionnement sur la base duquel il a été élu. Plus précisément, se pose le problème du responsable de l'administration, chargé de proposer au choix du responsable politique des stratégies cohérentes alternatives. Si on attend du responsable politique qu'il émette une opinion par rapport à l'évolution constatée de certains systèmes socio-économiques et, surtout, qu'il prenne toutes mesures permettant que, du point de vue sur lequel il a été élu, les choses aillent "mieux", il n'en reste pas moins vrai que la maîtrise des systèmes en cause lui échappe, en grande partie, et qu'il ne peut agir que via des incitations, contraintes et demandes de concertations à l'égard des multiples acteurs qui interviennent, directement, à des titres divers, dans le fonctionnement du système. Il s'agit bien d'un processus de décision organisationnelle.

L'étude des processus de co-construction des connaissances et des compétences dans la chaîne des procédures : données – informations – connaissances, prépare la prise de décision collective en situation de risques opérationnels et/ou organisationnels.

Une hypothèse centrale est que la maîtrise des décisions et le management des risques associés dépend de la qualité du couplage entre l'organisation de l'action collective et le cadre contextuel : en situation de confrontation à un environnement incertain, inconstant, imprévisible, l'efficacité d'un collectif humain ne naît pas d'une invariance organisationnelle - définition a priori du cadre de l'action - où l'écart à la norme est interprété comme une erreur ou une déviance.

Le processus de maîtrise du risque exige d'éviter les ruptures cognitives entre individus ou groupes placés devant la nécessité de construire une compréhension commune et de co-opérer.

Bien sûr, les technologies de l'information facilitent aujourd'hui la communication et l'échange de données. Cependant, les performances technologiques des systèmes d'information ne suffisent pas à l'émergence d'un réel mode de travail collectif, où l'ensemble des processus cognitifs d'apprentissage, d'argumentation et de décision, attribués à un groupe d'acteurs, atteindraient et dépasseraient un niveau de performance dans l'action non atteint isolément par chacun des individus.

Pour atteindre le niveau d'une intelligence collective, il est nécessaire d'instrumenter ce *sujet-collectif* qu'est une organisation constituée en vue de l'action afin de le doter de capacités analogues à celles « naturelles » d'un individu cognitif. On peut en effet parler d'une entité collective donnée, une entreprise par exemple, comme étant susceptible d'apprendre, d'avoir une mémoire, de faire des expériences, des choix, de prendre des décisions, d'agir, etc., c'est-à-dire, comme étant en quelque sorte dotée d'une cognition individuelle.

L'apport des sciences humaines et sociales à la décision joue ainsi un rôle considérable dès lors que celle-ci relève d'un processus collectif.

Ouvrir le champ des possibles

L'innovation est une issue possible des décisions publiques sans doute celle la plus à même de bousculer l'inertie psychologique collective, de susciter la controverse la plus enflammée.

De par le monde, il est question d'« innovation » aussi bien en matière d'industrie, de science, de technique, d'administration que de sport ou de loisir. Ce terme a progressivement remplacé ceux de « nouveauté » et d'« invention » pour devenir un mot d'usage courant désignant une amélioration ou un perfectionnement de tout ordre. Par conséquent, la généralité même du terme rend difficile une définition précise. L'innovation est ici considérée comme une forme



ou une résultante particulière du processus de décision : par leur caractère même de nouveauté, les innovations déstabilisent les modes de fonctionnement et les manières de faire traditionnelles. Elle induit donc une rupture dans les savoirs cités précédemment qui peut susciter des refus ou des réticences de la part de ceux dont l'existence et les habitudes sont perturbées, la notion d'acceptabilité de la décision devient alors essentielle.

Il ne suffit pas qu'une idée soit bonne, pratique et rentable pour qu'elle soit automatiquement applicable et qu'elle suscite une approbation générale. L'imagerie populaire laisse à penser qu'innover est un combat, une action héroïque, volontariste et solitaire. Mais cette image a peu de rapports avec la réalité :

l'innovation est d'abord le résultat d'un travail collectif où les apports de chacun sont difficiles à individualiser et où le succès tient moins à un éclair de génie qu'à un travail ingrat de mise au point, de développement et d'adaptation.

Alors que l'invention peut être gratuite et aléatoire, l'innovation suppose d'envisager une application et une mise en œuvre industrielle. C'est pour cela que le terme d'« innovation » s'est progressivement substitué à celui d'« invention », et que l'on parle désormais plus volontiers d'innovateur que d'inventeur. Ce dernier terme évoque le bricoleur de génie, contrairement aux innovateurs, véritables entrepreneurs, qui considèrent que trouver une idée n'est pas plus important que de la développer et de l'exploiter sous une forme industrielle.

On distingue traditionnellement trois sortes d'innovations. Les innovations de produit touchent aux fonctionnalités ou à la nature de l'objet : c'est, par exemple, l'apparition du disque compact. Les innovations de procédé portent sur les techniques de conception et de fabrication mais n'affectent pas nécessairement le produit réalisé : quand un fabricant d'automobiles se dote d'installations robotisées et automatisées, les véhicules ne changent pas pour autant, ils sont simplement fabriqués plus vite et moins cher. Enfin, les innovations de service concernent des biens immatériels tels que des modes de distribution ou d'organisation : McDonald's a innové en organisant rationnellement une chaîne de restaurants bon marché, au service rapide et à la qualité égale partout.

Réussir le développement d'une innovation nécessite d'informer très tôt toutes les parties prenantes et de négocier sur les conséquences à attendre du projet afin d'obtenir l'adhésion de tous. Les organisations les plus performantes sont souvent celles qui pensent simultanément au changement technologique, au contenu du travail et aux modifications des rapports sociaux internes à l'organisation.

Les théories de la décisions : hypothèse d'ordre versus hypothèse d'équivocité

Parmi les théories qui abordent la décision, il y a d'abord le courant dit des « théories de l'utilité espérée ». Ces théories admettent la subjectivité de la notion d'utilité pour le décideur : il s'agit de ne pas renoncer à la modélisation de la décision sans pour autant la ramener à une objectivité que l'on sait illusoire [Roy, 1985 ; Roy, 1992 ; Roy et al., 1993]. La décision est semi-normative.

L'autre courant dit des « théories cognitives » s'attache à étudier le processus de décision. On ne cherche plus à déterminer un critère d'utilité mais à comprendre les étapes qui s'enchaînent lors d'une prise de décision. On s'intéresse moins à la recherche d'un résultat optimal qu'aux fonctions psychologiques qui traitent et interprètent l'information. On ne cherche pas, dans un tel cas, à rejeter l'erreur comme un écart par rapport au choix optimal mais à l'inclure dans le processus de décision. Dans la prise de décision comme dans l'apprentissage d'un savoir-faire, certaines erreurs sont des étapes naturelles du cheminement de l'esprit. Ainsi se



constitue une approche cognitive de la prise de décision, c'est à dire liée à la connaissance du décideur—qui admet que la recherche de la décision optimale est une illusion.

Toute vision de la décision s'inspire aujourd'hui de l'une ou de l'autre approche et tous les développements qui suivent balancent entre les deux... En effet, l'idée d'utilité intervient dans une décision, les visions et les croyances du décideur aussi. La décision se prend à la rencontre de ces deux dimensions.

Pour l'approche des théories de l'utilité espérée, la décision prise doit être le résultat d'un choix comparatif entre les diverses solutions possibles. Le décideur et ses conseillers doivent mesurer avec soin les risques et les issues probables de chaque formule, peser leurs avantages et inconvénients pour retenir finalement celle qui représente le meilleur rapport « coût/efficacité ». Cette analyse, en terme de calcul rationnel, postule l'existence d'un acteur unique qui agirait en vertu de préférences hiérarchisées en fonction de la meilleure utilité.

Ce séduisant modèle théorique fait abstraction des aspects organisationnels souvent implicites dans un processus décisionnel. La vision organisationnelle de la décision postule que le décideur ne possède pas en fait une connaissance totale de la situation, d'où le terme de « rationalité limitée » cher à H.A. Simon. Les limitations dans la connaissance des faits et hypothèses proviennent des contraintes de l'organisation qui sélectionne ou favorise tel ou tel scénario en fonction de ses intérêts.

Les sciences humaines et sociales insistent en définitive sur les aspects cognitifs de la décision, l'acquisition et le traitement de l'information apparaissant comme plus importants pour prendre une « bonne » décision que la recherche fine illusoire d'une décision en apparence « la meilleure ». Elles offrent une alternative à la vision mathématique de la décision en particulier dans le cadre d'une décision collective. Elles peuvent tout aussi bien guider l'élaboration des outils, informatiques en particulier, utiles à l'organisation dans la décision collective, que proposer des modèles sophistiqués, même s'ils ne sont pas toujours formalisés, du fonctionnement de l'organisation, de la manière dont celle-ci gère l'information et la connaissance. Ces modèles constitueront autant de mises en garde sur la validité des approches mathématiques de la décision en organisation.

En effet, dans les approches mathématiques, objectifs et contraintes ont généralement été modélisés par le biais de modèles logico-mathématiques et la recherche de la décision optimale est une question de « problem solving » où le cognitif n'a que bien peu de place. La modélisation mathématique des objectifs et autres contraintes rend l'objectivité du modèle implicite puisque sa formulation formelle a été établie. Dans une approche cognitive, la décision repose sur une hypothèse *d'équivocité* : pour une situation décisionnelle donnée, plusieurs interprétations de cette situation sont possibles selon le point de vue, la fonction, les intérêts de l'acteur en jeu...

Tout ne peut être signifié dans le modèle mathématique sans hypothèse simplificatrice sur la représentation des aspects politico-psychosocials de la décision collective.

Les situations décisionnelles

Une situation décisionnelle est définie par la confrontation d'une organisation dotée d'un projet avec un système évolutif et réactif dans un environnement imparfaitement perçu. Les outils méthodologiques nécessaires à l'appréhension d'une situation décisionnelle dépendent de la nature de celle-ci.



Si les objectifs et contraintes du projet ne donnent pas lieu à un débat au sein même de l'organisation, le problème peut être objectivement défini et on peut espérer en établir une représentation formelle. On dira que la situation décisionnelle est structurée. L'exercice consiste alors à élaborer cette représentation formelle qui, si elle aboutit à une description mathématique, se ramène à un problème de recherche opérationnelle. Un problème n'a de sens en recherche opérationnelle que si l'on connaît au préalable la fonction à optimiser, l'attitude à tenir face au risque : il s'agit ensuite de formaliser mathématiquement la fonction à optimiser, la stratégie à suivre, les objectifs à atteindre, les contraintes à respecter, l'attitude souhaitée face au risque. La complexité du problème réside donc, dans ce premier cas, dans sa formulation mathématique et dans la résolution du modèle établi. Cette configuration donne de l'organisation une vision idéale où le collectif peut purement et simplement se ramener à un acteur unique, à un décideur monolithique.

Cette hypothèse est rarement valide dans un processus de décision collective. Les actions ou les mesures résultantes sont élaborées à travers un processus long et enchevêtré mobilisant de nombreux acteurs ; il n'y a pas de décision, un décideur, mais une série de stratégies et de compromis entre les points de vue, entre des groupes qui ne partagent pas la même solution. Dans sa théorie du surcode [Sfez, 1992], Sfez montre que derrière l'image trompeuse d'une décision consciente et unifiée, il y a en fait une multiplicité de rationalités différentes qui s'imbriquent, se superposent et se confrontent. Dès lors notamment que des ensembles humains sont concernés, les décisions sont de l'ordre du «non-programmable».

Gommer les phases d'intelligence et de conception du processus de décision est une réduction inacceptable. L'aide à la décision consiste alors plutôt à essayer de représenter et d'améliorer la façon dont les hommes utilisent leurs capacités de raisonnement et de traitement des informations. Au lieu de chercher à désigner une décision ambitionnant d'être optimale, il est plus modeste dans les objectifs, mais peut-être plus efficace pour le résultat, de chercher à user d'une procédure de traitement de l'information et de raisonnement plus satisfaisante. La rationalité «limitée» ou «procédurale» vient ainsi se substituer à la rationalité optimisante et «substantive»[Simon, 1997].

Cette analyse insiste en définitive sur les aspects cognitifs de la décision, l'acquisition et le traitement de l'information apparaissant comme plus importants pour prendre une «bonne» décision que la recherche fine illusoire d'une décision en apparence «la meilleure». Les différentes phases de la décision ne se présentent pas de façon linéaire, mais en boucles. De nombreuses itérations sont nécessaires, au vu de la faible capacité cognitive des hommes et de la complexité des problèmes de décision, avant qu'un terme puisse être apporté au processus de décision. Davantage encore, chacune des phases engendre des sous-problèmes qui, à leur tour, appellent des phases d'intelligence, de conception, de sélection et de bilan. Les phases sont ainsi des «engrenages d'engrenages». Le vocable «système de traitement de l'information», S.T.I, permet de désigner commodément la lignée des modèles issus de la pensée de H.A.Simon. Si le schéma linéaire du processus de décision est ici rejeté, il n'en reste pas moins que dans cette seconde vision de la décision la planification conserve encore une place privilégiée. Nous dirons de ces situations décisionnelles qu'elles sont semi-structurées.

Enfin, des chercheurs américains définissent à la fin des années 60 [Cohen, March et Olsen, 1972], deux nouvelles notions. La première est celle d'anarchies organisées dont les universités sont selon, eux, un parfait exemple. Cette expression désigne les organisations :

- sans objectifs cohérents et partagés par tous ;



- où le processus de production relève d'une technologie complexe, peu matérialisable ;
- dont les membres participent de façon active aux prises de décision.

La seconde est le modèle de la poubelle qui décrit le style de décision propre aux anarchies organisées. Ce modèle remet en cause les théories où les décisions résultent d'une confrontation entre des objectifs identifiés, des solutions disponibles et leurs conséquences et les théories où les décisions sont le résultat d'une négociation entre des groupes aux intérêts divergents. Dans les anarchies organisées, des choix sont à la recherche de problèmes, des questions cherchent des opportunités pour décider, des solutions cherchent des questions auxquelles elles pourraient être une réponse et des décideurs cherchent du travail...

Des décisions se produisent quand les flux de problèmes, de solutions, de participants et d'opportunité de choix se rencontrent. Toute prise de décision est ainsi assimilable à une poubelle où des types de problèmes et de solutions sont déchargés par les participants dès qu'ils sont générés et qui, se rencontrant, font émerger un choix. La simulation informatique de ce modèle, a priori chaotique, ne fait pourtant apparaître que trois styles de choix possibles : les décisions par inattention, par déplacement des problèmes et par résolution des problèmes. Les deux premiers étant plus fréquents que le dernier. Cette émergence d'ordre dans les processus de décision anarchique a inspiré des travaux sur la capacité d'apprentissage des anarchies organisées. Cette dernière configuration où il s'agit de faire du « sense-making » dynamique définit ce que l'on appellera les situations décisionnelles non structurées.

En résumé, pour un projet donné, une sélection pertinente des modèles, approches et outils mathématiques, organisationnels et informatiques, de la décision, passe d'abord par l'identification de la nature de la situation. Dans la pratique, s'il est vrai que les problèmes de décision évoqués par le Ministère des Transports paraissent souvent relever de situations non structurées, il n'en reste pas moins que la résolution locale ou partielle d'un problème peut nécessiter des modèles plus formels relevant de l'ingénierie des données, de la connaissance ou de la décision.



PARTIE I : PREPARER UNE DECISION COLLECTIVE

Le transport, quelle que soit l'échelle territoriale ou fonctionnelle considérée, est un problème qui concerne nombre d'acteurs. Cela peut aller du simple usager de la route, du train ou de l'avion jusqu'au constructeur automobile, au responsable du trafic autoroutier, aux pouvoirs publics ou encore aux représentants syndicaux de certains de ces acteurs.

Parmi eux, l'état et les collectivités, doivent s'assurer de la validité, de la cohérence globale, de la complétude et de la pertinence des solutions qui devront être mise en œuvre pour répondre aux différents besoins en terme de transport de personnes ou de marchandises.

Leur responsabilité dans ce choix est primordiale et ils doivent prendre en compte pour cela deux axes stratégiques incontournables :

- Respecter la cohésion sociale : en tant que représentant des pouvoirs publics, ils doivent en effet et avant toute chose, assurer le maintien de l'emploi, rechercher en permanence le dialogue, l'entente ou le consensus avec les acteurs tout en tentant d'améliorer la qualité de vie des usagers comme leur sécurité.
- Intégrer les concepts de développement durable et de respect de l'environnement dans ces solutions.

Pour cela, le Ministère des Transport souhaite connaître et pouvoir appliquer une panoplie de méthodes et d'outils permettant de l'aider tout au long de ce processus de décision. L'objectif final est de pouvoir sélectionner une ou plusieurs alternatives parmi celles proposées et de valider, avec un taux de confiance considéré comme suffisant, la solution qui sera ensuite mise en œuvre. Cependant, cette activité décisionnelle, si elle s'avère évidemment nécessaire au cours du processus auquel participent normalement tous les acteurs concernés, doit être implicitement précédée d'une suite d'activités de recherche et de mise en évidence des alternatives possibles. Cette phase peut être d'ores et déjà avoir été mal engagée et avoir des résultats trop parcellaires pour plusieurs raisons :

- Manque de temps, de disponibilité ou de ressources,
- Subjectivité volontaire due par exemple à une activité de lobbying de la part d'un groupe d'acteurs,
- Subjectivité involontaire : problème mal défini, mal perçu, communication difficile entre les acteurs ou encore richesse insuffisante des références (comme le dit H.Poincaré : « *parmi les combinaisons choisies, les plus fécondes seront souvent formées d'éléments provenant de domaines fort éloignés* »).

Les solutions envisagées seront alors d'autant moins nombreuses, moins pertinentes, moins créatives.

Il est donc essentiel de mettre en avant les principes essentiels de la démarche de créativité au sein d'un processus décisionnel. Quelques méthodes dont il est possible de s'inspirer pour mener à bien la formulation de problèmes et la recherche de pistes de solutions répondant à la problématique globale du transport sont présentées dans cette première partie du document. Enfin, quelques propositions sont faites pour organiser au mieux les étapes de créativité tout au long du processus décisionnel.



1. LA DECISION COMME ENSEMBLE D'ACTIVITES

L'objectif du processus de décision est de définir un espace de solution répondant à une problématique donnée, un besoin à satisfaire ou un souhait d'amélioration, de changement ou d'adaptation en tenant compte de contraintes diverses. Il peut ainsi se décrire sous forme d'un ensemble d'activités non obligatoirement séquentielles :

1.1. Recherche de l'information

Rechercher et regrouper l'ensemble des informations pertinentes et des acteurs concernés par cette problématique puis faire un certain nombre d'hypothèses afin de pouvoir par la suite reformuler de manière consensuelle le problème tel qu'il est perçu par ces acteurs. C'est ainsi, mais pas seulement, l'occasion de faire le bilan de l'existant en se posant quelques questions de base qui pourraient être les suivantes dans la problématique qui nous intéresse ici :

- Quels sont les acteurs actuels et potentiels du système de transport (utilisateurs, partenaires sociaux et économiques, etc.) ? Quels sont leurs rôles et leurs objectifs ?
- Quel est l'état du système actuel de transport ? On peut se borner à faire une suite de constatations objectives des avantages et des inconvénients, des réussites et des échecs du système en place et se demander à quels besoins actuels répond ce système actuel.
- Quels sont les problèmes posés par ce système en tenant compte des axes stratégiques fixés par le Ministère ? Les échecs passés peuvent-ils être associés à un ensemble de causes identifiables ?

Quels sont ou quels pourraient être les besoins futurs ? Le but est d'exprimer sans ambiguïté et sans a priori les besoins et les contraintes tant opérationnelles que structurelles, sociales ou économiques auxquelles la ou les solutions recherchées devront alors répondre.

La problématique risque tout d'abord d'être mal comprise, tout au moins mal interprétée. Dans de nombreux cas, elle reste difficile à cerner. Dans le domaine des transports, par exemple, le lancement des premiers TGV sur la liaison Paris-Lyon a été présentée et ressentie comme une avancée majeure. Pour autant, aujourd'hui, malgré la mise en fonctionnement de nouvelles lignes TGV, force est de constater que tous les problèmes de transport dans les régions desservies par ces lignes n'ont pas été résolus, loin de là. Il y a eu progrès technique indéniable, mais la mise en œuvre de cette nouvelle solution n'a pas été faite en intégrant toutes les caractéristiques et les contraintes socio-économiques [Rossel, 1998].

Faut-il, par exemple, aborder, globalement ou non, l'aspect concernant l'insécurité du (des) système(s) de transport (en terme alors de sûreté et de disponibilité par exemple) ou l'insécurité dans les systèmes de transport ? Faut-il étudier séparément les problèmes de congestion et d'environnement ? [Laugeau, 1999].

D'autre part, le nombre important d'acteurs concernés, leurs envies, leurs besoins propres, leurs savoirs nécessairement limités et leurs croyances, en d'autres termes l'ensemble des freins psychologiques auxquels sont soumis ces acteurs, rend illusoire le fait d'obtenir un consensus.

D'autres méthodes et d'outils d'aide à la réflexion, à l'ouverture, à la créativité, à la communication et à la négociation entre les acteurs semblent alors incontournables.

1.2. Elaboration des hypothèses

L'activité suivante consiste à synthétiser et à agréger ces informations pour pouvoir établir la liste des hypothèses retenues : contraintes à respecter, état actuel et état futur souhaité,



trajectoires envisageables pour passer d'un état à l'autre et états intermédiaires atteignables, indicateurs de performances, d'efficience et de pertinence que l'on retiendra pour juger de l'adéquation des solutions futures au problème. L'objectif est donc de formuler la bonne question, sans préjuger des résultats ni des solutions évidentes.

1.3. Recherche de solutions

La troisième activité consiste à rechercher quelle(s) solution(s) peut(vent) être apportée(s) à la problématique initiatrice en tenant compte de contraintes et besoins particuliers. La créativité est nécessaire pour imaginer, rechercher, élaborer et argumenter le cas échéant plusieurs pistes de solutions. Celles-ci, bien qu'éventuellement incomplètes, devront être pertinentes en regard du problème posé et nombreuses pour assurer ainsi une référence suffisante de choix possibles. Il s'agit alors de susciter l'émergence de ces solutions par la concertation de partenaires pouvant prendre en compte l'intégralité du contexte. Chacun devra être capable de créativité individuelle ou collective pour générer des idées nouvelles, associer des idées existantes ou farfelues pour en faire émerger de nouvelles ou encore étudier pour s'en inspirer des scénarios voisins qui sont déjà connus, identifiés et quelquefois solutionnés.

1.4. Sélection de solutions

La quatrième activité consiste ensuite à sélectionner la ou les solutions éligibles réellement. C'est dans cette étape que pourront être utilisées les méthodes d'aide à la décision qui vous seront présentées dans la suite du document. Elles mettront alors en œuvre les indicateurs d'adéquation choisis mais aussi des paramètres plus classiques comme la faisabilité technique, humaine et sociale, le risque encouru, etc. Cependant, il faut à ce moment encore un effort d'ouverture et de créativité. Par exemple, un décideur doit accepter d'étudier sérieusement toutes les solutions qui lui sont proposées, même « celles auxquelles il n'aurait pas pensé » ou celles qui sortent de la norme en vigueur à cet instant. De même, au moment des réflexions stratégiques, il doit pouvoir tenir compte des indicateurs choisis même lorsque ceux-ci ne sont pas en accord avec ses convictions profondes.

Le processus de décision peut donc être vu comme une alternance de phases de divergence et de phases de convergence [Gavriloff, 2001] comme le montre la Figure 1.

En aval de ce processus se situe une étape qui consiste à optimiser et à mettre en œuvre la ou les solutions choisies. Plusieurs type d'outils, de méthodes et de concepts sont utilisables et couvrent plusieurs niveaux opérationnels : stratégie globale, plans de progrès, plans d'action et gestion de projet proprement dit.

De même, toute préparation de décision repose sur une phase de créativité qui occupe l'esprit des décideurs en amont de l'étape de décision à proprement parler, car elle seule peut garantir une plus grande adéquation problème/solution et une plus grande richesse à la fois quantitative et qualitative des idées et des solutions.

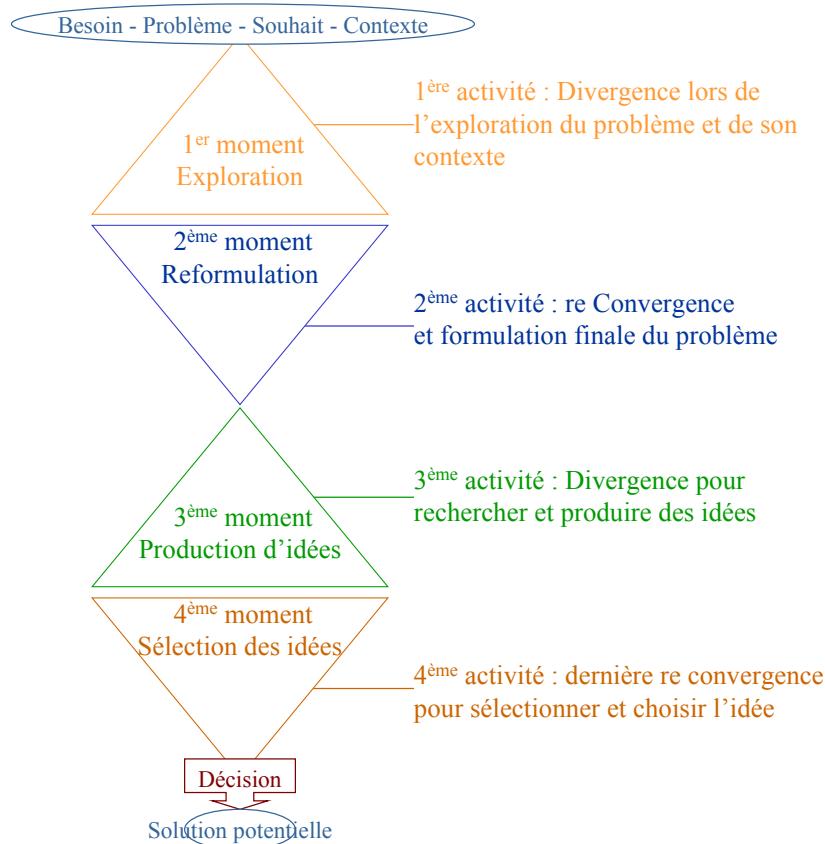


Figure 1 : Une première approche du processus décisionnel

2. L'APPROCHE SYSTEME DES SITUATIONS COMPLEXES

Dans un premier temps, il s'agit donc de procéder avant tout à l'analyse de la situation décisionnelle. Pour cela nous nous intéresserons à deux aspects de l'approche système, la point de vue systémique et l'analyse communicationnelle.

2.1. Approche systémique de la situation

Les politiques publiques et d'une manière générale les choix stratégiques ayant un impact au niveau sociétal sont au cœur de situations qui exigent la mise en place d'organisations et de stratégies de décision complexes.

L'intégration de composantes techniques, environnementales et humaines, en évolution permanente, crée des situations "incertaines" que l'on désigne souvent sous l'appellation de « situations complexes » et que l'on étudie parfois comme des « systèmes complexes »¹.

Le caractère de complexité fait d'abord référence à la difficulté pour un observateur désireux d'intervenir sur la situation de se doter d'une représentation de la situation pertinente pour guider et argumenter son action.

Dans cette première acception, la notion de complexité est mise en avant sans qu'une définition précise ne s'en dégage. Interrelations nombreuses, incertitudes et indéterminations, phénomènes aléatoires, imbrication des niveaux d'organisation, des niveaux de

¹ Le système est alors vu comme une abstraction métaphorique de la situation.



représentations, recours aux modélisations, hétérogénéité des connaissances et des savoir-faire... tels sont les "ingrédients" de la complexité.

On peut cependant préciser le propos en parlant de situations « où la relation homme – système » est complexe.

Par « système », on entend alors un système technique (et par réduction une technologie) ou un système socio-technique (et par extension, un système sociétal).

Par « homme », on désigne dans le cas général un collectif dont l'identité est définie par un certain nombre d'agencements organisationnels (intégration d'artefacts techniques et symboliques dans l'action) et une appartenance institutionnelle ou sociétale.

Par « complexité » de la relation homme–système, on désigne trois caractéristiques :

- émergence de phénomènes non entièrement prévisibles (dans les agencements organisationnels ou dans le système), car partiellement modélisables : boucles de régulation, boucles de causalité circulaires, couples ago-antagonistes, effets non-linéaires, effets combinatoires, effets de seuils, etc.
- dynamique d'évolution : la relation homme-système se transforme au cours du temps, comme se transforment la relation système-environnement et la relation homme-société.
- présence d'incertain : soit par méconnaissance (manque d'information, ou coût prohibitif de l'information ou coût prohibitif du traitement de l'information), soit par impossibilité d'appréhender directement les phénomènes déterminants (on ne peut les observer directement, les mesurer), l'homme est obligé de recourir à des représentations difficilement vérifiables (leur correspondance avec la réalité des phénomènes n'est qu'analogique ou métaphorique), voire même à plusieurs niveaux de représentations.

Ces trois caractéristiques de la relation homme – système conduisent à considérer que toute action en situation complexe comporte une part de risque.

Il arrive souvent que l'*homme* réponde à la complexité de sa relation avec le *système* par le développement d'une interface artificielle (en fait, un artefact technique, lequel peut être matériel ou symbolique) permettant de traiter l'information et/ou de gérer des représentations. Cette interface peut prendre la forme d'un véritable système d'information, avec lequel la relation peut bien sûr devenir également complexe (au sens que nous avons précisé plus avant) et prendre plusieurs formes : activités humaines médiatisées par le système, délégation de tâches au système, substitution du système à l'homme.

Ainsi, et cela est important à garder à l'esprit, il n'est pas nécessaire que le *système* soit complexe pour que la situation le devienne. Un agencement organisationnel ou une interface complexes suffisent à introduire dans la situation une part d'incertain et d'imprévisibilité. L'augmentation des contraintes environnementales sur le *système* (pression sociétale notamment) ou du niveau d'exigence relatif à l'*homme* (réactivité, adaptativité, inventivité) suffisent également à créer une dynamique d'évolution conduisant à la complexité.

La gestion de la situation repose sur le management du risque provenant des caractères d'imprévisibilité, de dynamique et de présence d'incertain. L'introduction de technologies de l'information dans la situation (c'est-à-dire d'interfaces et d'artefacts techniques) produit un impact positif (augmentation du niveau de maîtrise²), mais aussi négatif (apparition de nouveaux dangers).

² La maîtrise est comprise ici comme capacité à comprendre et anticiper les événements non souhaités



De telles situations mobilisent de la part des acteurs humains des capacités de communication accrues, suscitent des processus d'apprentissages collectifs et s'appuient de plus en plus sur des processus de traitement de l'information.

La maîtrise d'une situation, prise comme capacité d'un collectif humain à s'assurer de son évolution est étroitement dépendante de la maîtrise des données, des savoirs et des connaissances qui y sont associés.

La problématique de “l'action en situation complexe” est abordée ici au travers de deux questions générales :

- Comment définir, c'est-à-dire conceptualiser et décrire, une situation complexe ?
- Comment construire une représentation globale qui permette d'organiser l'action ?

Une approche systémique de la situation et la modélisation par les systèmes peut contribuer à fournir des réponses pertinentes à ces questions.

2.1.1. La forme canonique d'une situation

Le rapport privilégié qu'un sujet connaissant établit avec un phénomène perçu dans le cadre d'un projet d'action peut être observé comme une “situation cognitive”. Le sujet doit délibérer sur son action et produire une connaissance opérationnelle, c'est-à-dire utile à ses décisions. Pour l'observateur désireux d'intervenir, la situation est un « objet de connaissance ».

Dans la situation observée, convenons d'appeler “acteur” toute intelligence impliquée directement dans une action intentionnelle portant sur un objet qu'elle différencie de son environnement. On distingue donc l'action de l'acteur sur l'objet et l'intervention projetée par l'observateur sur cette relation. Autrement dit, on ne peut parler d'objet que relativement à des projets [d'action et d'intervention] envisagés par des “sujets” cognitifs [acteur et observateur]. Le caractère “projectif” de l'intervention (j'envisage un futur) le différencie du projet d'action qui se déroule “en temps et en lieu” de l'acteur.

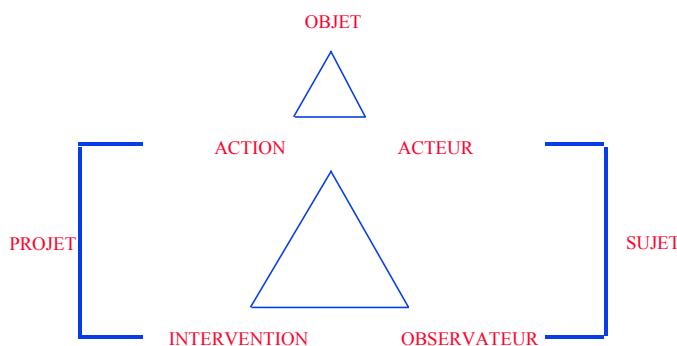


Figure 2 : la forme canonique d'une situation

Nous prendrons comme cas général les situations dans lesquelles l'acteur est pluriel et organisé (un groupe d'acteurs qui partagent la même mission forment une équipe, tandis qu'un groupe d'acteurs partageant les mêmes intérêts dans l'action constitue une organisation), et comme “passages à la limite”, les cas où l'acteur est un individu ou une collectivité (rassemblement de membres qui ne partagent qu'une culture, une idéologie ou simplement quelques mythes).



Dans le cas général, le groupe se distingue de l'organisation en ce que tous les membres sont en interaction dans l'action. Groupe et organisation supposent de chaque acteur des activités orientées vers les autres (négociation, reformulation, argumentation...), des activités collectives (délibération, construction de références communes...), toutes fondées sur la construction de représentations partagées [Teulier-Bourgine, 1997].

Un cas peut paraître particulier, où acteur et observateur sont confondus, cas d'un sujet qui désire intervenir sur la façon dont il agit dans une situation, et, en quelque sorte, observe donc sa propre action : c'est pourtant la situation de tout système intelligent, qui peut et doit observer son propre comportement dans l'action [Pitrat 1991], cette auto-référence étant la condition de son autonomie (J.-P. Dupuy cité dans [Mélèse 1979]). Un cas singulier serait par contre celui d'un sujet qui se prend pour objet, créant ainsi une situation "d'introspection" préjudiciable à l'action.

Contrairement à l'observateur dont la relation avec la situation est unique, le sujet peut se trouver en différentes positions : il est en position active quand il est à l'origine de l'action sur l'objet, mais l'objet peut exercer une action sur le sujet, ce qui met ce dernier en position réactive.

Une situation est donc prise ici comme un état du monde, au sein duquel existe un collectif humain (à la limite, un individu) animé de l'intention de rendre cette situation intelligible, de se préparer à d'éventuelles évolutions et d'acquérir un certain degré de maîtrise dans la transformation délibérée de cette situation [Liu 1993].

Si on peut à la limite créditer le sujet impliqué dans l'action d'une rationalité instrumentale (recherche de l'adéquation fin/moyens avec un recours minimal à un espace mental), l'intervention en situation complexe impose de créditer l'observateur d'une rationalité fondée sur l'adéquation entre les informations possédées et les représentations adoptées (rationalité cognitive).

2.1.2. Comprendre et anticiper une situation complexe

L'observateur-acteur est également un porteur d'enjeux, et à chaque enjeu correspond une dimension particulière de la situation. L'observateur peut privilégier certaines dimensions de la situation pour son projet d'intervention, ce qui délimite autant de champs perceptifs envisageables, dans lesquels sont reconnus des niveaux multiples d'imbrication d'actions, d'intentions et de représentations, dus aux différents acteurs.

Par ailleurs, on remarque aisément que le comportement des acteurs impliqués dans de telles situations et l'émergence de comportements globaux sont peu prévisibles, comme le sont également les conséquences d'actions et d'interventions. A la notion de complexité est donc également attachée la notion "d'imprévisible possible, d'émergence plausible du nouveau" [Le Moigne 1990], qui crée une difficulté d'anticipation de la part d'observateurs.

La notion d'anticipation est prise ici au sens de "se préparer" à faire face à l'évolution (prendre des dispositions à l'avance), selon une certaine idée que l'on s'en fait, mais sans disposer pour autant d'un modèle de prédiction. L'anticipation peut être considérée comme une adaptation des structures cognitives au champ des possibles envisageables dans la situation.

Dans les situations d'actions dites critiques, lorsque le décideur humain se trouve confronté à un risque de conséquences dommageables ou préjudiciables, cette imprévisibilité est traduite par la notion de (prise de) risque ([Guoguelin 1988]. La maîtrise de la situation repose sur un équilibre judicieux entre la satisfaction des objectifs et l'acceptation du risque encouru.



De façon générale, l'intervention en situation complexe ne comporte-t-elle pas toujours une part de risque³, dans la mesure où la portée des actes cesse d'être totalement mesurable (le risque comme "la part d'ombre de l'action" [Ricœur 1995]). Comme le soulignait [Mélèse 1972], la maîtrise des systèmes [complexes] passe en effet par l'acceptation d'un contrôle incomplet. La maîtrise est prise ici comme un état opposable à celui de crise, sans en être son contraire, qui serait la routine [Latour 1994].

La décision en situation complexe s'apparente donc au management du risque : traitement de l'incertain destiné à anticiper les crises (sensibilité à l'alea⁴) et à éviter les ruptures (déstabilisation dûe à l'émergence d'imprévu⁵).

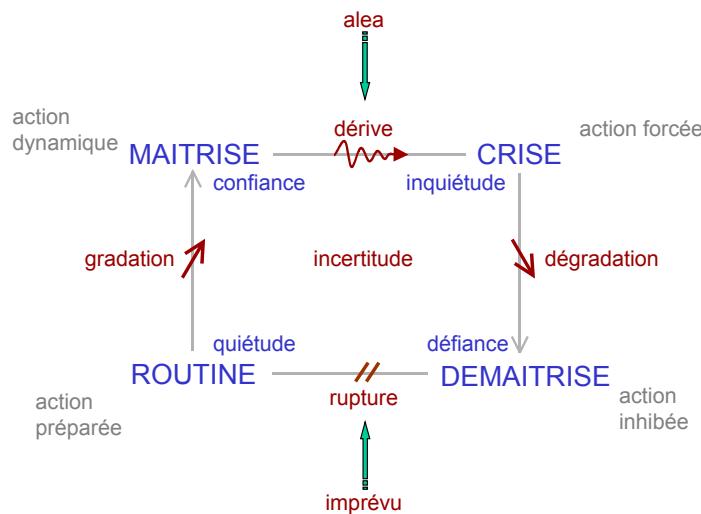


Figure 3 : les situations de management

2.1.3. prise de décision et management du risque

La confrontation d'une intelligence cognitive et d'un environnement évolutif est, d'une manière générale, une situation à risque, où les points de vue de l'observateur et de l'acteur (même s'il s'agit d'une seule et même personne) peuvent diverger.

L'observateur qui forme le projet d'intervenir sur la situation se trouve lui-même confronté au risque de provoquer des conséquences non souhaitées, tombant ainsi sous le coup des principes d'écologie de l'action avancés par [Morin 80, p. 82] :

- le niveau d'efficacité optimum de l'action se situe au début de son développement,
- les ultimes conséquences d'un acte donné sont non-prédictibles.

Dans les situations dynamiques, l'objet (ou sa relation avec son environnement) se transforme indépendamment de l'activité du sujet, ce qui introduit un facteur d'imprévisibilité dans les conséquences de l'action. Pour la prise de décision, les contraintes de temps peuvent minimiser fortement l'activité de compréhension et mettre au premier plan la faculté

³ Le risque n'est pas un défaut d'un projet, un manque d'efficacité, de rentabilité, de vérité, mais il est la nature même d'un projet. [Latour 1994]

⁴ L'alea est défini comme événement relatif à un phénomène connu mais dont l'occurrence ne peut être calculé avec précision

⁵ L'imprévu se distingue de l'alea par son caractère de nouveauté ; il peut s'agir d'un concours de circonstances ou bien d'un phénomène dont l'occurrence dans la situation était impensable ou encore d'un phénomène inconnu jusqu'alors



d'anticipation du sujet, celle-ci étant sollicitée par le caractère irréversible de la décision. Le risque est ici de laisser dans l'ombre des aspects de la situation qui peuvent se révéler essentiels pour la réussite de l'action.

La décision ne saurait ici se ramener à un calcul ; elle est fondée sur l'information mais n'en résulte pas [Jarrosson 1994]. D'autre part, la complexité de la situation conduit à considérer la rationalité de l'opérateur comme étant limitée. La difficulté d'anticipation du comportement du sujet se trouve encore aggravée si plusieurs acteurs sont concernés (et donc plusieurs rationalités) car il faut réaliser dans ce cas une agrégation des points de vues, premiers pas vers une intelligence collective, celle-ci pouvant éventuellement être assistée par des moyens artificiels de traitement d'information [Lévy 1994].

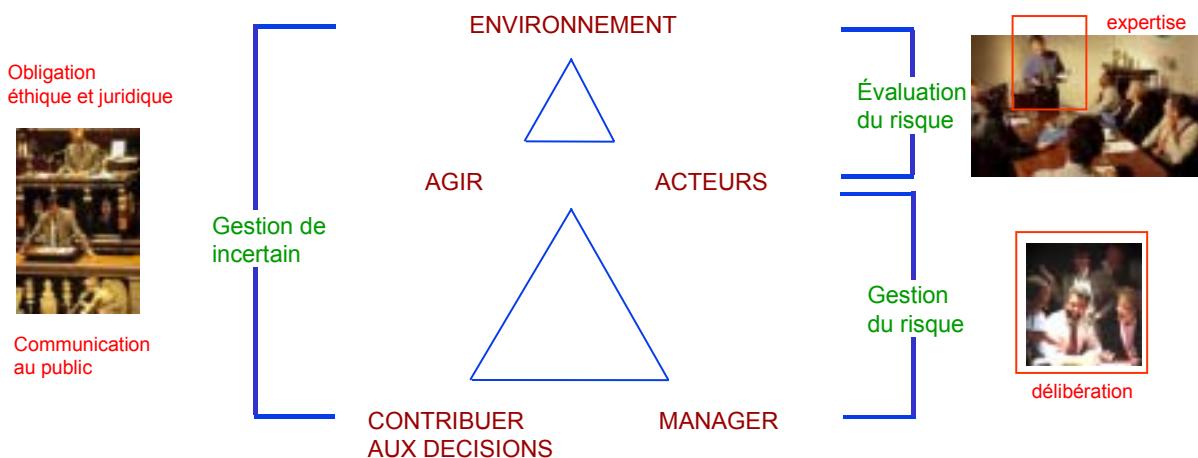


Figure 4 : le management du risque

Le management du risque se fonde sur l'expertise (prudence) pour la gestion des risques (prévention) et s'assure du respect des exigences sociétales (précaution). Le risque peut provenir de l'environnement (naturel, technique, sociétal), mais aussi de l'activité elle-même (activité dangereuse), ou encore du pilotage de l'activité (prise de risque).

L'incertitude dans l'évaluation renvoie à l'*'imprécision'* d'une estimation, à l'*'impossibilité de quantifier'* et à l'*'impossibilité de qualifier'* les impacts

L'incertitude dans la gestion du risque de la situation relève de l'*'imprévisibilité'* et de la *'dynamique'* de la situation.

Le management du risque est l'activité de pilotage de la décision en situation complexe ; il se fonde sur l'expertise (prudence) pour la gestion des risques (prévention) et s'assure du respect des exigences sociétales (précaution).

Ce type de management anticipatif et participatif peut être résumé par la notion de « gouvernance » et ne saurait faire partie de la « mythologie organisationnelle » [March, 1999], selon laquelle toute action, ou tout résultat d'une action est considérée comme un choix et rattaché à un décideur unique, qui est à même d'apporter la solution au problème considéré. Au contraire, on pose l'existence d'un morcellement des processus de décision, lesquels mobilisent des schémas cognitifs complexes (voir § 2.1.5) et qui ont une dimension collective (voir § 3.2).



2.1.4. cognition et situation

L'opérateur humain (le sujet connaissant) doit “s'adapter” à la situation, c'est-à-dire accorder ses structures cognitives aux circonstances du réel perçu afin de prendre des dispositions pertinentes pour obtenir le but visé, ou éviter des événements redoutés. Il combine un savoir contemplatif sur l'objet (il le perçoit au travers de caractères descriptifs qui lui donnent une forme particulière), un savoir faire (en tout premier lieu fabriquer des représentations de l'objet adéquates au projet) et un savoir pratique (un savoir-se-comporter dans l'action en utilisant ses connaissances).

L'objet de la situation représente en fait la partie de l'environnement global qui doit être intelligible pour l'accomplissement du projet. Ce dernier a donc pour fonction de désigner l'objet pertinent dans sa relation avec le sujet (enjeux). Ce dernier est porteur d'une stratégie décisionnelle (choix des finalités et des moyens de les satisfaire), qu'il détermine de façon autonome ou solidaire à l'organisation dont il est membre. Le processus d'adaptation de l'organisation à ses projets et à ses environnements est un comportement d'intelligence organisationnelle. (P. Dehaene, « Ingénierie organisationnelle et systèmes de symboles », in [Bartoli - Le Moigne, 96]).

Pour un observateur désireux d'intervenir dans la situation, les difficultés d'anticipation ou de maîtrise de la situation sollicitent ses capacités cognitives, telles que formulées par [Hofstadter, 1979] :

- capacité de réagir avec souplesse aux situations qui se présentent,
- capacité de tirer profit de circonstances fortuites,
- capacité de discerner le sens d'informations ambiguës ou contradictoires,
- capacité à attribuer une importance relative aux différents éléments d'une situation,
- capacité de trouver des similitudes entre les situations malgré les différences qui les séparent,
- capacité à établir des distinctions entre les situations malgré les similitudes qui les rapprochent,
- capacité à synthétiser de nouveaux concepts en assemblant différemment d'anciens concepts,
- capacité à trouver des idées nouvelles.

L'exercice de ces capacités cognitives correspond pour un observateur au passage d'un sens “brut” de la situation, c'est-à-dire fondé sur l'information qui la décrit, à une compréhension « globale ».

On peut assimiler le processus global d'interprétation d'une situation à un processus de compréhension de scène, où il s'agit de dégager d'une situation supposée convenablement imagée une hypothèse plausible sur la situation (nous proposerons plus loin de donner à cette hypothèse la forme d'un système). Comme dans le processus cognitif de reconnaissance de scènes visuelles auquel nous faisons ici référence, il n'est pas indispensable de supposer que le processus d'interprétation se place après le processus de description. Le travail d'interprétation fait en effet appel à de nombreuses informations sur la réalité perçue et met à contribution les fonctions cognitives, voire même les systèmes de croyance des acteurs [Fodor, 1983]. Parmi tous les mécanismes proposés pour expliquer ce processus d'interprétation deux principes essentiels ont montré leur utilité : la restriction du domaine de réalité observé et la construction de schémas de représentations.



La restriction du domaine est en effet nécessaire dans le cas d'imbrication de niveaux de situations (l'objet de la situation est lui-même une situation). La construction de représentations est l'objet même de la modélisation systémique.

2.1.5. La qualification d'une situation

La qualification d'une situation consiste à en organiser la description ; cette opération est décrite dans [Penalva, 2000] comme un processus cognitif Information – représentation – signification, décrit au §3.2.3.

La situation ainsi qualifiée devient réalité construite que l'observateur peut “interroger” pour accomplir son projet. Mais pour parvenir à cette “intelligence de la complexité”, selon l'expression de J. L. Le Moigne, de quels moyens d'exploration disposons-nous ?

En somme, l'observateur doit effectuer une sorte de saut qualitatif (de la barrière de complexité) pour concevoir un modèle caractérisé en terme de fonction, but, adaptation et considéré en terme d'impératifs plutôt que descriptifs [Simon, 1969]. On demandera à ce modèle de rendre compte de la situation ET de l'observateur. Rappelons avec J. Mélèze que tout individu ou groupe étant potentiellement porteur de sa propre représentation sur une situation, on se trouve devoir construire un “système des représentations”. Néanmoins, ce système n'a qu'une fonction d'interface avec les connaissances de la situation, sans prétendre les contenir. Dans ces conditions, il nous paraît raisonnable de faire l'hypothèse qu'il est possible aux acteurs concernés de construire une référence commune, et de la faire évoluer au cours d'un processus d'apprentissage collectif (formation d'hypothèses et argumentation).

Nous dirons alors que la description relative de la situation est effective lorsqu'elle permet de construire un système commun de représentation(s), qui permet à l'observateur d'agir.

Comme le suggère J. Ladrière, il subsistera toujours une distance [que la connaissance ne saurait abolir] entre la situation et le système qui la représente, ce qui nous amène à considérer le système comme “métaphore de la situation”⁶.

Pour agir en effet, on ne peut rester dans le cadre de cette réalité “objective”, extérieure à l'observateur, celle de tous les faits contenus dans le cadre particulier de la situation ; l'observateur opère dans le cadre d'une réalité “de second ordre” construite dans le cadre d'un projet qui donne au faits ordre, sens et valeur [Watzlawick, 1981].

Cette réalité de second ordre, nous l'appellerons “un système”, et l'opération de passage de la situation perçue au système sera nommée “recadrage”.

2.1.6. Le recadrage des situations complexes

“L'objectivité” de la situation est faible⁷, dans la mesure où son apparition [à la conscience] est dépendante du protocole d'observation, qui influe sur son surgissement et son devenir, voire sa structure [Miermont, 1995].

L'appréhension d'une situation doit donc être vue comme un processus de conceptualisation du réel, formalisée dans [Penalva, 1997] et inspiré de “l'épistémologie formelle” proposée par [Mugur-Schächter, 1995].

Les décisions d'intervention sur la situation dépendent du sens (connaissances

⁶ la métaphore suppose un changement de référentiel

⁷ Le champ réel de la connaissance n'est pas l'objet pur, mais l'objet vu, perçu et co-produit par nous. [Morin 1982]



opérationnalisées) et de la valeur (appréciation de l'utilité des informations) que l'observateur accorde au phénomène perçu. Dans ces conditions, il s'agit moins d'expliquer la situation en recherchant des lois fondamentales que de la "recadrer", c'est-à-dire en construire une représentation qui convienne à l'action⁸.

Illustrons la notion de recadrage avec l'exemple (très simplifié pour les besoins du propos) d'une situation de conduite automobile, où chaque conducteur doit intégrer son véhicule dans un environnement dynamique : le trafic automobile.

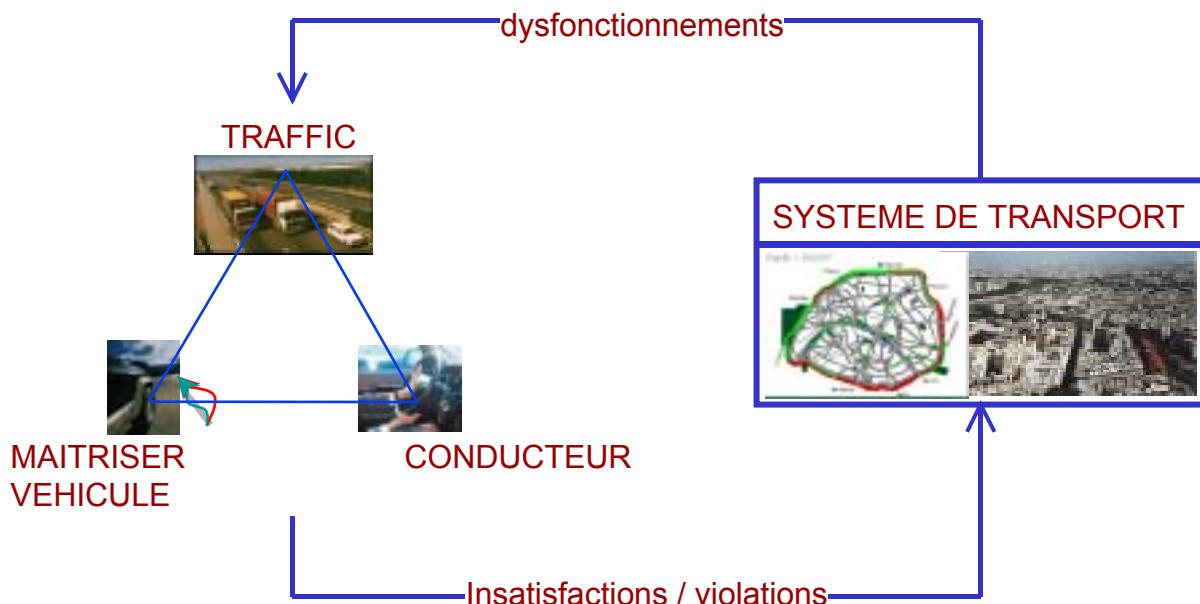


Figure 5 : Le recadrage de situation

Pour répondre aux problèmes de circulation, ou du moins améliorer la situation, l'analyste peut rechercher dans la situation les éléments qui en seraient les causes (événements accidentels, comportement inadéquat des conducteurs, contexte défavorable...).

Un systémicien va plutôt re-présenter la situation comme un "système de transport", apte à fonctionner et qu'il est raisonnable d'isoler pour l'étude, dont les dysfonctionnements ont pour symptômes la perturbation du traffic. Il faut alors agir sur la situation en fonction de cette nouvelle représentation - notamment en la communiquant aux acteurs impliqués - et non exclusivement sur le conducteur considéré comme un « homo violator » (qui n'obéit qu'à ses intérêts, parfois avec irresponsabilité).

Dans cette perspective, la quasi-totalité des violations n'est pas le fait d'individus irresponsables et pervers mais bien davantage le fait d'individus aux prises avec des incompréhensions, des lacunes, des problèmes de perception et d'interprétation dans l'action. On ne saurait non plus réduire le système à une simple « culture de sûreté » aux propriétés standardisantes ; ce concept exprime au contraire un agencement de solutions d'organisation productrices de comportements souhaités.

A l'inverse, le système peut être vu comme générateur « d'accidents systémiques » ou « d'accidents organisationnels ». Le point de vue est ici celui de l'existence d'un fait organisationnel en matière de risque [Bourrier et Laroche, 2001]. L'accident, le

⁸ Ce qu'on modifie en recadrant, c'est le sens accordé à la situation, pas ses éléments concrets.
[Watzlawick, 1973]



dysfonctionnement entraînant des conséquences dommageables, n'est que le révélateur d'une mauvaise défense du système dans son ensemble. Un système sûr est alors un système qui se protège contre le développement « d'histoires d'accidents » par une suite de défenses en profondeur dont l'empilement finit par conférer une fiabilité acceptable au système total [Amalbert et Malaterre, 2001].

2.1.7. Système de représentation et système de référence

La méthode qui consiste à représenter une situation complexe par un système se trouve fondée sur l'intelligence imaginative aussi bien que sur l'observation des faits de la situation. Elle identifie l'observateur-concepteur à un architecte de modèles doté d'un projet de connaissance.

“Un phénomène perçu comme complexe, donc irréductible à un modèle déterminant la prévision certaine de ses comportements, se représente par un système complexe” [Le Moigne, 1990].

La démarche proposée identifie l'observateur-concepteur à un architecte de système : il perçoit une situation qu'il représente comme un système.

La notion de système, telle qu'argumentée par [Bertalanffy, 1973] et auquel [Le Moigne, 1977] a associé le concept de modélisation, nous apparaît particulièrement convenante pour appréhender une situation perçue complexe : traiter un réseau de relations en tant que système, consiste à poser que la connaissance de cet ensemble est subordonnée à la connaissance des parties qui constituent le tout, des interactions entre les parties, des interactions avec l'environnement, et des objectifs de l'observateur-acteur de l'ensemble.

Il ne saurait cependant exister de modèle unique d'un système complexe. La notion de point de vue devient alors essentielle dans le processus de modélisation.

Par contre, le modèle ainsi exprimé doit être partagé par tous les locuteurs en présence : c'est là sa principale qualité.

Toute situation cognitive exige que l'observateur dispose d'une représentation mentale de l'objet de savoir qu'il vise au travers de son système de représentation.

Ce dernier est fondé sur une sorte de modèle de référence que J. L. Le Moigne a baptisé “système général”, artefact doté de propriétés théoriques, destiné à valider les représentations construites [Le Moigne, 1977].

Produire de la connaissance sur la situation revient à modéliser le système qui la représente.

On peut donc dire que le système est un modèle “primaire” construit pour exprimer une réalité perçue, au travers d'un domaine du réel conceptualisé dans un projet.

De façon générale, un discours de type théorique se présente comme une suite de propositions formulées dans un langage qui peut être entièrement artificiel ou consister en une extension du langage naturel.

Il constitue, pris dans un sens très général, une “micro-théorie” ou du moins un discours à caractère théorique, c'est-à-dire qui se réfère au moins à la théorie... des systèmes. Lorsqu'on exprime une situation vécue en des termes relevant de la théorie des systèmes, on constitue une sorte de théorie minimum du domaine destinée à faciliter l'expression des savoirs pertinents sur la situation .

Un système, dans cette première acception, peut être considéré comme une théorie partielle, c'est-à-dire limitée à une situation bien définie (quoique pouvant être complexe). Cela ne



suffit pas à lui donner un caractère formel, puisqu'en général les propositions constituant le modèle associé ne s'enchaînent pas par déduction.

Il n'est pas nécessaire d'ajouter que le modèle doit être un discours "logico-mathématique" comme le fait [Tiberghien, 1991] qui précise fort justement qu'un modèle ne saurait être un simple discours phénoménologique, ne saurait être réduit à un schéma, et n'est pas non plus assimilable purement et simplement à une théorie (laquelle doit posséder un degré de généralité élevé).

Pour "formuler un système", on peut se contenter d'étendre le langage naturel par ajout de catégories interprétatives ; le système constitue donc le support d'un discours argumentatif, dont la validation⁹ consiste à s'assurer que les informations sur la situation prise en compte pour sa construction sont cohérentes avec l'expérience des "sujets connaissants" dans le même domaine. C'est une procédure d'argumentation qui permet de relier les informations utilisées aux propositions sous forme d'énoncés. Ceux-ci doivent être intelligibles et communicables. Les propositions qui formulent le modèle peuvent être évaluées comme valides dans l'univers qu'elles décrivent mais le système ne saurait constituer une théorie complète - aucune autre proposition ne serait vraie - d'un univers considéré comme étant complexe.

On est ainsi amené à différencier le système - discours de type théorique - et le modèle - classe de propositions formulées dans un langage - dans le cadre d'un projet où il s'agit de concevoir le premier et construire le second. Le couple système-modèle ainsi constitué doit être assez riche pour permettre l'interprétation de tous les traits de la situation que l'on retient comme pertinents, c'est-à-dire comme des savoirs utiles au projet.

Ces savoirs doivent être exprimés au travers des différents « points de vue » que l'on peut avoir sur le système.

La notion de point de vue, prise littéralement au sens de "ce que l'on peut voir d'un certain endroit", peut introduire une confusion entre la vision singulière qu'un individu ou qu'un collectif de métier peut porter sur le système et une perspective générale d'examen de certaines propriétés du système.

C'est cette dernière acception que nous avons retenue pour la notion de point de vue. Pour lever l'ambiguïté, la position d'un observateur particulier est dénommée "angle de vue". On ne peut en effet accepter de varier les points de vue au hasard, sans risquer une redondance inutile ou une confusion dans la description du système.

On trouvera dans [Penalva, 1997] une définition formelle de neufs points de vue pour aborder les systèmes finalisés.

⁹ Valider ne signifie pas garantir que l'on dit le vrai, mais légitimer la pertinence de l'énoncé dans son contexte. [Avenier 1992].

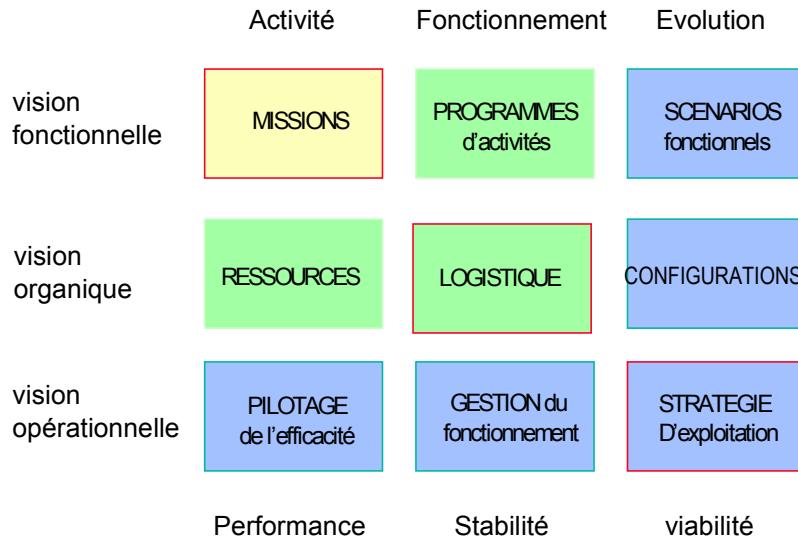


Figure 6 : La matrice des points de vue de modélisation d'un système

Dans chacun de ces points de vue, on peut développer un modèle (il faut alors utiliser un langage de représentation) que l'on peut raffiner jusqu'au niveau de détail utile au projet. Selon le caractère formel du langage utilisé, l'argumentation de l'action donnera lieu à calcul ou à un raisonnement cognitif de la part du collectif devant délibérer sur son action.

2.1.8. De la situation au projet

Modéliser en situation complexe suppose de la part d'un observateur la capacité "paradoxalement" de produire des modèles à propos d'un contexte d'action "qu'aucun modèle ne saurait épuiser". La résolution de ce paradoxe apparent nécessite de reformuler le problème : la complexité de la situation envisagée dans le cadre du projet d'intervention est irréductible à un principe simple ; dans ces conditions, comment argumenter une intervention de façon partageable et raisonnée ?

Rendre compte de la situation comme d'une action complexe, que l'on peut représenter comme un système pourrait être considéré comme une technique de contournement, voire même de réduction de la complexité, puisqu'il s'agit d'expliciter un "principe d'organisation".

Plutôt que d'attribuer cette complexité "de second ordre" à un niveau de complexité trop élevé, intrinsèque à la situation (acceptons ici l'avertissement de [Lévy-Leblond, 1984] "c'est moins les limites de validité ou de pertinence de la complexité qui nous intéressent ici, que celles de son utilité"), il nous semble plus juste de dire que le projet d'intervention doit lui-même être considéré comme complexe, ce qui justifie le déploiement d'une stratégie "par tâtonnements" [Avenier, 1997] fondée sur une rationalité "cognitive" [Walliser, 1992].

Projet et situation se construisent mutuellement sur le socle d'une ingénierie de la représentation systémique ouvrant la voie à "l'intelligence de la complexité".

La démarche mise en œuvre par le collectif de projet consiste d'abord à conceptualiser une situation pour construire une ontologie du domaine abordé : données, savoirs et connaissances élémentaires. La situation est alors interprétée par un système, l'investigation du modèle donnant lieu à argumentation

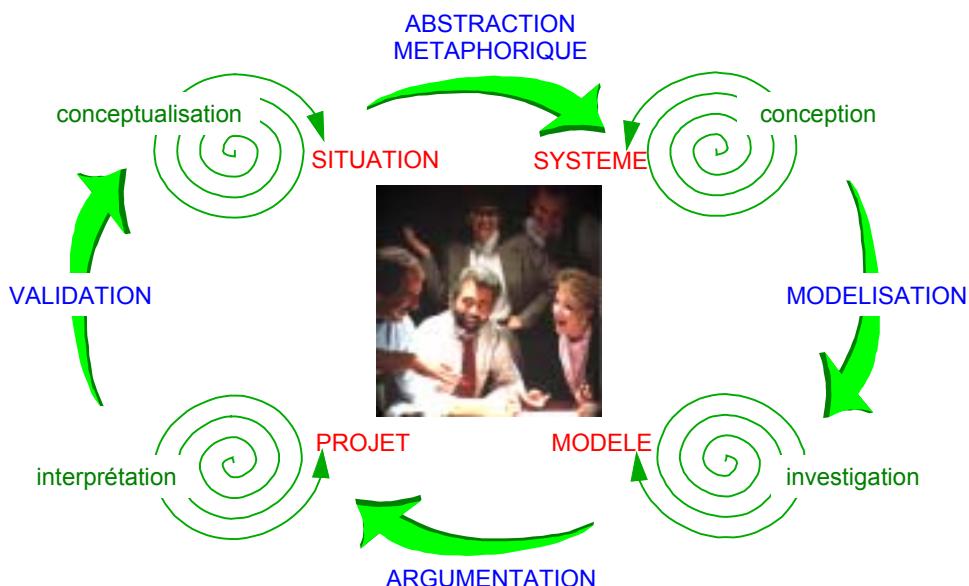


Figure 7 : la décision collective

On peut faire comme première remarque que l'entrée dans ce processus n'est pas déterminée a-priori. Si l'on peut considérer que la démarche théorique consisterait à partir de la situation pour concevoir un système que l'on modélise et qui donne lieu à controverse, la pratique est souvent tout autre. La controverse peut être lancée sur une situation mal définie, de nombreux projets considèrent le système comme un pré-requis, pour d'autres, c'est autour de modèles détaillés que naît une controverse...

Une seconde remarque porte sur le niveau de formalisation du modèle qui lui-même conditionne celui des énoncés. La démarche, que nous avons considéré dans le contexte le plus général - situation complexe, système non formel, modèle cognitif et énoncés discursifs - reste applicable dans les cas où l'argumentation peut produire des énoncés logico-mathématiques - cas d'un modèle à haut niveau de formalisation - dont la validation impose une réduction des dimensions prises en compte dans la situation.

Autrement dit, pour produire des énoncés facilement validables sur la situation, on doit restreindre le champ d'interprétation du modèle dont la formalisation poussée réduit les ambiguïtés, mais réduit aussi la complexité de la situation qu'il est possible de prendre en compte.

Le "détour par les systèmes" consiste à porter l'effort de modélisation sur le système de compréhension d'une situation perçue complexe en regard d'un projet d'intervention, et de laisser la liberté aux acteurs de ce projet de choisir le niveau de formalisation de la connaissance (au sens large de combinaison de données, savoirs et expertises).

Les connaissances sont alors considérées comme un construit collectif dans un système d'action, le point essentiel étant la capacité d'évolution du Système [de Connaissance [de la Situation... ainsi constitué. Les connaissances ne sont plus ici "vérités uniques et transcendantales" mais supports de l'action, donc contingentes et re-modelables, voire relativement éphémères (les connaissances sur le vol des oiseaux, utilisées par les premiers concepteurs d'engins volants, ont perdu de leur utilité pour les concepteurs d'avions...).

Pour générer ces connaissances, la conception d'un système est une méthode que nous avons rapprochée de la démarche de l'architecte ou de l'ingénieur qui délibère sur son action. L'association d'un modèle plus ou moins détaillé s'avère très utile pour "granulariser" la



connaissance, mais doit se prémunir du piège de “ l’élémentarisme ” qui consiste à confondre un élément du modèle avec un élément d’une “réalité révélée”.

Un modèle, qui par ailleurs est toujours réducteur puisqu’au moins rapporté à un champ de projet, est d’abord une façon de penser simplement dans la complexité.

Un modèle n’est ni vrai ni faux, ni bon ni mauvais ; il est plus ou moins utile…

2.2. Approche communicationnelle de la situation

Le projet nouveau est une composante essentielle de l’évolution des organisations. Au sein de celles-ci, mais aussi à l’extérieur, de nombreux acteurs se trouvent associés dans une quête avec un objectif principal : transformer des idées en de nouvelles solutions techniques, technologiques, de nouveaux services... sur lesquels l’organisation va asseoir son développement.

La démarche consiste à identifier, à l’aide d’une approche « communicationnelle », les mécanismes de la construction collective d’un projet. L’idée est d’éclairer la situation pour dégager les éléments essentiels à intégrer dans le processus d’aide à la décision. Le but : repérer comment il serait possible de procéder pour gagner du temps ou éviter certaines erreurs.

La démarche s’articule selon deux dimensions : sémio-contextuelle, et systémique pour dégager les éléments importants à partir d’une étude et de l’analyse des communications digitales et analogiques entre les acteurs de l’opération.

2.2.1. Approche sémio-contextuelle de la communication

L’objet de la « théorie des processus de la communication » [Mucchielli *et al.*, 1998] (ou théorie sémio-contextuelle de la communication [Mucchielli, 2000]) est de faire émerger, dans une approche systémique et constructiviste [Morin, 1994], le sens qui accompagne toute communication généralisée, expression d’intentionnalités explicites ou latentes, dans une situation d’échange par et pour des acteurs.

Le modèle situationnel pour l’analyse des communications, issu de cette théorie, constitue un véritable guide opératoire systémique. Il permet d’appréhender les processus de la communication qui interviennent pour transformer les contextes de la situation et de saisir le sens, résultat d’une mise en relation des acteurs en situation dans ces différents contextes.

Les processus de la communication, mis en évidence par Alex Mucchielli à partir de 1995 [Mucchielli, 1995], agissent sur sept contextes dans l’échange [Berger *et al.*, 1986].

Le contexte spatial

Le lieu de réunion, l’aménagement de la salle, la position des individus autour de la table, la distance entre les personnes, les moyens techniques, sont des composantes qui, évoquées ou manipulées par les acteurs d’une situation, influencent le sens des communications.

Il est réducteur de considérer que ces différentes composantes forment un cadre figé, une « mise en scène » statique. Le contexte spatial évolue en permanence pendant une rencontre. Il forme un contexte subjectif dynamique qui s’impose aux différents acteurs.



Le contexte physique et sensoriel

L'acteur communique par et à travers l'ensemble de ses organes sensoriels : parole, mais aussi vue, ouïe, odorat, ou toucher. Dans toute action de communication, il est immergé dans un univers physique et sensoriel qui contribue au sens de l'échange.

Des impacts sensoriels multiples (visuels, sonores, olfactifs, tactiles) en combinaison avec un contexte spatial organisé et des acteurs prédisposés, peuvent transformer les modalités de perception et la signification des communications [Hotier, 1997].

Le contexte temporel

Toute communication généralisée s'inscrit dans un contexte historique et dynamique. Pour faire émerger le sens d'une communication, il est nécessaire de la positionner dans son contexte temporel. La modification de ce contexte peut en changer le sens, de façon partielle ou complète.

Sur le plan historique, la communication vient toujours se positionner dans le cadre de ce qui a déjà été communiqué. Sur le plan dynamique, elle est par exemple caractérisée par un rythme, régulier ou en rupture, et par des appels au temps, explicites ou implicites.

Le contexte des positions respectives des acteurs

Dans son langage (tutoiement) mais aussi dans son paralangage (habillement), chaque individu propose sa vision du positionnement des acteurs de la situation. L'échange est une lutte permanente dans laquelle l'individu cherche à valoriser sa position.

Le processus de positionnement et de structuration des relations est particulièrement important dans le domaine des TIC compte tenu : de l'hétérogénéité des acteurs impliqués dans la construction des systèmes, et de l'impact des systèmes sur la place des acteurs dans l'entreprise.

Le contexte culturel de référence aux normes

Toute communication s'effectue dans un contexte normatif. Les normes sont multiples : culturelles et sociales dans la vie quotidienne, ou conduites en entreprise. Elles définissent un sens « *a priori* » partagé par un groupe d'individus.

Les normes et règles forment un « déjà là », par exemple sous la forme de sentences proverbes dictons ou maximes dans la vie quotidienne. Elles sont évoquées de façon explicite ou implicite, évoluent en permanence, et quelquefois émergent dans l'échange.

Le contexte relationnel immédiat

Dans la communication, les acteurs construisent une relation affective. Chaque individu utilise le langage et le paralangage pour séduire et influencer ses interlocuteurs. Pour faire émerger le sens, il est nécessaire d'identifier les composantes valorisantes et dévalorisantes mises en œuvre dans l'échange.

De nombreuses études ont montré que les relations dans l'entreprise sont structurées autour du phénomène de pouvoir. Aussi, il est essentiel de comprendre quelles sont les alliances, coopérations, compétitions, ou conflits en œuvre dans la situation.



Le contexte expressif des identités des acteurs

Tout individu est doté d'un système de pertinence qui en fonction de ses préoccupations forme sa vision du monde, c'est-à-dire une perception sélective des phénomènes de la vie [Schutz, 1987]. Il n'est pas possible de saisir le sens d'une communication sans comprendre la motivation des acteurs de la situation.

Les spécialistes distinguent chez l'acteur trois niveaux d'intentionnalité : une intentionnalité générale (faire carrière, investir dans un loisir, ...), une intentionnalité présente (obtenir une augmentation, être l'acteur incontournable sur un sujet, ...), et une intentionnalité réactionnelle (prendre la bonne décision).

Le modèle situationnel

Le modèle situationnel [Mucchielli *et al.*, 1998], issu de la théorie des processus, forme un complément naturel à la méthode des études de cas [Yin, 1984] [Stake, 1995] : « technique particulière de cueillette, de mise en forme et de traitement de l'information qui cherche à rendre compte du caractère évolutif et complexe des phénomènes concernant un système social comportant ses propres dynamiques » [Collerette, 1996].

Cette approche combinée, pour identifier et analyser des processus récurrents dans un ensemble de situations (étude de cas multiple) ou pour étudier une situation unique assez complexe (étude de cas intrinsèque), peut être mise en œuvre en deux étapes :

La première consiste à fixer le cadre de travail, recueillir et préparer les informations :

- Poser une problématique [Paillé, 1996] avec : un thème de travail, les considérants ou attendus, et une proposition de cadrage.
- Collecter les informations sur la ou les situations en utilisant une ou plusieurs méthodes qualitatives [Mucchielli *et al.*, 1996] : observation, observation participante, interviews non directives centrées, interviews de groupes ...
- Transcrire l'ensemble des informations, existantes et collectées, sous une forme pratique : réservier par exemple une colonne en partie droite des documents.
- Parcourir l'ensemble des documents mis au net et signaler dans la zone réservée à cet effet les contextes modifiés et processus de la communication.

La deuxième consiste à analyser les communications à l'aide du modèle situationnel :

- Découper le cas en « phases » : série de situations différentes compte tenu de l'intervention des processus de la communication.
- Choisir, parmi l'ensemble des processus de la communication qui sont interdépendants et dans une démarche d'analyse systémique, le processus émergent.
- Analyser la situation à l'aide de ce processus et ensuite compléter l'étude à l'aide des autres processus de la communication mis en œuvre dans la situation.
- Le sens, résultat d'un processus complexe, surgit alors progressivement de la mise en relation des acteurs en situation dans ces différents contextes.

2.2.2. Approche systémique des communications

L'objet de la « théorie systémique des communications » [Mucchielli *et al.*, 1999] est de faire émerger, dans une approche systémique et constructiviste [Morin, 1994], le sens qui accompagne toute communication-participation (segment d'une structure générale d'échange entre acteurs) pour accéder à la compréhension globale du système de communications



(ensemble récurrent régulier et repérable de formes d'échanges existant entre acteurs participant d'un cadre d'action) [Mucchielli, 2000].

La théorie systémique des communications est formée d'un ensemble de principes « premiers » et « auxiliaires » porteurs d'injonctions méthodologiques. De la communication au système, ou des systèmes aux communications, ils facilitent le repérage des causalités circulaires complexes dont la découverte va progressivement éclairer la compréhension de la situation.

Nous allons maintenant présenter de façon synthétique les principes premiers qui forment le noyau dur de la théorie systémique des communications.

Toute communication n'existe que dans un système

Un phénomène demeure incompréhensible tant que le champ d'observation n'est pas suffisamment large pour qu'y soit inclus le contexte dans lequel ledit phénomène se produit [Watzlawick *et al.*, 1972]. A l'image de la prestation du musicien dans l'orchestre [Winkin, 1981], la communication-participation (phénomène rationnel et relationnel) est insérée dans un système de communication (système qui forme son contexte).

La quête de l'analyste consistera à découvrir et expliciter le système de communication, qui ne peut pas ne pas exister. Dans de nombreux cas, l'identification de ce système conduira l'analyste à découvrir de nouveaux phénomènes de communication.

Le sens est porté par les catégories générales de communications

Le segment de communication-participation prend plusieurs formes : l'expression communicative brute (contenu explicite), la forme générale (construite par empilement des observations), et l'interaction (catégorie générale qui porte le sens).

A partir de la transcription des communications sous leur forme littérale (expression communicative brute), le travail de l'analyste consistera à faire émerger les concepts généraux (formes générales et interactions) qui sont les éléments clés de la modélisation systémique.

Tout système de communications forme un contexte

Le système dans lequel toute communication s'insère forme nécessairement un contexte à cette communication [Mucchielli, 2000]. Le sens surgit de la mise en relation d'une communication avec son contexte, c'est à dire l'ensemble des communications qui forme le système de communications.

La compréhension va émerger de l'explicitation du sens des communications dans le contexte systémique pertinent du point de vue des acteurs en situation. La démarche est complexe car elle nécessite un effort d'empathie indispensable pour appréhender l'ensemble des contextes individuels.



Les communications agissent à travers des boucles d'interaction

L'interaction est un classe d'événements qui ont lieu lors d'une présence conjointe et en vertu de cette présence conjointe [Goffman, 1974]. La situation est le siège de phénomènes communicationnels qui construisent par implication mutuelle une causalité circulaire complexe.

L'analyste s'attachera à rechercher les boucles d'interaction entre phénomènes de communication en privilégiant les formes générales d'échange. Il pourra s'appuyer pour ce faire sur l'identification des enjeux portés par les acteurs en situation.

Tout système de communications est régi par des règles

Tout système de communication est composé d'un ensemble de boucles d'interaction. Il possède une logique interne qui peut réguler son fonctionnement (homéostasie) mais aussi le détruire (dynamitage ou blocage progressif).

La découverte de la logique interne du système est un élément essentiel dans l'analyse de la situation. Cette logique peut être liée au système englobant (cadrage), ou portée par les acteurs en situation (elle est en général liée aux avantages que les acteurs attendent du jeu collectif).

Les émergences du système participent à sa restructuration

Le système, composé de boucles d'interaction, doté d'une logique interne, et d'une force homéostatique, est le siège de phénomènes appelés : « émergences systémiques ». Ces émergences systémiques naissent de la mise en relations des phénomènes dans le ou les contextes communicationnels.

L'identification des émergences systémiques est un exercice difficile pour l'analyste. Il s'agit en effet de mettre en évidence les effets propres au système et à son fonctionnement que seule une approche en compréhension permet d'identifier.

Le système est le siège de phénomènes paradoxaux

Les communications faites de langages et attitudes (digitales et analogiques) sont souvent contradictoires. Elles peuvent être paradoxaux et constituer un véritable piège, qui plonge le sujet dans un état de confusion mentale.

Pour identifier les communications paradoxaux l'analyste s'attachera à repérer les différences entre le contenu des communications et leur partie implicite et relationnelle. Partie relationnelle qui agit souvent par des processus non conscients.



Le modèle systémique

A l'image du modèle situationnel issu de la théorie des processus de la communication [Mucchielli *et al.*, 1998], le modèle systémique issu de la théorie systémique des communications [Mucchielli, 2000] forme un complément naturel à la méthode des études de cas [Yin, 1984] [Stake, 1995] : « technique particulière de cueillette, de mise en forme et de traitement de l'information qui cherche à rendre compte du caractère évolutif et complexe des phénomènes concernant un système social comportant ses propres dynamiques » [Collerette, 1996].

Cette approche combinée, pour identifier et analyser des phénomènes récurrents dans un ensemble de situations (étude de cas multiple) ou pour étudier une situation unique assez complexe (étude de cas intrinsèque), peut être mise en œuvre en deux étapes :

La première consiste à fixer le cadre de travail, recueillir et préparer les informations :

- Poser une problématique [Paillé, 1996] avec : un thème de travail, les considérants ou attendus, et une proposition de cadrage.
- Collecter les informations sur la ou les situations en utilisant une ou plusieurs méthodes qualitatives [Mucchielli *et al.*, 1996] : observation, observation participante, interviews non directives centrées, interviews de groupes ...
- Transcrire l'ensemble des informations, existantes et collectées, sous une forme pratique : réservier par exemple une colonne en partie droite des documents.
- Parcourir l'ensemble du corpus mis au net et signaler dans la zone réservée à cet effet : systèmes, interactions (catégories générales d'échange), contextes, boucles de rétroaction, logique des systèmes, émergences systémiques, et communications paradoxales.

La deuxième consiste à analyser les situations à l'aide du modèle systémique :

- Découper le cas en « situations » : définies à partir du recouplement des éléments culturels du système de pertinence des acteurs en présence.
- Construire pour chaque situation le modèle systémique (acteurs, communications-participations) en s'attachant à privilégier les formes générales d'échange (qui portent le sens) de préférence aux formes littérales (le contenu explicite).
- Repérer les éléments complexes (boucles de rétroaction, logique des systèmes, émergences systémiques) en effectuant des allers retours permanents entre corpus annoté et analyse.
- Le sens, résultat du processus, surgit alors progressivement de la mise en relation des acteurs en situation dans les différents contextes.

2.2.3. Conclusion

Dans une démarche d'aide à la décision, et en fonction de la situation naturellement différente pour chaque opération, l'analyse pourra à l'aide de l'une ou l'autre de ces deux approches



dégager un modèle situationnel, abstraction forcément simplifiée du système réel à intégrer dans un système destiné à faciliter la prise de décisions.

2.3. Approche organisationnelle de la situation

Ce chapitre présente un historique rapide des théories autour de l'information, principalement en sciences de l'information, en sciences de gestion et en sociologie des organisations.

Ces théories ne sont pas uniformes, se contredisent et ouvrent un débat qui sera repris aux chapitres suivants sur la gestion de l'information en pratique dans les entreprises et sur son lien avec les formes d'organisation.

2.3.1. Rappels sur la notion d'information

De l'information juridique à l'information binaire

L'origine étymologique du mot information est *Informatio* qui désigne en latin l'action de façonner, mettre en forme. Jusqu'à la seconde guerre mondiale, informer est surtout utilisé pour signifier « instruire une affaire » au sens juridique.

A la fin de la seconde guerre mondiale et en liaison avec les techniques de communication à distance, Claude Shannon pose les fondements de la théorie mathématique de l'information. Il définit l'information comme le degré d'incertitude contenu dans un message transmis d'un émetteur à un récepteur.

Le message est codifié en une suite de 0 et de 1 puis est transformé en signal électrique pour être transmis automatiquement. Arrivé à destination, le message est décodé. Codification et transmission provoquent des erreurs qu'il faut corriger. Cette correction correspond à une remise en ordre des 0 et des 1, l'information est la mesure du désordre ainsi provoqué. Lorsque plusieurs messages sont émis, l'information mesure également la quantité de nouveauté introduite volontairement cette fois par l'émetteur.

L'information, la mesure de l'entropie

Warren Weaver généralisera l'usage du terme « information » en faisant référence à la notion thermodynamique d'entropie qui signifie : degré de désordre d'un gaz. L'information est alors LA mesure de l'entropie (d'un message électrique) ou de son contraire (négentropie). Il assimile rapidement la mesure de désordre avec la prise en compte de la signification des messages et propose que l'information soit autant « forme » que « sens » [Bougnoux, 1993].

Cette confusion est toujours d'actualité même si les travaux récents sur la gestion des connaissances tentent de limiter l'information à une simple mise en forme de données (qui restent également difficiles définir ! [Vacher, 2000].



L'information, l'informatique et la logique formelle

A la même époque, Von Neumann combine les travaux de Turing sur les algorithmes¹⁰ et les fonctions logiques de l'algèbre de Boole¹¹ pour construire les premières machines de calcul¹² utilisées par l'armée américaine. La puissance des machines construites sur ce modèle de Von Neumann est telle que l'on croit avoir trouvé le langage idéal : « La logique est une valeur universelle, un outil de compréhension du monde autant qu'un moyen de le transformer » [Breton, 1987].

Vingt siècles plus tôt, la logique formelle était devenue l'instrument universel de la raison avec les syllogismes d'Aristote, “les règles nécessaires à l'édification d'un savoir”¹³. Or cette connaissance par la raison était l'apanage de l'être humain, étape primordiale vers la vérité, le plus grand bonheur à atteindre. Mais “la raison grecque, c'est celle qui de façon positive, réfléchie, méthodique, permet d'agir sur les hommes, non de transformer la nature”¹⁴.

Il est ainsi intéressant de noter comment, en très peu de temps, c'est-à-dire entre la seconde guerre mondiale et le début des années 50, les mots « information » et « logique » sont associés pour produire une nouvelle idéologie : résoudre les problèmes humains grâce aux machines créées par ces derniers¹⁵. Cette idéologie persiste même si elle est de plus en plus relativisée [Vacher, 2000].

2.3.2. Théorie de la décision et sciences de l'information

Si l'on considère encore souvent que « l'information est ce dont le décideur a besoin », c'est également en référence aux premiers travaux en sciences de l'information. L'auteur le plus connu est H. A. Simon qui a élaboré la théorie économique de la rationalité limitée [Simon, 1997 ; Vacher, 2000¹⁶] présentée en 1947 dans *Administration et processus de décision*. Il a également été un précurseur dans les sciences de l'information définies comme : l'étude des propriétés et des caractéristiques de l'information, c'est-à-dire les facteurs de flux, les moyens de traitement et les représentations des informations.

Sélectionner l'information, décider, planifier

Simon précise que la difficulté consiste à traiter l'information car elle est trop abondante. Pour cela, il met en avant les outils informatiques¹⁷ qu'il appelle des prothèses de l'homme au sens où ils aident ce dernier à :

¹⁰ Algorithmes = règles permettant la résolution de problèmes donnés

¹¹ L'algèbre de Boole est un formalisme pour écrire et traiter des expressions logiques.

¹² Computer en anglais (machine à calculer) a été traduit en français par ordinateur (machine à mettre de l'ordre), ce qui n'est pas anodin.

¹³ *La philosophie antique*, page 75

¹⁴ *Les origines de la pensée grecque*, page 133

¹⁵ il s'agit du mythe de la création artificielle parfaite, à l'image de son créateur (même si l'inventeur des premières machines, Von Neumann, ne croyait pas que les facultés humaines puissent se réduire à un simple mécanisme).

¹⁶ Pour résumer succinctement, les individus décident sur la base d'une information incomplète, en tenant compte d'un nombre limité d'alternatives et sans connaître exactement les résultats de leurs décisions. Ils cherchent donc des solutions satisfaisantes et non optimales (comme dans la théorie économique classique) aux problèmes qu'ils se posent.

¹⁷ La traduction de Computer Science est informatique qui signifie « traitement automatique des informations ». On utilise aussi souvent le terme « système d'informations ».



- poser plus rationnellement les problèmes,
- filtrer les informations et
- simuler et planifier l'action qui devra suivre.

Simon édicte également quelques principes pour la conception de ces outils dans *Le nouveau management* [Simon, 1980]. Il dit qu'il faut :

- comprendre la manière dont les décisions sont prises dans l'organisation,
- soulever les questions auxquelles l'information va répondre,
- adopter une approche arborescente et modulaire des problèmes.

Un modèle centralisateur et mécaniste de l'organisation

Plusieurs difficultés pratiques apparaissent face à ces assertions :

- la première est qu'aucune différence n'est faite entre l'homme et l'ordinateur concernant les capacités de traitement de l'information. Simon reste persuadé qu'un « *ordinateur pourra accomplir à notre époque n'importe quelle tâche cognitive dont est capable une personne* » [Simon, 1980],
- la seconde est que les qualités requises pour concevoir de tels systèmes d'information sont rarement réunies : à la fois bien connaître l'organisation et sa gestion, l'informatique et ses possibilités,
- la troisième suppose qu'il est possible et souhaitable de câbler toute l'organisation et son environnement pour résoudre tous les problèmes que se posent un décideur¹⁸,
- le fait d'assimiler le décideur au dirigeant d'entreprise, ou au moins d'un département de l'entreprise, laisse croire que tout autre individu ne décide pas mais agit d'après les décisions prises par le dirigeant. Celui-ci déciderait en fonction de ce qui lui semble être les buts de l'entreprise et le reste du personnel s'y conformerait,
- enfin, un tel modèle de l'organisation et de la place de l'information suppose un processus séquentiel et la possibilité d'anticiper avec une faible marge d'erreur : poser des problèmes (stratégiques, ie en fonction des buts de l'entreprise), chercher l'information pour résoudre les problèmes, sélectionner ces informations, décider, planifier, agir et éventuellement corriger l'énoncé des problèmes en fonction de l'évolution du contexte de l'organisation et recommencer la boucle.

Des liens ambigus entre l'information et la décision

Pour nuancer ces approches trop mécanistes, J.G. March rappelle la rationalité limitée dans *Décisions et organisation* [March, 1989] et insiste sur les conditions pratiques de la prise de décision et de l'élaboration de systèmes d'information. Il montre que :

¹⁸ Cette utopie reste d'actualité à travers notamment les offres d'outils logiciels pour reconfigurer « toute » l'entreprise ou, de façon moins mécanique, certaines propositions sur les réseaux de veille reliant tous les acteurs d'une organisation, d'un secteur, d'un marché ou d'une filière.



- on collecte toujours plus d'information qu'on n'en utilise (« on ne sait jamais... » est le discours associé),
- on utilise les informations pour s'affirmer, pour persuader les autres mais pas toujours pour diminuer une incertitude collective. L'auteur rappelle ici l'importance des conflits d'intérêt,
- on mobilise l'information comme un symbole de compétence, une marque d'efficacité sociale. L'information est alors un signal de légitimité même si elle ne sert pas directement à prendre de décision,
- les buts de l'entreprise ne sont jamais clairs car les préférences des uns et des autres sont vagues, contradictoires et évolutives, elles ne contrôlent pas les décisions. L'information mobilisée, même inutilisée pour la prise de décision, permet de construire des histoires cohérentes une fois la décision prise,
- les problèmes, les solutions et les actions ne se présentent que rarement de façon séquentielle. Les décisions se développent d'elles-mêmes au fur et à mesure que les actions prennent sens pour les personnes de l'entreprise. Les managers ne sont pas des décideurs tout puissants mais jouent un rôle pour aider leurs employés à donner du sens à leur expérience. L'information prend la forme d'histoires pour interpréter l'action.

2.3.3. L'action est située et la cognition distribuée

Théorie de la décision et action située

Lucy Suchman s'est opposée à H.A. Simon sur le modèle de la planification (ou de la programmation de l'information) où l'action est supposée être résolue par le plan et ne servirait éventuellement qu'à le raffiner. Elle montre au contraire que l'action est située, c'est-à-dire inscrite dans le contexte de circonstances concrètes et particulières. Le plan n'est alors qu'une ressource pour l'action : il est un construit social fortement dépendant de l'histoire et de l'expérience des individus qui l'élaborent. Le but du plan est de préparer à l'action pour mobiliser au mieux son savoir incorporé.

Dans la préface de son ouvrage *Plans and Situated Action* [Suchman, 1987], elle présente l'exemple de deux modèles de navigation : celui où chaque mouvement est prévu à l'avance et celui où seul l'objectif à atteindre est prévu. Dans le premier cas, on peut décrire exactement le chemin parcouru s'il a bien été modifié à chaque circonstance non prévue initialement. Dans le second cas, les efforts sont portés sur l'adaptation à la situation (nuages, étoiles, vagues, etc.) plutôt qu'au respect d'un plan mais il est impossible de décrire précisément a posteriori la route empruntée. Si notre culture occidentale nous incite à adopter le premier modèle, L. Suchman montre qu'il reste une utopie puisqu'il est impossible en pratique de modifier le plan pour l'adapter totalement à l'action. En revanche, le fait d'utiliser un plan comme prescription pour l'action permet de contrôler ou de justifier a posteriori des écarts en termes de moyens : retards, dépenses supplémentaires, etc., cela selon le point de vue de ceux qui ont élaboré et modifié le plan sans relation immédiate aux actions.

Incomplétude et imperfection de l'information

Selon cette approche, la place de l'information et de ses systèmes est radicalement différente de celle qu'elle avait selon H. Simon. Quels que soient les efforts fournis, l'information est non seulement incomplète (rationalité limitée des acteurs) mais également décalée, parfois inutile, au sens de « directement opérationnelle », pour l'action (même si elle joue un rôle



important socialement ou pour justifier telle ou telle action plus ou moins conforme aux prévisions, [Vacher, 2000]).

Ce point de vue est primordial pour comprendre les nombreux échecs en matière de gestion de l'information qui se voudrait exhaustive et parfaitement adaptée. Il permet de raisonner de façon plus pragmatique aux outils, non plus d'aide à la décision, mais de support de l'action (on remarque que le plan est un de ces outils, qu'il soit informatisé ou non).

Cognition distribuée et artefacts cognitifs

Un autre élément importe pour appréhender l'information et sa gestion dans une organisation. Non seulement l'information dépend de l'histoire des acteurs en présence mais elle se trouve dans de multiples supports qu'il est impossible de réunir : dans la tête des protagonistes, dans des documents et des instruments de travail, dans la situation dans laquelle se passe l'action (agencement des locaux, outils disponibles, etc.), dans le contexte de l'action (les contraintes de temps, de moyens ou autre), etc¹⁹.

On parle alors de cognition distribuée, dont une excellente illustration est donnée par E. Hutchins dans son article sur Comment le cockpit se souvient de ses vitesses (Hutchins, 1994). L'auteur présente l'aspect distribué de la mémoire entre différents supports, c'est-à-dire le pilote, le copilote, les outils de navigation à bord et les documents d'aide (check-list, tableaux de calculs de vitesse en fonction de la masse, cartes, etc.). Il montre l'intérêt d'utiliser de façon redondante ces supports, qu'ils soient visibles (l'horizon artificiel par exemple) ou auditifs (le fait de lire la check-list à haute voix par exemple) : cela permet de décharger les pilotes d'une partie de l'activité cognitive grandement sollicitée.

C'est la coordination de ces supports qui rend compte du processus de mémorisation et d'actions résultantes (le savoir en action) : "Les pilotes sont constamment en train de lire et d'écrire, de reconstituer et de reconstruire le sens et l'organisation des représentations internes et externes des vitesses".

Définitions - La cognition n'est plus ici seulement prise au sens du fonctionnement du cerveau humain ni même de celui des machines informatiques mais concerne un dispositif constitué de ressources composites (matérielles, humaines et symboliques) et reliées entre elles qui s'agencent mais ne peuvent se regrouper en une seule. C'est en ce sens que l'on parle de cognition distribuée. Le modèle centralisateur est remis en cause.

Les ressource symboliques ou matérielles (objets, documents, etc.) sont aussi appelées des *artefacts cognitifs*²⁰, c'est-à-dire des dispositifs artificiels créés pour amplifier les capacités de la pensée humaine. Ces artefacts modifient la nature de la tâche à effectuer pour chacun des individus et amplifient les performances du dispositif. La check-list par exemple permet aux pilotes de ne pas retenir tous les points listés mais les oblige à les énoncer. Les pilotes travaillent donc différemment avec ou sans check-list et le cockpit est plus ou moins performant selon la présence ou non de ce document.

¹⁹ On différencie le contexte de la situation : pour simplifier, la situation est le moment, le lieu, les objets, etc. de l'action et le contexte est ce qui est autour de l'action (avant, après, à côté, etc.).

²⁰ (Norman, 93)



3.1.4 - Donner du sens à l'action

De la décision à l'action : synthèse des théories

Avec la théorie de la décision, on supposait qu'il était possible de prévoir l'information utile (sous-entendu, dans le cadre de décisions mûrement réfléchies par les dirigeants). Le plus gros travail consistait alors à filtrer et traiter cette information, les systèmes d'information servant à réduire, voire à remplacer, l'homme dans ce travail.

Les critiques de J. March et de L. Suchman forcent à relativiser cette approche : il n'y a pas d'information indépendamment d'individus qui la produisent et l'utilisent dans le cadre de mises en situations particulières (actions situées). Pour l'organisation, l'information n'est plus seulement associée à une problématique d'efficience (moyen d'atteindre un objectif), c'est un élément indispensable d'un dispositif plus large qui rend l'action possible : des personnes qui se parlent, des objets que l'on manipule, des textes lus tout haut, etc. On parle alors de cognition distribuée pour signifier le fonctionnement de ce dispositif qui consomme et fabrique du savoir (qui n'est donc pas seulement dans le cerveau des individus).

Information, connaissance, compétence, savoir

La compétence regroupe la connaissance et l'information - Plus généralement, on a souvent tenté de définir séparément l'information et la connaissance, en précisant par exemple que :

- soit l'information est l'interprétation d'une donnée et la connaissance la mise en contexte de l'information pour son utilisation,
- soit l'information est le résultat d'une action à partir de données et la connaissance le fait d'avoir assimilé l'information pour renouveler l'expérience.

On place alors derrière la notion de connaissance un ensemble de savoirs : théoriques, organisationnels et situationnels - qui sait et fait quoi, pourquoi.

Dans ce cas, on définit également la compétence comme étant le savoir agir qui regroupe le savoir-faire, le savoir comment faire et avec qui. On y ajoutera le savoir être lié au comportement et aux motivations des individus [Durand, 2000].

La connaissance regroupe la compétence et l'information - En faisant référence à la théorie évolutionniste et à l'approche de l'organisation par l'information et la connaissance [Vacher, 2000], on peut aussi retenir un point de vue inverse : la connaissance regroupe les compétences et les habiletés des individus. Elle se construit dans l'interaction entre les individus.

La connaissance est :

- *tacite* ; il s'agit de la connaissance pratique, des routines et de l'expérimentation autant que la connaissance du contexte, de la culture d'entreprise,
- *explicite* ; elle correspond à de l'information interprétée (et notamment écrite) et mise en contexte.

Les auteurs de [Nonaka et Takeuchi, 1995] remarquent que la culture occidentale incite à privilégier la création de connaissance explicite en tentant de formaliser la connaissance tacite et de combiner les connaissances explicites de chacun des membres d'une entreprise. L'objectif est ainsi de rendre la connaissance moins personnelle et plus exploitable



indépendamment du contexte et des individus qui en sont à l'origine. A l'opposé, les Japonais préfèrent enfouir le savoir pour mieux se protéger (internalisation de la connaissance explicite) et privilégient la socialisation pour une meilleure appropriation et utilisation de la connaissance de chacun [Vacher, 2000].

S'affranchir de la distinction entre information et connaissance - L'évolution des théories prenant de plus en plus en compte la situation et la place des individus incitent à nuancer ces différences entre occidentaux et Japonais. La pratique confirme toutefois encore souvent ces différences.

Pour en revenir à la distinction tentée entre information et connaissance, il est très délicat en pratique de tenter une distinction claire entre ces termes car ils dépendent du point de vue de l'observateur [Mayère, 1997]. On remarquera simplement que l'utilisation du mot connaissance renvoie plus explicitement à une dynamique de création et d'échanges, dans lesquels les individus ont une position privilégiée.

Création de sens dans l'action collective et résultats tangibles

Les savoirs communs – Dans *Communication et intelligence collective*, Michèle Grosjean et Michèle Lacoste étudient longuement le travail à l'hôpital et les occasions de mobilisation de connaissance dans l'action [Grosjean et Lacoste, 1999]. Elles insistent sur le fait que la connaissance se construit dans l'interaction (en face à face ou à distance) entre les individus. Les savoirs deviennent communs par constitution de représentations compatibles à partir de référentiels communs et d'un modèle de l'interlocuteur que chacun se fabrique pour se comprendre mutuellement et agir en conséquence.

Ces savoirs ne peuvent pas être considérés comme un stock dans lequel il suffit de puiser. Ils « *se construisent au fur et à mesure de l'action commune, des échanges et des conflits, et de la contribution quotidienne à une même organisation* » [Grosjean et Lacoste, 1999].

L'information et le sens –

Si l'information (ou la connaissance) est distribuée entre les personnes et les artefacts cognitifs, si son lien à l'action n'est repérable qu'en situation et dans l'interaction, il faut rajouter une dimension à la gestion de l'information. Il ne s'agit plus seulement de filtrer, traiter l'information pour décider, ni encore de l'interpréter pour agir mais aussi de donner du sens à l'action, de comprendre ce que font les uns et les autres.

La théorie de l'interactionnisme propose également de passer d'une logique de stock à une logique de processus. Elle est liée au sensemaking, terme anglais pour désigner la création de sens. Elle s'est à nouveau constituée en complémentarité de la théorie de la décision. Un auteur important de ce courant est Karl Weick qui caractérise le sensemaking par les points suivants résumés par B. Journé dans son ouvrage *Les organisations à risques* [Journé, 1999] :

- le sens se construit a posteriori d'une action et dans le cadre d'interactions entre individus. Ces derniers partagent un minimum de référentiels communs mais pas forcément des valeurs, des croyances ou les mêmes représentations. L'action se situe dans un environnement sur lequel les individus peuvent agir.
- Le flux de l'action ne devient tangible et ne prend sens qu'à partir du moment où les acteurs lui prêtent une attention. Cette attention peut être stimulée par des événements inhabituels, des objets ou des symboles qui obligent à s'arrêter (remplir un document par exemple), etc.



- Le sens donné par les acteurs n'est pas la vérité mais « les gens découvrent ainsi ce qui se passe et ce qui doit être fait ».

La vigilance mutuelle et l'intelligence collective – dans leur article *Collective Mind in Organisations*, K.E. Weick et K.H. Roberts insistent très largement sur l'attention des individus dans l'action en faisant le lien avec l'intelligence collective [Weick and Roberts, 1993].

L'intelligence collective provient d'actions corrélées entre elle avec vigilance et dans un système social. L'action et l'intelligence sont plus performantes lorsqu'il y a chevauchement de savoir plutôt que spécialisation des tâches et la structure globale émerge d'interactions locales (l'organisation en réseau).

Les auteurs définissent la vigilance comme l'attention mutuelle qui existe lorsque les personnes agissent comme si elles étaient un groupe (anti-individualisme). Les actions convergent quand les représentations individuelles du social sont structurellement similaires (système solide), ce qui est une construction faite au cours de partages d'expériences, d'activités faites ensemble (contributions) et d'histoires racontées sur les événements, notamment aux nouveaux entrants.

Il y a baisse de vigilance (donc baisse d'intelligence collective et baisse de performance) quand il y a baisse d'interdépendance : l'attention est focalisée sur les situations locales ou les outils et non pas sur les liens entre les situations. Les individus agissent avec attention sur l'objet local mais sans respect les uns pour les autres, ce qui est un terrain favorable à l'amplification de petites erreurs qui se combinent pour provoquer des désastres (augmentation de l'incompréhension mutuelle sans que les intervenants ne s'en rendent compte).

Cette définition de la vigilance et sa relation à l'action et au partage d'information est au cœur des préoccupations actuelles des entreprises mais pose des problèmes de mise en pratique autant que de conceptualisation. Notre culture occidentale n'est pas étrangère à ces difficultés, qui n'a pas théorisé cette notion pourtant très ancienne. J. M. Oury insiste pourtant sur son importance et propose en 1983 une *Economie politique de la vigilance* pour associer l'activité vigilante à la confiance que doivent se porter les acteurs, à l'importance d'un projet collectif, à l'attention permanente des moindres détails de l'environnement, à l'agilité, la souplesse, la rapidité, la mobilité, etc [Oury, 1983]. La vigilance est l'art d'éviter la crise en prêtant une attention de tous les instants aux mouvements (les actions des autres acteurs, les paramètres des autres décisions, etc.).

Dans un langage économique, J.M. Oury mobilise les notions de contexte pour montrer la possibilité de percevoir le mouvement et de s'y adapter, non pas par une révolution globale mais par petits sauts successifs.

3. APPROCHE SYSTEME DES COMPORTEMENTS COLLECTIFS INTELLIGENTS

Dans un second temps, considérons la dimension collective de la décision en organisation, (sans préjuger ici de la nature de la cognition, artificielle ou humaine). On s'intéressera en particulier aux différents phénomènes qui font que d'un ensemble d'individus (au sens large du terme) isolés, on puisse parvenir à une organisation complexe.

3.1. Systèmes multi-agents

Les systèmes dynamiques sont constitués d'un grand nombre d'éléments interagissant entre eux et avec l'environnement. Dans cette section, ces éléments ne présentent pas ou peu de



capacité cognitive contrairement aux individus des sociétés humaines où chacun a la capacité de penser intentionnellement ses relations aux autres. Cependant des propriétés remarquables au niveau du système global peuvent émerger de la somme des comportements individuels et ceci d'une façon auto organisée, c'est-à-dire sans qu'aucun système de décision central ne coordonne les actions individuelles.

Ainsi, une approche centralisée de la gestion du trafic routier se heurte aux difficultés du recueil et du traitement d'une grande quantité de données, de la prise en compte de spécificités locales très diversifiées, et de l'évolution rapide de l'état du trafic au cours du temps. C'est pourquoi une approche originale telle que celle initiée en Allemagne dans le projet INVENT, tend à répartir les tâches d'acquisition de données sur les véhicules eux-mêmes et à délocaliser l'intelligence de la gestion du trafic afin d'augmenter la réactivité du service rendu aux usagers. Il serait alors possible d'évaluer l'état local de la circulation et d'informer si besoin les usagers concernés qui éviteraient ainsi des dangers ou au contraire exploiteraient une solution potentielle. Ainsi, considérant les infrastructures routières comme un ensemble de ressources distribuées et dynamiques dont l'utilisation doit être optimisée, par analogie avec le fonctionnement des organisations virtuelles (entreprises ou usines virtuelles), les usagers pourraient adapter leur parcours en temps réel, la répartition de ces ressources pouvant faire l'objet de négociations pour tenir compte des priorités et contraintes imposées aux différents types d'usagers. Cette approche est basée sur des concepts d'intelligence collective réactive où les acteurs interagissent et adoptent des comportements individuels qui tendent à la fois à maintenir l'intégrité et la survie du système et à permettre son adaptation aux changements de l'environnement.

A partir d'exemples de phénomènes collectifs de nature physique ou biologique dans ce qui suit, nous mettons en évidence des conditions d'émergence de comportement que nous qualifions d'intelligent car traduisant une capacité à résoudre des problèmes de façon générique.

3.1.1. Processus naturels d'auto-organisation

Les exemples proposés dans ce paragraphe 3.1.1 ont pour objectif d'illustrer les notions d'auto-organisation, de réactivité et d'adaptabilité, d'intelligence et de mémoire collectives, de distribution et de partage d'informations, de négociations et d'interaction entre les « individus » d'une organisation, sur des phénomènes naturels qui nous sont familiers et que nous ne regardons pas nécessairement avec le niveau d'abstraction pertinent ou nécessaire pour en comprendre le fonctionnement et en saisir toute la complexité. Le lecteur pressé pourra donc se contenter de se reporter directement aux conclusions partielles de ce paragraphe.

Phénomènes collectifs de nature physique

Présentation (d'après [Weisbuch, 1989])

Considérons un système physique dans lequel chacun des constituants peut-être caractérisé par son état, le nombre d'états accessibles étant fini. Une configuration du système décrit l'état des constituants à un instant donné. Sous l'influence des autres constituants ou de celle d'une perturbation extérieure, un état peut être modifié et une nouvelle configuration établie. On peut imaginer que ces changements d'état sont régis par des probabilités de transition.

Ainsi, la thermodynamique statistique permet d'expliquer les propriétés macroscopiques de tels systèmes à partir des propriétés microscopiques de leurs constituants. En particulier, dans le cas où le nombre de constituants est suffisamment grand, il peut être possible d'évaluer des



grandeur physiques macroscopiques par le calcul de moyennes locales. Un exemple classique de cette approche concerne l'étude des spins d'Ising, un spin étant un constituant ne pouvant prendre que deux états possibles.

Pour un tel système, il est possible de définir une énergie de configuration. L'interaction entre constituants est pris en compte par la notion de température et la probabilité d'observer une configuration d'énergie donnée est déterminée par le facteur de Boltzmann.

Selon la température, certaines configurations sont privilégiées. Ainsi, aux basses températures, les configurations de faible énergie sont favorisées alors que la probabilité d'observation des configurations de plus haute énergie augmente avec la température. A température nulle, le système évolue vers la configuration d'énergie minimale.

A température non nulle, le système évolue de façon à minimiser l'énergie libre du système. L'énergie libre est définie comme la différence entre l'énergie et un terme proportionnel à l'entropie. L'entropie est d'autant plus forte qu'un grand nombre de configurations sont accessibles au système. Ainsi l'augmentation de température favorise le désordre.

L'évolution dynamique du système est déterminée par les probabilités de transition entre configurations voisines. Une probabilité de transition est d'autant plus forte que la configuration finale est d'énergie inférieure à celle de la configuration initiale. De plus, pour une variation d'énergie négative donnée, la probabilité de transition augmente lorsque la température baisse.

Ce modèle permet d'expliquer les propriétés ferromagnétiques des matériaux sous forme cristal. Le ferromagnétisme caractérise des matériaux magnétiques dont les interactions entre constituants voisins sont identiques. A température nulle, le système évolue vers une des deux configurations où chacun des spins présente le même état binaire. Ces configurations d'énergie minimale constituent des attracteurs du système et décrivent la phase ferromagnétique du matériau. Lorsque la température monte, on observe davantage de changements d'états, l'aimantation diminue puis s'annule dans la phase paramagnétique. Le système se trouve alors à un col d'énergie, les configurations voisines d'énergie élevée sont stables. Le basculement dans l'autre état stable ferromagnétique peut alors se produire si la température baisse. Ce basculement peut être favorisé par un élément extérieur, par exemple un champ magnétique même faible influençant l'orientation des spins. La température de changement de phase s'exprime en fonction du champ magnétique moyen exercé par les voisins d'un spin donné. Ce champ moyen joue le rôle d'un facteur d'ordre qui est nul dans la phase désordonnée paramagnétique et maximal dans l'une des deux phases ferromagnétiques. Ainsi, la transition entre comportements ferromagnétique et paramagnétique se produit à une température critique proportionnelle à l'intensité du champ magnétique local entre constituants.

Dans le cas de matériaux de forme amorphe désordonnée tels que les verres, les interactions entre constituants ne présentent pas les propriétés de symétrie des cristaux. Par exemple, pour un modèle de verre présentant des interactions entre constituants fixes mais aléatoires, le nombre d'attracteurs varie comme une fonction exponentielle du nombre de spins. Parmi tous ces attracteurs, seul un petit nombre présente un bassin d'attraction suffisant pour piéger le système lors du refroidissement du matériau.

Conclusions

Cet exemple illustre comment des propriétés macroscopiques de systèmes constitués d'un grand nombre d'éléments en interactions, peuvent être modifiées sous l'effet de modification des interactions locales entre éléments. Ainsi, lors des transitions de phase les symétries sont brisées et le système désordonné se réorganise pour atteindre une autre configuration



d'énergie faible. Dans ses états de transition critiques, le système est sensible aux perturbations extérieures qui peuvent orienter le processus de reconfiguration. Ainsi un système s'adapte d'autant mieux à un environnement extérieur qu'il se trouve dans un état de transition critique. La maîtrise de cette faculté d'adaptation suppose d'une part de faire atteindre au système un état de transition critique au prix d'une dépense énergétique élevée et d'un désordre important et d'autre part d'influer efficacement sur le système afin de le faire évoluer vers une configuration jugée satisfaisante.

Le domaine de la biologie révèle des cas réussis d'adaptation de systèmes complexes à leur environnement. Nous présentons dans la suite l'exemple du système immunitaire.

Systèmes complexes adaptatifs : l'exemple du système immunitaire

Présentation

L'immunité peut-être définie comme l'ensemble des mécanismes biologiques permettant à un organisme pluricellulaire de maintenir la cohérence des cellules et des tissus qui le constituent et d'assurer son intégrité en éliminant ses propres constituants altérés et les substances étrangères ainsi que les agents infectieux auxquels il est exposé [Revillard, 1998].

Les mécanismes de défense des organismes pluricellulaires utilisent des cellules spécialisées, des systèmes de communication intercellulaires, des boucles de régulation et des processus de réparation tissulaire. Le nombre et la variété des agents infectieux provenant de l'environnement étant quasi infini, le système immunitaire doit présenter une grande flexibilité lui permettant de réagir aux agressions nouvelles. De plus, sous l'effet de l'action du système immunitaire lui-même, les agents infectieux évoluent selon un mécanisme de sélection naturelle, si bien que l'organisme est amené à adapter constamment ses réponses défensives. Le système immunitaire présente une capacité de 'reconnaissance' entre constituants de l'organisme, c'est à dire une capacité de distinction entre le soi et le non soi. Cette capacité est nécessaire pour éviter l'auto destruction des propres cellules de l'organisme hôte.

L'immunité naturelle ou "non spécifique" correspond à des fonctions de défenses primitives et immédiates qui mettent en jeu des protéines préformées, des signaux cellulaires d'activation et des médiateurs solubles de communication. Lors de la pénétration de l'agresseur, le processus immunitaire active un ensemble d'éléments défensifs formant le système "du complément". Une réaction inflammatoire se produit alors dans le but de limiter l'extension de l'infection. La réaction de complément provoque aussi l'opsonisation (recouvrement de l'intrus par des molécules de l'hôte) et la phagocytose (capture et ingestion de l'intrus) suivie d'une bactéricide (destruction de cellules). Lors d'une infection virale, les cellules de l'hôte infectées produisent des cytokines (médialeurs de communication intercellulaire) provoquant l'intervention de cellules naturelles 'tueuses' qui détruisent les cellules infectées et inhibent la réPLICATION virale.

L'immunité spécifique, acquise ou adaptative, se développe en quelques jours et est caractérisée par la mémoire immunologique. Cette mémoire permet une réaction plus rapide et plus intense lors d'une agression future identique (propriété utilisée dans la vaccination).

L'immunité spécifique met en jeu des cellules spécialisées appelées lymphocytes T et B. La reconnaissance de constituants de microorganisme appelés antigènes se fait par interaction de ces antigènes avec des molécules complémentaires des lymphocytes. L'interaction de l'antigène avec les récepteurs membranaires des lymphocytes induit la prolifération (expansion clonale) et la différenciation de ces derniers. Les lymphocytes B se différencient en plasmocytes sécrétateurs d'anticorps (molécules d'immunogloguline), et en lymphocytes B à mémoire. Les lymphocytes T se différencient en lymphocytes régulateurs (chargés de développer ou de limiter l'expression de l'immunité naturelle et de la production d'anticorps



par les lymphocytes B) et en lymphocytes effecteurs et cytotoxiques ou agissant sur les cellules de l'immunité naturelle. Les différents lymphocytes T exercent leurs activités par contact membranaire direct ou à distance par sécrétion de cytokines. Les anticorps générés par les lymphocytes B interagissent avec l'antigène pour faciliter l'opsonisation et la phagocytose du microorganisme parasite. Le phénomène de tolérance au soi est acquis au cours du développement des lymphocytes T et B, par élimination ou inactivation des lymphocytes dont le récepteur reconnaît les constituants du soi.

Les lymphocytes T et B sont fabriqués respectivement dans le thymus et dans la moelle épinière (organes lymphoïdes centraux) à partir de cellules souches. Ces organes sont le siège d'une intense prolifération et de la destruction de la plupart des cellules lymphoïdes selon un processus de sélection. Les lymphocytes matures gagnent alors les organes lymphoïdes secondaires où ils se distribuent en zones fonctionnelles avant de subir leur deuxième étape de différentiation induite par les antigènes comme évoqué ci-dessus. Les organes lymphoïdes secondaires sont organisés en fonction du mode de pénétration de l'antigène (dans la rate pour les antigènes pénétrant par voie sanguine, dans le ganglion lymphatique régional pour les antigènes pénétrant par voie cutanée ou dans certaines muqueuses). Les lymphocytes non fixés circulent du sang vers les ganglions puis dans la lymphe avant de revenir dans le sang.

Conclusions

Ainsi, le système immunitaire présente un niveau d'organisation très élaboré mettant à la fois en jeu des mécanismes d'auto-organisation puisque les différents constituants ne sont pas pilotés de l'extérieur. Il existe aussi une forme de contrôle centralisé et planifié qui assure que l'activité immunitaire se développe conformément au code génétique de l'organisme hôte.

De nombreuses transitions d'état se produisent non seulement au niveau des cellules impliquées mais aussi au niveau du système qui subit des transformations locales lors du passage de l'état de veille à l'état de défense (sollicitation du système du complément). Pour que le système puisse s'opposer efficacement à la diversité des influences de son environnement, il est nécessaire qu'il puisse mettre en œuvre une diversité supérieure de comportements. La réaction inflammatoire peut être interprétée comme une phase transitoire entre les états qui précèdent et succèdent à l'infection. Les phénomènes d'interactions entre composants confèrent au système global des propriétés remarquables de perception de l'environnement, de capacités d'actions finalisées, de communication et d'apprentissage. En effet le système est capable de percevoir et de distinguer son environnement constitué des cellules et microorganismes hôtes (constituants propres de l'organisme mais aussi substances ou microorganismes extérieurs présents par exemple dans la flore intestinale et utiles à l'organisme) et des agents infectieux extérieurs. Des actions adaptées sont systématiquement exercées en fonction du mode de pénétration des antigènes dans un but de survie c'est à dire de viabilité du système. Celui-ci est capable de réagir à une infinité de constituants pour les éliminer sans pour autant s'auto-détruire. Des dysfonctionnements graves peuvent cependant se produire provoquant des déficiences ou un emballement du système immunitaire.

Par analogie avec le système physique précédemment décrit, le risque de basculement dans un mode stable mais 'non désiré' car non viable est peut-être la conséquence de la grande faculté d'adaptation à l'environnement. Les activités locales peuvent faire l'objet de phénomènes d'amplification mettant en jeu des boucles de réaction positive et pouvant conduire soit à la destruction soit à l'expansion de cellules mais aussi des phénomènes de régulation par boucles de réaction négative (action des lymphocytes régulateurs). Un mode de communication basé sur la génération de médiateurs intercellulaires peut provoquer des actions à distance mais aussi le recrutement de constituants supplémentaires nécessaires à la défense. Le système immunitaire est de plus doté de capacité de mémorisation et d'apprentissage lui permettant de



reconnaitre une agression déjà subie et de réagir plus rapidement et plus fortement que la première fois.

En définissant l'intelligence comme la capacité d'interaction avec son environnement selon des boucles de perception/action de sorte à satisfaire des contraintes de type viabilité, on peut reconnaître au système immunitaire une forme d'intelligence.

Alors que les constituants du système immunitaire réalisent des fonctions précises, l'introduction d'un degré d'autonomie, c'est à dire de choix de l'action exercée par les éléments du système, confère au système une plus grande variété de comportements comme nous allons le voir dans les sociétés d'insectes.

Auto organisation dans les sociétés d'insectes.

Présentation (d'après [Bonabeau *et al.*, 1994]) :

Dans une fourmilière les comportements individuels semblent aléatoires et même absurdes. Pourtant l'activité résultant de ces comportements désordonnés présente un niveau d'efficacité suffisant à la survie de l'ensemble. Les conduites collectives les plus spectaculaires concernent la réalisation de réserves de nourriture.

Chez les fourmis l'activité de récolte peut s'effectuer de manière individuelle, mais la découverte d'une source de nourriture peut être suivie d'un recrutement sous l'une des trois formes suivantes :

- un recrutement d'un seul individu qui garde un contact antennaire permanent avec le recruteur
- un recrutement de groupe : lors de son retour au nid la recruteuse dépose une trace chimique évanescante qui sert de guide aux membres de la colonie qui répondent à l'invitation de la recruteuse.
- un recrutement de masse : la piste chimique suffit alors à provoquer le recrutement d'un grand nombre de fourmis.

Les mécanismes de récolte varient selon le mode alimentaire des fourmis, la taille de la colonie et la richesse de l'environnement en nourriture.

Dans certaines espèces de fourmi adoptant une stratégie de récolte individuelle sans recrutement on observe une répartition en deux catégories. La première est constituée de fourmis restant au voisinage du nid et récoltant ensemble dans des zones de forte densité de population. Dans l'autre on trouve des individus qui récoltent loin du nid, chacune dans une portion de territoire distincte des autres. Une telle organisation se produit sans qu'il y ait de communication directe entre individus. L'explication de cette organisation repose sur des principes d'apprentissage individuel et de compétitions indirectes entre individus. Le temps qui sépare le retour au nid d'un nouveau départ est plus long si la récolteuse revient bredouille que lorsqu'elle ramène une proie. De plus des succès répétés développent une mémoire à long terme, et la probabilité de retourner sur le même point d'approvisionnement augmente.

Ainsi en supposant que les probabilités d'explorer les différentes zones de territoire soient égales au départ, les chances pour que deux individus tombent sur une même source de nourriture éloignée diminue avec la distance au nid. La récolteuse ayant trouvé la première une source de nourriture éloignée y retournera plus probablement, et spécialise ainsi son territoire de recherche. Les autres récolteuses quittent plus rarement le nid, ne se spécialisent pas, leur présence dans une des zones restant aléatoire. Elles peuvent alors consacrer plus de temps à d'autres activités au sein de la fourmilière. Si le nombre de fourmis augmente, on



observe que sous l'effet de la compétition intrasociale, le phénomène de spécialisation disparaît car la probabilité qu'une autre fourmi découvre une source de nourriture éloignée augmente diminuant alors les chances de succès successifs d'un individu dans une même zone.

Pour d'autres espèces de fourmi, la récolteuse ayant découvert une source de nourriture dépose à son retour une piste chimique et invite un petit groupe de fourmis à la suivre. La piste devenant plus fortement marquée chimiquement provoque un recrutement de masse. C'est l'environnement et non pas les individus qui conservent la mémoire du chemin qui mène à la nourriture. Cependant certaines fourmis se perdent et contribuent à la découverte d'autres sources possibles de nourriture. Lorsque des sources de même richesse sont déposées simultanément près du nid, les deux commencent à être exploitées simultanément mais bientôt un phénomène de bifurcation se produit conduisant à l'exploitation de l'une des sources au détriment de l'autre. Par principe de dépôt d'une trace chimique évanescante, le recrutement par groupe est plus efficace dans le cas de colonies importantes.

Chez les abeilles, une récolteuse de retour à la ruche entame une danse indiquant la direction, la distance, la richesse et la facilité de récolte d'une source de nourriture donnée. Lorsque la qualité de la source de nourriture augmente, la vigueur avec laquelle la danse est effectuée augmente ainsi que la fréquence des visites à cette source. Les autres abeilles estiment indépendamment le rapport coût/bénéfice et ajustent leur réponse en tenant compte aussi d'autres facteurs tels que les besoins internes de la colonie. Ainsi une récolteuse de retour à la ruche a le choix d'abandonner la source de nourriture et d'attendre une invitation intéressante, de recruter des congénères ou de retourner directement à la source de nourriture pour continuer seule la récolte. L'organisation de la récolte qui en résulte est adaptée aux besoins de la colonie et à la richesse des sources de nourriture.

Conclusions

Le comportement collectif intelligent s'évalue à la capacité de résoudre des tâches génériques (comme ici la constitution de réserve de nourriture), à la faculté d'adaptation et d'apprentissage et à la propriété d'ajustement des comportements individuels à ceux des autres individus.

Les phénomènes d'intelligence collective ne nécessitent pas de régulateur central capable de détecter et de compenser des déviations globales et aucun des éléments impliqués ne possède de connaissance suffisante pour construire une représentation globale de la situation. Néanmoins, de tels systèmes sont capables de résoudre des problèmes de manière robuste distribuée et adaptative.

Les conditions d'existence d'un phénomène d'intelligence collective sont :

- la présence d'un collectif muni de possibilités d'interactions
- la possibilité d'interactions directes ou indirectes avec l'environnement
- l'intégration finalisée des comportements individuels pour atteindre un but

Les interactions se distinguent par :

- leur nature (stimulus ou modification du milieu)
- le mode de stimulation qui peut être soit direct (émanant d'un individu vers d'autres) soit indirect (via une trace laissée sur le milieu extérieur). Une conduite est dite socialisée lorsque sa réalisation implique l'échange de signaux interindividuels.



- leur effet : immédiat selon un processus de stimulus/réaction ou retardé lorsqu'il induit un changement de l'état réactionnel.

La société fonctionne à la fois comme un tampon protecteur vis à vis des perturbations extérieures et comme un milieu propre qui conditionne l'activité de chacun des individus présents dans la société. La structure d'une société est une forme transitoire qui peut changer, s'adapter, évoluer ou disparaître et non pas uniquement ce que le fonctionnement social tendrait à reconstituer ou à préserver. En fait sont présents de façon concomitante un processus de morphostase qui, par contre réactions négatives, tend à maintenir la forme, l'organisation ou l'état du système et un processus de morphogenèse qui, par contre réaction positive, tend à élaborer et à changer la forme, la structure ou l'état du système. Les deux processus ont pour but de satisfaire la contrainte de viabilité soit par compensation ou atténuation des perturbations extérieures soit par restructuration interne et adaptation aux modifications environnementales. L'activité supérieure émerge des interactions du collectif intelligent et est elle-même soumise à une contrainte de viabilité dont la satisfaction est nécessaire à la survie des agents individuels.

Les avantages d'une approche de résolution collective sur une approche centralisée sont les suivants :

- une plus grande fiabilité du collectif (la capacité de résolution émane du collectif et non pas d'un individu unique)
- une plus grande flexibilité (adaptation à l'environnement)
- une plus grande robustesse (l'erreur individuelle est tolérée, l'aléatoire et le bruit sont source d'adaptativité).
- affranchissement d'un contrôle central et d'une programmation explicite de tous les cas potentiellement rencontrés.

3.1.2. Apprentissage et systèmes distribués

L'apprentissage est un processus d'acquisition de nouvelles compétences. Après apprentissage, un individu ou un groupe d'individus est capable d'un comportement efficace dans des situations qu'il ne savait pas gérer préalablement. Nous avons mis en évidence dans les exemples précédents l'importance des processus d'apprentissage pour l'adaptation du collectif à son environnement.

Differentes formes d'apprentissage

On peut considérer trois modes d'acquisition de compétences :

- par construction ou transmission génétique : l'instinct est un savoir-faire non appris au cours de la vie. Il est inné et héréditaire. De même un système artificiel peut être conçu doté de certaines capacités comportementales.
- par instruction : le sujet apprend à partir d'exemples ou de conseils, il cherche à reproduire le comportement d'un 'modèle' ou d'un 'professeur'.
- par expérience : le sujet apprend par processus d'essais erreurs.

L'apprentissage est un processus dynamique. D'après G.Clergue, l'apprentissage apparaît comme une trajectoire dans l'espace des phases des systèmes cognitifs du sujet. Les concepts émergent de la convergence créée par les bassins d'attraction existants lorsque l'apprenant tente d'assimiler le réel aux schémas qu'il s'est déjà construit. De nouveaux attracteurs se



forment et remodèlent le paysage dans un désordre chaotique en s'accommodant aux variations imprévisibles de l'environnement. Pendant l'apprentissage, l'émergence d'un nouveau concept apparaît comme une transition de phase entre un état cognitif ancien et le nouvel état [Clergue, 1997].

Chez l'animal, l'apprentissage met en jeu des processus cérébraux plus ou moins complexes. Pour Pavlov, le conditionnement est à la base de tout apprentissage. Il permet une extension des réactions comportementales à un nombre plus grand de stimulations et révèle une capacité du cerveau à lier entre eux des messages qui n'activent pas la même zone cérébrale. Mais, apprendre n'est pas seulement étendre ses réactions à d'autres stimulations de l'environnement, pour Skinner c'est aussi sélectionner parmi toutes les réponses possibles celle qui procure un bénéfice : on parle alors de conditionnement opérant. Les apprentissages complexes mettent en évidence des processus cognitifs comme ceux impliqués par exemple dans des tâches d'orientation topographique (qui supposent de se forger une représentation du monde et de pouvoir l'exploiter) ou dans des expériences de conditionnement de type discriminatif (discrimination visuelle par exemple), ou de type probabiliste (sensibilité à la fréquence des stimuli).

L'élaboration d'un modèle d'apprentissage suppose de répondre à certaines questions :

- Qui est en train d'apprendre ?
- Qu'est ce qui est appris : un nouveau concept, un savoir, un savoir-faire, un savoir être ?
- Qu'elles sont les sources de l'apprentissage :
- si ce sont des exemples, ceux-ci sont-ils choisis au hasard, sont-ils positifs ou négatifs, simples ou difficiles, comment sont-ils décrits ?
- l'apprenant participe-t-il activement à son apprentissage : peut-il interagir avec son 'professeur' ou d'autres apprenants ? peut-il procéder à des expérimentations ?
- la source d'information est-elle bruitée ?
- Dans quel environnement se trouve l'apprenant ?
- De quelles connaissances ou compétences l'apprenant dispose-t-il au départ ?
- Quels sont les méthodes ou outils mis en œuvre pour l'apprentissage ? Sont-ils génériques ou spécifiques au domaine, à l'apprenant ?
- Quelles sont les critères et méthodes d'évaluation des résultats de l'apprentissage ?

Nous présentons ci-dessous un modèle général d'apprentissage par essais erreurs approprié au cas où l'environnement serait inconnu et changeant.

Apprentissage par renforcement (d'après [Sutton et al., 1998])

Ce type d'apprentissage convient lorsqu'il n'existe pas de modèle ou de professeur capable d'indiquer à l'apprenant ce qu'il doit faire. Cependant l'agent peut apprendre en interagissant avec l'environnement.



Nous dressons le tableau des caractéristiques de ce modèle d'apprentissage (tableau 1) :

Qui	Système artificiel programmable, doté d'une mémoire et d'une unité de calcul
Quoi	Séquences finies ou infinies d'actions continues ou discrètes,
Source de l'apprentissage	Signal de renforcement (pénalité ou récompenses) prenant des valeurs finies, éventuellement binaires ; éventuellement bruité.
Relations à l'environnement	<p>Le système est situé (dans l'espace), la connaissance de la situation peut être totale ou partielle.</p> <p>Le système est un processus de Markov : la probabilité d'une évolution à partir du temps t ne dépend que de l'état du système au temps t et non de son histoire.</p>
Connaissance initiale	En général nulle, modèle de l'environnement nécessaire dans certains cas
Méthode et outils	Algorithme d'optimisation maximisant la somme des gains à venir
Critères de performance	Convergence de l'algorithme, temps de calcul, ressources mémoires exigées

Tableau 1 : Caractéristiques de l'apprentissage par renforcement

Percevant sa situation 's' dans l'environnement, l'agent sélectionne et exécute une action 'a' et reçoit en retour de l'environnement un signal de renforcement 'r' de type pénalité récompense. Cette interaction entre l'apprenant et l'environnement qui caractérise l'apprentissage par renforcement est représenté par la Figure 8.

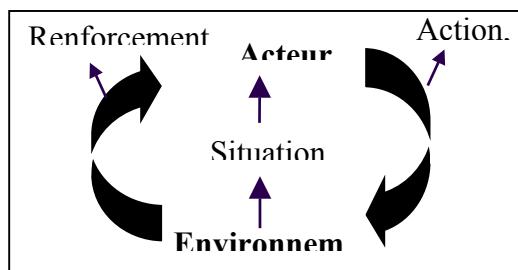


Figure 8 : Schéma de principe de l'apprentissage par renforcement

Le système de décision est supposé être un processus de décision de Markov, pour lequel la probabilité d'une transition d'état à partir de l'instant t ne dépend que de l'état du système à l'instant t et non de son histoire. L'agent exploite le signal de renforcement pour améliorer sa stratégie, c'est-à-dire pour trouver une séquence d'actions qui maximise le cumul des récompenses futures. L'apprentissage devient alors un problème d'optimisation.

On appelle 'valeur' associée à un état 's' (respectivement à un couple état action (s,a) , l'espérance mathématique de la somme des gains futurs depuis cet état 's' (respectivement depuis l'état 's' en exécutant l'action 'a')). La stratégie optimale correspond à la trajectoire passant par les états de valeur maximale.

Afin de maximiser ses gains, l'acteur doit trouver un équilibre entre exploration et exploitation. L'exploration consiste à tester de nouvelles actions pouvant conduire à des gains



supérieurs mais aussi inférieurs. L'exploitation consiste à appliquer la meilleure stratégie connue au cours de l'apprentissage.

Une difficulté de ce problème d'optimisation est qu'il n'est pas a priori possible d'anticiper les conséquences d'une action. Par conséquent la modification d'une stratégie est basée sur une évaluation même imprécise de la stratégie courante. L'évaluation et l'évolution d'une stratégie sont deux processus qui peuvent être imbriqués et concomitants.

Les algorithmes d'optimisation utilisés reposent sur un principe d'énumération des solutions de l'espace de recherche. On distingue trois grandes méthodes de résolution de ce problème d'optimisation : la propagation dynamique (PD), les différences temporelles (TD) et les méthodes de type Monte Carlo (MC) (Figure 9).

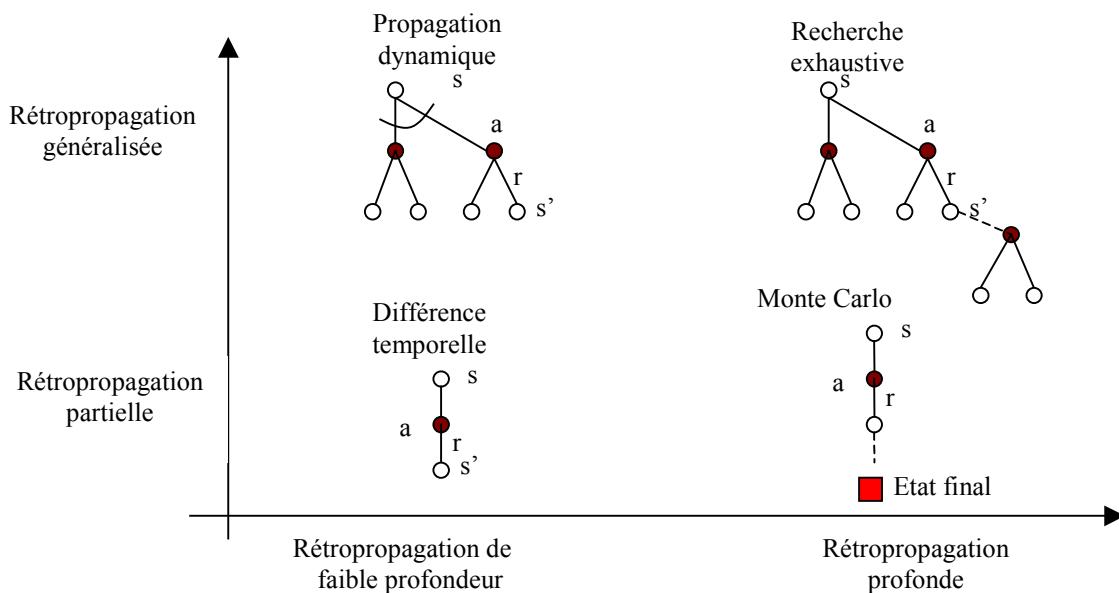


Figure 9 : Principe d'énumération dans les algorithmes d'apprentissage par renforcement

Les méthodes PD sont très performantes mais supposent la connaissance d'un modèle de l'environnement. Ce modèle peut être donné a priori ou construit au cours de l'apprentissage. Les méthodes MC reposent sur une estimation par calcul de moyenne de la valeur associée à un état ou à un couple état-action. Elles supposent donc la répétition d'un nombre élevé d'expériences similaires et ne sont pas adaptées au cas où ces expériences ne peuvent pas être rejouées. Les méthodes MC et TD procèdent par contre sur un horizon de temps limité. La méthode PD examine les conséquences immédiates de toutes les actions possibles à partir d'un état donné tandis que la méthode TD se limite à exploiter la sanction immédiate reçue lors de l'exécution de l'action selon la stratégie courante. Les méthodes MC et TD calculent itérativement les nouvelles estimations à partir des estimations anciennes et sont ainsi plus sensibles aux erreurs commises que les méthodes MC.

Les performances des méthodes d'apprentissage dépendent largement du choix du critère, c'est-à-dire de la fonction de renforcement ainsi que de l'équilibre entre les phases d'exploration et d'exploitation. Dans les cas extrêmes, s'il ne reçoit que des récompenses ou que des pénalités, l'apprenant ne peut pas discriminer les actions. De même, un sujet qui passerait son temps à explorer de nouveaux états ou au contraire qui se contenterait du premier succès obtenu ne pourrait pas apprendre. Il faut donc que le signal de renforcement change suffisamment fréquemment de nature. Pour compenser un nombre élevé d'échecs (surtout en début d'apprentissage), l'influence d'un succès peut être plus grande que celle d'un



échec. Différentes heuristiques ont été proposées pour faciliter la rapidité de convergence des algorithmes.

Notons aussi le risque d'explosion combinatoire dû au principe d'énumération lorsque la dimension de l'espace de recherche augmente.

Apprentissage et systèmes distribués

Nous avons soutenu l'idée qu'une intelligence collective pouvait émerger d'une population d'acteurs en interaction. L'activité générée tend à satisfaire les contraintes de viabilité de l'ensemble alors que chacun des individus agit localement sans perception globale du tout.

Il existe des situations où la recherche d'une satisfaction individuelle immédiate va à l'encontre de l'intérêt de tous.

- Un exemple célèbre connu sous le nom de 'Tragedy of the Commons (TOC)' a été étudié par Hardin en 1968 et montre comment une exploitation mal contrôlée de ressources a priori suffisamment abondantes peut provoquer rapidement leur épuisement. Cet exemple est généralement illustré par le comportement suivant d'un groupe d'éleveurs. Une prairie est partagée par plusieurs éleveurs et peut admettre sans dégradation un nombre limité de têtes de bétail. Un éleveur, pour améliorer son niveau de vie, est tenté d'augmenter la taille de son troupeau. Le gain qu'il peut tirer par la vente de bêtes supplémentaires est important au regard de la perte qu'il peut subir du fait que toutes les bêtes de la prairie se partageant la même quantité de fourrage deviennent moins productives. Le même raisonnement tenu par tous les éleveurs conduit à une surexploitation de la prairie qui se dégrade alors rapidement ne permettant plus à aucun éleveur d'en vivre. Les efforts développés actuellement pour promouvoir le développement durable tendent en partie à éviter ce type de tragédie.

- Une variante du TOC est connue sous le nom de paradoxe de Braess qui établit que l'ajout d'une liaison supplémentaire dans un réseau de communication peut augmenter le temps de trajet à l'équilibre. L'équilibre est atteint si le trafic est réparti de telle sorte qu'aucun utilisateur ne peut diminuer son temps de trajet en choisissant une route différente. Dans cet équilibre, les utilisateurs ont minimisé chacun leur temps de parcours.

- Les limites des stratégies individuelles ont été mises en évidence en théorie des jeux notamment dans le 'Dilemme du prisonnier'. Deux prisonniers ont le choix soit d'avouer leur culpabilité soit de dénoncer leur compère. Les peines encourues sont telles que, quel que soit le choix de son complice, un prisonnier a plus intérêt à trahir, alors que l'issue serait plus favorable pour les deux si tous deux optaient pour le silence. En trahissant simultanément ils obtiennent la peine maximale. Le choix de maintenir le silence, le plus irrationnel du point de vue d'un prisonnier seul, correspond à un équilibre de Nash pour lequel chacun des protagonistes serait perdant s'il changeait unilatéralement sa décision.

Plus généralement, les décisions d'un groupe d'individus est un optimum de Pareto si tout changement de décision d'un seul individu se traduit par une perte pour l'un des individus du groupe.

L'acquisition d'un comportement collectif intelligent suppose alors de pouvoir définir des objectifs individuels compatibles avec l'intérêt général. K. Tumer et D. Wolpert introduisent la notion de 'système factorisé' pour décrire la propriété d'un système dans lequel tout changement d'état d'un agent à l'instant t qui provoque une augmentation de ses gains ne peut diminuer les gains du système global [Tumer et al., 2000]. Les agents se situent alors toujours dans un équilibre de Nash. En mesurant l'intelligence comme la fraction des états possibles d'un agent à l'instant t qui ne conduiraient pas à une amélioration de ses propres gains,



l'équilibre de Nash atteint par chacun des agents correspond à la décision la plus 'intelligente'. Pour construire un système factorisé, les auteurs proposent de définir la fonction objectif de chaque individu à partir de la fonction globale selon le principe suivant : la fonction objectif d'un individu est la différence entre la fonction objectif globale et une fonction objectif virtuelle évaluée en calculant la fonction globale sur le système restreint dans lequel l'état de tous les agents influencés par l'individu est fixe. Cette stratégie, améliore aussi la facilité d'apprentissage (learnability) car un agent distingue plus aisément l'effet de ses propres actions des effets des actions des autres agents. Par cette approche les auteurs améliorent les performances de routage dans un système de télécommunications en évitant notamment le paradoxe de Braess [Wolpert *et al.*, 2000]. Cette approche est restreinte au cas où il est possible d'évaluer la fonction objectif globale du système dans des conditions d'activation différentes des agents.

3.1.3. Modélisation des phénomènes de comportement collectif intelligent

Modélisation mathématique

L'idée principale dans ce paragraphe est que l'activité d'un collectif se structure par le biais d'interactions entre individus et entre individus et l'environnement. Un modèle général d'un collectif intègre donc trois facteurs :

1) Un ensemble de N individus décrits sous une des trois formes suivantes :

- Chacun des individus i est défini par un état $X_i(t)$ évoluant au cours du temps t . L'état peut décrire à la fois un positionnement dans l'espace mais aussi des caractéristiques internes comme le niveau de satisfaction d'un besoin, la tendance à effectuer une action, la disponibilité ...).
- Les individus sont divisés en K types, chacun des types j présentent une concentration $n_j(x,t)$ où x est la position dans l'espace.
- On définit pour chacun des individus i une probabilité $p(X_i(t) = S)$ de se trouver dans un certain état s à l'instant t .

2) Un environnement dont certaines caractéristiques à la position x sont décrits par une grandeur $E(x,t)$ à l'instant t .

3) Un ensemble d'interactions qui se traduisent par des relations souvent non linéaires entre variables. Ces relations peuvent être de nature logique ou numérique. Les modèles reposent généralement sur :

- Des équations différentielles ordinaires d'interprétation souvent immédiate et qui se présentent sous une forme générale $dX/dt=f(X,t)$. La fonction f est une fonction non linéaire pouvant contenir par exemple des termes de produits de variables. La résolution numérique de ces équations est bien maîtrisée bien que certains systèmes d'équations peuvent présenter une grande sensibilité aux conditions initiales (cas des systèmes chaotiques).
- Des équations aux dérivées partielles, qui permettent de préserver le caractère continu du système. La résolution analytique est souvent difficile et, selon la complexité du modèle, la résolution numérique, peut s'avérer très coûteuse en ressource mémoire et en temps de calcul.



- Des équations maîtresses. Ces équations, sous forme discrète ou continue, décrivent l'évolution temporelle de la probabilité que le système soit dans un état donné à un instant t.
- Des équations différentielles stochastiques, qui sont introduites pour prendre en compte la notion de bruit dans une équation différentielle ordinaire. Le bruit décrit les phénomènes parasites superposés au comportement moyen du système.
- des règles logiques pour simuler des dynamiques de type automates tels que les automates cellulaires qui sont basés sur la discréétisation de l'espace et du temps. Chacune des cellules de l'espace est un automate à états finis qui évolue en fonction de l'état de ses voisins. Un des automates cellulaires les plus célèbres est celui du jeu de la vie et est dû à John Conway en 1970. Les automates cellulaires sont divisés en quatre classes [Langlois *et al.*, 1997]. La classe I correspond à des automates dont la dynamique converge vers un point fixe, la classe II à des automates dont la dynamique converge vers au moins un cycle attracteur de petite taille, la classe III à des automates à attracteur chaotique, la classe IV, plus difficiles à caractériser, à des automates capables de réaliser un niveau élevé de traitement de l'information (par exemple le jeu de la vie est un automate de classe IV équivalent à une machine de Turing universelle).

Les moyens informatiques permettent de concevoir et de simuler des modèles complexes comme les architectures de systèmes multi-agents que nous présentons ci-dessous plus en détail.

Systèmes multi-agents

Dans les systèmes de traitement de l'information centralisés, de gros programmes prennent en charge l'ensemble des aspects de la résolution de problèmes. De conception et de maintenance difficile, ces programmes ont des performances limitées du fait de leur isolement et du manque d'échange d'information avec l'extérieur. C'est pourquoi une approche décentralisée basée sur la distribution des tâches entre plusieurs entités ou agents logiciels autonomes et communicants connaît un grand développement. Ces agents logiciels, sont bien adaptés à la résolution de problèmes fonctionnellement distribués et hétérogènes qui se posent par exemple dans les activités de production ou de logistique.

Définition d'un agent

Selon Ferber [Ferber, 1995], un agent est une entité physique ou virtuelle qui :

- est capable d'agir dans un environnement
- peut communiquer avec d'autres agents
- possède des objectifs individuels
- est doué d'autonomie, c'est à dire qu'ils ne sont pas dirigés de l'extérieur
- possède des ressources propres
- est capable de percevoir même de façon limitée son environnement
- possède des compétences et offre des services
- peut éventuellement se reproduire



Le comportement de l'agent tend à satisfaire ses objectifs à partir de sa perception, ses représentations, des communications reçues compte tenu de ses ressources et de ces compétences.

Deux types d'agents peuvent être identifiés. Un agent est cognitif s'il dispose d'une base de connaissance comprenant les informations et savoir-faire utiles à la réalisation de ses tâches. Il possède une représentation symbolique même partielle du monde à partir de laquelle il peut raisonner. A l'inverse, un agent réactif dispose de facultés de traitement de l'information très limitées, ses actions sont des réponses immédiates à sa perception de son environnement. Entre ces deux catégories extrêmes d'agents peuvent être définis des agents intermédiaires aux capacités cognitives plus ou moins développées.

Systèmes multi-agents

Un système multi-agents est composé d'un ensemble d'agents opérant dans un environnement. La notion d'interaction est au centre de ces systèmes.

Plusieurs types d'interactions peuvent être définies selon que les buts des agents sont compatibles ou non, selon que les ressources sont partageables ou non et selon le niveau de compétences des agents pour la tâche à réaliser. A ces types d'interaction correspondent des situations d'indépendance, de collaboration, de coopération, d'encombrement (de gêne entre agents), de compétition ou de conflits.

Les agents entre en situation de coopération si l'ajout d'un nouvel agent permet d'accroître les performances du groupe ou bien si l'action des agents sert à résoudre des conflits potentiels ou actuels. Les méthodes de coopération sont :

- le regroupement ou la multiplication
- la communication permettant de bénéficier des informations et du savoir-faire des autres
- la spécialisation qui renforce l'aptitude de certains agents dans certaines tâches
- la collaboration qui consiste au partage des tâches et ressources entre agents
- la coordination d'actions qui regroupe l'ensemble des activités supplémentaires qu'il est nécessaire d'accomplir de façon à augmenter les performances du groupe ou à diminuer les conflits.
- la synchronisation qui définit l'enchaînement des actions pour qu'elles soient accomplies juste au bon moment
- la résolution de conflit par arbitrage et négociation

Les communications entre agents sont à la base des interactions. Elles peuvent être établies en mode point à point entre deux agents ou en mode diffusion pour toucher plusieurs destinataires. L'acheminement des messages, qu'il soit intentionnel ou non, peut être direct ou indirect (par exemple par l'intermédiaire de traces laissées dans l'environnement). Les communications font l'objet de différents types d'actes de langage : les assertifs donnent une information sur le monde, les directifs incitent le destinataire à une action, les promissifs engagent le locuteur, les déclaratifs entérinent un fait ou acte. Une séquence d'interactions qui s'établit lors d'une communication constitue une conversation dont la structure obéit à des règles prédéfinies.



Composantes de l'organisation multi-agents

Des interactions entre agents naît une organisation qui constraint elle-même en retour les interactions entre agents. Une organisation est définie ici comme un agencement de relations entre composants ou individus qui produit une unité dotée de qualités inconnues au niveau des individus. Une organisation peut être étudiée selon trois niveaux d'analyse. L'analyse fonctionnelle s'attache aux fonctions réalisées par l'organisation. L'analyse structurale différencie les formes d'organisation. L'analyse opérationnelle, s'intéresse à la manière dont le travail est réparti en définissant en particulier les degrés de redondances et de spécialisation des agents.

On peut identifier six grandes fonctions des organisations :

- la fonction représentationnelle regroupe l'ensemble des fonctions de modélisation et de représentation de l'environnement
- la fonction organisationnelle concerne la planification, l'allocation et le suivi des tâches
- la fonction conative décrit comment les buts sont formés, les contraintes établies et les décisions prises
- la fonction interactionnelle gère les communications entre agents et les interactions avec l'environnement
- la fonction productive définit les activités nécessaires à la résolution de tâches
- la fonction végétative tend à maintenir l'organisation

Chacune de ses fonctions peut être décomposée selon les diverses dimensions de l'organisation : physique en ce qui concerne son architecture, sociale pour ses capacités à résoudre des besoins généraux de l'organisation, relationnelle, environnementale pour ce qui concerne la perception et l'action dans l'environnement, et personnelle pour ce qui est de la représentation de soi.

Organisation	Analyse fonctionnelle	Fonctions	Représentationnelle, organisationnelle, conative, interactionnelle, productive, conservatrice
		Dimensions d'analyse	Physique, sociale, relationnelle, environnementale, personnelle
	Analyse structurale	Relations abstraites	D'accointance, communicationnelle, de subordination, opérative, informationnelle, conflictuelle, compétitive
		Structure de subordination	Hiérarchique, égailtaire
		Couplage	Fixe, variable, évolutif
	Analyse opérationnelle	Constitution	Prédéfinie, émergente
		Spécialisation	Hyperspecialisée, généraliste
		Redondance	Redondant, non-redondant

Tableau 2 : Eléments fonctionnels, structuraux ou opérationnels d'une organisation artificielle [Ferber, 1995]

L'analyse structurale décrit les relations organisationnelles qui régissent les rapports entre deux agents traduisant ainsi une relation de dépendance plus ou moins importante. Une



relation d'accointance indique une simple connaissance de l'existence de l'autre, une relation communicationnelle repose sur l'existence d'un canal de communication entre les deux agents, une relation de subordination soumet un agent à un autre, une relation opérative représente une dépendance entre tâches exécutées par les agents, une relation informationnelle traduit une dépendance envers un savoir supposé d'un autre agent, une relation conflictuelle apparaît lorsqu'une ressource ne peut être partagée simultanément, une relation compétitive caractérise des buts incompatibles.

Le tableau 2 rassemble selon Ferber les différents éléments fonctionnels, structuraux ou opérationnels d'une organisation artificielle d'un système multiagent.

Conclusions

Empruntant un grand nombre de concepts à la psychologie cognitive, à la sociologie, à la dynamique des groupes, à la biologie, à l'éthologie, les systèmes multi-agents s'avèrent des outils puissants de modélisation et de simulation de systèmes distribués. Les systèmes multi-agents offrent des solutions à la problématique de prise de décision dans les organisations, chaque acteur qui intervient dans des processus opérationnels étant modélisé sous la forme d'un agent. Dans ce contexte chaque agent à un point de vue différent et participe à la définition des objectifs et aux performances globales. Par exemple, les systèmes multi-agents fournissent un contexte théorique à la conception d'architectures logicielles pour la gestion de systèmes distribués tels que les systèmes de production. En effet, dans un marché fortement concurrentiel et versatile, imposant des contraintes de temps, de coût et de qualité de service toujours plus sévères, les entreprises doivent faire preuve de réactivité et de flexibilité. Sous la pression du marché, les entreprises s'organisent en réseau à l'échelle de plusieurs pays.

Une approche centralisée de la gestion de tels systèmes de production se heurtent aux problèmes : de l'acquisition d'informations réparties sur plusieurs sites, de l'identification des informations pertinentes (utiles à la résolution du problème et non obsolètes), du traitement de l'information en vue d'une prise de décision, de l'exécution de la décision et du suivi.

Une approche distribuée, basée sur des agents autonomes traitant localement l'information et adoptant une organisation adaptée au but poursuivi présente une plus grande réactivité car les décisions sont prises au bon niveau sans dépendre systématiquement d'un agent centralisateur et une plus grande flexibilité car ses architectures acceptent facilement une modification du nombre d'agents.

3.1.4. l'approche des problèmes de transport par des simulations multi-agents ?

Un système de transport met en relation des acteurs sociaux jouant des rôles :

- d'usagers. Ils empruntent les moyens de transport dans le but de déplacer des biens ou des personnes selon des critères de sécurité, d'efficacité, de confort.
- de clients. Ils payent pour bénéficier du service de transport. Ils peuvent être eux mêmes usagers ou utilisateurs indirects tels que les entreprises des différents secteurs industriels, des collectivités locales. Ces clients ont des exigences de fiabilité, de disponibilité, de flexibilité de moindre coûts.
- d'exploitants. Soumis à des contraintes de rentabilité, ils ont en charge l'entretien et la gestion des infrastructures.



- de producteurs. Ils ont en charge la conception et la réalisation des infrastructures, des véhicules, des systèmes de gestion et d'exploitation.
- de riverains. Ce sont les personnes ou collectivités qui bénéficient ou subissent les effets directs ou indirects du système de transport (valorisation du territoire, emploi, nuisances sonores, pollution)
- de consultants. Ce sont les bureaux d'études qui proposent les solutions techniques aux problèmes d'extension, de gestion, de démantèlement des infrastructures.
- de décideurs. Ce sont les organismes qui décident des politiques de transport, d'aménagement du territoire, de grands travaux.

Certains acteurs tels que les collectivités locales jouent plusieurs rôles.

L'environnement dans lequel évoluent ces acteurs est constitué de zones naturelles et de zones d'activité humaine à l'échelle locale, nationale ou internationale. Plusieurs systèmes de transports possédant leur spécificité propre peuvent offrir des services complémentaires ou peuvent entrer en concurrence.

Nous avons insisté tout au long de cette section sur l'importance des interactions entre constituants et entre constituants et environnement dans les systèmes complexes. L'analyse, la modélisation et la maîtrise de ces interactions locales peut permettre d'agir efficacement sur les performances du système global.

Une approche basée sur des concepts d'intelligence collective réactive peut trouver des applications par exemple pour l'étude et la régulation des flux dans les systèmes de transport routier ou pour la résolution de problème de gestion des services de transport.

En effet, les usagers ont tendance à maximiser leur propre fonction objectif sans prendre forcément en compte les actions des autres usagers (par ignorance, comportement individualiste, ou incapacité à en tenir compte) et ceci au risque de paralyser le système tout entier (cf 'Tragedy of the Commons' ou TOC au § *Apprentissage et systèmes distribués*). Il est remarquable que les moyens de communications actuels entre usagers soient aussi peu développés. La perception des autres usagers est visuelle ou auditive. Les seuls signaux généralement échangés sont des signaux tout ou rien lumineux ou sonores. Le manque de visibilité rend les signaux lumineux inefficients au détriment de la sécurité des usagers. En développant des systèmes d'échange de message adaptés, peut-être serait-il possible de pallier certaines lacunes des systèmes existants et d'améliorer l'utilisation des infrastructures routières en zone urbaine ou rurale au moment des périodes de pointe de trafic.

Une autre application des concepts d'intelligence collective concerne l'extension des modèles d'organisation virtuelle aux systèmes de transports. Le but des organisations virtuelles est d'utiliser d'une manière optimale les différentes ressources disponibles. Ainsi, les entreprises manufacturières, pour faire face efficacement à la pression du marché, sont amenées à se regrouper dans des entreprises virtuelles qui sont des organisations dynamiques pouvant être reconfigurées selon les besoins du marché. Le bon fonctionnement de ces organisations repose sur des systèmes de communication et d'information permettant de partager et d'échanger des données techniques et commerciales afin d'offrir la réponse la plus satisfaisante possible à la demande des clients.

Néanmoins cette approche basée sur les concepts de systèmes multi-agents réactifs trouve une limite lorsque la granularité des agents augmente. Ainsi la modélisation des comportements d'acteurs impliqués dans des tâches nécessitant des capacités cognitives importantes, suppose



une analyse plus fine des fonctionnalités susceptibles de caractériser les capacités représentationnelles et décisionnelles de ces acteurs. C'est l'objet de la section suivante consacrée à l'intelligence humaine collective.

3.2. Intelligence humaine collective

La gestion des connaissances est considérée aujourd'hui comme un enjeu majeur de toute organisation, enjeu économique, et plus encore, condition de sa viabilité dans un environnement évolutif complexe, en somme, condition de sa survie. Les nouvelles technologies de l'information et de la communication y sont naturellement associées et accréditent l'idée d'un déploiement rapide *sur le terrain*. Cependant, ce domaine que l'on rattacherait par sa dénomination aux sciences de la gestion, relève bien plus des sciences cognitives et des sciences sociales. Du coup, les problématiques qui sont abordées ne sauraient être réduites à de simples considérations de gestion de ressources – stock, flux d'échanges, péremption - et surtout si l'on s'intéresse à l'impact attendu de ce processus sur l'organisation : passer de l'expertise individuelle à l'intelligence collective. La conception même des outils censés susciter ce changement majeur doit être fondée sur des concepts non triviaux et exige de véritables expérimentations en grandeur réelle.

3.2.1. Mémoire collective

Les technologies de l'information facilitent aujourd'hui la communication ou l'échange de données, indépendamment des localisations géographiques des serveurs et des acteurs. La vague Internet, qui en constitue la partie la plus visible avec son arsenal de technologies conviviales, a remis au premier plan des objectifs ambitieux pour les organisations comme la promotion de l'innovation, l'appropriation par la collectivité des expériences et des savoirs individuels, ou encore l'argumentation et la décision collectives.

Cependant, ces performances technologiques ne suffisent pas à l'émergence d'un réel mode de travail collectif, où l'ensemble des processus cognitifs - apprentissage, délibération, argumentation, décision - attribués à un groupe d'acteurs atteindrait et dépasserait un niveau de performance dans l'action non atteint isolément par chacun des individus.

Curieusement, il semble que l'objectif d'une meilleure performance collective nécessite de mieux reconnaître l'existence et l'intérêt des actions et interactions individuelles, alors que domine généralement une « norme organisationnelle » composée de principes stratégiques, procédures et consignes pré-déterminés. A l'opposé, l'innovation que l'on tend parfois à considérer comme issue du génie individuel repose sur la performance collective tout autant que sur des construits individuels.

Pour atteindre le niveau de performance collective pouvant être qualifié d'intelligence collective, il semble raisonnable d'instrumenter ce « sujet-collectif » qu'est une organisation constituée en vue de l'action (une équipe de projet par exemple) afin de le doter de capacités analogues à celles – naturelles – d'un individu cognitif. On peut en effet parler d'une entité collective donnée - une entreprise, par exemple - comme étant susceptible d'apprendre, d'avoir une mémoire, de faire des expériences, des choix, de prendre des décisions, d'agir, etc. - c'est-à-dire, comme étant en quelque sorte dotée d'une cognition individuelle [Andreevski, 1998].

En tout premier lieu, il s'agit de doter le collectif d'une « mémoire commune », et ce, quels que soient les métiers concernés, les spécialités, les fonctions et la nature des connaissances



nécessaires à l'exécution des tâches. C'est sur la base de cette mémoire partagée que pourront s'édifier des processus collectifs d'apprentissage, de conception, de décision. Comme le remarque très généralement Pierre Lévy [Lévy, 1991], « le savoir de la communauté pensante n'est plus un savoir commun, car il est désormais impossible qu'un seul humain, ou même un groupe, maîtrise toutes les connaissances, toutes les compétences, c'est un savoir collectif par excellence ».

Mais la nature de cette mémoire collective mérite cependant d'être discutée.

Comme le rappelle Joëlle Proust [Proust, 1997], la métaphore la plus courante de la mémoire est celle de l'entrepôt : y sont rangées dans un ordre particulier les données relatives aux expériences vécues et aux connaissances acquises. Platon nous a transmis deux autres métaphores qui continuent à exercer une influence profonde

La première est l'image de la cire : les souvenirs sont imprimés dans l'esprit par l'impact de la réalité, sensible ou abstraite, à laquelle le sujet est exposé. La notion d'empreinte suggère une conservation de la forme dans le temps.

La seconde, plus dynamique, voit dans la formation du souvenir et sa récupération deux manières de « chasser ». L'esprit ressemble à une volière, où sont rassemblées les connaissances après leur prise initiale ; elles devront ensuite être recapturées une par une pour être utilisées dans le jugement.

Mémoire-entrepôt, mémoire-cire et mémoire-volière, ces trois métaphores peuvent être mises en correspondance avec trois processus de la cognition :

- Structuration de l'information relative aux caractères identifiables des objets considérés,
- Formalisation et représentation, opérations dépendantes des structures internes du sujet connaissant,
- Reconstruction dynamique.

Mais il faut distinguer ici la cognition individuelle, de ce que l'on peut appeler la cognition collective.

Dans le premier cas –l'individu– il est établi que l'activité mémorielle n'est pas totalement consciente, c'est-à-dire qu'elle ne se réduit pas à des processus de remémoration délibérée et intentionnelle. Certains mécanismes *automatiques* de l'esprit humain sont à l'œuvre, accordant par exemple une préférence aux processus de traitement de l'information proches de ceux qui ont permis l'encodage de l'information (on se souvient mieux d'un mot si on a eu à l'utiliser auparavant). L'effet de cette mémoire automatique, fondée sur un processus *d'apprentissage incident*, est cependant pondéré par d'autres mécanismes : la construction de l'image mémorielle est tributaire des conditions de sa production (les faits inhérents au contexte d'action sont associés aux faits initialement mémorisés).

Dans le cas d'un « sujet collectif », la situation est tout autre ; il n'existe pratiquement plus de mécanismes automatiques, sauf à postuler un inconscient collectif, et la phase de reconstruction dynamique est alors prédominante. Et ceci d'autant plus que l'encodage de l'information n'est plus réalisé par le même individu. Il s'agit maintenant pour un acteur de ce collectif d'utiliser « les souvenirs des autres » et de participer à un *apprentissage collectif*.

Les remarques précédentes conduisent ainsi à la question fondamentale, qui est d'ailleurs au cœur des approches cognitives : pour agir efficacement, comment construisons nous des



connaissances pertinentes ? Ou, de façon plus appliquée : comment instrumenter la construction collectives de connaissances utiles à l'action commune ?

Là encore, il est intéressant de considérer les paradigmes actuels relatifs aux processus de représentation des connaissances. La thèse cognitiviste (ou computationnaliste) met l'accent sur les systèmes de représentation symbolique, c'est-à-dire les modèles nécessaires au traitement de l'information. La thèse de l'émergence montre que cette image mentale n'est pas simplement "reçue" mais interprétée (c'est cela qui justifie qu'un même stimulus peut être perçu différemment par un même sujet). Enfin, la thèse de l'enaction montre que l'apprehension du monde est intimement liée à l'action envisagée, c'est-à-dire dans le cas d'une action distribuée, à la construction du référentiel collectif.

En résumé, la cognition, et encore moins la cognition collective, ne peut donc pas être réduite à un stockage symbolique des informations sous forme brute (mémoire entrepôt) ou sous forme de représentation (mémoire-cire). Elle se fonde toujours sur une boucle d'actions et c'est sur cette boucle d'actions qu'il faut agir pour l'optimiser. Nous nous intéresserons donc aux connaissances relatives à l'action, que l'on peut désigner du néologisme *connaissances actionnables*.

3.2.2. Connaissances actionnables

Le néologisme "connaissance actionnable", comme le rappelle J.L. Le Moigne [Le Moigne, 1998], a été introduit dans la littérature organisationnelle par D. Schön en 1983 ("actionable knowledge"), afin de dépasser le distinguo habituel entre savoir et savoir-faire, c'est-à-dire la séparation entre la composante épistémique (la connaissance) et la composante pragmatique (l'action). C. Argyris, qui dès 1978 a proposé avec D. Schön des modèles organisationnels pour l'action [Argyris *et al.*, 1978], parle en 1995 de savoir pour agir ("knowledge for action").

Il ressort de ces travaux que c'est par la réflexion sur ses actions, sur ses savoirs d'expérience, que le praticien (le sujet connaissant engagé dans l'action) pourra mieux prendre conscience des théories d'actions qu'il a élaborées, et donc pourra les corriger. La traduction de ses savoirs tacites en savoirs d'action constitue le cœur du processus d'apprentissage (Figure 10).

En développant le concept dual d'Action Intelligente, H.A. Simon avait attiré l'attention sur la légitimité d'une rationalité procédurale, qui ne se réduit pas à la logique et au calcul fondé sur des faits ayant valeur de vérité, et qui tient pour tout aussi satisfaisants les modes cognitifs que sont la délibération et l'argumentation. Il s'agit alors de concevoir et construire des connaissances qui soient représentations d'expériences, d'actions, de réflexions, de délibérations.

Relier le savoir et le faire, tel est l'enjeu de la représentation de la « connaissance actionnable » dont il faut souligner qu'elle porte sur des "Faire" plutôt que sur des "Faits" et qui subit en conséquence une permanente transformation.

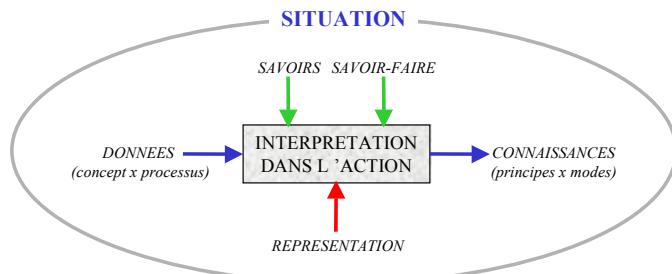


Figure 10 : Des données aux connaissances

Dès lors, la gestion des connaissances actionnables n'est pas une encapsulation du savoir, mais un suivi dynamique d'un corpus de connaissances en expansion. La question n'est plus alors de savoir transmettre des connaissances présumées efficaces mais de susciter les conditions cognitives et organisationnelles par lesquelles s'exerce l'action intelligente de sujets connaissants : délibération, choix, décision.

3.2.3. Expertise individuelle

Le rapport privilégié qu'un sujet connaissant établit avec un phénomène perçu dans le cadre d'un projet d'action peut être observé comme une "situation cognitive". Le sujet doit délibérer sur son action et produire une connaissance opérationnelle, c'est-à-dire utile à ses décisions [Penalva, 1997].

Une situation est donc prise ici comme un état du monde, au sein duquel un individu cognitif animé de l'intention de rendre cette situation intelligible, se prépare à d'éventuelles évolutions et acquiert un certain degré de maîtrise dans la transformation délibérée de cette situation [Liu, 1993].

Dans la situation observée, convenons d'appeler "acteur" toute intelligence impliquée directement dans une action intentionnelle portant sur un objet qu'elle différencie de son environnement. Dans le cas d'un acteur individuel, on ne peut distinguer l'action de l'acteur sur l'objet et l'intervention projetée par l'acteur lui-même sur cette relation. Autrement dit, le caractère "projectif" de l'intervention (j'envisage un futur) ne peut être différencié du projet d'action qui se déroule "en temps et en lieu" de l'acteur.

L'opérateur humain (le sujet connaissant) doit "s'adapter" à la situation, c'est-à-dire accorder ses structures cognitives aux circonstances du réel perçu afin de prendre des dispositions pertinentes pour obtenir le but visé, ou éviter des événements redoutés. Il combine un savoir contemplatif sur l'objet (il le perçoit au travers de caractères descriptifs qui lui donnent une forme particulière), un savoir faire (en tout premier lieu fabriquer des représentations de l'objet adéquates au projet) et un savoir pratique (un savoir-se-comporter dans l'action en fonction de connaissances acquises).

La qualification d'une situation cognitive consiste à en déterminer les classes descriptives. Nous en avons proposé neuf [Penalva, 1997] rappelées dans le schéma ci-après (Figure 11).

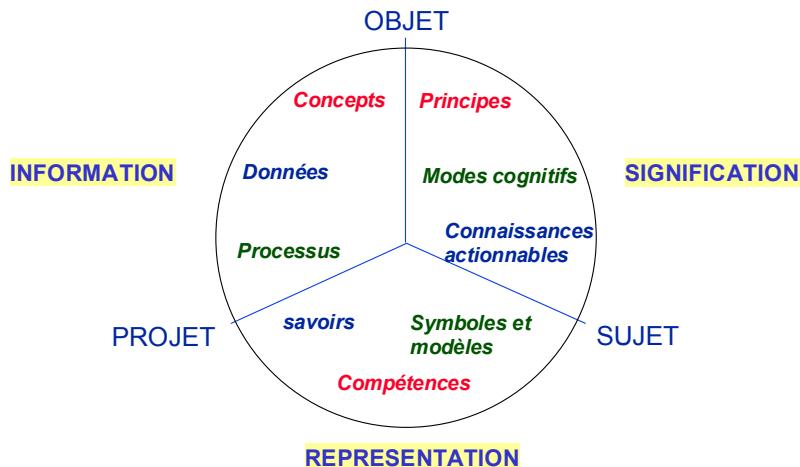


Figure 11 : La forme générale d'une situation

Les trois dimensions de la situation, Objet-Projet, Projet-sujet, et Sujet-Objet, correspondent à trois processus :

- l'objet prend forme dans le projet. Ce processus d'**INFORMATION** rassemble les données de base relatives à la conceptualisation de l'objet et aux processus concernant l'objet dans son environnement.
- le projet mobilise le sujet. L'objet est évoqué par le sujet (à la fois désigné et interprété par le sujet), processus que l'on qualifie de **REPRESENTATION**, c'est à dire de construction de systèmes de symboles (pouvant devenir des modèles) correspondant aux savoirs et savoir-faire du sujet sur l'objet.
- l'objet prend sens dans le cadre du projet. On donne sens à la relation objet-sujet par un processus de **SIGNIFICATION**²¹ qui explicite les connaissances élémentaires (actionnables) construites par le sujet, utilisés dans ses raisonnements soumis au respect de principes de référence (normatifs, paradigmatiques, stratégiques...).

Parmi les éléments descriptifs de chacune de ces dimensions, certains sont dotés de conditions d'existence et de validité :

- données : faits objectifs, informations primaires sur l'objet, que l'on peut structurer en systèmes d'informations plus ou moins complexes,
- savoirs : généralement formalisés - à l'extrême sous forme théorique - ces agrégats complexes d'informations peuvent être codifiés (relatifs à un domaine), et exprimés sous une forme indépendante des actions qui pourraient les utiliser,
- connaissances actionnables : expressions d'une expérience, d'une action, d'une délibération, sur lesquelles peuvent influer les croyances, les perspectives et les valeurs.

Cette vision ontologique de la situation peut être opposée à une vision phénoménologique correspondant à la façon dont l'objet se manifeste à la conscience du sujet :

- les processus d'évaluation de l'objet dans son environnement,

²¹ Le terme SIGNIFICATION pourra paraître bien impropre ou insuffisant à certains... En voulant désigner « l'action de faire sens pour l'action », d'autres néologismes tels qu'opérationnalisation, actionnalisation, contextualisation ne nous ont pas non plus satisfaits.



- les agencements symboliques (images mentales ou modèles) par lesquels l'objet est représenté pour l'action,
- les modes cognitifs (ou états) par lesquels le sujet va composer les éléments de connaissances utiles à ses décisions.

Il ressort des deux visions précédentes (ontologique et phénoménologique) que l'apprehension de la situation par le sujet peut être vue comme un processus d'assimilation-accommodation, qui vise à transformer dans le temps :

- Les concepts par lesquels l'objet est appréhendé,
- les compétences développées par le sujet,
- les principes directeurs de l'action.

Ceci fait bien sûr écho à Jean Piaget qui en 1970 énonçait : « la connaissance ne saurait être conçue comme prédéterminée, ni dans les structures internes du sujet, puisqu'elles résultent d'une construction effective et continue, ni dans les caractères préexistants de l'objet, puisqu'ils ne sont connus que grâce à la médiation nécessaire de ces structures... En d'autres termes, toute connaissance comporte un aspect d'élaboration nouvelle. »

Ainsi posée, la situation cognitive devient un cadre de référence permettant d'organiser l'explicitation des « connaissances » relatives à l'action et au sujet qui y est impliqué.

3.2.4. Intelligence collective

Nous prendrons comme cas général une situation cognitive dans laquelle l'acteur est pluriel et peut être organisé en deux niveaux :

- l'équipe, groupe où les membres construisent un référentiel opératif commun,
- l'organisation, le collectif finalisé, dont les membres partagent des objectifs communs mais construisent leurs propres représentations et opposent leurs stratégies décisionnelles.

Nous considérerons comme "passages à la limite" les cas où l'acteur est un individu (intervention de mécanismes non intentionnels) ou un collectif (dont les membres ne partagent qu'une culture commune et ne construisent que des principes d'action).

Dans le cas général, le groupe se distingue de l'organisation en ce que tous les membres sont en interaction dans l'action. Groupe et organisation supposent de chaque acteur des activités orientées vers les autres (négociation, reformulation, argumentation...) et des activités collectives (délibération, construction de références communes...) mais toutes fondées sur la construction de représentations partagées [Teulier-Bourgine, 1997].

Le cas d'un collectif qui désire intervenir sur la façon dont il agit dans une situation, et, en quelque sorte, observe donc sa propre action, revient bien à considérer le collectif comme un sujet-collectif individualisé [Andreevski, 1998] et intelligent : comme le signalait Jacques Pitrat, « tout système intelligent peut et doit observer son propre comportement dans l'action », cette auto-référence étant la condition de son autonomie. Il paraît donc légitime de considérer l'intelligence collective en s'inspirant du processus de cognition individuelle, mais en gardant à l'esprit une différence importante : si l'on peut à la limite créditer l'individu impliqué dans l'action d'une rationalité instrumentale (recherche de l'adéquation fin/moyens) avec un recours minimal à un espace mental), l'intervention en situation complexe d'un collectif organisé nécessite une rationalité cognitive fondée sur l'adéquation entre les informations possédées et les représentations adoptées.



Le rapport entre les différentes rationalités, l'usage des savoirs et la notion d'intelligence est un vieux débat dont on trouve les traces dans la mythologie grecque et à travers quelques allusions faites dans les écrits de Platon et Aristote : « Dans la navigation, indique Aristote, il n'y a pas de savoir général de tous les cas particuliers, pas de connaissance certaine de tous les souffles qui sillonnent les eaux de la mer. Pour le plus expérimenté des pilotes, Pontos demeure toujours l'Inconnu. L'excellence du navigateur ne se mesure pas à l'étendue de son savoir, elle se reconnaît à sa capacité de prévoir et de découvrir à l'avance les pièges de la mer qui sont aussi les occasions qu'elle offre à l'intelligence du pilote » (M. Detienne et J.-P. Vernant, p. 214).

Les ruses de l'intelligence

La mètis des Grecs à laquelle (J.M. Oury) fait référence est aussi appelée Ruses de l'intelligence pour en signifier ses caractéristiques retorses et conjecturales propres à l'artisanat, la navigation ou tout ce qui est de l'ordre du quotidien, du jamais certain. Mètis est à la fois la première femme de Zeus et une qualité que l'on retrouve chez d'autres divinités ou quelques rares hommes, légendaires, (en particulier Ulysse associé à Athéna, fille de Mètis et de Zeus).

Pour ce faire, vivacité d'esprit et recherche intellectuelle doivent être combinées avec « l'habileté manuelle et la connivence avec le réel » (p. 29). Dans ces contextes mouvants, on remarque l'importance de la mobilité, du déguisement, de la ruse, de l'art de se transformer, de la polymorphie, et surtout de l'art de la combinaison : “Quand deux hommes marchent ensemble, si ce n'est l'un, c'est l'autre, à sa place, qui voit l'avantage à saisir. Seul, on peut voir aussi, mais la vue est plus courte et la mètis plus légère” (p. 24).

Les auteurs qualifient Mètis ou sa fille Athéna, par une intelligence rusée, pratique, ou prudente, une puissance qui lie, encercle et jamais ne va droit, ne connaît de frontière. C'est-à-dire une puissance qui combine harmonieusement ruse et raison au nom de l'action pratique. Mais c'est une puissance qui fait peur : “Sa parole a valeur hypothétique ou problématique ; [...] elle dit le futur, non comme ce qui est déjà fixé, mais comme malheur ou heur possibles, en livrant son astucieux savoir pour le faire tourner au mieux plutôt qu'au pire” (p. 105). A l'opposé, une autre puissance divinatoire, Thémis, qui prit la succession de Mètis aux côtés de Zeus, est plus proche de la conception du monde que nous présente le plus souvent la science : “L'omniscience de Thémis a rapport en effet à un ordre conçu comme déjà instauré, définitivement fixé et établi” (p. 105), celui que l'on chercherait à maîtriser et que Thémis connaît : “Son rôle est de marquer les interdits, les frontières à ne pas franchir, les préséances à respecter pour que chacun soit à jamais maintenu dans les limites de son domaine et de son rang”.

L'efficacité occidentale vs orientale

Cette intelligence est souvent citée par les auteurs faisant la promotion de la vigilance et de la connaissance pratique mais n'est théorisée dans aucun modèle occidental (ce qui n'est pas le cas des japonais - cf. chap. 2.3 sur l'information et la connaissance dans l'entreprise et 3.1.4).

En comparant les partis pris occidentaux et chinois à travers les écrits philosophiques des uns et des autres, F. Jullien explique pourquoi dans son **Traité de l'efficacité**.

Aristote, explique-t-il, pense la praxis (au sens de la conduite) dans le décalage entre la pratique et la théorie. L'efficacité se définit en deux étapes : (1) la conception, le plan sont les formes idéales posées comme but, et (2) un effort de volonté pour rentrer la réalité dans le



modèle²². Aristote n'oublie pas la prudence (*phronesis*) et l'intelligence pratique (*mètis*) mais ne les théorise pas, de même qu'aucun philosophe par la suite.

En Chine, c'est le contraire : l'efficacité se définit comme la détection de ce qui peut être mobilisé, c'est-à-dire le potentiel inscrit dans une situation. Il n'y a donc pas comme en Europe cette double notion de « moyens/fins » et « actions/occasions ». Pour les Chinois, tout n'est que processus pensé sous forme « conditions/conséquences » : il faut aménager les conditions en amont pour que les effets découlent naturellement et indirectement. En d'autres termes, le chinois ne cherche pas à forcer les faits, ce qui revient à « biner autour des pousses au lieu de tirer sur les brins d'herbes ».

En Occident, c'est le thème de l'action que l'on retrouve dans l'épopée (action héroïque) et le théâtre (le drame). La philosophie chinoise est au contraire celle du « non-agir » en vue de l'efficacité, ce qui revient à limiter les efforts et les résistances pour éviter les risques. Si en Occident il existe toujours quelque chose qui échappe à l'entendement (Dieu, le hasard, le destin), ce n'est pas le cas pour les chinois car il suffit que les conditions soient réunies pour que les conséquences découlent. Le sage n'agit pas, il transforme. La transformation est globale, dans la durée, ne se démarque pas, on n'en voit que les effets ; contrairement à l'action qui est locale, assignée à un sujet, spectaculaire et de court terme. Le plein régime de l'efficacité c'est de ne pas forcer, de savoir faire basculer l'ennemi de son côté (et non pas le détruire) : « Les troupes victorieuses sont celles qui ont gagné avant d'engager le combat ». Il ne s'agit pas d'attendre le moment du combat, tout est dans la préparation, contrairement à la conception grecque du « *Kairos* », l'occasion, le moment juste et décisif.

Le sage chinois est celui qui voit le danger avant qu'il ne se manifeste, c'est une philosophie de l'anticipation et non de la projection, il faut alors savoir profiter d'une tendance inscrite dans une situation pour intervenir. Par exemple, le terme « porteur » est central en Chine alors qu'il est inexistant en Occident : si rien n'est porteur, le Chinois se met sur la touche, et attend un facteur favorable pour intervenir, ce qui arrivera forcément puisque tout est transformation. Les penseurs en stratégie occidentale (Machiavel, Clausewitz) ne pensent pas à l'efficacité mais au sujet, à la gloire, au plaisir, au prince. Pour le Chinois, il y a disparition de la notion de risque, d'audace, de plaisir. Le prince éclairé se fond dans la situation, se sert de l'immanence. « L'idéal de l'efficacité c'est la facilité ». Il n'y a plus de héros mais un art de gérer discrètement le cours des choses.

La conception grecque de l'efficacité consiste à édifier un modèle (une abstraction de formes idéales) projeté sur le monde, projet qu'il faut réaliser grâce à la volonté. « C'est la tradition du plan dressé d'avance et de l'héroïsme de l'action » (moyens-fin ou théorie-pratique).

L'efficacité entendue en Chine consiste à recueillir l'effet du déroulement d'une situation pour mieux réussir dans le monde : c'est la logique de l'implication que F. Jullien appelle Stratégie.

S'il fallait changer nos partis pris théoriques, c'est-à-dire opérer un décalage de pensée pour passer de « faire la réalité » à « laisser advenir la réalité », il faudrait revoir nos organisations basées sur le savoir codifié plus que sur le savoir actionnable.

Si l'on cherche à penser ensemble les savoirs conjecturaux de la *mètis* (non programmables) et les savoirs formels et logiques, sans pour autant entrer dans un système centralisateur comme dans la théorie de l'information, il s'agit de mettre à disposition des collectifs des systèmes de gestion dynamique des connaissances.

²² On retrouve bien dans ce schéma les éléments clés de la théorie de la décision.



3.2.5. Les Systèmes de Gestion Dynamique des Connaissances (SGDC)

La problématique que nous souhaitons aborder est celle de la gestion des connaissances en particulier scientifiques et techniques utilisées par les organisations pour leurs décisions d'action. Parmi toutes les démarches envisageables en gestion des connaissances, on constate deux tendances :

- la capitalisation des savoirs et du savoir-faire, dont l'objectif principal est de consigner les connaissances stratégiques, et qui porte donc l'effort sur la sélection et la structuration des connaissances,
- le partage dynamique des connaissances que l'on cherche à référencer sans préjuger de leur utilisation future, mais aussi sans éliminer des connaissances dont l'intérêt pourrait être révélé plus tard.

Parmi les sciences relatives à la cognition²³, c'est la discipline appelée "Intelligence Artificielle" qui a développé la première voie avec une ingénierie des connaissances fondée sur la constitution de structures de données exploitables et validables dans les domaines techniques où une expertise stable peut être dégagée.

L'utilisation massive des réseaux d'information et de communication a relancé récemment l'intérêt de constituer des corpus d'éléments de connaissances dynamiques partageables entre acteurs humains. Mais cette fois, c'est la capacité des systèmes à suivre l'évolution des connaissances (actionnables) qui est recherchée.

Les SGDCs explorent en priorité cette seconde voie et reposent sur une organisation de l'information à deux niveaux : les sources d'information et les connaissances actionnables, CAs. Un corpus de CAs recouvre un domaine de connaissances qui est aussi un domaine d'action (de décision).

La CA (souvent tacite) est une donnée informative jugée utile, une trace de raisonnement (mode cognitif respectant des principes stratégiques), une unité de sens partageable et utilisable (reliant savoir, savoir-faire et représentation symbolique).

Dans le domaine de l'ingénierie des connaissances, on appelle connaissance élémentaire une "donnée informative qui prend du sens dans un contexte". Fondé sur l'identification d'une information dégagée d'un ensemble de données, un savoir élémentaire se distingue donc d'une simple information par le fait qu'il comporte une part *d'interprétation* liée à la personne qui l'énonce. Si cette interprétation tend à rendre l'information utile à l'action (on lui associe une pertinence et une utilité), on peut parler alors de connaissance actionnable.

Nous avons développé des SGDC baptisés Référentiels des Connaissances où la CA est désignée par Élément de Connaissance (EC) et a la structure particulière suivante (Figure 12).

²³ au sens restreint de processus mettant en œuvre les couches supérieures de la pensée humaine, c'est-à-dire manipulant des concepts, des schémas, des connaissances et des processus explicites (raisonnement, activité langagière).

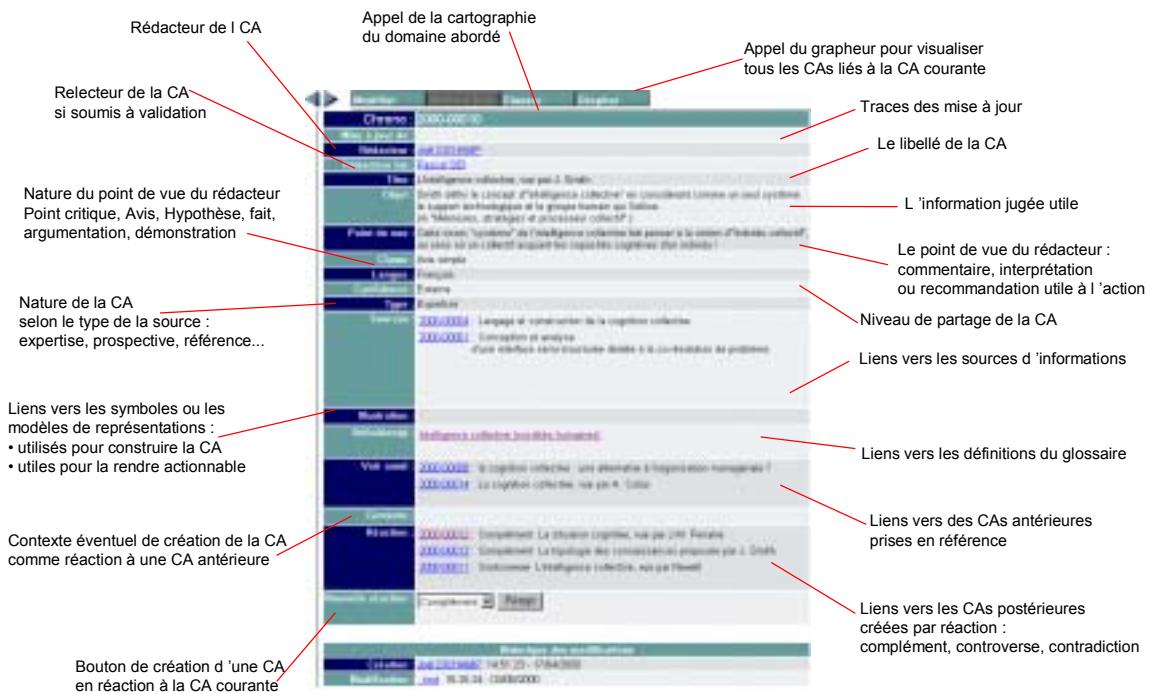
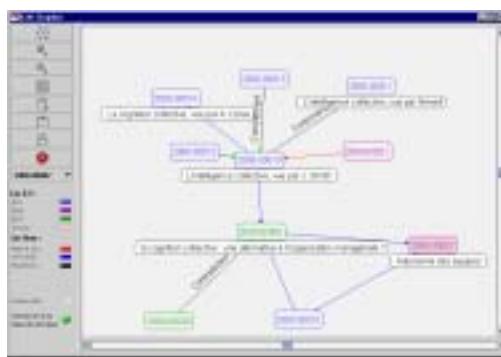


Figure 12 : Fiche de suivi d'un EC

La CA est ainsi identifiée par un triplet daté CA_t {objet informatif, source, rédacteur}, où l'objet est l'information issue d'une source documentaire ou donnée directement par le rédacteur. Cet objet d'information est interprété par le rédacteur de la CA_t pour son action, sous forme d'un commentaire où il exprime un point de vue ou une recommandation. Sous sa forme la plus simple, il s'agit "d'un paragraphe factuel - avec référence à la source d'information - qui porte l'information technique que le rédacteur estime devoir porter à la connaissance des acteurs du domaine. On peut si nécessaire lui donner une structure plus formelle, en référence à un *modèle* de connaissance.

Le corpus de connaissances est ainsi géré de façon dynamique sous forme de graphes se développant dans le temps, dont les nœuds sont les éléments de connaissance et dont les arcs figurent les relations entre les CA_t (Figure 13).





Pour chaque domaine de connaissance, on construit un référentiel, mais selon la problématique abordée dans le projet, on parlera de référentiel de sûreté, de référentiel qualité, de référentiel technique ou encore de référentiel de conception. Chaque domaine sera visualisé sous forme d'une grille (processus d'évaluation x concepts examinés) où les CAs seront référencés. Le rédacteur d'une CA est l'autorité légitime pour désigner quelles sont les cases (critère i ; concept j) de la grille du domaine concernées par sa CA et pour en estimer l'utilité et la pertinence. En pratique, la valeur de la CA (fonction de l'utilité et de la pertinence) est considérée comme un vote pour un couple (critère i ; concept j). L'agrégation de ces votes permet de qualifier l'ensemble du domaine (Voir section 2.5. *Le couplage d'un SIADG avec un système de gestion des connaissances*).

Selon la nature de la dynamique collective (processus de conception, réseaux de compétences...) un référentiel comportera différents niveaux organisés autour du niveau fondamental qu'est la base d'éléments de connaissances (Figure 14).

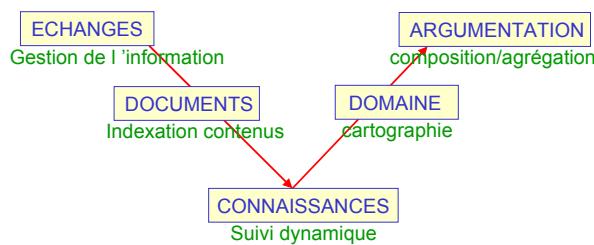


Figure 14 : Les niveaux d'un référentiel de connaissances

3.2.6. Retour d'expériences : une typologie

Afin d'expérimenter sur le terrain les concepts relatifs à l'intelligence collective, nous avons conçu un environnement informatique de travail collaboratif²⁴ autour de l'outil Référentiel des Connaissances fondé sur les objectifs suivants :

- Transformer les échanges d'informations en argumentation utile à l'action (aide à la décision),
- Organiser le partage des connaissances (corpus dynamique d'éléments de connaissances),
- Catalyser l'émergence d'une intelligence collective et l'amplifier.

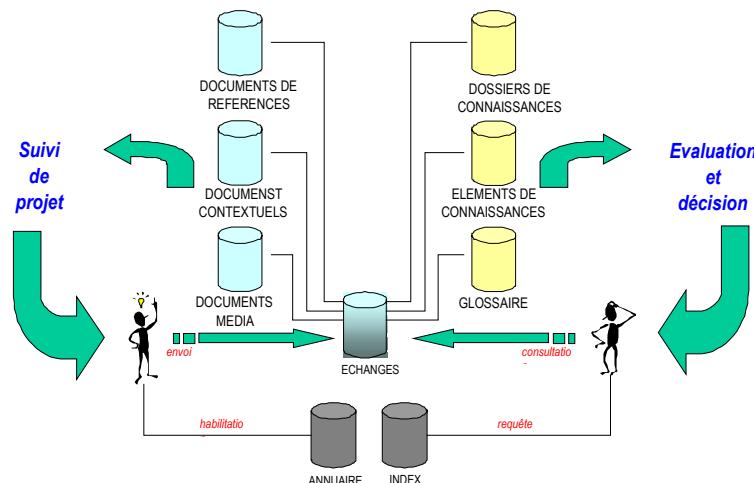


Figure 15 : La structure du Référentiel des Connaissances

²⁴ Cet environnement a reçu le prix INTRANET'99 décerné par le Ministère de l'Industrie.



Cet environnement a été utilisé dans des contextes différents de travail collectif où chaque acteur a des activités orientées vers les autres (négociation, reformulation) pour contribuer à des activités collectives (recherche, conception, décision...), mais toutes fondées sur la construction de représentations partagées [Teulier-Bourgine, 1997].

Dix-huit plates-formes au total ont été déployées dans huit organisations différentes, mais qui ont en commun les caractéristiques suivantes :

- Travail distribué entre équipes ou individus dispersés géographiquement,
- Projets, cycles industriels ou technologiques de durée notable (un an au moins),
- Connaissances produites dans un contexte organisationnel peu structuré.

Cet ensemble d'expériences en grandeur réelle (collectifs de 15 à plus de 200 personnes, serveurs hébergeant jusqu'à 3000 documents partagés) a permis de construire une typologie des modes de travail collectif, et une différenciation fonctionnelle en trois classes de référentiels. Dans chaque classe, des distinctions peuvent être faites selon le critère de cohésion.

Pour chaque mode de travail collectif, d'importantes variations dans le niveau de partage et même dans la nature de ce qui est partagé ont pu être constatées : une identité, une culture, des biens et des services, des connaissances, des informations, des expériences, des risques ou bien simplement... de l'estime.

Travail en groupe

Le groupe est pris ici comme rassemblant des individus fortement dépendants, c'est-à-dire en interaction (parfois continue) dans l'action. L'identité du collectif est ici donnée par le contexte (l'institution), et on peut distinguer trois sous-classes.

Le mode-projet correspond au travail coopératif dont la cohésion repose sur l'adhésion de chacun des acteurs. Un cas particulier réside dans le mode-plateau qui regroupe tous les acteurs en un même lieu et donc minimise les fonctions de gestion de communauté virtuelle (forums de discussion, identification des acteurs).



Figure 16 : un référentiel de projet au Commissariat à l'Energie Atomique
(120 ingénieurs sur 4 sites)

Plus général est le *mode-équipe* dont les membres peuvent contribuer à des projets différents, mais qui partagent une culture et des ressources communes ; le critère *d'appartenance* fonde ici la cohésion. Dans le travail de groupe, les fonctions de sécurisation des échanges sont minimales, les règles de partage étant explicitement fixées par l'institution.



Travail en communauté

Une communauté regroupe des individus autonomes engagés dans un processus de socialisation, volontaire ou provoqué. L'identité commune doit ici être construite et émerge de l'interaction des acteurs. Le critère de cohésion est ici très différenciant.

Fondé sur la *reconnaissance* de la valeur de chaque membre, le travail collectif s'effectue alors en *mode concurrentiel* : partage de ressources communes (financements, ressources humaines, matériel). Les fonctions de gestion du partage, de gestion des accès et de sécurisation doivent être particulièrement développées.



Figure 17 : un référentiel de programme scientifique au CEA (270 chercheurs sur 5 sites)

Le travail en communauté peut également procéder du mode *associatif* où la *motivation* des membres soutient l'œuvre commune. Les fonctions de synthèse des connaissances, et de cartographie dynamique des contributions sont ici très recherchées.

Le seul critère de *convivialité* peut suffire pour constituer des communautés émergentes en *mode interactif* (les forums internet en sont le meilleur exemple). La règle est ici du partage total, et le processus de socialisation y est le plus apparent.

Travail en réseau

Le travail en réseau repose sur la mise en relation d'individus à la fois autonomes (ils ont chacun leur sphère d'action propre) et indépendants, visant des objectifs individuels.

Le critère de regroupement peut être la *cooptation*, caractéristique des réseaux d'experts fonctionnant sur le *mode collégial*. Paradoxalement, le niveau de partage des connaissances peut être dans ce cas assez faible, l'échange d'informations restant privilégié. Le sujet collectif laisse la place à une collection d'individus opérant en fonction de leurs représentations personnelles.

Le réseau d'échanges adopte le critère de *valorisation* pour fonctionner sous le *mode de la reciprocité*. Les échanges interpersonnels finissent par former un construct collectif. Ce type de réseau suppose des processus d'authentification stricte des acteurs.

Enfin, le rassemblement d'acteurs réunis par des intérêts communs constitue le cas particulier du *mode coalition* (exemple des consortiums, ou des programme d'intérêt collectif). Les acteurs adoptent un objectif général, mais fonctionnent selon le principe de *subsidiarité*. L'objectif général n'est pas censé être atteint par la planification centrale et détaillée des tâches, mais grâce à la contribution des projets particuliers, avec leurs objectifs respectifs.



3.2.7. Conclusion

Les technologies avancées de l'information sont utilisées dans les SGDCs (le référentiel de connaissance en particulier) pour faciliter le *parcours* du corpus de connaissances (navigation) mais aussi pour suivre l'évolution du corpus de connaissances (graphes dynamiques), améliorer la *visibilité* du domaine abordé (cartographie des connaissances) ou encore améliorer l'*utilisation* des connaissances pour l'action (composition des connaissances).

Un corpus de connaissances dynamiques ne contient pas en lui-même de solutions préfabriquées aux problèmes que se pose l'organisation ; il constitue plutôt un espace de confrontation de différentes stratégies décisionnelles.

Aujourd'hui, les SGDC, déployés en intranet, sont expérimentés au sein de grands projets de conception collective au sein de grandes entreprises du monde industriel. La diversité des expérimentations menées a permis de dégager quelques principes généraux, mais également de "bousculer" certaines idées reçues. Parmi celles-ci, la plus évidente est que la collaboration ne se crée pas spontanément même lorsqu'on partage un même espace de travail. La seconde est que la seule interaction entre acteurs n'est pas suffisante pour l'émergence d'une identité collective, et la réalisation d'un *construit collectif* constitue le premier pas vers l'intelligence collective.

4. LA CREATIVITE

La prise de décision et la résolution de problèmes avec créativité doivent souvent être associées. Il s'agit alors d'ouvrir de nouvelles voies à la façon de penser et d'augmenter l'habileté à trouver des solutions créatives face aux problèmes rencontrés quotidiennement. Par un processus logique, il faut utiliser la pensée créative pour identifier les problèmes, leurs causes ou racines, et trouver les solutions adéquates, utiliser des outils spécifiques comme le brainstorming ou la pensée latérale, introduire les changements et surmonter les résistances naturelles. C'est ce dont nous allons discuter dans cette section dédiée à la créativité.

4.1. Quelques définitions

Une définition de la créativité couramment acceptée est la suivante : « *faculté à organiser les éléments du champ de perception de façon originale et susceptible de donner lieu à des réalisations ou des conceptions nouvelles* ».

Toutes les activités intellectuelles mises en œuvre par un utilisateur ou un collectif d'utilisateurs face à une situation problématique font appel à cinq formes ou niveaux de créativité que l'on décrit habituellement comme suit :

- Celui de l'utilisation de schémas de pensée prédéfinis et de la modélisation. La réponse à cette nouvelle situation surgit alors automatiquement car elle est déjà identifiée, mémorisée et mise en pratique donc validée par le passé. A ce niveau, bien évidemment, la créativité s'exprime extrêmement peu. Par exemple, améliorer la sécurité des achats de billets sur Internet implique de faire appel à des techniques et à des outils informatiques (cryptage, sécurisation des échanges et des données, mots de passe, etc.) connus et maîtrisés. Il suffira d'étudier ces systèmes existants pour en tirer une solution adaptée.
- Celui de la réparation du modèle. En effet, un dysfonctionnement peut se traduire par un écart entre la situation actuelle et la situation normalement souhaitée ou attendue au même instant. La créativité s'exprimera alors par la capacité de cet utilisateur ou de ce collectif à



modifier le modèle de pensée avec lequel il raisonne pour se retrouver dans cette nouvelle situation. La démarche de créativité est ainsi orientée sur les effets de cette situation et non exclusivement sur ses causes. Par exemple, une décision des pouvoirs publics qui a pourtant été mûrement réfléchie et argumentée se trouve mise à mal suite au mécontentement de nouveaux acteurs dont il n'a pas été tenu compte auparavant. Ce peut être les riverains d'une voie de chemin de fer en extension ou les utilisateurs d'un réseau d'autoroute sur lequel des péages automatiques sont mis en place. Il faudra alors mettre en œuvre un processus de créativité pour trouver des moyens de négociation et palier ce mécontentement (insonorisation des habitations ou gratuité pendant un certain temps des péages) sans pour autant remettre en cause la décision appliquée.

- Celui de l'analyse de l'origine du dysfonctionnement et de l'écart. La démarche de créativité a alors pour but de formuler les causes potentielles du problème ayant fait apparaître cette situation afin de mieux les maîtriser. Elle nécessite évidemment une ouverture d'esprit, une remise en cause de l'existant et une prise de recul plus importante que dans le cadre des deux premiers niveaux.
- Celui de la vision dans le futur qui nécessite une capacité à caractériser la situation de manière objective, à projeter et à discerner vers quel état le système à l'origine de la situation doit évoluer afin de résoudre le problème donné. Par exemple, le transport des marchandises par les airs peut-il solutionner une partie du problème de l'engorgement des autoroutes ?
- Celui de la pensée innovante qui nécessite une capacité supplémentaire à trouver les moyens de résoudre un problème en utilisant les ressources disponibles, même cachées. Dans le même ordre d'idées que précédemment, comment faire accepter aux commerçants et artisans vivant et travaillant dans une agglomération, la solution de transport de marchandise par des dirigeables aptes à manœuvrer et à se poser sans installation préalable particulière ?

Dans le cadre qui nous intéresse ici, l'individu confronté à des problèmes complexes de transport (qu'il soit utilisateur ou qu'il offre un service auprès d'autres utilisateurs) attend des pouvoirs publics qu'ils lui fournissent une vision à long terme et, si possible, qu'ils mettent en œuvre des solutions pertinentes et nouvelles dont l'intérêt économique est assuré, c'est à dire des solutions innovantes. Il leur demande donc de faire preuve de créativité.

4.2. La créativité : une approche physiologique

Afin de mieux, cerner le processus créatif lui-même, il peut être intéressant de rappeler les mécanismes physiologiques en jeu [Brabandère, 1998]. Le cerveau humain est constitué, d'une part, de deux hémisphères qui ont des fonctionnalités différentes :

- L'hémisphère gauche est le centre de la pensée analytique, il permet une approche logique, séquentielle des problèmes.
- L'hémisphère droit est celui de l'intuition, de la vision globale des problèmes, de la synthèse et de l'association.

Les deux hémisphères fonctionnent en relation mais chacun d'entre nous a tendance à privilégier un mode de pensée, c'est à dire à utiliser davantage un de ses hémisphères. Une façon d'augmenter sa capacité créative est alors de s'entraîner à faire travailler ses deux hémisphères simultanément [Guntern, 2001].

D'autre part, les spécialistes ont mis en évidence une structure en couches du cerveau, correspondant à l'évolution au cours du temps de cet organe : ainsi, le « cerveau reptilien » est



la couche la plus ancienne, c'est lui qui gère les réflexes. Nous utilisons notre « cerveau limbique » pour les émotions, la mémoire, et notre « cerveau cortical » pour le raisonnement, la conscience.

A partir de ces constatations, Ned Hermann a défini quatre grands types de caractères :

- « l'hyperlogicien » (cortical gauche) qui privilégie le raisonnement logique, analytique,
- « le praticien » (limbique gauche) qui planifie, vérifie, et cherche à tout contrôler,
- « le rêveur » (cortical droit) qui se laisse guidé par son intuition, par des associations d'idées plutôt que par le raisonnement,
- « le communicateur » (limbique droit) qui est attentif aux autres, porté sur la communication.

Nous utilisons au maximum notre potentiel créatif lorsque nous arrivons à faire fonctionner les deux hémisphères de notre cerveau face à une question posée. Selon les concepts introduits par E. De Bono, cela revient à ne pas privilégier la pensée « verticale » par rapport à la pensée « latérale » [Bono, 1991] : la première est celle du processus logique conventionnel, lorsque l'on se focalise sur l'action qui doit résulter de la décision, et que l'on emprunte le chemin le plus probable pour rejoindre l'étape suivante du processus de décision. La pensée latérale, au contraire, consiste à assembler et à structurer différemment les modèles habituels. Pour cela, le processus mental doit accepter des points de divergences, en relâchant le contrôle du raisonnement logique, et même éventuellement en acceptant le hasard. Cela est rendu possible par l'utilisation de l'évasion et de la provocation : l'acteur doit accepter de prendre du recul, de regarder autrement le problème, mais aussi parfois de formuler une idée a priori mauvaise qui peut aboutir à une solution meilleure, plus originale .

Cette classification en quatre types est certes trop simpliste pour représenter fidèlement la réalité de la complexité des processus intellectuels et émotifs, cependant elle donne quelques pistes à suivre afin d'améliorer les capacités créatives de chacun, et également le fonctionnement en groupe.

Par exemple, un travail de recherche de solutions mené de manière idéale devrait reposer sur des acteurs disposant de connaissances et de savoirs différents, mais aussi de caractères différents afin de créer des synergies. Cela permet aussi de limiter les facteurs de blocages inhérents à chaque type de caractère.

4.3. L'attitude créative

Lors du déroulement d'un processus décisionnel tel que nous l'avons décrit précédemment en termes de processus créatif , l'ensemble des acteurs doit adopter une attitude créative, à la fois individuellement et collectivement. Nous allons voir à présent les moyens pour favoriser une telle attitude :

4.3.1. Vaincre l'inertie psychologique des acteurs

L'inertie psychologique est un phénomène qui bloque les possibilités de l'acteur à décrire le problème lui-même, à s'exprimer et à générer des idées réellement nouvelles et indépendante de toute contrainte technologique, économique ou humaine au moins durant la phase de divergence. Elle limite ainsi la prise de recul et ne permet pas de sortir des sentiers battus.

Elle est liée aux savoirs (acquis, pratiques, innés), aux croyances, aux habitudes et aux connaissances souvent parcellaires ou incomplètes, volontairement ou non, que l'on peut avoir d'un phénomène, d'un système, d'un domaine ou d'un métier. Par exemple, un



fabricant de moteur atmosphériques ne pourra pas facilement concevoir, pour ne pas dire accepter, de remettre en cause le principe du moteur à explosion pour étudier un nouveau moteur à air comprimé. De même, un responsable de collectivité ne pourra pas facilement passer outre les recommandations et exigences (quelquefois suite à certaines formes de lobbying) de ses électeurs.

Il est intéressant de revenir sur les classifications vues plus haut, afin de comprendre les modes de blocage auxquels est plus particulièrement soumis chacun des types de caractères :

- Les « hyperlogiciens » peuvent se retrouver bloqués dans une démarche créative par leur trop grande foi en la logique, les statistiques ou en se fixant sur la solution apparente.
- Les « praticiens » ont du mal à percevoir les relations inhabituelles, et se focalisent à l'extrême sur le côté pratique et économique.
- Les « rêveurs » ont des difficultés à découper un problème complexe en sous tâches, à fournir des précisions.
- Les « communicateurs » quant à eux, peuvent être bloqués par leur désir de conformisme, leur dépendance par rapport à l'opinion d'autrui.

4.3.2. Utiliser des référentiels

Toute activité intellectuelle se réfère à un univers du discours donné. Il paraît souhaitable que les acteurs du collectif connaissent et comprennent un certain nombre de concepts et de termes ce qui ne devrait en aucun cas préjugé de leur autonomie de pensée.

Cet univers du discours peut être représenté comme un référentiel ayant deux rôles :

- Un rôle normatif des relations entre acteurs : il permet de définir sans ambiguïté un ensemble de termes, d'idées et de règles communément admises par tous les acteurs pour décrire un domaine. Par exemple, le cas du transport nécessite de définir et de s'accorder sur le sens de finalité, de sécurité ou encore de confort d'utilisation d'un système de transport. Ce référentiel doit faire émerger une sorte de consensus ou au moins permettre de signaler les interprétations éventuellement divergentes qui pourraient être faites de certaines informations ou connaissances. Il ne doit pas non plus limiter l'univers du discours à ce seul consensus et empêcher tout apport éventuellement intéressant d'un autre domaine. Il est par exemple commode d'utiliser un référentiel d'idées apparues ailleurs ou dans d'autres domaines, de connaissances et de savoirs transverses, de solutions connues et usitées [Terninko, 1998] pour améliorer tant en terme de durée qu'en terme de pertinence la phase de créativité.
- Un rôle de capitalisation, de gestion et de communication doté de (ou basé sur une) capacité de représentation, de traitement, de mise en forme et de navigation d'une information qualifiée de pertinente. Il se doit d'être riche et contenir plus qu'une simple concaténation des savoirs parcellaires des différents acteurs. Pour cela, il doit faire appel à des concepts de représentation et des mécanismes de manipulation de ces concepts les plus génériques possibles car nécessairement indépendants d'un domaine donné.

4.3.3. Impliquer, inciter et piloter les acteurs de la créativité

Un collectif d'acteurs durant une phase de créativité se doit d'adopter certaines règles que l'animateur du groupe est chargé de faire respecter :



- Les mots sont très importants, ils véhiculent tout un contexte propre à chacun. Pour amorcer une phase de divergence, il faudra donc choisir avec soin des mots « neutres » permettant l'écueil des associations d'idées et des schémas de pensée automatiques.
- Il faut accepter de séparer les phases de proposition d'idées (divergence) et celles d'évaluation de ces idées (convergence) : cela doit se faire au moins dans le temps (laisser tout le monde s'exprimer sans émettre de critique), ou même en changeant les acteurs (les décisions se prennent en deux temps, pour éviter d'être à la fois juge et partie).
- Il vaut mieux penser « solution idéale » [Alshuller, 2002 ; Terninko, 1998] plutôt que d'accepter dès le début une situation de compromis. Cela revient aussi à séparer les phases de divergence de celles de convergence (au cours desquelles sont évalués les avantages et inconvénients)
- Il faut pouvoir écouter les autres. La coopérativité lectorielle [Gavriloff, 2001] permet d'améliorer la phase de divergence par un mécanisme d'écoute et de prise en considération (essentiellement par absence de critique et par respect du temps de parole) des autres acteurs. Elle nécessite cependant une tolérance et un contrôle raisonné de la personnalité et de certains réflexes plus ou moins incontournables voire conditionnés dus à la formation acquise, au rang social, aux compétences reconnues, à l'appartenance éventuelle à un groupe restreint voire élitiste, etc. Elle consiste à accepter le fait que tout ce qui se produit ou se dit autour de nous, quelle que soit la source de l'information ou de l'action, contient un sens dont l'interprétation peut amener de nouvelles idées ou de nouvelles formulations d'un problème. Le but est donc de disposer d'un maximum de sens différents, sources d'autant de nouvelles situations. C'est cette richesse de références et de points de vue qui amènera les acteurs du processus à prendre en considération de nouvelles solutions qui leur paraîtront dans un premier temps étranges, inutilisables selon leur point de vue voire dénuées d'intérêt pour l'ensemble du projet car en dehors de leur métier ou de leurs domaines de compétences
- Il faut enfin laisser le temps aux acteurs de pouvoir s'approprier les différents points de vue, idées et solutions proposées de manière à permettre à chacun à son tour d'enrichir à nouveau cette information.

4.4. Classification des méthodes d'aide à la créativité

Il existe nombre d'approches, de méthodes et de techniques permettant de mettre en œuvre et/ou d'améliorer de manière indiscutable la créativité. Elles favorisent l'exploration de pistes encore inconnues, l'expression, la communication, le rêve, l'ouverture d'esprit, la remise en cause ou l'abstraction de soi-même. En d'autres termes, elles facilitent, si elles sont correctement et objectivement employées, le fait de passer outre le domaine de compétences, d'engagement ou de souhait des acteurs face à une problématique donnée.

Ces méthodes sont utilisables, à défaut seul ou, mieux, au sein d'un collectif. Elles peuvent être plus ou moins spécialisées selon les cas, pour répondre à des besoins de levée de l'inertie psychologique, de (re) formulation de problèmes à résoudre ou encore la recherche pure et simple d'idées et de solutions nouvelles ou différentes de celles pratiquées actuellement.

Les méthodes essentiellement utilisées dans la levée de l'inertie psychologique, ou de divergence sont d'autant plus profitables que chacun s'entraîne à les utiliser dans la vie de tous les jours, afin d'acquérir l'habitude de penser « autrement ». Elles sont souvent très simples à mettre en œuvre et nécessitent finalement très peu de moyens.



Les méthodes de re formulation, ou de re convergence, sont nécessaires après la phase de divergence pour préciser ou modifier les hypothèses à l'origine du problème à résoudre.

Les méthodes permettant enfin d'organiser les idées (il n'est effectivement pas souhaitable de guider la recherche d'idées ce qui serait parfaitement illusoire et opposé à la notion même de créativité) sont souvent plus complexes et adaptées à certains domaines de compétences : systèmes techniques, systèmes sociaux, etc.

Enfin, certaines méthodologies (appelées cependant méthodes par abus dans la bibliographie) associent et font collaborer plusieurs méthodes et techniques différentes.

Lorsque cela s'avère nécessaire, essentiellement dans le cas de méthodologies plus lourdes à utiliser, des outils informatiques supports complètent alors l'éventail de moyens à la disposition de l'acteur désireux d'améliorer sa créativité.

Ces nombreuses méthodes d'aide à la créativité peuvent être répertoriées de la façon suivante.

4.4.1. Méthodes combinatoires

Le mode opératoire consiste dans un premier temps à décomposer le système en différents sous systèmes, éléments. On peut par exemple faire une décomposition selon les principes de l'analyse multidimensionnelle (carte morphologique), ou bien encore en déclinant différents points de vue, ou ressources (matrices de découverte), etc. [Brabandère, 1998]. Il faut ensuite de croiser, opposer ou combiner ces composants les uns par rapport aux autres de manière plus ou moins systématique pour faire émerger de nouvelles perspectives ou idées. L'outil de représentation employé est essentiellement les matrices.

4.4.2. Méthodes antithétiques

Le problème étant formulé, on cherche à le déformer de manière à exprimer exactement le contraire voire même à le nier. Le but est alors d'explorer le nouvel univers de possibilités qui s'offre et de raisonner ainsi par l'absurde. On trouve ainsi des techniques comme le concassage, ailleurs et pas maintenant, l'optimisation d'une fonction ou les sept chemins [Fustier, 2001].

4.4.3. Méthodes associatives

Elles consistent à raisonner par association d'idées à partir d'une première formulation du problème et, ainsi, à enrichir le champ d'investigation. Les cas les plus connus sont ceux du brainstorming, du brainwriting [Gavriloff, 2001] ou encore du diagramme pieuvre ou des cartes mentales [Buzan, 1998].

4.4.4. Méthodes analogiques

Le problème étant formulé, on peut tenter de rechercher des idées ou des pistes de solutions en raisonnant sur des analogies de ce problème, éventuellement en dehors de tout contexte et de tout domaine spécifique, mais paraissant semblable ou très voisin [Fustier, 2001]. Il peut s'agir d'une analogie fonctionnelle (un système dans un autre champs d'application remplissant une fonction identique) ou d'aspect (un système ressemblant physiquement évoquant des pistes de solutions). Les moyens de transport aérien, fluvial, maritime ou terrestre utilisent par exemple une énergie en majorité d'origine fossile. Les ressources naturelles tendent à s'amenuiser et leur transformation induit de plus des rejets dans l'atmosphère que l'on souhaite voir diminuer. Solutionner ce problème de dépendance et d'effet transport / ressource pourrait trouver un parallèle dans les systèmes vivants qui doivent



nécessairement s'auto-organiser et s'auto-adapter au milieu dans lequel ils vivent sans le détruire pour pouvoir y survivre.

4.4.5. Méthodes empathiques

Ces méthodes nécessitent que les acteurs s'approprient la situation ou le problème en s'identifiant à l'une de ses composantes et se comportent en tant que tel. Par exemple, la résolution d'un problème de congestion de la circulation à l'intérieur d'une ville peut être abordée en s'identifiant à un des véhicules circulant. Chaque acteur s'efforcera de simuler et de ressentir les différents événements, situations et scénarios auxquels est confronté chaque type de véhicule. Un exemple de méthode empathique est celle dite « du Sénat » ; dans ce cas, les participants sont invités à se mettre « à la place » d'une personnalité connue, afin de donner ce qu'ils imaginent être son point de vue [Gavriloff, 2001].

4.4.6. Méthodes aléatoires

Si le fait de raisonner sur des analogies au problème ou sur le contraire de ce problème ne permet pas d'aboutir à des idées, il peut être intéressant, bien que beaucoup moins probant, de raisonner sur des concepts étrangers, éloignés, voire sans aucun lien avec la réalité du problème [Fustier, 2001]. L'idée est de briser l'inertie par la provocation.

4.4.7. Synthèse

En règle générale, face à un problème, il est recommandé de croiser, de mélanger ces différents types de méthodes, en vue d'obtenir une efficacité maximum. De toute manière, il serait illusoire de vouloir dresser une liste exhaustive ainsi qu'une classification des techniques en fonction de l'étape du processus décisionnel au cours de laquelle elles peuvent être appliquées. Cependant, en se référant au déroulement du processus décisionnel tel que nous l'avons établi plus haut, on peut tenter de les classer en fonction de la phase à laquelle elles s'appliquent le plus souvent ou de la manière la plus adéquate.

4.5. Mise en œuvre des méthodes

4.5.1. Approche directe ou indirecte

Comme nous l'avons vu précédemment, le processus décisionnel est caractérisé par trois états particuliers :

- L'état source qui correspond à une formulation incomplète, ambiguë, entachée de non-dits ou de sur-spécifications. C'est cet état qui est à l'origine du processus.
- L'état intermédiaire qui correspond à une formulation acceptée et validée de tous les acteurs du problème à résoudre. Cet état est ainsi le résultat de la première phase divergence/convergence.
- L'état final qui correspond à une ou plusieurs solutions argumentées et pouvant ensuite être mises en œuvre. Cet état est alors le résultat de la deuxième phase de divergence/convergence.

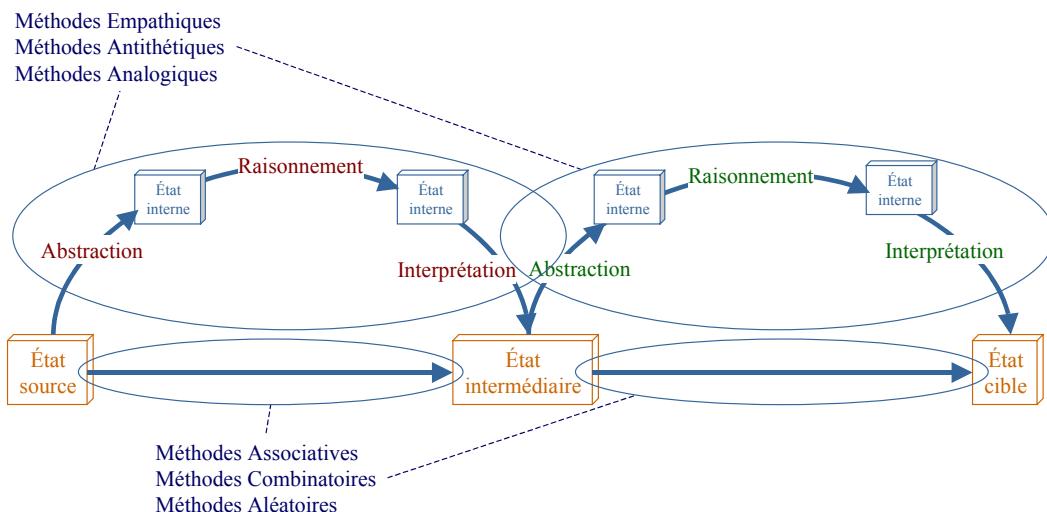


Figure 18 : Approches directes et indirectes

Il existe alors deux types d'approches permettant de passer de l'état source à l'état cible telles que le montre la Figure 18.

- Des approches directes : ces approches permettent souvent d'atteindre rapidement un état résultant mais assez souvent au détriment de la pertinence ou de l'exhaustivité du résultat atteint. Ces approches sont essentiellement basées sur des méthodes associatives, combinatoires ou encore aléatoires.
- Des approches indirectes : sensiblement plus lourdes à utiliser mais plus exhaustives quant à la qualité du résultat, elles mettent en œuvre les trois phases suivantes en faisant alors apparaître des états internes:
 - L'abstraction pour se doter d'une représentation suffisante et fiable, de la réalité et du problème. On obtient alors un modèle formel ou semi-formel (mathématisable, exécutable ou autre) ou informel (en langage naturel ou autre) contenant les hypothèses et informations de base nécessaires à la compréhension de la situation.
 - Le raisonnement sur le modèle obtenu qui permet de sortir de la réalité d'un domaine, d'un métier et de limiter ainsi les effets de bord néfastes provoqués par les savoirs, les croyances, certains mots au sens définit avec ambiguïté et par d'autres freins.
 - L'interprétation des résultats du raisonnement dans le monde réel. Le but est alors d'identifier les solutions concrètes, applicables ou de reformuler certains faits issus du modèle en réemployant les termes consacrés, réservés et connus du domaine.

Ce deuxième type d'approches fait appel essentiellement aux méthodes analogiques, empathiques, antithétiques ou encore aléatoires.

4.5.2. Les phases de divergence

Formulation du problème

Toute phase de divergence commence par une appropriation du sujet traité. Pour cela, les acteurs doivent tenter de cerner la problématique, d'établir les limites de leur champ d'action et les contraintes auxquelles ils sont soumis. Afin de prendre du recul, il faut accomplir la



démarche indiquée par William Gordon : « rendre familier l'étrange et rendre étrange le familier ».

Un premier travail peut consister à acquérir des informations d'origines diverses pouvant être utilisées a priori. Selon le type de problème, cela peut aller d'une analyse bibliographique scientifique jusqu'à la recherche et l'analyse de documents techniques (brevets, rapports d'expériences, publications, enquêtes, statistiques, standards, normes, etc.). Le risque est de limiter le domaine d'étude et il faut alors ouvrir le champ exploratoire afin de se faire une idée sur des problèmes connexes, identiques ou éloignés. Pour cela les techniques de créativité présentées ci-dessous peuvent être utiles. Il faut ensuite procéder en alternant des phases de travail de recherche d'informations et des phases de créativité pure [Fustier, 2001 ; Gavriloff, 2001].

Le martien, l'habitant de Sirius, l'enfant, le stagiaire, l'attitude du débutant

Cette approche, aux noms évocateurs, consiste à s'étonner de tout : les réponses aux questions, la situation actuelle, l'existence ou l'absence de telle chose, etc. Elle entraîne alors des questions auxquelles doivent accepter de répondre les autres acteurs. Chacun peut tenir le rôle d'étonné mais chacun doit faire en sorte, pour mieux se faire comprendre et mieux partager son point de vue, de reformuler ses hypothèses et de répondre aux questions des autres acteurs mêmes si elles paraissent étonnantes, dénuées en apparence d'intérêt ou de sens voire même impertinentes.

Pourquoi ?

Cette méthode, a priori très simple à mettre en œuvre, reste assez contraignante pour l'interviewé qui peut se sentir agressé par l'interviewer. Elle consiste en effet à enchaîner les questions « Pourquoi ... ? » après chacune des réponses données par l'interviewé. Par exemple, voici une suite de questions : « pourquoi transporter ces matières par camion ? » puis « pourquoi construisez-vous des camions ? » puis enfin « pourquoi ne pensez-vous pas pouvoir construire autre chose que des camions ? ». Cela peut alors amener les deux acteurs de cette discussion, l'interviewé comme l'interviewer, à reconsiderer le problème sous d'autres angles, à mieux comprendre sans doute l'origine même du problème ou le point de vue de l'autre.

La défectuologie

Le point de vue du martien ou de l'enfant cherche à comprendre un système et/ou un problème d'un œil externe en se démarquant ainsi de toute considération partisane ou tout préjugé technique, idéologique ou autre. A l'opposé, la défectuologie est une démarche qui consiste à (s'autoriser à) critiquer le système vu alors de l'intérieur et à détecter l'insatisfaction que ce système entraîne. En d'autres termes, elle permet à un utilisateur, avec son propre point de vue, quelquefois de manière très subjective, de mettre en avant les défauts qu'il perçoit du système. Une fois le problème ciblé, il apparaît alors deux possibilités au poseur de problème. Il peut perfectionner le système de manière à voir disparaître le défaut, le dysfonctionnement ou le problème. Le processus qui est alors en jeu pousse souvent ce poseur de problème à raisonner et à trouver une solution dans le même domaine que celui du problème originel. [Fustier, 2001] souligne le cas du moteur à explosion qui a évolué et s'est grandement perfectionné. Cependant, toute amélioration de ce système est passé par une remise en cause essentiellement des aspects mécaniques ou thermiques. La deuxième voie possible est alors de solliciter le poseur de problème pour sortir du domaine originel et rechercher alors des solutions à ses critiques ailleurs. L'exemple du moteur à explosion, vu maintenant comme un système dont il faut réduire les rejets polluants, permet de chercher



d'autres types de systèmes qui assurent évidemment les mêmes fonctions : l'amélioration pourrait ainsi venir des moteurs électriques toujours plus puissants, petits et faciles à produire et surtout non polluants.

L'amont de l'amont

Dans le même ordre d'idées que la technique du pourquoi, cette technique consiste à poser des questions sur les faits, événements ou situations qui sont à l'origine du système et du problème étudiés. Il faut alors forcer les acteurs à formuler l'origine de ce problème ou de l'origine même du système dans son entier de manière à sortir des frontières fixées a priori. Ainsi, un constructeur automobile devant faire évoluer un modèle existant vers un modèle possédant des caractéristiques non-polluantes, pourrait formuler son problème comme suit : «Est-il nécessaire ou souhaitable de produire ce nouveau moyen de transport qui corresponde effectivement à ce nouveau besoin ? ». Le fait de formuler une question plus large englobant la finalité du futur véhicule permet de raisonner non plus sur un nouveau modèle de véhicule mais de remettre en cause d'autres parties du métier même de ce constructeur.

Plus –Moins – Intéressant

Cette technique, très simple à mettre en œuvre de manière collective, est adaptée pour décrire, échanger et classer des points de vue sur le problème et non sur la solution [Bono, 1985]. Elle consiste à établir une liste de points positifs sur le système au sein duquel le problème est perçu par les acteurs et une liste des points négatifs qui semblent être réellement à l'origine du problème mais aussi qui font émerger d'autres problèmes apparemment passés inaperçus aux yeux de certains acteurs. Ensuite une troisième liste est constituée contenant les points intéressants, qui devront être étudiés plus en détail : ce peut être les points qui apparaissent à la fois comme positifs et négatifs, mais aussi cela permet de passer outre au jugement de valeur, et de classer sous cette rubrique une idée « qu'on n'aime pas », qui dérange, mais qui semble avoir quelques avantages. Cela permet donc une dynamisation de la réflexion.

Par exemple, le nombre de véhicules qui sont arrêter pour des délais plus ou moins longs au bord de la chaussée, et qui gênent donc la circulation à l'intérieur d'un centre ville, devient trop important. Les acteurs peuvent alors être amenés à lister les points positifs d'une telle façon de faire : on peut laisser sa voiture le moteur en marche quelques instants pour faire des courses rapides, cela gène le voisin irascible, cela permet de garder un œil sur son véhicule, de s'arrêter n'importe où, etc. Les points négatifs seront alors : il faut laisser le moteur tourner et cela consomme et pollue, le véhicule est dangereux là où il se trouve, les véhicules gênent la circulation, personne de peut faire circuler un véhicule sans chauffeur, le temps de parking est incontrôlable, etc. Deux points intéressants que l'on peut trouver dans ces deux formulations concerne la sécurité du véhicule et le temps de parking qui doit être le plus court possible. Des solutions diverses peuvent alors être avancées comme mettre à disposition des places sécurisées de parking. Celle-ci peuvent être payantes (éventuellement très cher en fonction du temps passé) ou encore utiliser un composant de la voiture elle-même pour limiter le temps de parking (comme la durée de vie d'une batterie par exemple lorsque le véhicule doit rester phares allumés).

Le cœur du sujet

Cette technique consiste à formuler l'objectif ayant permis de déterminer l'objectif initial du système ou de la situation en cours d'étude.

Par exemple, il faut décider du tracé d'une nouvelle autoroute en fonction de contraintes à déterminer. L'objectif de ce travail est donc de choisir un tracé idéal. L'objectif de cet objectif n'est-il pas de faire en sorte que l'existence même de cette autoroute ne soit pas remise en cause d'ici quelques années par le fait qu'elle traverse une région donnée à un endroit donné



ou qu'elle ne donne accès qu'à une catégorie de véhicules dont la technologie risque d'être obsolète et donc d'évoluer très différemment d'ici peu (par exemple, par le remplacement d'un système de conduite classique par un système de guidage automatique nécessitant une infrastructure particulière) ? Répondre à ce deuxième objectif amènera les acteurs à déterminer de manière plus exhaustive de nouvelles contraintes à respecter aujourd'hui pour déterminer une autoroute adaptée aux futurs besoins.

L'ambition de l'objectif

Plus un objectif est ambitieux, plus il nécessite une remise en cause profonde de l'existant et plus il poussera les acteurs à être créatif. Il paraît donc souhaitable de formuler des objectifs très ambitieux voire inatteignables mais cela doit évidemment se faire avec une formulation très adroite et très motivante pour les acteurs participants à cette phase de reconvergence. En effet, il peut être préférable de dire que l'on doit « supplanter le transport par camion au profit du rail » plutôt que dire que l'on doit « augmenter de 50% la capacité du trafic ferroviaire d'ici trois ans ».

L'avocat de l'ange

Cette approche consiste à positiver ou à se réjouir d'un échec de manière à en faire ressortir la possibilité de succès ou tout au moins d'en faire émerger de nouveaux faits résultant de cette situation d'échec qui, si ils ne correspondent pas à l'objectif initial qui était fixé, décrivent un environnement éventuellement profitable si on sait le reconnaître et se l'approprier. Il faut cependant faire attention à la formulation de l'échec à partir duquel on tente de partir et ne pas se borner à une constatation passive et fermée des faits. Enfin, il faut ensuite reprendre chaque fait pour l'analyser en détail et y détecter des pistes potentielles de 'sortie d'échec'.

Ainsi que nous l'avons précédemment montré, la recherche d'idées peut se faire « directement », en partant de la situation réelle, ou bien en passant par des étapes d'abstraction (de modélisation), de raisonnement puis d'interprétation. On peut classer les techniques d'aide à la divergence selon le type d'approche utilisée ;

La recherche d'idées : approche directe

Les méthodes associatives ou combinatoires suivantes sont alors mises en œuvre.

Le Brainstorming

Il s'agit d'un travail de groupe pour formuler une série d'idées qui pourront ensuite être triées, examinées et servir ensuite d'orientation vers la solution d'un problème. Pour une bonne efficacité, le problème posé doit être suffisamment précis et spécifique. En cas de sujet trop vaste, il faut identifier des sous problèmes et consacrer à chacun une séance distincte. L'animateur a pour rôle de présenter le problème sous forme qu'une question bien définie, et de noter toutes les idées sous forme de phrases complètes. Il doit veiller à ce que le groupe reste dans une phase de production d'idées, sans critique. Il doit également pouvoir relancer cette production en fournissant quelques pistes. Les membres du groupe (en moyenne une quinzaine) peuvent utiliser toutes sortes d'associations, de détournements ou de contradictions pour fournir des réponses, et ils sont invités à réagir et à répondre aux propositions du groupe, afin d'obtenir une sorte de « réactions en chaîne ».

L'étape de classement et d'évaluation de ces idées doit être distincte dans le temps, et même être réalisée par d'autres intervenants.



Le Brainwriting

Il s'agit de la version « papier » du brainstorming : chaque intervenant écrit une proposition sur une fiche, puis récupère la fiche d'un autre participant et tente d'associer une nouvelle idée à cette idée. Cette méthode a pour avantage entre autres de faciliter l'expression de personnes introverties à l'oral et la formalisation d'idées par des représentations visuelles (schémas, ..).

Les cartes mentales

Il s'agit là encore de noter des associations d'idées, mais contrairement au brainstorming, on établit au fur et à mesure une structuration des idées grâce à une représentation visuelle : la problématique est résumée par un mot, placé dans une bulle au centre de la carte. A partir de ce mot, on note les premières associations d'idées autour de la bulle, puis on note les associations autour de ces mots, en oubliant le contenu de la première bulle, et ainsi de suite [Buzan, 1998].

L'analyse morphologique

La première étape consiste à lister toutes les composantes du système qui posent problème, en déclinant toutes les variantes possibles. Ensuite, on considère le système résultant de la combinaison d'une variante donnée pour chacune des composantes, et on évalue son intérêt. Cette combinaison peut être systématique (on croise toutes les possibilités de chacune des composantes) ou alors faite aléatoirement, par tirage au sort ou autre.

La matrice des neuf écrans

C'est un des outils intégrés dans la démarche d'aide à la résolution de problèmes dans les systèmes techniques TRIZ [Alstshuller, 2002]. Elle reste cependant applicable dans d'autres domaines et est destinée à aider l'utilisateur à réfléchir à l'évolution passée et future de son système. Elle amène en effet l'acteur à prendre conscience puis à analyser les raisons (humaines, techniques, technologiques, etc.) qui ont été à l'origine d'une évolution passée du système visé. Les mêmes concepts, phénomènes ou idées appliqués dans le passé pourraient-ils à nouveau être appliqués au système actuel pour qu'il évolue encore ? Si oui, lesquels et comment ?

Il est ainsi nécessaire de bien positionner le système dans une hiérarchie schématisée Figure 19 :

- vis-à-vis d'un référent de plus haut niveau appelé ici super système dont le système étudié peut-être un composant qui entretient des relations particulières avec lui. A chaque point de vue peut ainsi correspondre un super système donné différent. Par exemple, pour un système 'voiture', on peut considérer aussi bien le super système composé de l'ensemble des acteurs concernés par la fabrication et la maintenance d'un véhicule mais aussi par l'ensemble de l'infrastructure routière (routes, chemins, autoroutes) qui permettent à l'utilisateur d'utiliser sa voiture.
- vis-à-vis des sous-systèmes composants et interagissant au sein du système pour atteindre sa finalité.

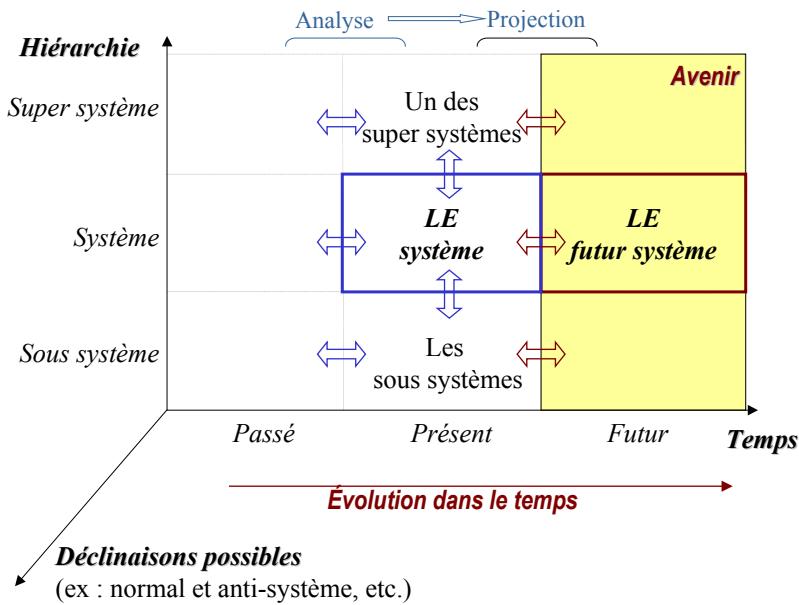


Figure 19 : Principe de la matrice des neuf écrans

Un troisième axe peut être proposé permettant de positionner des déclinaisons possibles du système visé et ainsi de ne pas représenter le système lui-même et isolément, mais aussi son anti-système et d'ouvrir ainsi considérablement le champ exploratoire. L'anti-système se définit alors comme le contraire exact du système selon un point de vue donné c'est à dire ayant un objectif inverse de celui du système ou une fonctionnalité opposée par exemple. Les méthodes antithétiques présentées plus loin utilisent elles aussi ce principe. Dans l'exemple précédent, si le système reste la voiture qui est essentiellement un moyen de transport personnel et familial, un anti-système possible serait un transport en commun.

Une technique d'inventaire systématique

Cette technique permet d'élaborer le besoin fondamental à l'origine de la création ou de l'amélioration, voire de la remise en cause pure et simple d'un système. Elle se décompose en 3 étapes :

- **Rechercher et valider le besoin fondamental du système** : comme le schématisé la Figure 20, cette première étape consiste à élaborer le besoin fondamental du système en répondant à une suite de questions autour desquelles les acteurs se devront d'obtenir un consensus. Il faut tout d'abord positionner le système dans son environnement, lister ses entrées / sorties, l'interaction (l'effet) du système sur cet environnement et la place du système dans une structure plus large. Il faut ensuite lister le plus exhaustivement possible les buts qui paraissent être à l'origine du système, les acteurs auquel ce système est destiné mais aussi ceux avec lesquels ce système interagira (par exemple en les regroupant en fonction de leurs rôles, compétences et attentes) et les différentes caractéristiques et paramètres des éléments constitutifs de l'environnement sur lesquels ce système va agir. La validation de ce besoin se fera alors en précisant à nouveau l'origine du système dont les caractéristiques viennent d'être listées, son but et si le besoin fondamental qui a été retenu sera sujet à évolution dans le futur.

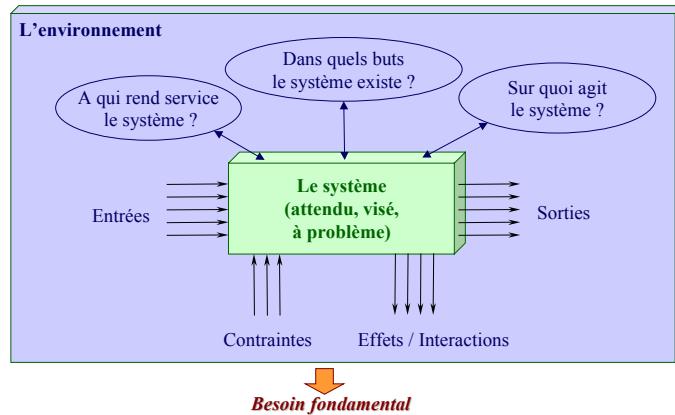


Figure 20 : Recherche du besoin fondamental du système

- Rechercher les éléments extérieurs au système : ces éléments matériels, immatériels ou humains peuvent être en contact avec le système interagir d'une manière ou d'une autre (physiquement, socialement, sentimentalement, etc.). Ce nouvel inventaire doit permettre d'isoler plus finement chaque élément et, en particulier, de décrire leur comportement vis-à-vis du système.
- Rechercher les différentes fonctions assurées par le système et les contraintes auxquelles il doit répondre : certaines techniques vues ci-dessus comme les cartes mentales ou encore un diagramme pieuvre (mettant en évidence le rôle joué par le système sur chacune des interactions entre les éléments extérieurs du système comme illustré dans la Figure 21), peuvent alors être employés pour déterminer les fonctions d'usage du système (à quoi sert-il ?), les fonctions d'estime (comment tester si la fonction d'usage est bien remplie ?) et les contraintes (de déploiement du système, architecturales, de fonctionnement, de sûreté, de sécurité, etc.).

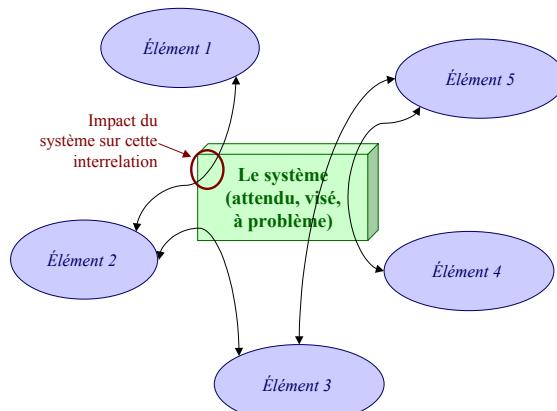


Figure 21 : Exemple de diagramme pieuvre

La recherche d'idées : approche indirecte

Il s'agit dans ce cas d'aider les acteurs à se familiariser avec le problème en utilisant un modèle qui leur permet de s'affranchir de leur domaine de compétences ou de leurs pratiques routinières. Toutes les méthodes analogiques ou empathiques sont bien sûr à positionner à cet endroit de notre liste, mais aussi les méthodes antithétiques, où on va s'efforcer de déformer le problème pour pouvoir changer le point de vue.



Le portrait chinois

Cette méthode amène les participants à identifier l'objet du problème à un animal, une ville, une matière, une musique, etc.. L'idée est de permettre une représentation à la fois plus précise et plus large en faisant appel à des sensations, à des images que peuvent partager les acteurs [Brabandère, 1998].

Les petits hommes

Cette approche repose sur un principe d'empathie introduit par la synectique de William Gordon dans les années 50 et repris par G. Altshuller [Altshuller, 2002]. D'après W. Gordon, un acteur doit pouvoir s'imaginer comme étant lui-même le système à l'origine du problème ou le problème lui-même. C'est, à la base, le changement de point de vue provoqué par cette approche qui peut effectivement permettre d'exprimer ce problème différemment voire de le solutionner. Cependant, il apparaît qu'un acteur humain aussi objectif soit-il, a de nombreuses difficultés pour respecter la réalité des faits et a tendance à gommer, à abstraire une partie des défauts, problèmes ou limitations des systèmes à l'origine du problème. Par exemple, il est difficile de demander à un transporteur routier de s'imaginer à la place d'un camion transportant des matières dangereuses. Le transporteur aura tendance à considérer que l'outil de transport utilisé, en l'occurrence un camion, possède certaines caractéristiques de freinage qui en font un moyen de transport sûr qui peut rouler par tout temps. Ce point de vue est alors préjudiciable pour repenser le problème du transport de matières dangereuses en fonction de la sûreté de fonctionnement de ce camion.

Pour lever cette ambiguïté possible dans l'interprétation, G. Altshuller propose, non pas de se représenter à l'intérieur de l'objet ou de la situation posant problème, mais à les représenter sous forme d'une multitude de petits bonshommes pouvant prendre n'importe quelle forme, taille ou spécificité, pouvant assurer n'importe quel rôle ou n'importe quelle fonction. On ne se considère plus comme une partie de l'objet ou de la situation à l'origine du problème mais comme un chef d'orchestre qui va percevoir le problème à résoudre seulement au travers des yeux de petits bonshommes qui représentent ce problème. Ceux-ci vont en effet sentir, vivre et agir. Il faudra les observer, les guider, leur donner des ordres qu'ils devront exécuter. Il faudra ensuite analyser leur comportement, les effets de leurs actes et accepter ainsi ce filtre déformant entre le problème et nous-mêmes.

Ainsi le transporteur routier pourra représenter le camion comme un ensemble de petits hommes fournissant de l'énergie, un autre assurant le transfert de cette énergie vers un troisième ensemble de petits hommes représentants le contact au sol. D'autres petits hommes peuvent alors représenter les conditions de route (pluie, verglas, etc.) et imaginer comment ces derniers petits hommes vont gêner le fonctionnement normal des trois premières catégories vues ci-dessus.

Les opérateurs Dimension - Temps - Coût (DTC)

Il s'agit ici aussi, d'une approche exigeant de la part de l'acteur une formulation différente de la finalité, des fonctions et du comportement du système et du problème en cause. La démarche consiste à poser et à répondre à six questions mettant en évidence des caractéristiques de temps, de dimension et de coût puis à s'obliger à raisonner sur le nouveau système ou problème perçu dans chacun des cas. Qu'adviendrait-il si ...

- Le système était minuscule ? Le système était immense ? Les réponses permettent d'imaginer un système aux dimensions différentes de ce qu'elles sont dans la réalité et, éventuellement, un système dans lequel le problème n'apparaît plus. Par exemple, considérons l'infrastructure routière qui semblera très rapidement inadaptée en terme de sécurité et de respect de l'environnement aux futurs besoins de déplacements.



Qu'adviendrait-il si une route se réduisait à une simple bande très étroite – auquel cas il paraît aisément d'en rajouter à l'infini mais il reste alors nécessaire de remettre en cause le principe du guidage des véhicules qui pourraient alors ne plus avoir de roues - ou au contraire si l'ensemble de la planète n'était qu'une seule et même route – auquel cas l'environnement en souffrirait ?

- Le système opérait en un rien de temps ? Le système opérait en un temps infini ? Les réponses permettent d'imaginer un système dont les temps de réponse sont très différents de ce qu'ils sont dans la réalité. On obtient alors deux ‘modèles temporels’ à partir desquels il est possible de se poser d'autres questions voire de formuler des hypothèses restreignant le problème originel. Qu'adviendrait-il si nous arrivions à nous déplacer d'un point à un autre en un temps infiniment court – auquel cas tout autre système de transport s'avèrerait obsolète très rapidement – ou, au contraire infiniment long – auquel cas, on peut se demander pourquoi se déplacer soi-même alors qu'un simple système de communication remplacerait avantageusement tout système de transport ?
- Le système avait un coût nul ? Le système était très onéreux ? Les réponses évoquent alors une possibilité de ne pas prendre en compte l'aspect financier, ce qui permet de lever l'inertie qui règne quand aux coûts élevés qu'une solution apparente pourrait entraîner.

Le concassage

Il existe dans notre environnement direct un certain nombre de faits, d'objets ou de situations, c'est à dire des référents, auxquels nous sommes tellement habitués que nous n'y prenons plus garde. Ces choses représentent notre univers et nous ne pensons pas toujours à les remettre en cause implicitement. Or, créer nécessite avant tout d'adopter une attitude de refus de l'existant et de concasser celui-ci de manière à en faire resurgir les bases et hypothèses mêmes qui permettront de réfléchir avec plus de liberté sur un sujet. [Fustier, 2001] propose ainsi de procéder comme suit :

- Représenter le référent en question en distinguant trois facettes :
 - La facette technique : l'ensemble des composants, des matières ou des événements qui une fois structurés ensemble font émerger le référent.
 - La facette fonctionnelle : la description des fonctions et des usages que l'homme a de ce référent
 - La facette sociologique : les circonstances dans lesquelles l'homme en fait effectivement un usage et les événements qui en sont soit la cause, soit un effet.
- Appliquer ensuite au référent un certain nombre de verbes qui permettent alors d'entrevoir complètement différemment le référent en question. Pour plus de facilité, un guide appelé table de concassage regroupe ces verbes et aide l'utilisateur pris isolément ou au sein d'un collectif, à formuler ses impressions. Les verbes à appliquer sont :
 - Augmenter : l'acteur peut chercher à augmenter, améliorer, perfectionner différentes propriétés du référent et faire alors apparaître d'autres référents. Ces propriétés peuvent être le poids, le volume, les dimensions, le prix, le temps d'utilisation, la période d'utilisation, l'intensité, l'usage, l'utilisabilité, la fonctionnalité, la portée sociale, l'intérêt, le sens, la perfection, l'adéquation, l'historique, l'avenir, la fiabilité, la sûreté, la beauté (subjective !), la sécurité, la simplicité, la commodité, la cohérence, la cohésion, la structure.
 - Diminuer : l'acteur appliquera alors le sens contraire de l'augmentation à chacune des propriétés vues ci-dessus.



- Combiner : l'acteur peut alors chercher à faire émerger un nouveau référent en associant son référent avec un autre référent remplissant une fonction amont ou aval, avec un autre référent très voisin (géographiquement, physiquement, structurellement, de situation très proche, etc.) ou au contraire avec un référent qui n'entretient normalement aucune relation avec le référent d'origine.
- Inverser : l'acteur peut alors tenter d'inverser complètement sa vision du référent. Il peut alors utiliser les verbes renverser (cela pourrait consister à mettre le référent tête en bas s'il s'agit d'un objet physique par exemple ou positiver son point de vue face à une situation apparemment catastrophique), inverser une des propriétés vues ci-dessus, prendre une chronologie à rebours, penser de façon remontante et non plus descendante jusqu'à imaginer ou réaliser l'anti-système (comme nous le verrons dans le cas de la matrice des neuf écrans).
- Modifier : l'acteur doit alors changer le référent et/ou son environnement : changer le temps (délai, durée ou période d'utilisation), le lieu (position géographique, événements à l'origine de la cause ou circonstances), l'usage, l'usager, les matières, les formes, les couleurs, les sons, les odeurs, l'énergie, le degré de liberté (mobile / immobile), les composants ou constituants, la configuration à l'origine de la situation elle-même, chercher un autre référent remplissant le même usage ou y ressemblant.,
- Sensualiser : l'acteur doit alors chercher à rendre son référent plus attristant ou plus excitant pour les sens humain (ouïe, odorat, vue, toucher, goût) ou au contraire, ce qui le rendrait moins attristant.

Un exemple d'application au système de transport urbain de type métro pourrait alors permettre d'exprimer des points de vue à partir desquels on pourrait reconSIDérer ce référent :

- Limiter le temps d'arrêt à chaque station des rames
- Augmenter le nombre de places dans une rame
- Diminuer la longueur d'une rame
- Mettre en place des rames réservées au transport de marchandises voire de déchets
- Vaporiser de l'air pur ou chargé d'une certaine odeur dans chaque station en fonction de la situation ou du lieu
- Mettre en musique l'arrivée et le départ des rames
- Multiplier les stations jusqu'à ce qu'elles se touchent quasiment et soient alors accessibles l'une à l'autre par un moyen pédestre sous-terrain
- Assurer le passager d'un temps maximum pour aller d'un point à un autre et non pas l'assurer d'une heure de départ régulière
- Faire un système aérien et configurable dans ses trajets mais utilisant l'énergie électrique tel qu'un bus électrique

Se passer de ...

Les acteurs de la première phase de divergence doivent alors s'obliger à répondre à la question : qu'adviendrait-il si nous devions nous passer du système, de sa fonction, de l'organisation qui l'utilise, voire même qu'adviendrait-il ou que serions-nous devenus si le système n'avait jamais existé [Fustier, 2001]? Par exemple, si l'on se passe d'une voiture, du train ou du réseau de garagistes ? Qu'est-ce qui, au vu et au su de notre société moderne, de



ses besoins, des pouvoirs politiques en place et autre caractéristique, remplacerait l'automobile si personne ne l'avait inventée ?

Le paradoxe

Comme la méthode des neuf écrans qui permet de mettre en avant l'anti-système, cette approche consiste à retourner complètement le problème, c'est à dire à le formuler et à raisonner sur le problème exactement contraire [Gavriloff, 2001]. Par exemple, supposons un décideur désirant obtenir un consensus entre plusieurs acteurs opposés sur la nécessité d'augmenter le trafic ferroviaire d'ici à trois ans de 50%. Le problème pourrait alors être formulé comme suit : « Il faut réduire de moitié le trafic ferroviaire d'ici fin 2005 ! ». L'exercice suivant consistera alors à relever tous les arguments et mettre en avant toutes les incongruités et dysfonctionnements majeurs qu'engendrerait cette solution. Une nouvelle formulation du problème initial peut alors émerger en employant d'autres termes et faisant appel à d'autres points de vue.

Le décalage

Cette approche consiste à ne pas considérer le problème ou la situation de manière frontale et directe pour en rechercher des solutions [Gavriloff, 2001]. Au contraire, elle consiste à contourner cette situation pour essayer de la voir sous un autre angle plus facile, mieux outillé ou mieux documenté à partir duquel on puisse raisonner différemment. Cette méthode s'applique généralement aux problèmes et situations qui ne semblent plus avoir de solutions ou au contraire, avoir des solutions connues mais dont les effets peuvent ne pas être acceptés par tous les acteurs. Par exemple, le coût kilométrique d'un camion deviendra tôt ou tard si important qu'il faudra alors réfléchir à un mode combiné de transport de marchandise. Le décalage permettrait de ne pas s'arrêter à cette première idée de solution et à constater qu'un camion peut être conduit de différentes manières dont certaines sont moins coûteuses, moins polluantes, plus sûres. Ainsi, la formation des jeunes conducteurs ou la reformation des conducteurs actuels rendues obligatoires, pour acquérir de nouvelles techniques de conduite réputées adaptées. On ne travaille plus seulement sur l'aspect technique du camion ou organisationnel d'un futur réseau, mais on tente d'employer un moyen détourné, décalé, pour trouver une piste de solution.

4.5.3. Les phases de convergence

Il s'agit à présent de trier et d'ordonner les idées qui ont émergé lors de la phase de divergence précédente. Le but ultime de cette étape reste l'évaluation des idées, c'est-à-dire la quantification de leur adéquation avec le problème. Il faut pour cela disposer d'une présentation facilement « manipulable » de la somme d'idées issues de la phase de divergence précédente [Brabandère, 1998].

L'arbre conceptuel

Il s'agit d'ordonner les idées selon une arborescence, en dégageant quelques concepts fondamentaux, puis pour chacun de ces concepts, quelques concepts « secondaires » et ainsi de suite, jusqu'à parvenir aux différentes pistes de solutions proposées. Il sera ainsi plus facile par la suite de rapprocher cette liste de solutions d'une liste de problèmes ou de contraintes.

L'abaque de Régnier

Elle facilite l'analyse et la synthèse de données qualitatives en offrant un support visuel. Chacun des acteurs est invité à exprimer son opinion sur une idée grâce à une échelle de couleurs (par exemple du vert signifiant l'approbation totale au rouge indiquant un rejet absolu, et jusqu'au noir du refus de répondre). La mise en commun de ces représentations



colorées constitue une image précise de l'opinion du groupe, pouvant être réactualisée à tout moment et qui permet d'orienter les discussions [Chapuy et al., 1998].

4.6. La résolution de problèmes inventifs : la méthode TRIZ

La méthode TRIZ [Altshuller, 2002 ; Terninko, 1998] a été développée pendant environ 35 ans par un russe spécialiste à l'origine des brevets puis, peu à peu, expert en innovation technique, G. Altshuller. L'acronyme russe TRIZ peut se traduire par théorie de la résolution de problèmes ‘inventifs’. Si elle ne s’applique essentiellement, tout au moins dans la forme de base qui est présentée ci-après, qu’aux systèmes techniques, il est à noter que la philosophie qui s’en dégage a inspiré de nombreux chercheurs et consultants qui se la sont appropriés et l’ont appliquée très largement à d’autres domaines comme nous le verrons en conclusion.

TRIZ est par définition une méthode indirecte et intégrée dont l’idée d’origine est basée sur deux constats issus de l’expérience de l’inventeur, sur l’analyse d’un très grand nombre de documents (quelques millions de brevets, de la bibliographie, etc.):

- Les principes de résolution de problèmes sont très peu nombreux en soi mais ils sont appliqués différemment dans chaque domaine ou chaque corps de métier.
- Une démarche directe, bien que souvent plus rapide et plus utilisée dans le monde industriel qui doit posséder des temps de réaction les plus courts possibles, amène souvent au choix d’une solution de compromis. Celle-ci, bien que solutionnant un problème donné, pose ou posera à terme d’autres problèmes qui devront à leur tour être solutionnés. De plus, la résolution ne fait généralement appel qu’à des domaines de compétences très proches ou identiques à celui du poseur de problème lui-même. Par exemple, un mécanicien aura tendance à rechercher des pistes de solutions dans le domaine qu’il maîtrise et connaît le mieux.

Ainsi, une abstraction du problème est nécessaire pour sortir d’un domaine donné avant de rechercher une solution réelle, c’est-à-dire d’appliquer quelques uns des principes génériques de résolution.

TRIZ fournit alors les outils et concepts nécessaires pour aborder la modélisation du problème, s’extraire très rapidement du domaine réel d’application en levant ainsi tous les freins psychologiques et autres que cela va entraîner et, enfin, interpréter des principes de résolutions génériques pour rechercher des pistes de solution. La démarche proposée se présente comme le montre la Figure 22.

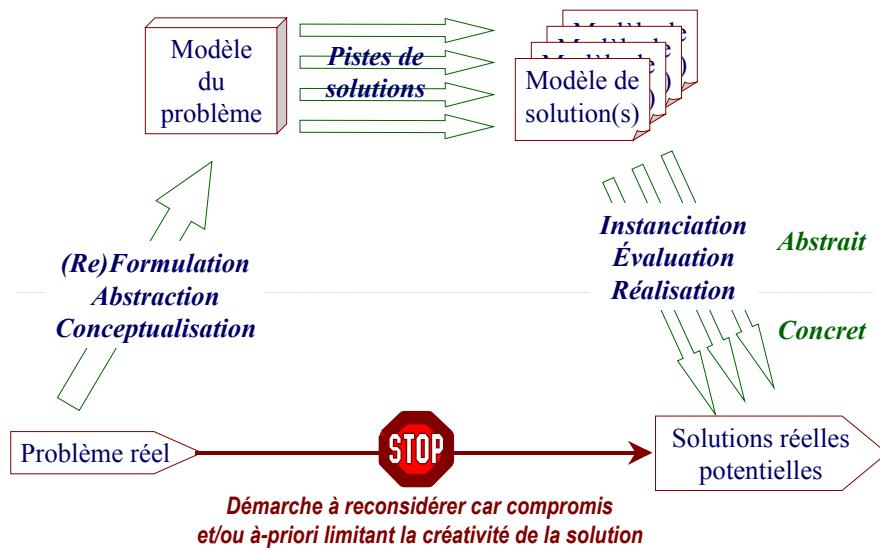


Figure 22 : Vue synthétique de la démarche

On voit alors apparaître quatre classes d'outils :

- Outils de levée d'inertie psychologique : les petits hommes, la matrice neuf écrans et les opérateurs DTC qui ont été vus plus haut.
- Outils et concepts retenus pour la modélisation des problèmes : contradictions et vépoles présentés ci-dessous.
- Outils d'aide à l'interprétation des modèles obtenus et au raisonnement sur des solutions : la matrice de résolution des contradictions et les standards de résolution présentés dans la suite.
- Outil méthodologique : il s'agit de l'algorithme ARIZ dont l'objectif est d'aborder de manière systématique la résolution de problèmes particulièrement difficiles.

Ces outils sont alors utilisés comme le montre la Figure 23.

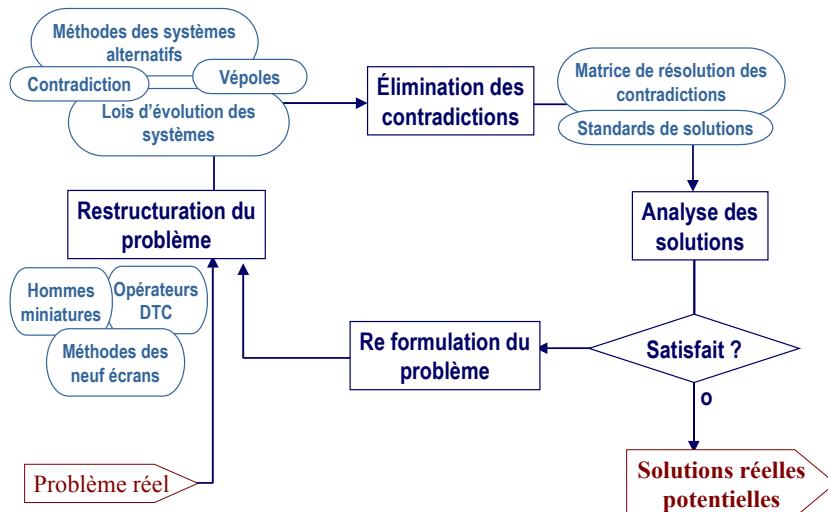


Figure 23 : Une vision simplifiée de la démarche

Les caractéristiques essentielles de ces outils sont synthétisées dans le Tableau 3.



TRIZ	Outil	Principe	Composé de
Formuler le problème et commencer à le résoudre : Vaincre l'inertie psychologique	Petits hommes	Modéliser simplement le problème (synectique)	Modèle graphique
	Méthode des neufs écrans	Système vis-à-vis de son environnement, de ses composants et de leur évolution dans le temps	Matrice 3x3
	Opérateurs Dimension-Temps-Coûts (DTC)	Forcer l'utilisateur à se poser les bonnes questions	6 questions de base
Modéliser et résoudre le problème	Matrice d'Altshuller (de résolution des conflits)	Résoudre les Contradictions Techniques	39 paramètres de conception 40 principes d'innovation
	Méthode de résolution des contradictions physiques	Résoudre les Contradictions Physiques	11 principes de résolution
	Méthode des systèmes alternatifs	Coupler les avantages de plusieurs systèmes pour résoudre un problème	Table de comparaison
	Vépoles	Modélisation de problème par Substance/Champs	6 types de champs 5 classes de vépoles
	Standards	Résolution de problèmes par réduction des vépoles	76 standards de résolution
Comment procéder pour utiliser les autres outils ?	ARIZ	Algorithme	9 étapes

Tableau 3 : Les caractéristiques des outils de TRIZ

4.6.1. Principes

La méthode TRIZ repose sur quelques principes de base.

Le premier définit ce qu'est un système technique (ST) afin de limiter les prétentions de la méthode en terme de résolution de problèmes. Il s'agit bien, dans sa forme originale, d'une méthode dédiée à l'aspect technique et technologique des systèmes. Ainsi, un système technique est défini comme l'interrelation de composants qui assurent chacun une fonction particulière au sein du système mais partagent une même finalité. Ces éléments, décrits dans la Figure 24, sont :

- L'élément moteur qui transforme une énergie externe en une nouvelle forme d'énergie utilisable en interne pour les besoins du système
- L'élément de transmission qui assure, si possible sans pertes, la transmission de cette énergie interne vers l'organe effecteur
- L'élément de travail qui assure la fonction essentielle du système et lui permet ainsi de remplir sa fonction utile
- L'élément de contrôle qui assure, en fonction d'objectifs de pilotage et de contraintes à respecter, la cohérence de l'ensemble du système et la synchronisation des divers éléments entre eux.

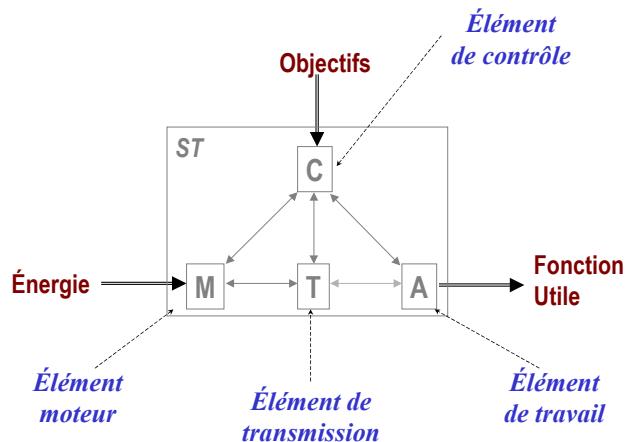


Figure 24 : Le système technique

Le deuxième principe consiste à observer le système technique dans le temps. Chaque système qui nous entoure a évolué, s'est transformé, s'est amélioré et s'est adapté à de nouveaux besoins et à de nouvelles technologies au cours du temps. Nous sommes ainsi passés d'un mode de propulsion des trains à vapeur aux moteurs à explosion Diesel puis à propulsion électrique. Cependant, nous pouvons constater que cette évolution dans le temps devrait toujours respecter un certain nombre de lois (dites lois statiques, lois dynamiques et lois cinématiques) exprimées hors de tout champ d'application ou de tout domaine d'expertise. La méthode TRIZ préconise alors d'appliquer ou de garder à l'esprit ces lois au cours du processus de résolution. En effet, contrevenir à l'une d'elles peut signifier pour le système une perte de performance, une inadéquation vis-à-vis d'un besoin, une évolution imparfaite ou encore remettre en cause sa conception voire une partie de lui-même.

Le troisième principe permet de penser différemment de manière à ne pas se précipiter vers des solutions de simplicité et donc très souvent de compromis susceptibles de poser d'autres problèmes plus tard. Est-ce que le fait de mettre des ceintures de sécurité à l'arrière des véhicules a résolu le problème des accidents mortels sur la route ? Si cela a effectivement permis d'en diminuer le nombre, cela a cependant provoqué l'augmentation, même minime, du coût du véhicule, la remise en cause de sa conception, etc.

TRIZ propose alors de réfléchir en terme de Résultat Idéal Final (RIF) c'est à dire de pousser les acteurs du processus à formuler une solution ou un nouveau système :

- Qui n'a pas de coût,
- Qui n'a pas de volume,
- Qui n'a pas de surface,
- Qui maximise sa capacité de travail,
- Qui maximise ses fonctionnalités.

Cette formulation permet alors de s'abstraire plus facilement de la réalité et de percevoir le problème en laissant de côté toute référence à la machine, au processus ou à la situation étant le siège de ce problème. Le but peut être, par exemple, de faire disparaître l'objet automobile tel qu'on le perçoit et le côtoie chaque jour du conscient des acteurs du processus. Ces derniers se représenteraient plutôt un système de transport de personnes et de biens qui ne coûterait rien à utiliser, qui ne nécessiterait pas de place pour se mouvoir ni pour s'arrêter, qui assurerait une parfaite sécurité des usagers comme de l'environnement et qui resterait disponible tout le temps.



4.6.2. Levée d'inertie psychologique

Formuler le problème peut nécessiter, comme nous l'avons vu plus haut, de lever l'inertie psychologique dont font preuve les acteurs.

Si les quelques règles de base et les quelques techniques vues précédemment ne permettent pas de s'abstraire de la réalité pour percevoir et obtenir une formulation suffisamment claire et précise du problème à traiter, TRIZ préconise alors aux acteurs l'utilisation spécifique des petits hommes, des opérateurs DTC ou encore de la matrice neuf écrans.

4.6.3. Modélisation du problème

L'approche TRIZ repose sur le constat que tout problème dans un système dit ‘technique’ vient d'une contradiction entre certaines propriétés caractérisant les éléments (vus alors comme des composants) ou certains phénomènes se déroulant dans ce système.

Une contradiction s'exprime par le fait que ces propriétés ou ces phénomènes s'opposent (par valeurs ou par leurs effets) à un instant ou à un endroit donné du système. Par exemple, un avion doit être le plus grand possible pour accueillir le plus grand nombre de passagers mais être aussi le plus petit possible pour ne pas nécessiter trop de carburant en vol. La même propriété de dimension de l'appareil doit donc prendre deux valeurs opposées : c'est ce que l'on appelle contradiction physique. De même, la vitesse d'un véhicule peut augmenter mais cela se fera au détriment de la sécurité des passagers et du respect de l'environnement (pollution plus importante dans les conditions actuelles d'existence des moteurs à explosion en tout cas). Les deux propriétés vitesse et sécurité ou respect de l'environnement sont maintenant opposées deux à deux. Si l'une d'elles s'améliore, ce sera toujours au détriment de l'autre. Ce phénomène s'avère être réversible et l'on parle alors de contradiction technique.

TRIZ propose alors de projeter les propriétés mises en cause dans la formulation de la contradiction sur un ensemble de 39 paramètres usuels dits ‘de conception’ dont la liste est donnée dans le Tableau 4. Par exemple, la vitesse d'un véhicule est associée directement au paramètre n°9 (vitesse) et la sécurité des passagers, interprétable de façons différentes, pourrait être associée aux paramètres n°15 (longévité d'un objet mobile), n°22 (gaspillage d'énergie), n°23 (gaspillage de substance), n°24 (perte d'information), n°26 (quantité de substance) ou encore n°30 (facteurs nuisibles agissant sur l'objet).

1 Masse d'un objet mobile	21 Puissance
2 Masse d'un objet immobile	22 Gaspillage d'énergie
3 Longueur d'un objet mobile	23 Gaspillage de substance
4 Longueur d'un objet immobile	24 Perte d'information
5 Surface d'un objet mobile	25 Perte de temps
6 Surface d'un objet immobile	26 Quantité de substance
7 Volume d'un objet mobile	27 Fidélité
8 Volume d'un objet immobile	28 Précision de la mesure
9 Vitesse	29 Précision de l'usinage
10 Force	30 Facteurs nuisibles agissant sur l'objet
11 Tension, Pression	31 Facteurs nuisibles annexes
12 Forme	32 Usinabilité
13 Stabilité de l'objet	33 Facilité d'utilisation
14 Résistance	34 Aptitude à la réparation
15 Longévité d'un objet mobile	35 Adaptabilité
16 Longévité d'un objet immobile	36 Complexité de l'appareil
17 Température	37 Complexité de contrôle
18 Brillance	38 Degré d'automatisation
19 Energie dépensée par l'objet mobile	39 Productivité
20 Energie dépensée par l'objet immobile	

Tableau 4 : Les 39 paramètres de conception



Un autre outil de modélisation, appelé Vépoles (ou SuField ou encore substance-champs), permet à l'acteur de représenter un système comme un ensemble de composants (ou substances) en présence, possédant chacune des caractéristiques propres. Chaque substance peut être la source d'un effet physique, mécanique, ou autre (champs) qui peut provoquer une modification de certaines des caractéristiques des autres substance. Chaque interaction peut alors être qualifiée d'utile (donc devant être préservée : un pneu supporte la voiture), nuisible (devant donc être détruite ou annihilée d'une manière ou d'une autre : un pneu abîme la route) ou insuffisante (devant donc être renforcée : la route guide le pneu). Il est important de noter qu'un même résultat de modélisation de problème au moyen de vépoles peut effectivement correspondre à plusieurs problèmes très différents les uns des autres, voire même appartenir à des domaines différents. La Figure 25 montre la modélisation unique de deux problèmes sans aucun rapport entre eux. On démontre ainsi l'indépendance des vépoles de tout domaine réel.

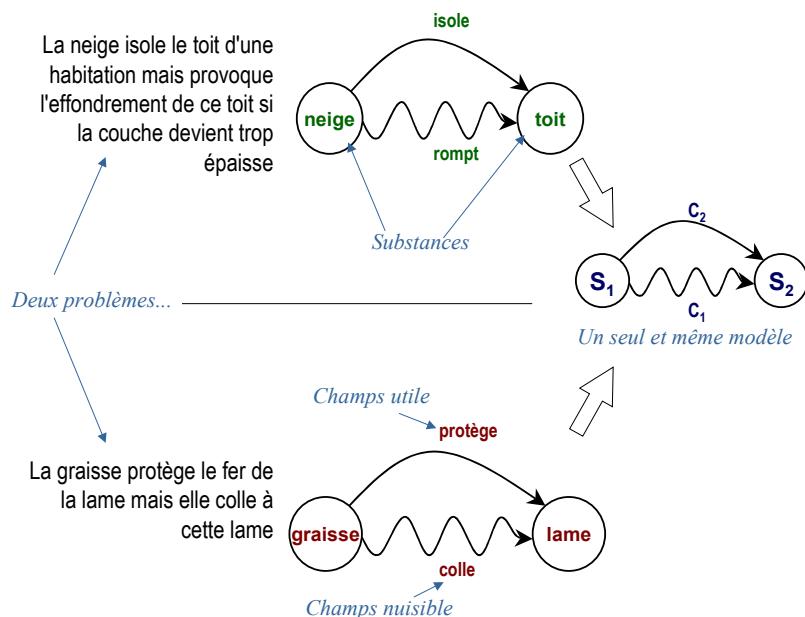


Figure 25 : Exemple de modélisation vépole

Une fois le système modélisé, il est alors nécessaire de chercher des solutions permettant d'annuler tous les effets nuisibles tout en respectant les effets utiles et en renforçant si possible les effets insuffisants.

4.6.4. Rechercher des pistes de solution

TRIZ propose des outils d'aide au solutionnement adaptés au type de modélisation employé : contradiction technique, physique ou vépole.

La matrice de résolution des contradictions techniques

Une analyse statistique du contenu d'un très grand nombre de brevets existant à l'époque a permis de mettre en relation certains types de problèmes à résoudre et le type de solutions qui ont été généralement retenues. Le type de problème est alors représenté par la mise en opposition deux à deux des paramètres de conception vus ci-dessus. Le type de solution est alors représenté par une liste de 40 principes génériques de résolution indépendants de tout domaine applicatif. Ces principes de résolution sont donnés Tableau 5.



1 La segmentation	21 L'action rapide
2 L'extraction	22 La transformation d'un plus en moins
3 La qualité locale	23 L'asservissement
4 L'asymétrie	24 L'intermédiaire (l'insertion)
5 Le groupement (la combinaison)	25 Le self service
6 L'universalité	26 La copie
7 Le placement interne (la poupée russe)	27 L'éphémère et l'économique
8 Le contrepoids	28 La substitution d'un système mécanique (la reconception)
9 L'action opposée préventive	29 Le système hydraulique ou pneumatique
10 L'action préliminaire	30 La membrane flexible et le filù de faible épaisseur
11 La prévention (la compensation)	31 La porosité d'un matériau
12 L'équipotentialité	32 La changement de couleur
13 L'inversion	33 L'homogénéité
14 La sphéricité	34 Le rejet et la régénération des parties
15 Le dynamisme (le degré de dynamisme)	35 Le changement des paramètres physiques et chimiques
16 L'action partielle ou excessive (Le surplus)	36 La transition de phase
17 Le changement de dimension	37 L'expansion thermique
18 La vibration mécanique (l'oscillation)	38 L'oxydation
19 L'action périodique	39 L'environnement inerte
20 La continuité d'une action utile	40 Le matériau composite

Tableau 5 : Les 40 principes de résolution

La mise en opposition deux à deux de chacun des paramètres de conception et leur mise en relation avec un ou plusieurs des principes de résolution applicables à chaque couple prend alors la forme d'une matrice de contradiction dont le processus d'utilisation est représenté sommairement dans la Figure 26.

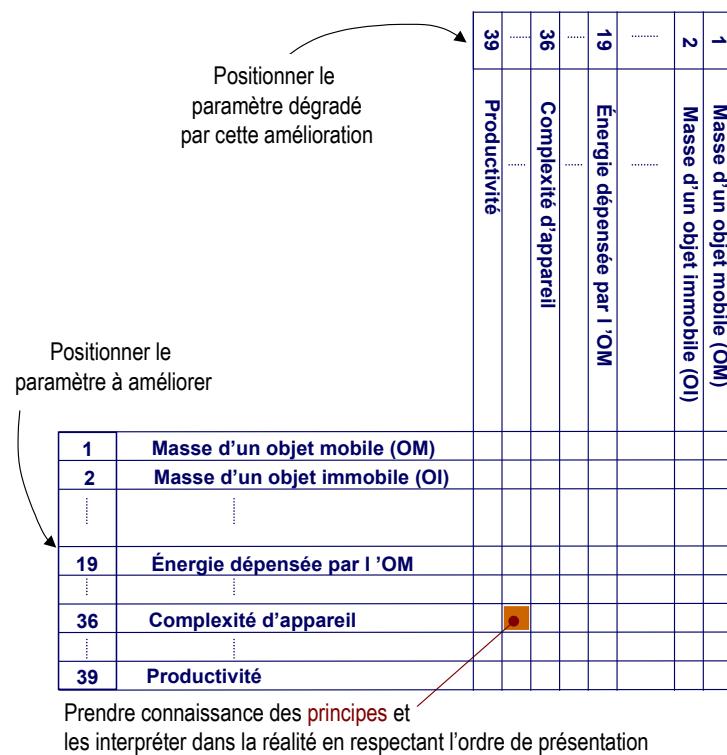


Figure 26 : La matrice des contradictions ou Matrice d'Altshuller

Reprenons l'exemple vu précédemment concernant la contradiction technique opposant la vitesse d'un véhicule à la sécurité des passagers en supposant que le paramètre sécurité soit ici interprété comme un ensemble de facteurs nuisibles (pouvant alors nuire aux passagers à bord du véhicule, au conducteur ou aux autres usagers de la route). La Figure 27 montre alors la partie de cette matrice y correspondant.



		Worsening Feature →	Volume of moving object	Paramètre détérioré					Changement d'un paramètre → harmful factors	Ease of operation	Ease of repair	Device complexity	Difficulty of detecting and measuring
		Improving Feature ↓		Speed	Force (dynamy)	Stress or pressure	Shape	Volatility					
			7	9	10	11	12	27	31	33	34	36	37
9	Speed	7, 29, 24	13, 28, 15, 19	6, 18, 26, 40	35, 24, 15, 24	27, 26, 18, 21,	33, 21, 10, 35,	2, 24, 9, 35,	32, 28, 13, 12	24, 2, 26, 27	10, 28, 4, 24	3, 34, 27, 16	
	Paramètre à améliorer		15, 9, 2, 37	13, 28, 13, 12	+	18, 21, 11	10, 35, 40, 24	9, 35, 15, 21	18, 3, 30, 24	1, 28, 3, 23	15, 1, 11	26, 35, 10, 18	36, 37, 10, 19
11	Stress or pressure	6, 35, 10	6, 35, 26	34, 35, 21	+	35, 4, 13, 10	10, 13, 19, 35	2, 33, 27, 16	11	2	19, 1, 33	2, 36, 37	
12	Shape	14, 4, 15, 22	35, 13, 24, 18	33, 10, 37, 40	34, 13, 10, 14	+	18, 40, 18	28, 1	32, 33, 26	2, 13, 1	16, 29, 1, 28	15, 13, 29	
15	Duration of action of moving object	30, 2, 19, 30	5, 35, 9	19, 2, 16	19, 3, 27	14, 20, 28, 25	11, 1, 16, 22	21, 39, 1	12, 27, 27	29, 30, 29, 15	30, 4, 29, 29		39, 29
33	Ease of operation	3, 19, 35, 15	18, 13, 34	26, 13, 35	2, 32, 13, 34	17, 27, 8, 30		+	32, 29, 1, 32	32, 26, 12, 17			39, 35

Principes à appliquer classés statistiquement

Figure 27 : Extrait de la matrice

Les principes à mettre en avant qui devront alors être interprétés dans la réalité (techniquement faisable ou pas, ce qui permet de juger de la créativité de celui-ci) par l'acteur sont donnés dans le Tableau 6, classés en fonction de leur fréquence d'utilisation comme suit :

Principes	Aide à la compréhension et première interprétation possible (liste évidemment non exhaustive)
2 L'extraction	Enlever ou séparer de l'objet la partie perturbatrice ou Enlever seulement la partie ou propriété nécessaire - limiter drastiquement la vitesse par construction - ne plus mettre de passagers à bord - entourer les passagers d'un système de protection (airbags par exemple, etc.)
24 L'intermédiaire (l'insertion)	Utiliser un objet intermédiaire pour transmettre ou transférer l'action Joindre temporairement à l'objet un autre objet facile à éliminer - mettre des absorbeurs de chocs - diffuser autour de la voiture un champs magnétique protecteur, etc.
35 Le changement des paramètres physiques et chimiques	Modifier l'état de phase de l'objet, sa concentration, sa consistance, son degré de flexibilité, sa température - pourquoi de pas faire des voitures souples pouvant encaisser les chocs avec déformation maximale mais absorption de l'énergie ?
21 L'action rapide	Réaliser le processus ou certaines de ses étapes (nuisibles ou dangereuse) à très haute cadence - faire des autoroutes pour véhicule auto-guidés mais circulant à très grande vitesse sans intervention humaine

Tableau 6 : Application des principes à l'exemple



Les 11 méthodes de résolution des contradictions physiques

Lorsque la contradiction est d'ordre physique (la même caractéristique doit posséder deux valeurs opposées) et non technique comme ci-dessus, le principe de résolution est un peu différent.

En effet, il est peut-être souhaitable de séparer dans le temps cette contradiction : la caractéristique aura effectivement deux valeurs opposées mais pas au même moment. Par exemple, un avion décollant et lourdement chargé pourrait avoir une très grande surface portante pour pouvoir s'arracher du sol mais diminuer cette surface lorsqu'il atteint son altitude de croisière afin de diminuer sa traînée et donc sa consommation de carburant. Cette solution est maintenant connue et mise en œuvre de plusieurs façons différentes : voilure rétractable, voilure à géométrie variable, ré injection de gaz pris en sortie des réacteurs vers les surfaces de fuite des ailes, etc.

Il est aussi possible de séparer cette contradiction non pas dans le temps mais dans l'espace : ce n'est effectivement pas au même endroit du système qu'une même caractéristique doit posséder deux valeurs opposée. Les autres principes, dont il serait un peu long ici de présenter les caractéristiques, sont donnés dans le Tableau 7.

Division des propriétés contradictoires	
1	Dans l'espace
2	Dans le temps
Transition des systèmes (1)	
3	Assemblage des systèmes homogènes ou hétérogènes en un supersystème
4	Transition d'un système à un antisystème
5	Le système est doté de la propriété C, ses parties aussi
Transition des systèmes (2)	
6	Passage à un système fonctionnant au micro niveau
Transition de phase	
7	Remplacement de l'état de phase d'une partie du système
8	Etat de phase 'dualiste' d'une partie du système (en fonction des conditions de travail)
9	Utilisation des phénomènes accompagnant la transition de phase
10	Remplacement d'une substance à une phase par une substance à deux phases
Transition physico-chimique	
11	Apparition / Disparition de la substance suite à la décomposition / combinaison

Tableau 7 : Les 11 principes de résolutions des contradictions physiques

Les 76 standards de solutions

Enfin, une modélisation sous forme de vépoles nécessite une troisième classe de principes de résolution qui sont alors appelés Standards. Au nombre de 76, il ont pour objectif de donner des idées concernant la simplification, la modification ou la destruction (dans le cas de champs nuisibles par exemple) des vépoles qui modélisent le problème. Le principe appliqué et le nouveau vépole alors obtenu doivent être interprétés par l'acteur une fois de plus dans la réalité physique.

Les systèmes alternatifs

Il peut être difficile de tenter de formaliser directement un problème. Il peut alors être intéressant de comparer les systèmes ou les solutions existantes afin d'en détecter les points faibles (les problèmes que chacun d'eux posent), les points forts (le résultat atteint). La notion de Résultat Idéal Final pousse alors l'acteur à tirer de cette comparaison un certain nombre d'idées pouvant l'amener à réfléchir sur une solution qui posséderait alors tout les avantages des solutions existantes sans pour autant posséder leurs inconvénients. Cette approche, baptisée Méthode des systèmes alternatifs, se base sur l'utilisation pure et simple d'une



matrice où l'ensemble des systèmes et solutions existantes est portée en abscisse et leur avantages/inconvénients portés en ordonnée comme le montre l'exemple du Tableau 8.

	voiture	camion	métro
plus	Rapide Manoeuvrable	Capacité de transport	Indépendance des engorgements en surface Capacité Peu de pollution
moins	Quantité	Pollution Risques Manoeuvrabilité	Adapté seulement au transport de passager Infrastructure lourde
intérêt	Rapide Manoeuvrable	Capacité de transport	Peu de pollution

Tableau 8 : Exemple de matrice pour l'exploitation de systèmes alternatifs

L'acteur sera alors à même de choisir les points forts de chaque solution, de raisonner sur les raisons qui font que cette solution possède cet attrait particulier (cela vient-il des matériaux utilisés ? de leur forme, de leur organisation dans l'espace ? etc.) et des raisons qui font que cette solution ne correspond pas à son résultat idéal final. Poser cette question revient alors à sélectionner dans chaque solution existante les parties pertinentes et de les associer, les combiner ensuite pour converger vers une nouvelle solution. L'exemple montré dans le Tableau 8 pourrait par exemple amener l'acteur à se poser la question : qu'est ce qui pourrait être manœuvrable, peu polluant et avoir pourtant une bonne capacité de transport sans pour autant nécessiter de grosses infrastructures ? Par exemple, les améliorations sur les ballons dirigeables ont fait de cet outil de transport une solution à examiner.

4.6.5. Conclusion

Une méthode dans la méthode

TRIZ n'est pas une méthode simple au premier abord bien qu'elle soit supposée aider l'acte créatif. Il est vrai que la plupart des cas à traiter dans le domaine technique ne nécessitent pas toujours d'utiliser tous les outils. Mais que faire s'il s'agit d'un problème relativement hermétique et difficile à formuler, à détailler, à comprendre puis, évidemment, à solutionner ?

Il existe alors un algorithme baptisé ARIZ qui est en fait une méthode au sein de la méthode TRIZ. Cet algorithme guide l'acteur dans sa démarche et lui impose plus ou moins des passages obligés avant d'utiliser à proprement parler les outils de TRIZ vus ci-dessus.

Apport possible de TRIZ

Une telle approche est-elle généralisable à d'autres domaines que le technique ? Oui certainement... A titre d'exemple, l'idée même de contradiction entre deux propriétés décrivant un système complexe, vue comme une cause potentielle de problème, a été largement reprise et mise en œuvre depuis peu par des chercheurs et des consultants travaillant dans des secteurs autres que techniques.

On voit alors apparaître différentes approches basées sur la même philosophie que TRIZ mais dédiées à l'organisation en entreprise, à l'éducation, à la communication, aux systèmes agro-alimentaires, à la maîtrise des risques ou encore aux systèmes d'information [Mann, 2002 ; Verduyn, 2000 ; Winkless, 2000 ; Mann, 1999 ; Rea, 1999a ; Rea, 1999b].



Dans chacun de ces nouveaux domaines, il apparaît alors très intéressant de rechercher et de formuler ces contradictions au cours de la phase d'abstraction et de les projeter sur d'autres types de paramètres de conception. Les moyens utilisées et formalismes usités sont alors bien évidemment différents de ceux proposés par TRIZ mais s'en inspirent largement.

4.7. L'innovation

Les étapes d'acquisition de connaissances sur un contexte donné et de choix entre les alternatives fournies entraînent la nécessité de comprendre les mécanismes de l'innovation. Les pouvoirs publics ont en effet un rôle important à la fois comme partenaires (par l'intermédiaire des aides à l'innovation), comme acteurs (grâce aux entreprises publiques) mais aussi comme initiateurs (par exemple, par l'effet des réglementations). Il s'agit alors de pouvoir tirer des enseignements de ce qui a déjà été fait, et ensuite de se donner les moyens de faire fonctionner ce que l'on propose.

Nous allons préciser quelques notions à ce sujet.

L'innovation peut être vue comme l'ensemble des processus allant de l'invention ou de la découverte d'un produit nouveau jusqu'à sa diffusion et son utilisation dans le tissu social. Elle permet le passage entre deux mondes : celui de la découverte ou de l'invention, et celui du marché, avec ses contraintes inhérentes.

Le résultat d'une innovation est donc un changement concret de la réalité, alors que la créativité a pour résultat un changement de la « perception » de la réalité.

L'innovation est alors une notion qui recouvre celles d'invention ou de découverte, mais qui les dépasse sur deux plans :

- La dimension économique fait partie intégrante de l'innovation : on peut dire qu'une innovation est une invention qui a « réussi », c'est-à-dire qui, après diffusion, a été acceptée par son marché.
- Il ne s'agit pas d'un événement ponctuel que l'on peut dater, mais d'une succession d'étapes. La notion d'invention est donc souvent étudiée en termes de « trajectoire ». Cette trajectoire peut être de longue durée. Par exemple, la généralisation de l'utilisation de la charrue a été effective plusieurs siècles après l'invention de la charrue proprement dite. Les différentes étapes correspondent aux étapes de l'appropriation par le tissu social. Cette appropriation doit être vue comme un mécanisme collectif.

On distingue classiquement les innovations de produit (mise à la disposition des utilisateurs d'un produit ou d'un service nouveau) et celles de process (mise en œuvre d'une nouvelle façon de faire ou amélioration, optimisation d'un facteur de production) [Broustail, 1993]. Cependant, cette distinction peut souvent s'avérer artificielle et peu adéquate, lorsque l'on observe une interdépendance de ces deux mécanismes. Pour être complet, il faut de toute façon y rajouter la notion d'innovation organisationnelle.

Une innovation peut être une rupture brutale dans les habitudes, correspondre à un changement radical des moyens de production, des structures d'une organisation ou des habitudes des consommateurs. On parle alors d'« innovation majeure ».

Elle peut être aussi une modification locale, sans remise en cause de l'ensemble d'un système : dans ce cas, il s'agit d'une « innovation incrémentale » ou « mineure ». D'une façon tout à fait générale, le poids économique des innovations mineures est souvent largement supérieur à celui des grands bouleversements entraînés par les innovations majeures.



4.7.1. Modélisation des processus d'innovation

Sociologues et économistes se penchent sur le problème de la modélisation de l'innovation [Perrin, 2001]. Jusque dans les années 80 prévalait un modèle linéaire de type « poussé par la science » : une découverte scientifique permet une invention (c'est à dire la solution à un problème concret) qui après une phase de production et de diffusion devient ensuite une innovation. Ce modèle s'inscrivait dans la continuité des travaux de Schumpeter, qui fut l'un des premiers économistes à s'intéresser à ce sujet.

Ce modèle a l'avantage d'être simple, et également de fournir une justification économique à la recherche scientifique. Il reste souvent bien adapté pour la description d'innovations majeures, dans lesquelles la part de la science est importante.

Mais il ne permet pas de prendre en compte la diversité des processus d'innovation, en particulier lorsque l'innovation apparaît en réponse à un besoin. C'est le cas par exemple des innovations « mineures » qui ne modifient pas en profondeur le tissu social, mais s'y insèrent facilement. Dans ce cas, la demande et le marché peuvent effectivement jouer un rôle moteur, bien plus que d'éventuelles avancées techniques.

Les modèles actuellement utilisés proposent une approche multiple. Par exemple, celui de Kline et Rosenberg [Perrin, 2001] met en évidence les interactions et rétroactions entre le marché, les activités de conception, de production et les activités de recherche. Une innovation « réussie implique une confrontation permanente entre les points de vue techniques mais aussi économiques, commerciaux, sociaux ».

Ce modèle insiste sur l'existence de plusieurs processus d'innovation, interconnectés. L'accumulation de connaissances grâce à la recherche scientifique est essentielle, voire nécessaire, mais ne peut pas être considérée comme l'étape initiale.

D'autre part, le modèle doit prendre en compte la dimension dynamique d'une innovation : au cours du temps, à la fois le produit (ou le service) évolue grâce aux avancées scientifiques et techniques, mais aussi son marché, ses clients potentiels, ses concurrents.

4.7.2. L'innovation dans les services

Ce sujet a fait l'objet de relativement peu de travaux par rapport au nombre de ceux dédiés à l'innovation au sens industriel [CGP, 1999].

Une des raisons est sans doute le flou qui entoure la notion même de « service », souvent défini uniquement par la négative : les services sont tout ce qui n'est pas agriculture ou industrie. On peut donc y trouver beaucoup de choses, des activités financières à l'éducation en passant par les administrations et bien sûr, les transports.

Cependant, les quelques études théoriques ou enquêtes qui y ont été consacrées permettent de faire ressortir quelques spécificités :

- Il est souvent difficile de séparer les phases de conception, production, commercialisation, et les entreprises de services sont rarement structurées avec un service de R&D bien identifié.
- Les projets sont le plus souvent initiés par la demande plutôt que « poussés par la science ».
- La part d'innovation organisationnelle entièrement non technologique est plus importante que dans l'industrie.



Ainsi, dans le domaine du transport routier de marchandises [Djellal, 2001], où des enquêtes ont mis en évidence plusieurs types d'innovation :

- Technologiques matérielles
- Technologique informationnelles
- Organisationnelles

Ces dernières sont loin d'être négligeables, et vont jusqu'à l'apparition de nouveaux métiers.

4.7.3. Le management de l'innovation

La capacité à innover d'une organisation est améliorée lorsque les mécanismes d'acquisition et de transformation de connaissances y sont mieux compris, exploités et mis en valeur. La notion d'« entreprise apprenante » est désormais au cœur des études sur l'innovation. Cette notion ne doit pas limiter à sa composante technologique représentée par les systèmes d'information. Elle recouvre également une composante stratégique : la gestion du portefeuille de connaissances et une composante organisationnelle. Ces trois composantes doivent être prises en compte simultanément si l'on désire réellement améliorer la capacité d'apprentissage collectif. Il faut enfin s'intéresser à l'identité même de l'organisation, à sa culture, à l'orientation des relations entre ses membres [Métais, 2001].

Chacun de ces points peut être utilisé efficacement comme facteur d'amélioration. Mais on constate que, mal employés, ou poussés à l'extrême, ils peuvent aussi conduire à des situations de blocage.

Les systèmes d'information

Parallèlement à l'augmentation des potentialités matérielles rendues possible par les avancées technologiques, on constate la tendance à l'intégration toujours plus poussée de ces systèmes dans l'entreprise, et au-delà dans son environnement.

Ces systèmes constituent certes un formidable outil de diffusion et de partage de connaissances. Le risque peut venir justement d'un excès de confiance dans leur utilisation. De par leur nature, ils ne permettent de traiter que les informations explicitées. L'erreur est alors de croire que grâce à eux, le contrôle complet des connaissances de l'organisation est possible, en négligeant la part non formalisée de ces connaissances, qui est souvent très importante.

Les structures innovantes

L'innovation n'est pas le fruit du hasard. Une structure innovante est une structure qui favorise la germination d'idées nouvelles, et qui est capable d'en assurer le développement.

La structure même d'une organisation ou d'une entreprise n'est pas sans influence sur sa capacité à innover : elle peut l'augmenter, ou bien la freiner considérablement.

Il n'existe pas de modèle de structure idéale, unique. Mais il existe quelques règles très générales, dont l'application peut améliorer la capacité d'innovation.



Le modèle de Burns et Stalker fait la distinction entre les structures « mécanistes » et « organiques » [Broustail, 1993]

Les premières correspondent généralement à une structure fonctionnelle (les spécialités sont regroupées par fonction, par exemple : le marketing, la production, ...) ou à une structure par divisions (où la spécialisation se fait plutôt par produit). Dans ce cas, les activités sont fractionnées en tâches élémentaires, la structure est fortement hiérarchisée, centralisée et formalisée. Ce type de structure est a priori contraire à l'expression et à la circulation des idées nécessaires à toute innovation. Seules les innovations mineures, incrémentales, peuvent se réaliser.

Les structures organiques sont au contraire fortement décentralisées et adaptables, faiblement hiérarchisées. La coordination entre les différents acteurs ne nécessite pas de formalisation très poussée et se fait plutôt par ajustement mutuel. Un exemple de cette structure est la structure par projets, où les équipes de travail sont constituées de façon temporaire pour répondre à un objectif précis et limité dans le temps. L'intérêt de cette structure est de favoriser les contacts et les échanges entre des personnes évoluant dans des univers de savoirs différentes.

Par principe, ce type de structure est celui qui semble le plus innovant. Mais dans son principe même, il ne facilite pas la capitalisation de l'expérience, et donc par là, ne permet pas un développement optimal de la connaissance.

En réalité, si les phases de génération d'idées y sont effectivement facilitées, il n'est cependant pas très adapté aux phases de mise en œuvre, plus routinières. Il peut conduire à une certaine dilution de la responsabilisation, et concourir à complexifier les circuits de décisions.

En fait, une organisation pouvant être à la fois « mécaniste » et « organique » en fonction de l'étape du processus d'innovation est ce qui se rapproche le plus d'une structure optimale.

La gestion du portefeuille de connaissances

On s'accorde à reconnaître les connaissances disponibles au sein d'une entreprise comme une source d'avantages concurrentiels très importante. La volonté d'innover doit donc s'accompagner à la fois :

- d'une cartographie précise des connaissances disponibles à un instant donné, afin de pouvoir par exemple mettre en place des synergies efficaces et originales.
- du choix des connaissances à acquérir à plus ou moins long terme : il s'agit d'une décision stratégique, qui conditionne l'évolution de l'offre proposée par l'entreprise.

Cette gestion de biens « immatériels » n'est pas facile. Les connaissances ne sont pas toujours explicites, formalisées et directement utilisables dans un système d'informations. D'autre part, la vision à long terme doit permettre de repérer les domaines dans lesquels l'entreprise souhaite se développer : le risque est dans ce cas de projeter dans le futur la situation présente, et de favoriser les compétences reconnues indispensables actuellement



L'identité apprenante

Quand tous les outils sont mis à la disposition, encore faut-il que chacun ait l'envie et la possibilité de les utiliser. La mise en place de systèmes d'information performants, de structures type projet doit s'accompagner de pratiques managériales adaptées. Là encore, il n'existe pas de recette miracle, mais quelques grands principes :

- il faut valoriser l'apprentissage mutuel, le partage de connaissance
- il faut également reconnaître le droit à l'erreur, et savoir en tirer les enseignements utiles,
- il faut instaurer dans la mesure du possible des relations de confiance, afin d'éviter les risques de dérapage par rétention d'informations par exemple.

L'acceptabilité d'une innovation

Au sein de tout groupe ou société, l'attitude collective face à une nouvelle situation passe habituellement par une phase d'incrédulité, puis de rejet avant d'être éventuellement enfin acceptée. Des chercheurs se sont intéressés à ce mécanisme d'acceptation. Ils ont mis en évidence deux facteurs :

- le niveau de « formalisation » d'une innovation, c'est à dire la précision apportée dans la présentation des nouveaux objectifs et des moyens à mettre en œuvre pour les atteindre
- le niveau de « contextualisation interne » de cette innovation, c'est à dire à la perception par le groupe de l'état qu'on lui propose d'atteindre, par rapport à ce qu'il perçoit de son état présent.

Les techniques managériales devront s'adapter à ces deux paramètres afin d'atteindre le maximum d'efficacité [David, 1996].

4.7.4. Les méthodes d'aide à l'innovation

On a vu que la capacité à innover se jouait à tous les étages d'une organisation. Pour cette raison, la liste des méthodes d'aide à l'innovation peut devenir extrêmement longue puisqu'elle va concerner chacun des acteurs.

- au niveau stratégique : Les décideurs ont à établir des stratégies qui engagent leur organisation pour de longues périodes. Pour cela, ils ont à prendre des décisions à partir d'indicateurs dont il faudra s'assurer de leur pertinence. Abernathy et ses successeurs ont mis au point un modèle décrivant le cycle de vie d'une activité, c'est à dire à la fois les lois régissant l'évolution d'une technologie, d'un marché. Les méthodes de prévision, de prospective et de veille technologique permettront alors de diminuer les risques stratégiques. D'autre part, les outils de gestion permettent de choisir entre les différentes voies d'accès possible aux technologies innovantes, que ce soit par la sous-traitance, l'achat de licence, ou le développement en interne.
- au niveau technologique : Les outils de conception comme l'analyse fonctionnelle et l'analyse de la valeur, et les méthodes de veille technologique [Broustail, 1993] permettent une approche rationalisée des innovations techniques.



- au niveau structurel : On a vu qu'il n'existe pas de remède miracle ou de d'organisation idéale. Une amélioration passe alors avant tout par un état des lieux. Pour cela, il existe différents types de modélisation de systèmes socio-techniques. Certaines sont axées sur les mécanismes décisionnels, d'autres sur les flux d'informations [Vernadat, 1999]. Des travaux ont également portés sur la quantification de l'innovation elle-même au sein d'une entreprise.
- à tous les niveaux : La créativité doit être recherchée, encouragée, aidée. Elle apparaît comme la condition locale de l'innovation. Finalement, toutes les méthodes d'aide à la créativité vues précédemment peuvent être employées ici.

4.8. Conclusion

Le processus de décision repose sur, pour ne pas dire oppose, des partenaires tout à fait légitimes (utilisateurs, industriel lobbyistes ou non, pouvoirs publics locaux et régionaux, responsables divers, etc.) et représentatifs de la société actuelle. En tant que tel, leur présence, leurs compétences et leurs avis sont nécessaires.

Cependant, une décision engage ces différents partenaires et doit donc être argumentée de façon indiscutable. Pour cela, elle doit partir d'une vision partagée et validée par tous d'un problème, d'un besoin ou d'une exigence. Cela suppose de comprendre et d'accepter les points de vue des autres puis de rechercher des solutions. Certaines pourraient être qualifiées d'idéales si elles satisfont à tous les critères et contraintes ce qui augurerait d'une excellente capacité de la part des acteurs à la créativité. Cependant, la majorité des solutions restent des compromis acceptables.

Pour sortir du cadre usuel des solutions basées sur un compromis qui ne satisfait finalement personne dans le temps, il paraît nécessaire de stimuler et d'impliquer d'avantage les acteurs du processus de décision. En effet, la créativité d'un individu repose déjà sur son bon vouloir et sur sa plus ou moins bonne utilisation de techniques éprouvées comme celles décrites ci-dessus. Elle peut être améliorée par les propositions suivantes :

- Associer des acteurs venant de plusieurs domaines différents de manière à enrichir la base de connaissance accessible durant les phases de divergence puis de reconvergence nécessaires.
- Piloter le processus de décision ce qui suppose un management des acteurs participants allant de la formation aux techniques à utiliser jusqu'à l'obligation d'emploi de ces techniques.
- Bâtir une vision systémique et partagée, validée par l'ensemble des acteurs, de la problématique du transport.
- Bâtir un référentiel de problèmes du transport en France et en Europe par la mise en commun des problématiques et le benchmarking des systèmes de transport.
- Bâtir un référentiel de solutions à ces problèmes.
- Développer des méthodes d'aide à la gestion du risque décisionnel.



PARTIE II : THEORIES DE LA DECISION

1. LES APPROCHES MATHEMATIQUES DE LA DECISION

1.1. Extraction de Connaissances et Fouille de Données

La complexité des systèmes dans lesquels l'homme est impliqué aujourd'hui conduit à l'émergence de systèmes dynamiques de traitement de l'information (STI) de plus en plus sophistiqués et incontournables où la prise de décision est de plus en plus difficile. Le premier objectif que l'on peut alors assigner aux STI est d'analyser, classifier et structurer les données manipulées dans le système. Ces tâches relèvent de l'ingénierie des données.

L'un des challenges des systèmes d'information interactifs réside dans la sélection, l'organisation et la présentation dynamique de l'information ; la performance de l'ensemble dépend de l'efficacité de la communication établie par l'interface (au sens large du terme). La conception de STI coopératifs repose sur l'analyse du système homme-machine afin d'établir les besoins informationnels (modèles au niveau symbolique).

Nous avons vu au début de ce document que le couple homme-machine définit une situation cognitive où l'opérateur humain doit s'adapter à la situation, c'est-à-dire accorder ses structures cognitives aux circonstances du réel perçu (l'objet de son action) afin de prendre des dispositions pertinentes pour atteindre le but visé (le projet d'action). Dans un premier temps, l'objet de l'action doit prendre forme relativement au projet d'action. Ce processus d'information mobilise le sujet pour conceptualiser l'objet, identifier les données informatives et les qualifier au travers de processus d'évaluation. A cet effet, nous proposons de passer en revue différentes méthodes d'extraction de connaissances, de fouille de données, de regroupement de données dont la qualité principale est l'interprétabilité par l'opérateur humain. Il s'agit de mettre en évidence des groupements de données à partir de bases de données, produisant ainsi des représentations concises de la base et permettant de découvrir de nouvelles relations entre les données.

1.1.1. Introduction

Cette section présente des méthodes, techniques et outils liés à l'informatique décisionnelle. De l'entrepôt de données qui définit un support aux systèmes d'information décisionnels, aux outils de fouille de données permettant d'extraire de nouvelles connaissances, de nombreux moyens informatiques sont aujourd'hui mis en œuvre pour aider la prise de décision.

Le mot d'ordre principal est : "fournir à tout utilisateur reconnu et autorisé, les informations nécessaires à son travail". Ce slogan fait naître une nouvelle informatique, intégrante, orientée vers les utilisateurs et les centres de décision des organisations. Tout utilisateur de l'entreprise ayant à prendre des décisions doit pouvoir accéder en temps réel aux données de l'entreprise, doit pouvoir traiter ces données, extraire l'information pertinente de ces données pour prendre les "meilleures" décisions. Il se pose des questions du type : "Quels sont les résultats des ventes par gamme de produit et par région pour l'année dernière ?"; "Quelle est l'évolution des



chiffres d'affaire par type de magasin et par période ?" ; ou encore "Comment qualifier les acheteurs de mon produit X ?". Le système opérationnel ne peut satisfaire ces besoins pour au moins deux raisons : les bases de données opérationnelles sont trop complexes pour pouvoir être appréhendées facilement par tout utilisateur et le système opérationnel ne peut être interrompu pour répondre à des questions nécessitant des calculs importants. Il s'avère donc nécessaire de développer des systèmes d'information orientés vers la décision. Il faut garder un historique et restructurer les données de production, éventuellement récupérer des informations démographiques, géographiques et sociologiques.

Les entrepôts de données et la fouille de données sont les éléments d'un domaine de recherche et de développement très actifs actuellement : l'extraction de connaissances à partir de données ou Knowledge Discovery in Databases (KDD). Par ce terme, on désigne tout le cycle de découverte d'information. Il regroupe donc la conception et les accès à de grandes bases de données, tous les traitements à effectuer pour extraire de l'information de ces données, l'un de ces traitements est l'étape de fouille de données ou Data Mining (DM).

Bien que le terme de fouille de données recouvre à lui seul la découverte de connaissances, il ne constitue qu'une seule des étapes du processus général de l'ECD [Fayad *et al.*, 1996] (Extraction de Connaissances dans les Bases de données), qui comprend globalement trois phases :

- Préparation des données : l'objectif de cette phase consiste à sélectionner uniquement les données potentiellement utiles de la base. L'ensemble des données est ensuite soumis à un pré-traitement, afin de gérer les données manquantes ou invalides (opération de nettoyage). L'étape suivante dans cette phase consiste à formater ces données, pour les rendre compréhensibles au processus de fouille de données (opérations de transformation et réduction).
- Extraction : en appliquant des techniques de fouilles de données, l'objectif de cette phase est de proposer des modèles ou motifs représentatifs du contenu de la base.
- Interprétation des résultats : le but de cette dernière phase est d'interpréter la connaissance extraite lors de la phase précédente, pour la rendre lisible et compréhensible par l'utilisateur.

Les techniques de fouille de données sont utilisées dans de nombreux domaines d'applications. Les exemples les plus courants sont les compagnies d'assurance, les compagnies bancaires (crédit, prédition du marché, détection de fraudes), le marketing (comportement des consommateurs, "mailing" personnalisé), la recherche médicale (aide au diagnostic, au traitement, surveillance de la population sensible), les réseaux de communication (détection de situations alarmantes, prédition d'incidents), l'analyse de données spatiales ou encore statistiques. Nous proposons un rapide survol en nous intéressant plus particulièrement à celles développées pour manipuler de grandes bases de données.



Les Règles d'Association

Le problème de recherche de règles d'association a fait l'objet de nombreux travaux de recherche ces dernières années. Introduit dans [Agrawal *et al.*, 1993] à partir du problème du panier de la ménagère ("Market Basket Problem"), il peut être résumé ainsi : étant donné une base de données de transactions (les paniers), chacune composée d'items (les produits achetés), la découverte d'associations consiste à chercher des ensembles d'items, fréquemment liés dans une même transaction, ainsi que des règles les combinant. Un exemple d'association pourrait révéler que "75% des gens qui achètent de la bière, achètent également des couches". Ce type de règles concerne un grand champ d'applications, telles que la conception de catalogues, la promotion de ventes, le suivi d'une clientèle, etc.

Les Motifs Séquentiels

Introduit dans [Agrawal *et al.*, 1995], les motifs séquentiels peuvent être vus comme une extension de la notion de règle d'association, intégrant diverses contraintes temporelles. La recherche de tels motifs consiste ainsi à extraire des ensembles d'items, couramment associés sur une période de temps bien spécifiée. En fait, cette recherche met en évidence des associations inter-transactions, contrairement à celle des règles d'association qui extrait des combinaisons intra-transaction. Dans ce contexte, et contrairement aux règles d'association, l'identification des individus ou objets est indispensable, afin de pouvoir suivre leur comportement au cours du temps. Par exemple, des motifs séquentiels peuvent montrer que "60% des gens qui achètent une télévision, achètent un magnétoscope dans les deux ans qui suivent". Ce problème, posé à l'origine par souci marketing, intéresse à présent des domaines aussi variés que les télécommunications (détection de fraudes), la finance ou encore la médecine (identification des symptômes précédant les maladies).

Les Dépendances Fonctionnelles

L'extraction de dépendances fonctionnelles à partir de données existantes, est un problème étudié depuis de nombreuses années, mais qui a été abordé récemment avec une "vision" fouille de données. Les résultats obtenus par ces dernières approches sont très efficaces. La découverte de dépendances fonctionnelles est un outil d'aide à la décision à la fois pour l'administrateur de la base, les développeurs d'application, les concepteurs et intégrateurs de systèmes d'information. En effet, les applications relèvent de l'administration et du contrôle des bases de données, de l'optimisation de requêtes ou encore de la rétro-conception de systèmes d'information.

La Classification

La classification, appelée également induction supervisée, consiste à analyser de nouvelles données et à les affecter, en fonction de leurs caractéristiques ou attributs, à telle ou telle classe prédéfinie. Bien connus en apprentissage, les problèmes de classification intéressent également la communauté Base de Données qui s'appuie sur l'intuition suivante : "plus importants sont les volumes de données traités, meilleure devrait être la précision du modèle de classification". Les techniques de classification sont par exemple utilisées lors d'opérations de "mailing" pour cibler la bonne population et éviter ainsi un nombre trop important de non-réponses. De la même manière, cette démarche peut permettre de déterminer, pour une banque, si un prêt peut être accordé, en fonction de la classe d'appartenance d'un client.



La Segmentation

La technique de segmentation ressemble à celle de la classification, mais diffère dans le sens où il n'existe pas de classes prédefinies : l'objectif est de grouper des enregistrements qui semblent similaires dans une même classe. De nombreux algorithmes efficaces ont été proposés pour optimiser les performances et la qualité des classes obtenues dans de grandes bases de données. Les applications concernées incluent notamment la segmentation de marché ou encore la segmentation démographique par exemple en identifiant des caractéristiques communes entre populations.

Les Séries Chronologiques

L'objectif est de trouver des portions de données (ou séquences) similaires à une portion de données précise ou encore de trouver des groupes de portions similaires issues de différentes applications. Cette technique permet par exemple d'identifier des sociétés qui présentent des séquences similaires d'expansion ou encore de découvrir des stocks qui fluctuent de la même manière.

Cette section est organisée en cinq parties. Dans la première partie, les grandes étapes du processus d'extraction de connaissances sont exposées. Les trois chapitres suivants sont consacrés aux principales techniques de fouille de données : classification, règles d'association et motifs séquentiels. Enfin, dans une dernière partie, nous présentons quelques applications des techniques de fouilles de données.

1.1.2. Le processus d'Extraction de Connaissances à partir de données

Nous présentons dans cette partie le processus complet d'extraction de connaissances [Gilleron *et al.*, 2000; Denis *et al.* 2000]. Ce processus (représenté sur la figure 28) peut se décomposer en trois grandes phases : Préparation, Extraction et Interprétation.

La phase de préparation consiste à partir de la formulation d'un projet à : récolter, préparer et organiser les données permettant de traiter ce projet. Les grandes étapes constituant cette phase seront présentées plus tard dans le document. La phase d'extraction des connaissances permet, à l'aide de méthodes adaptées aux données fournies par la phase précédente, de mettre en évidences des caractéristiques et des modèles contenus intrinsèquement et implicitement dans les données. Cette phase sera abordée en détail dans la suite de cette section. La phase d'interprétation, consiste à intégrer les caractéristiques et modèles mis en évidence par les méthodes de fouille de données dans le processus de décision.

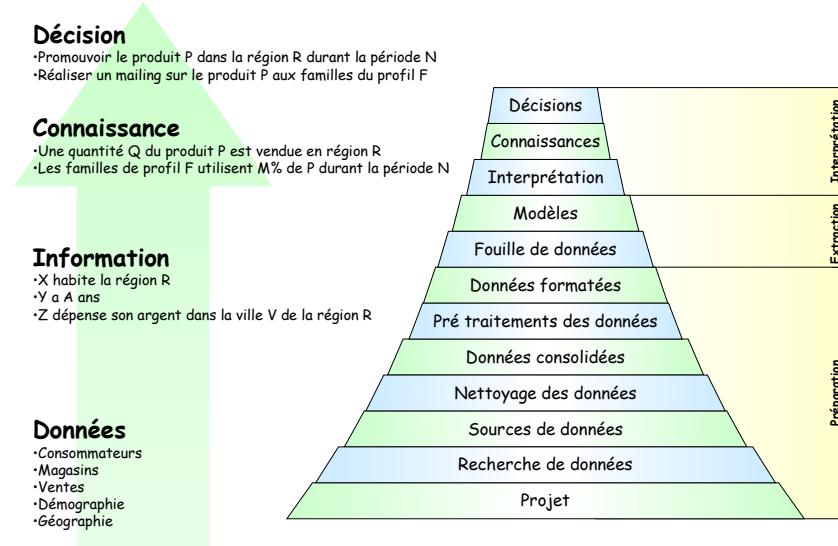


Figure 28 : Processus général d'extraction de connaissances à partir de données

Le déroulement d'un projet n'est pas linéaire. On peut constater dans l'étape d'interprétation que les performances obtenues ne sont pas suffisantes ou que les utilisateurs du domaine jugent les modèles inexploitables, il s'agira alors de choisir une autre méthode de fouille ou de remettre en cause les codages ou de chercher à enrichir les données. On évalue dans un projet le temps passé à l'étape de fouille de données (qui est l'étape de découverte d'informations proprement dite) à moins de 20% du temps. Par conséquent, plus de 80% du temps est dédié aux phases de préparation des données et d'interprétation des modèles.

Nous présentons maintenant l'exemple qui nous servira pour illustrer chaque étape du processus. Cet exemple est issu du livre de P. Adriaans et D. Zantige [Adriaans *et al.*, 1996].

Un éditeur vend 5 sortes de magazines : sport, voiture, maison, musique et BD. Il souhaite mieux étudier ses clients pour découvrir de nouveaux marchés ou vendre plus de magazines à ses clients habituels. Les questions qu'il se pose sont :

1. "Combien de personnes ont pris un abonnement a un magazine de sport cette année ?"
2. "A-t-on vendu plus d'abonnements de magazines de sport cette année que l'année dernière ?"
3. "Est-ce que les acheteurs de magazines de BD sont aussi amateurs de sport ?"
4. "Quelles sont les caractéristiques principales de mes lecteurs de magazines de voiture ?"
5. "Peut-on prévoir les pertes de clients et prévoir des mesures pour les diminuer ?"

Les questions qui sont proposées sont de natures différentes et mettent en jeu des processus différents. De simples requêtes SQL sont suffisantes pour répondre aux deux premières questions. Les données recherchées dans ce cas ne sont que le résultat d'un calcul simple sur un ou des groupes d'enregistrements. Ce qui distingue ces deux premières questions, c'est la



notion de temps et la comparaison. Pour établir cette comparaison, les données doivent être présentes dans la base, ce qui n'est pas toujours le cas pour les bases opérationnelles.

Nous pouvons donc supposer que la requête 1 est réalisable sans technique particulière à partir des données opérationnelles sous réserve d'indexations suffisantes des tables concernées.

La requête 2 nécessite de conserver toutes les dates de souscription même pour les abonnements résiliés. Nous pouvons imaginer aussi que l'utilisateur émettra une grande variété de requêtes de ce type. Nous supposons alors que ces questions seront résolues à l'aide d'outils de création de requêtes multidimensionnelles.

La question 3 est un exemple simplifié de problème où l'on demande si les données vérifient une règle. La réponse pourrait se formuler par une valeur estimant la probabilité que la règle soit vraie. Souvent, des outils statistiques sont utilisés. Cette question peut être généralisée. On pourrait chercher des associations fréquentes entre acheteurs de magazine pour effectuer des actions promotionnelles. On pourrait également introduire une composante temporelle pour chercher si le fait d'être lecteur d'un magazine implique d'être, plus tard, lecteur d'un autre magazine.

La question 4 est beaucoup plus ouverte. La problématique ici est de trouver une règle et non plus de la vérifier ou de l'utiliser. C'est pour ce type de question que l'on met en œuvre des outils de fouille de données.

La question 5 est également une question ouverte. Il faut pour la résoudre disposer d'indicateurs qui pourraient sur notre exemple être : les durées d'abonnement, les délais de paiement. Ce type de question a une forte composante temporelle et nécessite des données historiques.

Le processus de KDD utilise des outils de création de requêtes, des outils statistiques et des outils de fouille de données. Là encore, nous nous apercevons qu'une très grande partie des problèmes de décision se traite avec des outils simples. La fouille de données quand elle est nécessaire, suit souvent une analyse des données simple ou statistique.

Phase de préparation des données

Dans notre exemple, nous avons identifié quelques objectifs précis, exprimés sous forme de questions. La préparation des données consiste dans un premier temps à obtenir des données en accord avec les objectifs que l'on s'impose. Ces données proviennent le plus souvent de bases de production ou d'entrepôts. Pour illustrer cette phase, nous partons d'une liste des souscriptions d'abonnements avec cinq champs (tableau 9).

No client	Nom	Adresse	Date d'abonnement	Magazine
23134	Bemol	Rue du moulin, Paris	7/10/96	Voiture
23134	Bemol	Rue du moulin, Paris	12/5/96	Musique
23134	Bemol	Rue du moulin, Paris	25/7/95	BD
31435	Bodinoz	Rue verte, Nancy	11/11/11	BD
43342	Airinaire	Rue de la source, Brest	30/5/95	Sport
25312	Talonion	Rue du marché, Paris	25/02/98	NULL
43241	Manvussa	NULL	14/04/96	Sport
23130	Bemolle	Rue du moulin, Paris	11/11/11	Maison

Tableau 9 : Obtention des données



L'obtention des données est souvent réalisée à l'aide d'outils de requête (OLAP : On-Line Analytical Processing, SQL : Structured Query Language...).

Il faut, dès à présent, noter que l'on ne peut résoudre des problèmes que si l'on dispose des données nécessaires. Il semble, par exemple, impossible de s'attaquer à la question cinq de notre exemple avec les données dont nous disposons.

On peut avoir recours à d'autres bases, achetées ou produites en un autre lieu, pour enrichir nos données. L'opération va se traduire par l'ajout de nouveaux champs en conservant souvent le même nombre d'enregistrements. Une première difficulté ici est de pouvoir relier des données qui parfois sont hétérogènes. Des problèmes de format de données apparaissent et des conversions sont souvent nécessaires. Une deuxième difficulté est l'introduction de nouvelles valeurs manquantes.

Pour notre exemple, supposons que nous ayons accès à des informations sur les clients données dans le tableau 10.

Client	Date de naissance	Revenus	Propriétaire	Voiture
Bemol	13/1/50	20 000 F	Oui	Oui
Bodinoz	21/5/70	12 000 F	Non	Oui
Airinaire	15/06/63	9 000 F	Non	Non
Manvussa	27/03/47	15 000 F	Non	Oui

Tableau 10 : Enrichissement des données

Nettoyage des données

Avant de pouvoir utiliser les données à travers les outils de fouille, un certain nombre de vérifications et de transformations sont nécessaires pour garantir la qualité des données.

Consolidation

Selon les choix des unités pour les dimensions, des opérations de consolidation devront accompagner le chargement des données (par exemple sommer les ventes pour obtenir et enregistrer un total par jour et non pas toutes les transactions).

Uniformisation d'échelle

Pour éviter de trop grandes dispersions dans les valeurs numériques, une homogénéisation des échelles de valeurs est utile. Ne pas la réaliser peut pénaliser les outils d'analyse et de visualisation et peut-être simplement remplir inutilement les supports de stockage.

Doublons, erreurs de saisie

Les doublons peuvent se révéler gênants parce qu'ils vont donner plus d'importance aux valeurs répétées. Une erreur de saisie pourra à l'inverse occulter une répétition. Dans notre exemple, les clients numéro 23134 et 23130 sont certainement un seul et même client, malgré la légère différence d'orthographe.



Intégrité de domaine

Un contrôle sur les domaines des valeurs permet de retrouver des valeurs aberrantes. Dans notre exemple, la date de naissance du client 23130 (11/11/11) semble plutôt correspondre à une erreur de saisie ou encore à une valeur par défaut en remplacement d'une valeur manquante.

Informations manquantes

C'est le terme utilisé pour désigner le cas où des champs ne contiennent aucune donnée. Parfois, il est intéressant de conserver ces enregistrements car l'absence d'information peut être une information (ex: détection de fraudes). D'autre part, les valeurs contenues dans les autres champs peuvent être utiles. Dans notre exemple, nous n'avons pas le type de magazine pour le client 25312. Il sera écarté de notre ensemble. L'enregistrement du client 43241 sera conservé bien que l'adresse ne soit pas connue.

No client	Nom	Adresse	Date d'abonnement	Magazine
23134	Bemol	Rue du moulin, Paris	7/10/96	Voiture
23134	Bemol	Rue du moulin, Paris	12/5/96	Musique
23134	Bemol	Rue du moulin, Paris	25/7/95	BD
31435	Bodinoz	Rue verte, Nancy	NULL	BD
43342	Airinaire	Rue de la source, Brest	30/5/95	Sport
43241	Manvussa	NULL	14/04/96	Sport
23130	Bemol	Rue du moulin, Paris	NULL	Maison

Tableau 11 : Données après nettoyage

Pré traitement des données

A ce stade du processus, les choix sont particulièrement guidés par l'algorithme de fouille utilisé et des ajustements des choix de codage sont souvent nécessaires.

Regroupements

Certains attributs prennent un très grand nombre de valeurs discrètes.

C'est typiquement le cas des adresses. Lorsqu'il est important de considérer ces attributs pour la fouille de données il est obligatoire d'opérer des regroupements et ainsi obtenir un nombre de valeurs raisonnable. Dans l'exemple, nous regroupons les adresses en deux régions : Paris, province.

Attributs discrets

Les attributs discrets prennent leurs valeurs (souvent textuelles) dans un ensemble fini donné. C'est le cas de la colonne magazine dans notre exemple qui peut prendre les valeurs Sport, BD, Voiture, Maison, Musique. Deux représentations sont possibles pour ces données : une représentation verticale telle qu'elle est présentée en table 9 ou une représentation horizontale ou éclatée table 12.

No	Sport	BD	Voiture	Maison	Musique
23134	0	1	1	0	1
31435	0	1	0	0	0
43342	1	0	0	0	0
43241	1	0	0	0	0

Tableau 12 : Données après nettoyage



La représentation horizontale est plus adaptée à la fouille de données et certains calculs sont simplifiés. Par exemple, la somme de la colonne sport que divise le nombre d'enregistrements calcule le pourcentage de clients ayant souscrit un abonnement à un magazine de sport. Notons que la date d'abonnement dépend du type de magazine. De façon générale, la modification présentée en table 12 peut induire une perte d'information pour tous les champs en dépendance fonctionnelle avec le champ transformé. Nous choisissons de transformer le champ date d'abonnement en date du plus vieil abonnement. Il est à noter que cette transformation ne nous permet plus d'espérer générer des règles de la forme : un acheteur de BD s'abonne à un magazine de musique dans les deux ans qui suivent.

Dans notre exemple, le même codage en deux valeurs 0 et 1 sera réalisé avec les champs Oui/Non issus de l'enrichissement.

Changements de type

Pour certaines manipulations, comme des calculs de distance, des calculs de moyenne, il est préférable de modifier les types de certains attributs. Dans notre exemple, la date de naissance et la date d'abonnement ne permettent pas d'effectuer simplement des comparaisons, des différences. Nous pouvons les convertir en âge ou en durée.

Uniformisation d'échelle

Certains algorithmes, comme la méthode des plus proches voisins, sont basés sur des calculs de distance entre enregistrements. Des variations d'échelle selon les attributs sont autant de perturbations possibles pour ces algorithmes. Des échelles très "étirées" vont artificiellement rendre des dimensions trop "vides".

C'est typiquement le cas pour le champ Revenus dans notre exemple. Les centaines de francs ne sont pas significatives. Nous convertissons donc les revenus en les divisant par mille. L'intervalle de valeurs est alors dans la même échelle que les dates de naissance et les durées d'abonnement.

No	Sport	BD	Voiture	Maison	Musique	Date naissance	Revenus	Propriét aire	Voiture	Paris?	Durée abonnement
23134	0	1	1	0	1	50	20	Oui	Oui	1	4
31435	0	1	0	0	0	30	12	Non	Oui	0	NULL
43342	1	0	0	0	0	37	9	Non	Non	0	5
43241	1	0	0	0	0	53	15	Non	Oui	NULL	4

Tableau 13 : Données après nettoyage

Phase de fouille des données

La fouille de données est le cœur du processus, elle permet d'extraire les caractéristiques et les modèles à partir de données formatées. C'est souvent une étape difficile à mettre en œuvre, dont les résultats doivent être interprétés et validés.

Nous allons dans cette partie présenter l'approche générale de cette phase, les méthodes et techniques de fouille de données seront abordées plus en détail dans les sections suivantes.



L'Approche fouille de données

La fouille de données est une approche particulière de l'analyse de données. Elle se différencie des approches classiques, dites " confirmatives " par son caractère " exploratoire ". En effet, dans les méthodes d'analyse de données classique la démarche généralement mise en œuvre est la suivante :

1. Description des données – Mesures et graphiques issus des statistiques descriptives
2. Hypothèse d'un modèle ou d'une caractéristique
3. Choix et mise en œuvre d'une méthode d'analyse
4. Validation
5. Itération des étapes 3 et 4 jusqu'à l'obtention des performances souhaitées.

Dans une approche plus " Data Mining ", toutes les étapes existent à l'exception de l'étape 2. L'idée clé de la fouille de données, étant d'essayer de faire apparaître durant l'étape 3 d'analyse, des caractéristiques et/ou des modèles inconnus et " insoupçonnés ".

La qualité du modèle obtenu se mesure selon les critères suivants :

1. Rapide à créer
2. rapide à utiliser
3. compréhensible pour l'utilisateur
4. les performances sont bonnes
5. Le modèle est fiable
6. les performances ne se dégradent pas dans le temps
7. Il évolue facilement

Il va de soi qu'aucun modèle n'aura toutes ces qualités. Il n'existe pas de meilleure méthode de fouille. Il faudra faire des compromis selon les besoins dégagés et les caractéristiques connues des outils. Pour une utilisation optimale, une combinaison de méthodes est recommandée. On peut rapidement donner 3 catégories de méthodes de fouille de données:

- **Classification supervisée, régression** : Il s'agit de trouver une classe ou une valeur selon un ensemble de descriptions. Ce sont des outils très utilisés. Les algorithmes reposent sur des arbres de décision, des réseaux de neurones, la règle de Bayes, les k plus proches voisins.
- **Classification non supervisée, segmentation, regroupement, clustering** : La problématique est de trouver des groupes homogènes dans une population. On utilise souvent les techniques d'agrégation autour de centres mobiles ou de classification



ascendante hiérarchique. La difficulté essentielle dans ce type de construction est la validation.

- **Association, sequencing** : Il s'agit de trouver des similarités ou des associations. Le sequencing est le terme anglais utilisé pour préciser que l'association se fera dans le temps. Par exemple, si j'achète un couffin aujourd'hui, j'ai trois fois plus de chance dans trois mois d'acheter un lit bébé (sequencing) ou encore si j'achète des pâtes et de la purée de tomates, j'ai deux fois plus de chance d'acheter aussi du parmesan (association).

Validation

Les méthodes de validation vont dépendre de la nature de la tâche et du problème considéré. Nous distinguerons deux modes de validation : statistique et par expertise. Pour certains domaines d'application (le diagnostic médical, par exemple), il est essentiel que le modèle produit soit compréhensible. Il y a donc une première validation du modèle produit par l'expert, celle-ci peut être complétée par une validation statistique sur des bases de cas existantes.

Pour les problèmes d'apprentissage non supervisé (segmentation, association), la validation est essentiellement du ressort de l'expert. Pour la segmentation, le programme construit des groupes homogènes, un expert peut juger de la pertinence des groupes constitués. Là encore, on peut combiner avec une validation statistique sur un problème précis utilisant cette segmentation.

Pour la recherche des règles d'association, c'est l'expert du domaine qui jugera de la pertinence des règles, en effet, l'algorithme, s'il fournit des règles porteuses d'information, produit également des règles triviales et sans intérêt. Pour la validation statistique, la première tâche à réaliser consiste à utiliser les méthodes de base de statistique descriptive. L'objectif est d'obtenir des informations qui permettront de juger le résultat obtenu ou d'estimer la qualité ou les biais des données d'apprentissage.

1. Calculer les moyennes et variances des attributs.
2. Si possible, calculer la corrélation entre certains champs.
3. Déterminer la classe majoritaire dans le cas de la classification.

Pour la classification supervisée, la deuxième tâche consiste à décomposer les données en plusieurs ensembles disjoints. L'objectif est de garder des données pour estimer les erreurs des modèles ou de les comparer. Il est souvent recommandé de constituer :

- Un ensemble d'apprentissage.
- Un ensemble de test.
- Un ensemble de validation.

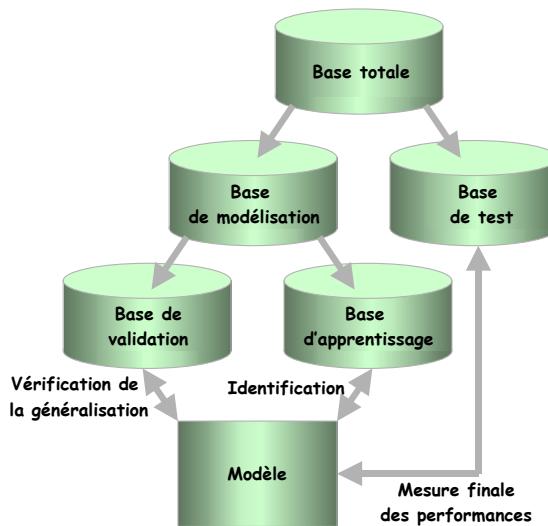


Figure 29 : Processus de validation par le test

Au moins deux ensembles sont nécessaires : l'ensemble d'apprentissage permet de générer le modèle, l'ensemble test permet d'évaluer l'erreur réelle du modèle sur un ensemble indépendant évitant ainsi un biais d'apprentissage. Lorsqu'il s'agit de tester plusieurs modèles et de les comparer, on peut sélectionner le meilleur modèle selon ses performances sur l'ensemble de validation et ensuite évaluer son erreur réelle sur l'ensemble de test.

Lorsqu'on ne dispose pas de suffisamment d'exemples on peut se permettre d'apprendre et d'estimer les erreurs avec un même ensemble par la technique de validation croisée ou du Bootstrap.

Validation croisée

La validation croisée ne construit pas de modèle utilisable mais ne sert qu'à estimer l'erreur réelle d'un modèle selon l'algorithme suivant :

```
Validation croisée (S;x)
// S est un ensemble, x est un entier
Découper S en x parties égales {S1, ..., Sx}
Pour i de 1 à x
    Construire un modèle M avec l'ensemble S - Si
    Evaluer une mesure d'erreur ei de M avec Si
Fin Pour
Retourner l'espérance mathématique des mesures des erreurs
```

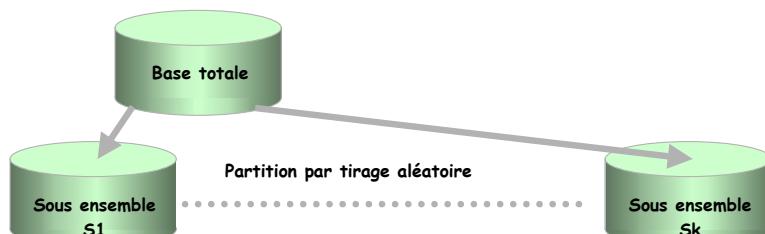


Figure 30 : Processus de validation croisée



Leave one out

Si la taille des Si est de un individu, on parle alors de validation par " leave one out "

Bootstrap

Etant donné un échantillon S de taille n, on tire avec remise un ensemble d'apprentissage de taille n (un élément de S peut ne pas appartenir à l'ensemble d'apprentissage, ou y figurer plusieurs fois), l'ensemble test est S.

L'estimation de l'erreur réelle est alors l'espérance mathématique des erreurs obtenues pour un certain nombre d'itérations de l'algorithme d'apprentissage.

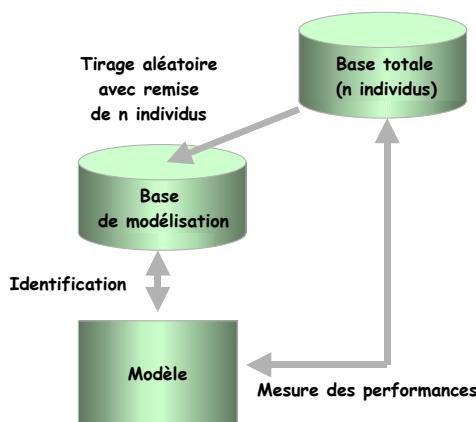


Figure 31 : Processus de Bootstrap

Ces trois méthodes fournissent de bons estimateurs de l'erreur réelle mais sont très coûteuses en temps de calcul. Elles sont très utiles pour les petits échantillons.

1.1.3. Classification

Problématique [Denis et al., 2000], [Gilleron et al., 2000]

Les méthodes de classification ont pour but d'identifier les classes auxquelles appartiennent des objets à partir de certains traits descriptifs. Elles s'appliquent à un grand nombre d'activités humaines et conviennent en particulier au problème de l'aide à la prise de décision. Il s'agira, par exemple, d'établir un diagnostic médical à partir de la description clinique d'un patient, de donner une réponse à la demande de prêt bancaire de la part d'un client sur la base de sa situation personnelle, de déclencher un processus d'alerte en fonction de signaux reçus par des capteurs. Une première approche possible pour résoudre ce type de problème est l'approche " systèmes experts ". Dans ce cadre, la connaissance d'un expert (ou d'un groupe d'experts) est décrite sous forme de règles. Cet ensemble de règles forme un système expert qui est utilisé pour classifier de nouveaux cas. Cette approche, largement utilisée dans les années 80, dépend fortement de la capacité à extraire et à formaliser les connaissances de l'expert. Nous considérons ici une autre approche pour laquelle la procédure de classification sera extraite automatiquement à partir d'un ensemble d'exemples. Un exemple consiste en la description d'un cas avec la classification correspondante. Par exemple, on dispose d'un historique des prêts accordés avec, pour chaque prêt, la situation personnelle du demandeur et



le résultat du prêt (problèmes de recouvrement ou non). Un système d'apprentissage doit alors, à partir de cet ensemble d'exemples, extraire une procédure de classification qui, au vu de la situation personnelle d'un client, devra décider de l'attribution du prêt. Il s'agit donc d'induire une procédure de classification générale à partir d'exemples. Le problème est donc un problème inductif, il s'agit d'extraire une règle générale à partir de données observées. La procédure générée devra classifier correctement les exemples de l'échantillon mais surtout avoir un bon pouvoir prédictif pour classifier correctement de nouvelles descriptions.

Les méthodes utilisées par les systèmes d'apprentissage sont très nombreuses et sont issues de domaines scientifiques variés. Les méthodes statistiques supposent que les descriptions des objets d'une même classe se répartissent en respectant une structure spécifique à la classe. On fait des hypothèses sur les distributions des descriptions à l'intérieur des classes et les procédures de classification seront construites à l'aide d'hypothèses probabilistes. La variété des méthodes viendra de la diversité des hypothèses possibles. Ces méthodes sont appelées semi-paramétriques. Des méthodes non paramétriques (sans hypothèse a priori sur les distributions) ont été également proposées en statistiques. Les méthodes issues de l'intelligence artificielle sont des méthodes non paramétriques. On distingue les méthodes symboliques (la procédure de classification produite peut être écrite sous forme de règles) des méthodes non symboliques ou adaptatives (la procédure de classification produite est de type " boîte noire "). Parmi les méthodes symboliques, les plus utilisées sont basées sur les arbres de décision ou la logique floue. Pour les méthodes adaptatives, on distingue deux grandes classes : les réseaux de neurones et les algorithmes génétiques. L'apprentissage automatique, dans une définition très générale, consiste en l'élaboration de programmes qui s'améliorent avec l'expérience. Les applications sont nombreuses et concernent des domaines très variés. On peut citer, par exemple, la reconnaissance de formes avec, en particulier, la reconnaissance de la parole et du texte écrit, le contrôle de processus et le diagnostic de pannes, les programmes de jeu. Les méthodes d'apprentissage à partir d'exemples sont très utilisées dans la recherche d'informations dans de grands ensembles de données. En effet, l'évolution de l'informatique permet de nos jours de manipuler des ensembles de données de très grande taille (entrepot de données).

Citons, parmi d'autres, quelques applications et leur domaine :

- analyse financière : prévision d'évolution de marchés,
- marketing : établir un profil client, mailing,
- banque : attribution de prêts,
- médecine : aide au diagnostic,
- télécoms : détection de fraudes.

Jusqu'à présent, nous avons supposé que l'ensemble des classes était défini : ce n'est pas toujours le cas. Un enfant apprend à catégoriser, c'est-à-dire à associer une classe à un objet alors même que la classe en question n'a pas de définition bien précise. Comment définir ce qui relève de la catégorie " chaise " ? Le fait de supposer qu'il existe une application Y associant une classe à chaque individu de la population revient à éliminer ce problème. Même lorsque l'espace des classes est bien défini, il est parfois naturel de supposer que le système apprenant doit les retrouver au cours du processus d'apprentissage. On peut par exemple



espérer qu'un enfant mis en présence de mammifères (à pattes), de reptiles, d'oiseaux et de poissons saura faire les regroupements adéquats et apprendra d'un coup les classes et le moyen de les différencier. On parle dans ce cas d'apprentissage non supervisé, autrement dit, sans professeur. Nous allons maintenant étudier les principales techniques de classification supervisée et non supervisée.

Techniques de classification supervisée

Classifieur de Bayes

[Denis *et al.*, 2000], [Gilleron *et al.*, 2000] Établir un diagnostic dans le domaine médical, signifie être capable d'associer le nom d'une maladie à un certain nombre de symptômes présentés par des malades. On repère, dans ce problème, trois objets essentiels : les malades, les maladies et les symptômes. Les malades forment la population, les symptômes sont les descriptions des malades, les maladies sont les classes. On suppose qu'il existe un classement correct, c'est-à-dire qu'il existe une application qui associe à tout malade une maladie. Apprendre à établir un diagnostic, c'est associer une maladie à une liste de symptômes de telle manière que cette association corresponde au classement défini ci-dessus.

Pour formaliser notre propos, nous utiliserons les notations suivantes :

- Π est la population, D est l'ensemble des descriptions et l'ensemble des classes est $\{1, \dots, c\}$.
- $X : \Pi \rightarrow D$ est la fonction qui associe une description à tout élément de la population.
- $Y : \Pi \rightarrow \{1, \dots, c\}$ est la fonction de classement qui associe une classe à tout élément de la population.
- Une fonction $C : D \rightarrow \{1, \dots, c\}$ sera appelée fonction de classement ou procédure de classification.

Le but de l'apprentissage est alors de rechercher une procédure de classification C telle que $C \circ X = Y$ ou, de manière plus réaliste, telle que $C \circ X$ soit une bonne approximation de Y . Supposons que nous ayons un ensemble de patients dont nous souhaitons savoir s'ils sont malades ou pas. Un patient est décrit par une liste de symptômes et de mesures. Il est alors possible que les descriptions ne permettent pas toujours de différencier un patient malade d'un patient sain. En effet, il se peut que deux patients, l'un malade, l'autre pas, aient les mêmes descriptions. Dans ce cas, on ne peut espérer trouver une procédure de classification exacte, le but sera donc de trouver une "bonne" procédure de classification.

Dans la pratique, on dispose souvent d'un ensemble d'attributs A_1, \dots, A_n logiques, symboliques ou numériques qui prennent leurs valeurs dans des domaines D_1, \dots, D_n . Décrire un élément de la population consiste alors à attribuer une valeur à chacun de ces attributs. L'espace de description D est alors égal au produit cartésien $D_1 \times \dots \times D_n$. Par exemple, on décrira un patient par un ensemble de symptômes (mal de tête, douleurs abdominales, ...)



une suite de mesures (tension, température, ...) ; on décrira un client par un ensemble de données que l'on possède sur lui (âge, sexe, catégorie socioprofessionnelle, ...).

Comment exprimer le fait que $C \circ X$ doit être une bonne approximation de Y ?

Intuitivement, cela signifie que $C \circ X$ est rarement différent de Y . Une manière de formaliser cela consiste à supposer l'existence d'une distribution de probabilité sur l'ensemble Π et à dire que $C \circ X$ est une bonne approximation de Y s'il est peu probable que ces deux fonctions diffèrent. On supposera donc que l'ensemble Π est probabilisé. Nous supposerons également, pour simplifier la présentation, que l'ensemble D est discret.

Soit P la probabilité définie sur la population Π . On peut alors définir les probabilités et notations suivantes :

On note $P(d)$ la probabilité qu'un élément de Π ait d pour description, soit encore $P(d) = P(X^{-1}(d))$.

On note $P(k)$ la probabilité qu'un élément de Π soit de classe k , soit encore $P(k) = P(Y^{-1}(k))$.

On note $P(d/k)$ la probabilité qu'un élément de classe k ait d pour description, soit encore $P(d/k) = P(X^{-1}(d)/Y^{-1}(k))$. Cette probabilité n'est définie que si la probabilité pour un élément de Π d'être de classe k est non nulle.

On note $P(k/d)$ la probabilité qu'un élément ayant d pour description soit de classe k , soit encore $P(k/d) = P(Y^{-1}(k)/X^{-1}(d))$. Cette probabilité n'est définie que si la probabilité pour un élément de Π d'avoir d pour description est non nulle.

La formule de Bayes s'écrit alors :

$$P(k/d) = \frac{P(d/k) \times P(k)}{P(d)}$$

Supposons que nous soyons dans la situation idéale où nous pouvons évaluer les quantités $P(d)$, $P(k)$ et $P(d/k)$ pour toutes les valeurs d de D et k de $\{1, \dots, c\}$.

Comment choisir la fonction C ?

Règles de choix des fonctions de classement.

1. Règle majoritaire : C_{maj} associe à tout élément d de D la classe k de $\{1, \dots, c\}$ telle que $P(k)$ soit maximum.
2. Règle du maximum de vraisemblance : $C_{vraisemblance}$ associe à tout élément d de D la classe k de $\{1, \dots, c\}$ telle que $P(d/k)$ soit maximum.



3. Règle de Bayes : C_{Bayes} associe à tout élément d de D la classe k de $\{1, \dots, c\}$ telle que $P(k/d)$ soit maximum, soit encore telle que $P(d/k) \times P(k)$ soit maximum.

On peut facilement vérifier que la règle de Bayes se ramène à la règle du maximum de vraisemblance lorsque les classes sont équiprobables.

Pour comparer des procédures, nous allons définir l'erreur de classification.

Soit C une fonction de classement, l'erreur $E(d)$ pour une description d est la probabilité qu'un élément de la population Π de description d soit mal classé par C .

$$E(d) = P(Y \neq C(d) / x = d)$$

L'erreur de classification $E(C)$ d'une fonction de classement est la moyenne pondérée des erreurs sur les descriptions d .

$$E(C) = \sum_{d \in D} E(d) \times P(X = d)$$

On peut démontrer que : l'ensemble Π étant probabilisé, la règle de décision de Bayes est celle dont l'erreur de classification est minimale.

La règle majoritaire (la plus grossière) et la règle de Bayes (la plus fine) constituent donc deux bornes naturelles en apprentissage. S'il existe une fonction de classement correcte, i.e. qui classifie sans erreur tous les individus au vu de leur description, on a alors $E(C_{\text{Bayes}}) = 0$.

Autrement dit, la fonction de classement définie par la règle de Bayes est correcte. On remarque qu'une fonction d'erreur nulle existe si et seulement si la probabilité que des individus appartenant à des classes différentes aient des descriptions identiques est nulle. On dit dans ce cas que le problème est déterministe. Cette situation est très exceptionnelle en pratique. En effet, il est rare que les paramètres descriptifs dont on dispose soient suffisants pour classifier correctement tous les individus de la population. Par exemple, deux patients peuvent avoir les mêmes symptômes et des maladies différentes, deux clients ayant les mêmes profils peuvent ou non répondre à un même mailing. De plus, on dispose rarement de données exactes et il suffit d'ajouter un peu de "bruit" à un problème déterministe pour le transformer en un problème non déterministe.

Nous avons donc montré dans cette section qu'il existe une règle optimale au sens de l'erreur de classification. Cependant, pour pouvoir appliquer cette règle, il faut que l'apprenant puisse disposer de probabilités que dans la plupart des problèmes réels, il est difficile d'estimer. Néanmoins, de nombreuses méthodes statistiques reposent sur la règle de Bayes, en se basant sur différentes techniques d'estimation des probabilités utiles à son application. Il est important de remarquer que ces méthodes sont parmi les plus performantes en "text mining" (recherche d'informations dans les documents texte).



Classifieur naïf de Bayes

Soit D un ensemble de descripteurs, soit $\{1, \dots, c\}$ l'ensemble des classes, sous les hypothèses usuelles d'existence de lois de probabilité, la règle de classification de Bayes est la procédure qui, à toute description d de D associe :

$$C_{\text{Bayes}}(d) = \arg \max_{k \in \{1, \dots, c\}} [P(k/d)] = \arg \max_{k \in \{1, \dots, c\}} [P(d/k) \times P(k)]$$

Mais, en règle générale, les quantités $P(d/k)$ et $P(k)$ ne sont pas connues. On suppose que D est un produit cartésien de domaines, on suppose également disposer d'un échantillon S d'exemples $(\bar{x}, c(\bar{x}))$. On souhaite classer un élément $\vec{d} = (d_1, \dots, d_n)$

La règle de classification de Bayes s'écrit : $C_{\text{Bayes}}(\vec{d}) = \arg \max_{k \in \{1, \dots, c\}} [P((d_1, \dots, d_n)/k) \times P(k)]$

Pour rendre la méthode effective, on souhaite remplacer $P((d_1, \dots, d_n)/k)$ et $P(k)$ par des estimations faites sur l'échantillon S . Pour toute classe k , on estime $P(k)$ par $\hat{P}(k)$ qui est la proportion d'éléments de classe k dans S . Par contre, l'estimation $P((d_1, \dots, d_n)/k)$ est difficile car le nombre de descriptions possibles peut être très grand et il faudrait un échantillon S de taille trop importante pour pouvoir estimer convenablement ces quantités. On fait donc l'hypothèse simplificatrice suivante : les valeurs des attributs sont indépendantes connaissant la classe.

Cette hypothèse permet d'utiliser l'égalité suivante :

$$P((d_1, \dots, d_n)/k) = \prod_{i \in \{1, \dots, n\}} P(d_i/k).$$

Maintenant, il suffit d'estimer, pour tout i et toute classe k , $P(d_i/k)$ par $\hat{P}(d_i/k)$ qui est la proportion d'éléments de classe k ayant la valeur d_i pour l'ième attribut.

Finalement, le classifieur naïf de Bayes associe à toute description d la classe

$$C_{\text{NaiveBayes}}(\vec{d}) = \arg \max_{k \in \{1, \dots, c\}} \left[\prod_{i \in \{1, \dots, n\}} \hat{P}(d_i/k) \times \hat{P}(k) \right]$$

expression dans laquelle les probabilités sont

estimées sur l'échantillon S . Cette méthode est simple à mettre en œuvre. Bien qu'elle soit basée sur une hypothèse fausse en général (les attributs sont rarement indépendants), elle donne cependant de bons résultats dans les problèmes réels. Elle fournit un seuil de performance pour d'autres méthodes.

k plus proches voisins [Denis et al., 2000], [Gilleron et al., 2000]

La méthode des k plus proches voisins (k PPV, nearest neighbor en anglais ou k NN) est une méthode dédiée à la classification qui peut être étendue à des tâches d'estimation. La méthode PPV est une méthode de raisonnement à partir de cas. Elle part de l'idée de prendre des décisions en recherchant un ou des cas similaires déjà résolus en mémoire. Contrairement aux autres méthodes de classification qui seront étudiées dans les sections suivantes, il n'y a pas d'étape d'apprentissage consistant en la construction d'un modèle à partir d'un échantillon d'apprentissage. C'est l'échantillon d'apprentissage, associé à une fonction de distance et d'une fonction de choix de la classe en fonction des classes des voisins les plus proches, qui constitue le modèle.



L'algorithme générique de classification d'un nouvel exemple par la méthode PPV est :

Algorithme de classification par k-PPV

donnée : un échantillon de m enregistrements classés $(\vec{x}, c(\vec{x}))$

entrée : un enregistrement \vec{y}

1. déterminer les k plus proches enregistrements de \vec{y}

2. combiner les classes de ces k exemples en une classe c

sortie : la classe de \vec{y} est $C(\vec{y}) = c$

Nous allons, par conséquent, présenter les différents choix possibles pour la définition de la distance et pour le mode de sélection de la classe du cas présenté.

Définition de la distance

Le choix de la distance est primordial au bon fonctionnement de la méthode. Quoique les distances les plus simples permettent d'obtenir des résultats satisfaisants (lorsque c'est possible), il faut être attentif à différents pièges. Tout d'abord, rappelons qu'une distance doit satisfaire les propriétés suivantes :

$$d(A, A) = 0$$

$$d(A, B) = d(B, A) \quad \text{pour tous points } A, B \text{ et } C \text{ de l'espace.}$$

$$d(A, B) \leq d(A, C) + d(C, B)$$

Dans notre problématique, les points sont des enregistrements d'une base de données. D'après la première de ces propriétés, le plus proche voisin d'un enregistrement est lui-même. Par la suite, nous désignons par plus proche voisin un point le plus proche de l'enregistrement autre que lui-même.

Pour définir la fonction de distance, on définit d'abord une distance sur chacun des champs, puis on combine ces distances pour définir la distance globale entre enregistrements. Commençons par étudier les choix possibles selon le type du champ.

Champs numériques

la distance entre deux valeurs numériques x et y peut être choisie égale à :

- $d(x, y) = |x - y|$, ou
- $d(x, y) = \frac{|x - y|}{d_{\max}}$ où d_{\max} est la distance maximale entre deux points.

Lorsque nous allons combiner les différentes distances sur les champs, il faut éviter que les valeurs de distance aient des ordres de grandeurs trop différents. Si les champs ont été préalablement normalisés sur l'intervalle [0,1], on prend la première distance, sinon on choisit la seconde qui est une distance normalisée.



Champs discrets

- les données binaires : 0 ou 1. On choisit $d(0,0) = d(1,1) = 0$ et $d(0,1) = d(1,0) = 1$
- les données énumératives : la distance vaut 0 si les valeurs sont égales et 1 sinon.
- les données énumératives ordonnées : elles peuvent être considérées comme des valeurs énumératives mais on peut également définir une distance utilisant la relation d'ordre. Par exemple, si un champ prend les valeurs 1, 2, 3, 4 et 5, on peut définir la distance en considérant 5 points de l'intervalle [0,1] avec une distance de 0,2 entre deux points successifs, on a alors $d(1,2) = 0,2$; $d(1,3) = 0,4$; ...; $d(4,5) = 0,2$

Autres champs

Une distance peut être définie sur de nombreux types de champs comme des textes, des images, des informations géographiques.

Il reste à combiner les distances entre champs pour définir une distance entre enregistrements.

Soit $\vec{x} = (x_1, \dots, x_n)$ et $\vec{y} = (y_1, \dots, y_n)$ deux enregistrements, d_1, \dots, d_n sont les distances définies sur les différents champs, la distance entre deux enregistrements peut être définie par :

- la distance euclidienne : $d_e(\vec{x}, \vec{y}) = \sqrt{d_1(x_1, y_1)^2 + \dots + d_n(x_n, y_n)^2}$,
- la sommation : $d_s(\vec{x}, \vec{y}) = d_1(x_1, y_1) + \dots + d_n(x_n, y_n)$

La distance euclidienne favorise les voisins dont tous les champs sont assez voisins, la distance par sommation permet de tolérer une distance importante sur l'un des champs.

Nous avons présenté différents choix possibles, selon le type des données, pour définir une distance sur un champ. Des variantes de ces définitions existent et, parfois, une distance spécifique peut être définie en fonction d'une expertise. De même, il existe de nombreuses variantes pour combiner les distances entre champs. Il est également possible de pondérer l'importance des différents champs.

Sélection de la classe

L'idée de la méthode est la recherche de cas similaires au cas à résoudre et d'utiliser les décisions des cas proches déjà résolus pour choisir une décision. La méthode la plus simple est de rechercher le cas le plus proche et de prendre la même décision. C'est la méthode 1-PPV (1-NN) du plus proche voisin. Si cette méthode peut fournir de bons résultats sur des problèmes simples pour lesquels les points (les enregistrements) sont bien répartis en groupes denses d'enregistrements de même classe, en règle générale, il faut considérer un nombre de voisins plus important pour obtenir de bons résultats. Nous supposons avoir déterminé les k voisins $((\vec{x}_1, c(\vec{x}_1)), \dots, (\vec{x}_k, c(\vec{x}_k)))$ d'un enregistrement \vec{y} auquel on souhaite attribuer une



classe $c(\vec{y})$. Une première façon de combiner les k classes des k voisins les plus proches est le vote majoritaire. Elle consiste simplement à prendre la classe majoritaire. Dans le cas de deux classes, on choisit une valeur de k impaire. Une seconde façon est le vote majoritaire pondéré. Chaque vote, c'est-à-dire chaque classe d'un des k voisins sélectionnés, est pondéré. Soit \vec{x}_i le voisin considéré, le poids de $c(\vec{x}_i)$ est inversement proportionnel à la distance entre l'enregistrement \vec{y} à classer et \vec{x}_i .

Dans les deux cas précédents, il est possible de définir une confiance dans la classe attribuée égale au rapport entre les votes gagnants et le total des votes. Lorsque la technique est appliquée à une tâche d'estimation, donc à prédire la valeur d'un attribut continu, la notion de vote perd tout son sens. Une première solution pour combiner les réponses est l'interpolation, c'est-à-dire de calculer une moyenne pondérée des réponses. Un défaut de cette solution est de "lisser" les données. Une deuxième solution est de considérer les k enregistrements avec la valeur prédite correspondante et d'utiliser les techniques de régression linéaire pour estimer la valeur en \vec{y} .

Mise en place de la méthode

La méthode ne nécessite pas de phase d'apprentissage. Le modèle sera constitué de l'échantillon d'apprentissage, de la distance et de la méthode de combinaison des voisins. Il faut choisir l'échantillon, c'est-à-dire les attributs pertinents pour la tâche de classification considérée et l'ensemble des enregistrements. Il faut veiller à disposer d'un nombre assez grand d'enregistrements par rapport au nombre d'attributs et que chacune des classes soit bien représentée dans l'échantillon choisi.

Le choix de la distance par champ et du mode de combinaison des distances se fait en fonction du type des champs et des connaissances préalables du problème. Il est possible d'affiner le choix de la distance en utilisant un ensemble test ou la validation croisée.

Le choix du nombre k de voisins peut être déterminé par utilisation d'un ensemble test ou par validation croisée. Une heuristique fréquemment utilisée est de prendre k égal au nombre d'attributs plus 1. La méthode de vote ou d'estimation peut aussi être choisie par test ou validation croisée.

Les arbres de décision [Denis *et al.*, 2000], [Gilleron *et al.*, 2000]

Dans cette section, nous étudions les algorithmes de génération d'arbres de décision à partir de données. Les deux algorithmes les plus connus et les plus utilisés (l'un ou l'autre ou les deux sont présents dans les environnements de fouille de données) sont CART [Adrians *et al.*, 1996] et C5 (version la plus récente de ID3 et C4.5 [Adrians *et al.*, 1996]). Ces algorithmes sont très utilisés pour leur performance et par le fait qu'ils génèrent des procédures de classification exprimables sous forme de règles.

Qu'est-ce qu'un arbre de décision?

Un arbre de décision est une représentation graphique d'une procédure de classification. Les nœuds internes de l'arbre sont des tests sur les champs ou attributs, les feuilles sont les classes. Lorsque les tests sont binaires, le fils gauche correspond à une réponse positive au test et le fils droit à une réponse négative. Pour classer un enregistrement, il suffit de descendre dans l'arbre selon les réponses aux différents tests pour l'enregistrement considéré.



Soit l'exemple d'arbre de décision suivant qui permet un diagnostic simplifié d'un patient suivant sa température et l'irritation de sa gorge.

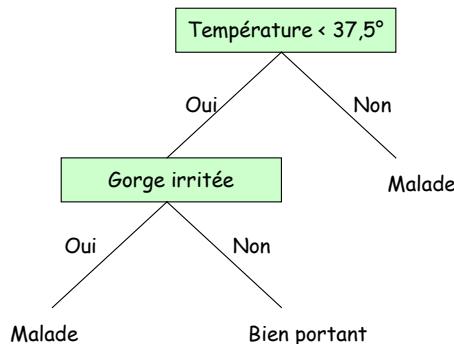


Figure 32 : exemple d'arbre de décision

Soit l'enregistrement \bar{x} défini par : nom = Yard, prénom = Drareg, âge = 38, Température=37,2°, Gorge Irritée = oui. Cet enregistrement sera classé Malade car Température <37,5° et Gorge irritée = oui.

On peut déjà remarquer quelques propriétés importantes des arbres de décision :

- la procédure de classification associée est compréhensible par tout utilisateur,
- la classe associée à un enregistrement particulier peut être justifiée,
- les attributs apparaissant dans l'arbre sont les attributs pertinents pour le problème de classification considéré.

On peut également remarquer qu'un arbre de décision est un système de règles. Il est immédiat de transformer l'arbre de la Figure 33 en :

Si Température > 37,5° alors Malade

Si Température < 37,5° et Gorge irrité = oui alors Malade

Si Température < 37,5° et Gorge irrité = non alors Bien portant

Les systèmes de règles construits sont particuliers. En effet, pour tout enregistrement une et une seule règle s'applique, c'est-à-dire que les règles sont exhaustives et mutuellement exclusives.

Apprentissage des arbres de décision

Dans cette section nous allons voir : comment à partir d'enregistrement sur des individus générer automatiquement un arbre de décision optimal ?

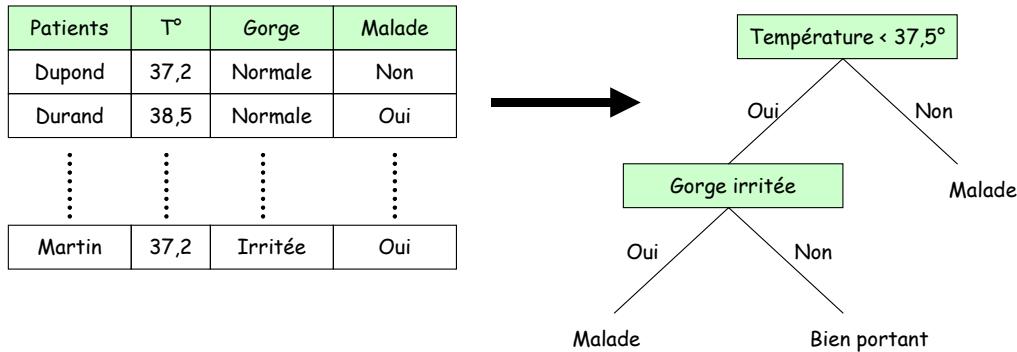


Figure 33 : Génération automatique d'arbre de décision

Avec en entrée, un échantillon de m enregistrements classés $(\vec{x}, c(\vec{x}))$, un algorithme d'apprentissage doit fournir en sortie un arbre de décision. La plupart des algorithmes procèdent de façon descendante, c'est-à-dire qu'ils choisissent la racine de l'arbre (en général un test) puis, récursivement, choisissent l'étiquette des fils. Pour simplifier la présentation, nous nous limitons, dans cette section à des problèmes où les attributs sont discrets et le nombre de classes égal à 2.

L'algorithme générique peut s'écrire :

Algorithme d'apprentissage par arbres de décision

donnée : un échantillon S de m enregistrements classés $(\vec{x}, c(\vec{x}))$

initialisation : arbre vide ; nœud courant : racine ; échantillon courant : S

répéter

décider si le nœud courant est terminal

si le nœud courant est terminal alors

étiqueter le nœud courant par une feuille

sinon

sélectionner un test et créer le sous-arbre

finsi

nœud courant : un nœud non encore étudié

échantillon courant : échantillon atteignant le nœud courant

jusque production d'un arbre de décision

élaguer l'arbre de décision obtenu

sortie : arbre de décision élagué

Nous allons préciser les différents points de cet algorithme en mettant en évidence les particularités des différents algorithmes.

Décider si le nœud courant est terminal

Le nœud courant est terminal si :

- il n'y a plus d'attributs disponibles, c'est-à-dire que sur le chemin menant de la racine au nœud courant tous les tests disponibles ont été utilisés ;
- tous les exemples de l'échantillon courant sont dans une même classe.



- Les critères précédents sont indiscutables, les autres critères sont spécifiques aux différents algorithmes et sont souvent paramétrables. Des exemples de critères sont :
- la proportion d'exemples d'une classe est supérieure à un seuil prédéfini. Par exemple, on décide de l'arrêt si une des classes contient plus de 95% des exemples.
- C5 utilise le critère suivant : s'il n'existe pas de test ayant au moins k éléments sur deux branches alors le nœud est terminal. L'objectif de ce critère est d'éviter une croissance trop grande de l'arbre par l'exploration de branches comprenant trop peu d'exemples. La valeur de k est, par défaut, égale à 2, elle peut être modifiée par l'utilisateur.

Etiqueter le nœud courant par une feuille

On étiquette le nœud courant par la classe majoritaire. Par exemple, si le nœud courant est terminal et s'il y a 5 exemples de classe 0 et 20 exemples de classe 1, on étiquette par 1. Cependant, pour certains problèmes, il se peut que les erreurs de classification d'une classe vers l'autre aient des conséquences différentes. C'est le cas, par exemple, pour un diagnostic médical pour lequel classer un individu malade comme sain ou classer un individu sain comme malade n'a pas les mêmes conséquences. Dans ce cas, il est possible de définir des coûts de mauvaise classification et la classe choisie le sera en fonction des coûts attribués.

Sélectionner un test

On suppose que le nœud courant n'est pas terminal. Soit S l'échantillon associé au nœud courant. Pour introduire les possibles critères de sélection du test, considérons l'exemple suivant : S contient 100 exemples, 60 de classe 0 et 40 de classe 1. Le nœud courant sera étiqueté par le couple (60,40). Supposons que deux tests soient disponibles, et que ces deux tests déterminent les répartitions suivantes :

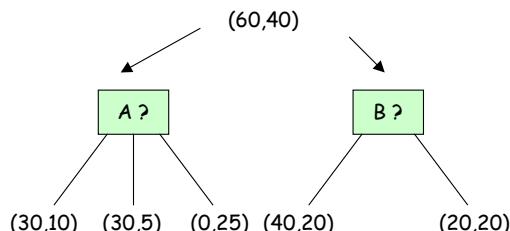


Figure 34 : Exemple de sélection d'un test

Pour choisir le test, on utilise des fonctions qui mesurent le "degré de mélange" des différentes classes. Pour les problèmes à deux classes, on peut utiliser une des fonctions suivantes :

- la fonction de Gini : $Gini(x) = 4x(1-x)$
- la fonction entropie : $Entropie(x) = -x \log x - (1-x) \log(1-x)$

où x désigne la proportion d'éléments dans l'une des deux classes. Ces deux fonctions sont à valeurs dans l'intervalle réel [0;1], prennent leur minimum pour x = 0 ou x = 1 (tous les exemples sont dans une même classe) et leur maximum lorsque x = 1/2 (les exemples sont également répartis entre les deux classes).



Choisissons, par exemple, la fonction de Gini.

Pour le nœud courant, $x = 60/100$ et $\text{Gini}(x) = 4 \times 60/100 \times (1 - 60/100) = 0.96$

Pour le test A :

- pour le premier fils (le plus à gauche), $x = 30/40 = 3/4$ et $\text{Gini}(x) = 0.75$,
- pour le second fils $x = 30/35 = 6/7$ et $\text{Gini}(x) = 0.49$,
- pour le troisième fils, $\text{Gini}(x) = 0$.

Pour le test B :

- pour le premier fils (le plus à gauche), $x = 40/60 = 2/3$ et $\text{Gini}(x) = 0.89$,
- pour le second fils $x = 20/20 = 1$ et $\text{Gini}(x) = 1$.

Pour comparer les deux tests, on estime le "degré de mélange espéré" en pondérant les degrés de mélange des fils par la proportion des exemples allant sur ce fils, on obtient:

- pour A : $40/100 \times 0.75 + 35/100 \times 0.49 + 25/100 \times 0 = 0.47$
- pour B : $60/100 \times 0.89 + 40/100 \times 1 = 0.93$

On choisit alors le test qui fournit le degré de mélange espéré minimum (pour l'exemple le test A). Souvent, on introduit le Gain qui est égal au degré de mélange du nœud courant diminué du degré de mélange espéré par l'introduction du test, on choisit alors le test qui apporte le gain maximal.

Elaguer l'arbre de décision obtenu

Il est possible de poursuivre la croissance de l'arbre jusqu'à obtention d'un arbre d'erreur nulle (si c'est possible : si il n'existe pas d'exemples ayant la même description mais des classes différentes) ou d'un arbre d'erreur mesurée sur l'ensemble d'apprentissage la plus petite possible. Cependant, l'objectif d'une procédure de classification est de bien classer des exemples non encore rencontrés, on parle de pouvoir de généralisation. Si l'algorithme fournit en sortie un arbre très grand qui classe bien l'échantillon d'apprentissage, on se trouve confronté au problème de sur-spécialisation : on a appris "par cœur" l'ensemble d'apprentissage, mais on n'est pas capable de généraliser. L'objectif de la phase d'élagage est d'obtenir un arbre plus petit (on élague des branches, c'est-à-dire que l'on détruit des sous-arbres) dans le but d'obtenir un arbre ayant un meilleur pouvoir de généralisation (même si on fait augmenter l'erreur sur l'ensemble d'apprentissage).

Les algorithmes d'apprentissage

CART

CART a été défini dans les années 80. Depuis, il a été intégré à de nombreux environnements de fouille de données sous de nombreuses variantes. Nous donnons ici ses principales particularités. A l'origine, l'algorithme ne considérait que des tests binaires. La fonction qui



mesure le degré de mélange (et donc le gain) est la fonction de Gini (les versions diffusées proposent d'autres choix). Pour l'élagage, on effectue un parcours ascendant de l'arbre construit. Pour décider si un sous arbre peut être élagué, on compare l'erreur réelle estimée de l'arbre courant avec l'arbre élagué. L'estimation de l'erreur réelle est mesurée sur un ensemble test ou par validation croisée.

C5

C5 est la version la plus récente d'un algorithme ID3 développé par R. Quinlan. L'algorithme peut prendre en compte des attributs d'arité quelconque. La fonction qui mesure le degré de mélange (et donc le gain) est la fonction entropie. Cette fonction a tendance à privilégier les attributs possédant un grand nombre de valeurs. Pour éviter ce biais, une fonction gain d'information est également disponible. L'élagage est effectué avec l'ensemble d'apprentissage par une évaluation pessimiste de l'erreur.

Techniques de classification non supervisée[Lebart, 2002], [Jain et al., 1999], [Bezdek, 1981], [Hartigan, 1975]

Principes de base

La classification non supervisée (*clustering en anglais*) consiste à segmenter un ensemble de vecteurs non labellisés (la classe n'est pas fournie) en groupes (*clusters*) qui possèdent les propriétés suivantes :

- Homogénéité dans les groupes, i.e. les données appartenant à un même cluster doivent être les plus similaires possibles.
- Hétérogénéité entre groupes, i.e. les données appartenant à différents clusters doivent être les plus dissemblables possibles.

Les notions de similarité / "dissimilitude" seront abordées dans la section suivante.

Un exemple illustratif d'une classification non supervisée est donné sur la figure 35. Les vecteurs d'entrée sont décrits sur la figure 35(a) et les clusters désirés sur la figure 35(b).

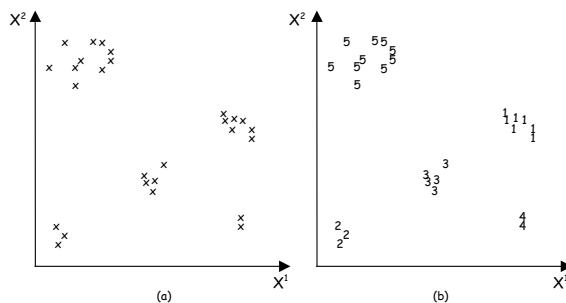


Figure 35 : Principes du Clustering

La classification non supervisée est une branche de l'analyse des données qui a donné lieu à des publications nombreuses et diversifiées. Nous nous bornerons ici aux principes de base des méthodes les plus largement utilisées.

Pour l'essentiel, les techniques de classification font appel à une démarche algorithmique et non aux calculs formalisés usuels. La définition des classes ne se fera qu'à partir d'une



formulation algorithmique : une série d'opérations est définie de façon récursive et répétitive. Il en découle que la mise en œuvre de la plupart des techniques de classification non supervisée ne nécessite que des notions mathématiques relativement élémentaires.

Il existe plusieurs familles d'algorithmes de classification non supervisée : les algorithmes conduisant directement à des partitions comme les méthodes d'agrégation autour de centres mobiles ; les algorithmes ascendants (ou encore agglomératifs) qui procèdent à la construction des classes par aggrégations successives des objets deux à deux, et qui fournissent une hiérarchie de partitions des objets; enfin les algorithmes descendants (ou encore divisifs) qui procèdent par dichotomies successives de l'ensemble des objets, et qui peuvent encore fournir une hiérarchie de partitions. On se limitera ici aux deux premières techniques de classification non supervisée:

- les groupements peuvent se faire par recherche directe d'une partition, en affectant les éléments à des centres provisoires de classes, puis en recentrant ces classes, et en affectant de façon itérative ces éléments. Il s'agit des techniques d'agrégation autour de centres mobiles, apparentées à la méthode des "nuées dynamiques", ou méthode "k-means", qui sont particulièrement intéressantes dans le cas des grands tableaux ;
- les groupements peuvent se faire par aggrégation progressive des éléments deux à deux. C'est le cas de la classification ascendante hiérarchique qui est présentée ici suivant plusieurs critères d'agrégation. Nous envisagerons d'une part la technique "du saut minimal" (single linkage) équivalente, d'un certain point de vue, à la recherche de l'arbre de longueur minimale, et d'autre part la technique d'agrégation "selon la variance", intéressante par la compatibilité de ses résultats avec certaines analyses factorielles.

Ces techniques présentent des avantages différents et peuvent être utilisées conjointement. Il est ainsi possible d'envisager une stratégie de classification basée sur un algorithme mixte, particulièrement adapté au partitionnement d'ensembles de données comprenant des milliers d'individus à classer.

Mesures de similarité

Toutes les techniques de *clustering* que nous allons étudier font référence à la notion de mesure de similarité entre deux vecteurs. La diversité des types et des échelles des dimensions des vecteurs donne une importance fondamentale à la mesure de similarité utilisée.

Les mesures de similarité les plus utilisées sont les mesures de distance.

Nous allons plus particulièrement nous intéresser aux mesures de distance pour des champs continus. Pour ce type de données, la distance euclidienne reste la plus utilisée.

$$d_2(x_i, x_j) = \left(\sum_{k=1}^d (x_{i,k} - x_{j,k})^2 \right)^{1/2} = \|x_i - x_j\|_2$$

avec $x_i = (x_{i,1}, \dots, x_{i,d})$ et $x_j = (x_{j,1}, \dots, x_{j,d})$ deux vecteurs de dimension d.

Cette distance est un cas particulier ($p=2$) de la métrique de Minkowski.

$$d_p(x_i, x_j) = \left(\sum_{k=1}^d (x_{i,k} - x_{j,k})^p \right)^{1/p} = \|x_i - x_j\|_p$$



Ces mesures posent un problème lorsque les échelles des données ne sont pas homogènes. En effet les grandes échelles sont favorisées dans le calcul de la distance.

Pour palier le problème, les données peuvent être normalisées (part rapport à la variance ou à une plage fixée).

La corrélation linéaire entre caractéristiques peut distordre les distances, cette distorsion peut être évitée en utilisant la distance de Mahalanobis dans le cas où les densités suivent une loi normale.

$$d_M(x_i, x_j) = (x_i - x_j) \Sigma^{-1} (x_i - x_j)^T \text{ avec } \Sigma \text{ matrice des covariances.}$$

Des recherches récentes dans le domaine de la reconnaissance de formes préconise la distance de Hausdorff $d_H(A, B)$ pour déterminer des similarités entre groupes (A et B) de vecteurs.

$$d_H(A, B) = \max\{d_S(A, B), d_S(B, A)\}$$

$$\text{avec } d_S(A, B) = \max_{a \in A} \left[\min_{b \in B} (\|a - b\|) \right] \text{ et } d_S(B, A) = \max_{b \in B} \left[\min_{a \in A} (\|a - b\|) \right]$$

Algorithme de calcul de h distance de Hausdorff

1. $h = 0$
2. Pour tous les points a_i de A,
 - 2.1 $\text{lePlusCourt} = \text{Infini}$;
 - 2.2 Pour chaque point b_j de B
 - $d_{ij} = d(a_i, b_j)$
 - Si $d_{ij} < \text{lePlusCourt}$ then
 $\text{lePlusCourt} = d_{ij}$
 - 2.3 if $\text{lePlusCourt} > h$ then
 $h = \text{lePlusCourt}$

Certaines mesures de similarité tiennent compte des points voisins. Ces points voisins sont appelés Contexte. La similarité entre deux points x_i et x_j dans un contexte donné s'exprime par : $s(x_i, x_j) = f(x_i, x_j, C)$ avec C ensemble des voisins.

La métrique de *Mutual neighbor distance* (MND) utilise cette notion de contexte :

$$MND(x_i, x_j) = NN(x_i, x_j) + NN(x_j, x_i) \text{ avec } NN(x_i, x_j) \text{ le nombre de voisins de } x_j \text{ par rapport à } x_i.$$

Sur la figure 36(a) :

- $NN(A, B) = NN(B, A) = 1$ et donc $MND(A, B) = 2$.
- $NN(B, C) = 1$ et $NN(C, B) = 2$ donc $MND(B, C) = 3$.

La figure 36(b) a été obtenue en rajoutant les points D, E et F. Sur cette figure : $MND(A, B) = 5$ et $MND(B, C) = 3$. Le voisinage (Contexte de A) a augmenté la dissimilarité entre A et B, et donc la similarité entre B et C.

La MND n'est pas une distance (elle ne satisfait pas l'inégalité triangulaire) mais elle a été appliquée avec succès dans plusieurs algorithme de *clustering*.

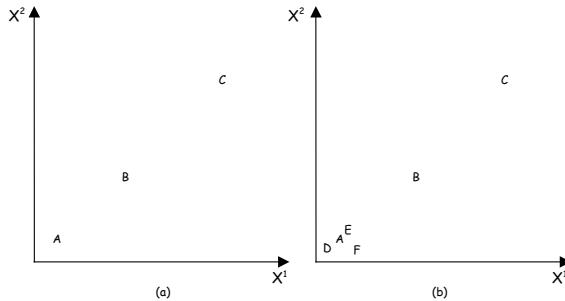


Figure 36 : Mutual neighbor distance

D'autre mesure de similarité introduise en plus du contexte la notion de concepts.

La mesure de similarité peut alors s'exprimer par $s(x_i, x_j) = f(x_i, x_j, C, Concepts)$.

Cette notion de concept est illustrée sur la figure 37. La distance euclidienne entre les points A et B est beaucoup plus petite que la distance entre B et C. En utilisant cette distance comme mesure de similarité, A serait considéré comme beaucoup plus similaire à B qu'à C. En introduisant les concepts d'appartenance à un rectangle et une ellipse la similarité s'inverse alors.

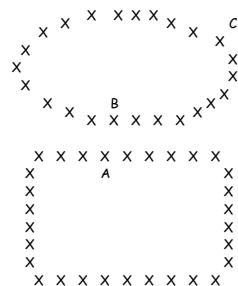


Figure 37 : Mesure de distance avec concept

Centres mobiles

Bien qu'elle ne fasse appel qu'à un formalisme limité et que son efficacité soit dans une large mesure attestée par les seuls résultats expérimentaux, la méthode de classification autour de centres mobiles est probablement la technique de partitionnement la mieux adaptée actuellement aux vastes recueils de données ainsi que la plus utilisée pour ce type d'application.

Elle est particulièrement intéressante pour les gros fichiers numériques car les données sont traitées en lecture directe : le tableau des données, conservé sur une mémoire auxiliaire (disque, CD-ROM), est lu plusieurs fois de façon séquentielle, sans jamais encombrer de zones importantes dans la mémoire centrale de l'ordinateur.

Bases théoriques de l'algorithme

Soit un ensemble I de n individus à partitionner, caractérisés par p caractères ou variables. On suppose que l'espace \mathbb{R}^p supportant les n points-individus est muni d'une distance appropriée notée d . On désire constituer au maximum q classes. Les étapes de l'algorithme sont illustrées par la figure 38.

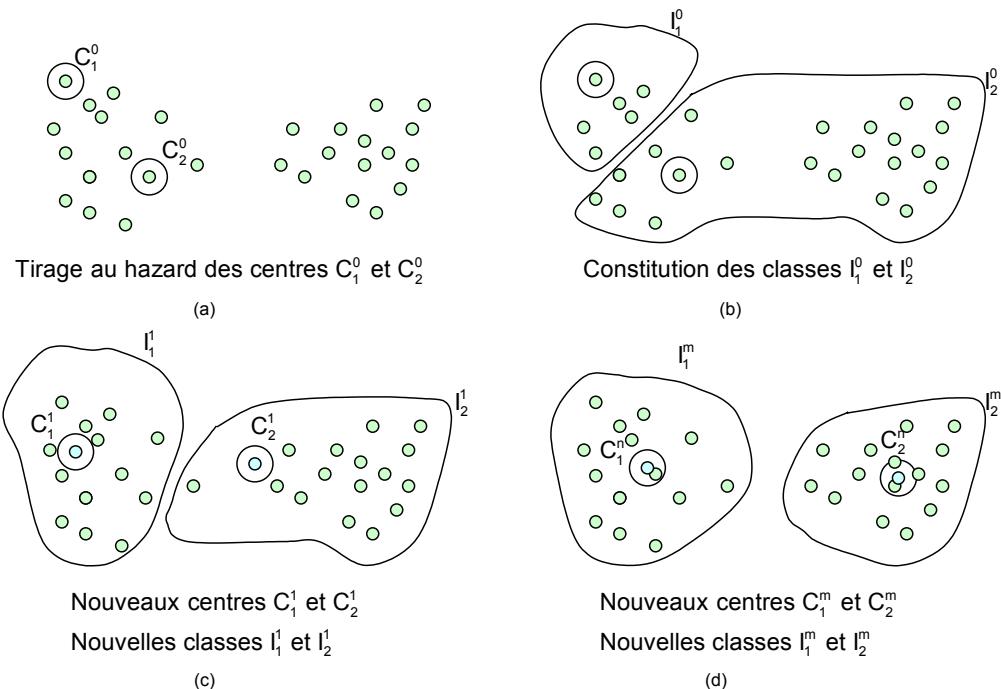


Figure 38 : Etapes de l'algorithme des centres mobiles

Étape 0 : figure 38(a)

On détermine q centres provisoires de classes (par exemple, par tirage aléatoire sans remise de q individus dans la population à classifier). Les q centres : $\{C_1^0, \dots, C_k^0, \dots, C_q^0\}$ induisent une première partition P^0 de l'ensemble des individus I en q classes : $\{I_1^0, \dots, I_k^0, \dots, I_q^0\}$. Ainsi l'individu i appartient à la classe I_k^0 s'il est plus proche de C_k^0 que de tous les autres centres.

Étape 1: figure 38 (b) et (c)

On détermine q nouveaux centres de classes : $\{C_1^1, \dots, C_k^1, \dots, C_q^1\}$ en prenant les centres de gravité des classes qui viennent d'être obtenues : $\{I_1^0, \dots, I_k^0, \dots, I_q^0\}$. Ces nouveaux centres induisent une nouvelle partition P^1 de I construite selon la même règle que pour P^0 . La partition P^1 est formée des classes notées : $\{I_1^1, \dots, I_k^1, \dots, I_q^1\}$

Étape m : figure 38(d)

On détermine q nouveaux centres de classes : $\{C_1^m, \dots, C_k^m, \dots, C_q^m\}$ en prenant les centres de gravité des classes qui ont été obtenues lors de l'étape précédente, $\{I_1^{m-1}, \dots, I_k^{m-1}, \dots, I_q^{m-1}\}$.



Ces nouveaux centres induisent une nouvelle partition P^m de l'ensemble I formée des classes :

$$\{I_1^m, \dots, I_k^m, \dots, I_q^m\}$$

Le processus se stabilise nécessairement et l'algorithme s'arrête soit lorsque deux itérations successives conduisent à la même partition, soit lorsqu'un critère convenablement choisi (par exemple, la mesure de la variance intra-classes) cesse de décroître de façon sensible, soit encore parce qu'un nombre maximal d'itérations a été fixé *a priori*.

Généralement, la partition obtenue finalement dépend du choix initial des centres.

Techniques connexes

Il existe de nombreux algorithmes dont le principe général est voisin de l'algorithme d'agrégation autour de centres mobiles mais qui en diffèrent cependant sur certains points.

Ainsi, dans la technique des *nuées dynamiques*, les classes ne sont pas caractérisées par un centre de gravité, mais par un certain nombre d'individus à classer, dénommés "étalons", qui constituent alors un "noyau" ayant pour certaines utilisations un meilleur pouvoir descriptif que des centres ponctuels. Ce formalisme a permis plusieurs généralisations de la méthode.

La méthode dite des *k-means* (*k-moyennes*) commence effectivement par un tirage pseudo-aléatoire de centres ponctuels. Cependant la règle de calcul des nouveaux centres n'est pas la même. On n'attend pas d'avoir procédé à la réaffectation de tous les individus pour modifier la position des centres : chaque réaffectation d'individus entraîne une modification de la position du centre correspondant. En une seule itération, cette procédure peut ainsi donner une partition de bonne qualité. Mais celle-ci dépendra de l'ordre des individus sur le fichier, ce qui n'est pas le cas pour la technique exposée précédemment .

Classification hiérarchique

Principes

Les principes généraux communs aux diverses techniques de classification ascendante hiérarchique sont également extrêmement simples. Le principe de l'algorithme consiste à créer, à chaque étape, une partition obtenue en agrégeant deux à deux les éléments les plus proches. On désignera alors par élément à la fois les individus ou objets à classer eux-mêmes et les regroupements d'individus générés par l'algorithme. Il y a différentes manières de considérer le nouveau couple d'éléments agrégés, d'où un nombre important de variantes de cette technique.

L'algorithme ne fournit pas une partition en q classes d'un ensemble de n objets mais une hiérarchie de partitions, se présentant sous la forme d'arbres appelés également dendrogrammes et contenant n-1 partitions. L'intérêt de ces arbres est qu'ils peuvent donner une idée du nombre de classes existant effectivement dans la population.



#Point	x1	x2
1	1.0000	2.0000
2	2.5000	4.5000
3	2.0000	2.0000
4	4.0000	1.5000
5	4.0000	2.5000

Distance entre les points

	1	2	3	4	5
1	0	2.9155	1.0000	3.0414	3.0414
2	2.9155	0	2.5495	3.3541	2.5000
3	1.0000	2.5495	0	2.0616	2.0616
4	3.0414	3.3541	2.0616	0	1.0000
5	3.0414	2.5000	2.0616	1.0000	0

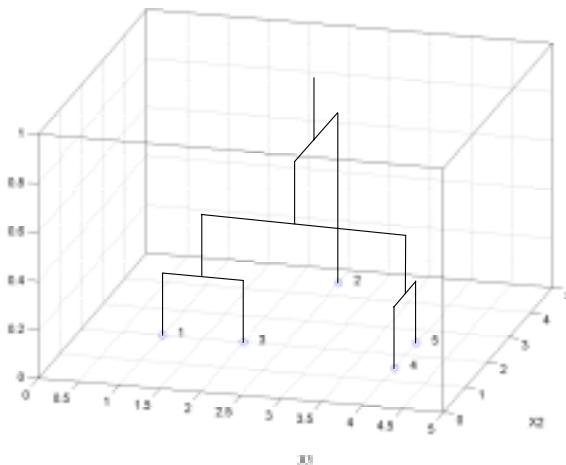


Figure 39 : Dendrogramme ou arbre hiérarchique

Chaque coupure d'un arbre fournit une partition, ayant d'autant moins de classes et des classes d'autant moins homogènes que l'on coupe plus haut.

Distances entre éléments et entre groupes

On suppose au départ que l'ensemble des individus à classer est muni d'une distance²⁵. Ceci ne suppose donc pas que les distances soient toutes calculées au départ : il faut pouvoir les calculer ou les recalculer à partir des coordonnées des points-individus, celles-ci devant être accessibles rapidement. On construit alors une première matrice de distances entre tous les individus.

Une fois constitué un groupe d'individus, il convient de se demander ensuite sur quelle base on peut calculer une distance entre un individu et un groupe et par la suite une distance entre deux groupes. Ceci revient à définir une stratégie de regroupements des éléments, c'est-à-dire se fixer des règles de calcul des distances entre groupements disjoints d'individus, appelées critères d'agrégation. Cette distance entre groupements pourra en général se calculer directement à partir des distances des différents éléments impliqués dans le regroupement.

Par exemple, si x, y, z sont trois objets, et si les objets x et y sont regroupés en un seul élément noté h, on peut définir la distance de ce regroupement à z par la plus petite distance des divers éléments de h à z :

$$d(h,z) = \text{Min} \{d(x,z), d(y,z)\}$$

²⁵ Il s'agira parfois simplement d'une mesure de dissimilarité. Dans ce cas, l'inégalité triangulaire $d(x,y) \leq d(x,z) + d(y,z)$ n'est pas exigée).



Cette distance s'appelle le saut minimal (single linkage) et constitue un critère d'agrégation.

On peut également définir la distance du saut maximal (ou diamètre) en prenant la plus grande distance des divers éléments de h à z :

$$d(h,z) = \text{Max} \{d(x,z), d(y,z)\}$$

Une autre règle simple et fréquemment employée est celle de la distance moyenne ; pour deux objets x et y regroupés en h :

$$d(h,z) = \frac{\{d(x,z) + d(y,z)\}}{2}$$

Plus généralement, si x et y désignent des sous-ensembles disjoints de l'ensemble des objets, ayant respectivement n_x et n_y éléments, h est alors un sous-ensemble formé de $n_x + n_y$ éléments et on définit :

$$d(h,z) = \frac{\{n_x d(x,z) + n_y d(y,z)\}}{n_x + n_y}$$

Algorithme de classification

L'algorithme fondamental de classification ascendante hiérarchique se déroule de la façon suivante :

Soit n éléments à classer

- Étape 1

On construit la matrice de distances entre les n éléments et l'on cherche les deux plus proches, que l'on agrège en un nouvel élément. On obtient une première partition à $n-1$ classes;

- Étape 2

On construit une nouvelle matrice des distances qui résultent de l'agrégation, en calculant les distances entre le nouvel élément et les éléments restants (les autres distances sont inchangées). On se trouve dans les mêmes conditions qu'à l'étape 1, mais avec seulement ($n-1$) éléments à classer et en ayant choisi un critère d'agrégation. On cherche de nouveau les deux éléments les plus proches, que l'on agrège. On obtient une deuxième partition avec $n-2$ classes et qui englobe la première;

- Étape m

On calcule les nouvelles distances, et l'on réitère le processus jusqu'à n'avoir plus qu'un seul élément regroupant tous les objets et qui constitue la dernière partition.

Nous illustrons cette procédure en prenant comme objets à classer cinq points (figure 40).

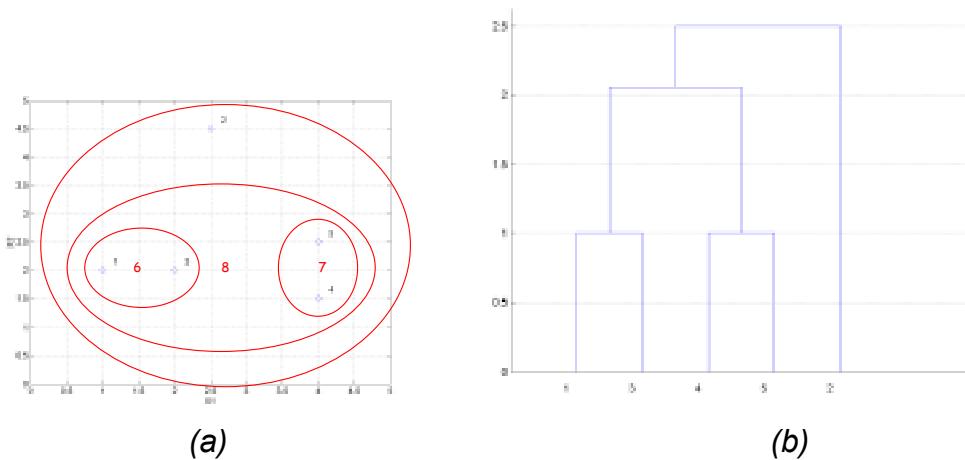


Figure 40 - Agglomération progressive de 5 points

Les regroupements successifs peuvent être représentés par un arbre ou dendrogramme, comme le montre la figure 40(b).

Approche floue de la classification

Modèles flous

Les modèles flous ou systèmes d'inférence floue (SIF) permettent de représenter le comportement de systèmes complexes. Pour comprendre le fonctionnement et surtout comment identifier ces SIF nous devons dans un premier temps définir les notions de sous-ensembles flous et de logique floue.

La théorie des sous-ensembles flous

Les sous-ensembles flous permettent une représentation simple des incertitudes et imprécisions liées aux informations et aux connaissances. Leur principal avantage est d'introduire le concept d'appartenance graduelle à un ensemble alors qu'en logique ensembliste classique cette appartenance est binaire (appartient ou n'appartient pas à un ensemble). Pour illustrer ces concepts, considérons la figure suivante sur laquelle nous avons représenté trois sous-ensembles flous "pH Acide", "pH Neutre" et "pH Basique". A chacun de ces sous-ensembles, nous associons une fonction d'appartenance dont les valeurs varient de 0 à 1. Ces fonctions d'appartenance permettent de définir le degré d'appartenance d'un pH donné (par exemple pH 9) aux différents sous-ensembles flous. Pour l'exemple choisi, la valeur du pH 9 n'appartient pas au sous-ensemble flou "pH Acide". Par contre, elle appartient aux sous-ensembles "pH Neutre" et "pH Basique" avec des degrés d'appartenance respectifs de 0.15 et 0.6.

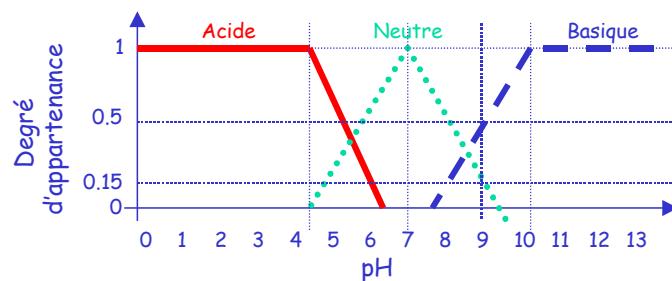


Figure 41 - Exemple de sous ensembles flous



Logique floue - Modèle Flou - Système d'Inférence Floue (SIF)

La logique floue permet de décrire des relations d'implication entre des sous-ensembles flous sous la forme de règle floue de la forme :

SI "pH est Acide" ALORS "danger est Très_Elevé".

La prémissse de la règle est constituée de la partie gauche ("pH est Acide") et la conclusion de la règle de la partie droite ("danger est Très_Elevé")

Ph et danger sont des variables floues.

Acide et Très_Elevé sont des qualificatifs linguistiques.

Un ensemble de règles de ce type forme un modèle flou ou SIF.

Règle 1 : SI "pH est Acide" ALORS "danger est Très_Elevé"

Règle 2 : SI "pH est Neutre" ALORS "danger est Faible"

Règle 3 : SI "pH est Basique" ALORS "danger est Elevé"

A partir de l'énoncé de ces règles et la connaissance de faits précis et/ou flous sur les entrées du SIF (pour l'exemple le pH mesuré) nous sommes capables de définir une valeur précise et/ou floue de la ou les sorties du SIF (pour l'exemple une mesure du danger).

Il existe plusieurs types de SIF suivant le type de règles utilisées. Les modèles du type Takagi-Sugeno sont les plus souvent utilisés. Dans ces modèles, la sortie de chaque règle est une combinaison linéaire des variables d'entrée.

Une représentation classique d'un SIF de Takagi-Sugeno avec deux entrées et une sortie est donnée sur la figure 42.

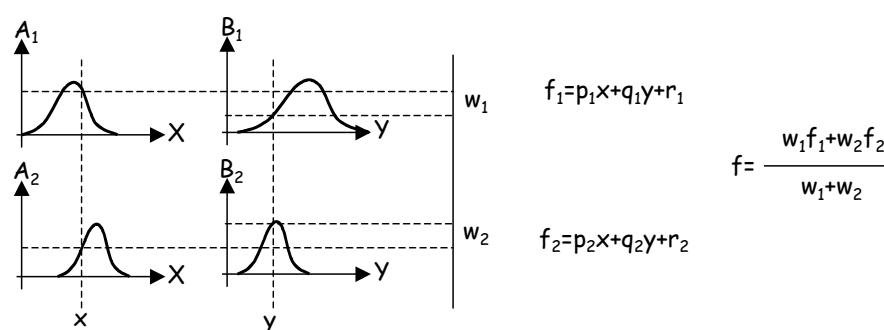


Figure 42 - Exemple d'un SIF de Takagi-Sugeno

Le SIF représenté sur cette figure possède deux règles de la forme :

Règle 1 : SI x EST A_1 ET y EST B_1 ALORS $z_1 = p_1 x + q_1 y + r_1$

Règle 2 : SI x EST A_2 ET y EST B_2 ALORS $z_2 = p_2 x + q_2 y + r_2$

Avec : A_j et B_j les fonctions d'appartenance, p_j , q_j , et r_j les paramètres linéaires de sortie pour chaque règle.

Le processus de raisonnement peut être décomposé en quatre étapes :



1. Calcul des degrés d'appartenance de chaque entrée de chaque règle (a₁, a₂, b₁, b₂ pour l'exemple de la figure 42)
2. Application de l'opérateur flou ET (multiplication pour l'exemple) pour la partie prémissie.
3. Calcul de la partie conclusion de chaque règle (combinaison linéaire des entrée plus une constante)
4. Agrégation des sorties des règles (moyenne pondérée pour l'exemple de la figure 42)

L'expression des règles composant le SIF peut s'obtenir soit à partir de connaissances d'expert soit à partir d'analyse de données on parlera alors de méthodes d'identification de SIF. Les méthodes d'identification sont nombreuses, nous présentons dans ce document les plus courantes basées sur la notion de classification floue présentée dans le paragraphe suivant.

La classification floue

Le groupement (clustering) de données numériques forme la base de nombreux algorithmes de classification et de modélisation de systèmes. L'objectif du clustering est de mettre en évidence des groupements de données à partir de bases de données importantes, produisant ainsi une représentation concise du comportement du système.

Méthode du Fuzzy C-means

Cette méthode fait partie des méthodes de groupement par réallocation. Elles consistent à optimiser par itération la position des groupements à partir d'une solution initiale constituée du nombre de groupements et de leurs positions (centres des clusters). Ces méthodes sont généralement basées sur la minimisation d'une fonction caractérisant l'objectif du groupement.

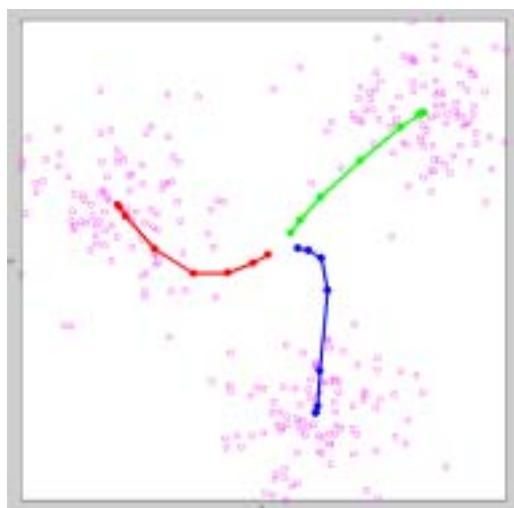


Figure 43 - Exemple d'application du Fuzzy c-means

La méthode de groupement par réallocation la plus connue est la méthode du K-means. En 1981, Bezdek a proposé une version floue du K-means : l'algorithme du Fuzzy C-means (FCM). La figure précédente illustre l'application de cette méthode sur un nuage de points en deux dimensions (X, Y). A l'état initial les trois centres sont positionnés au centre du nuage. Puis après quelques itérations, ils convergent vers trois positions d'équilibre qui sont les barycentres des trois clusters constituant le nuage de points.



Méthode du " Subtractive clustering "

L'inconvénient majeur de l'algorithme du FCM est que l'on doit connaître a priori le nombre et la position des centres des clusters. La méthode du " subtractive clustering " [Chiu, 1994] permet d'estimer et de positionner automatiquement les clusters initiaux. C'est une méthode itérative basée sur une pondération des points à traiter. Plus un point aura de voisins proches, plus son potentiel sera élevé. A partir du calcul initial des différents potentiels, l'algorithme permet de sélectionner les meilleurs représentants du nuage de points et de définir ainsi les clusters principaux.

La figure 44 illustre le principe de formation des clusters pour un espace trois dimensions (Y_1 , Y_2 , Z). La figure 44a décrit le nuage de points initial. Les figures suivantes illustrent l'évolution des potentiels des points, le choix des centres et l'affectation des fonctions d'appartenance aux centres.

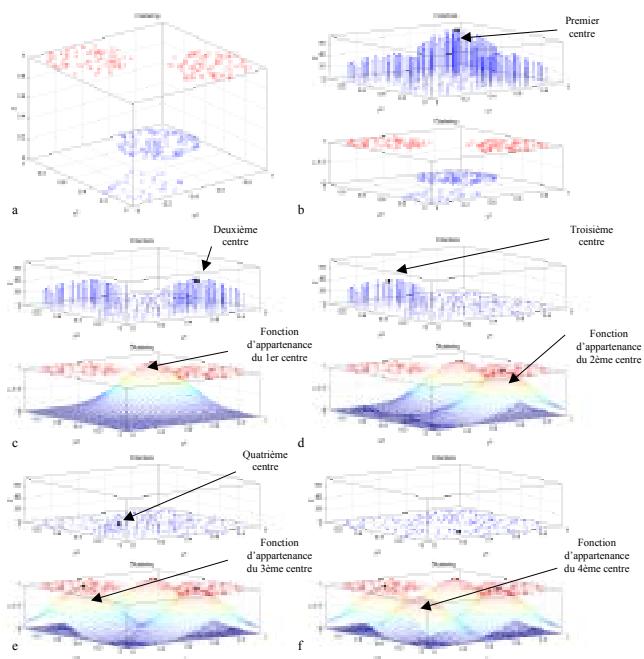


Figure 44 - Etapes de l'algorithme du Subtractive Clustering

ANFIS : Adaptive-Network-Based Fuzzy Inference System

ANFIS est une méthode permettant de régler les paramètres d'un SIF en utilisant une technique d'optimisation basée sur la rétro propagation du gradient de l'erreur.

Cette méthode consiste à représenter un SIF par un réseau adaptatif et y appliquer une procédure d'apprentissage pour régler les paramètres par rapport à un ensemble d'entrées et de sorties.

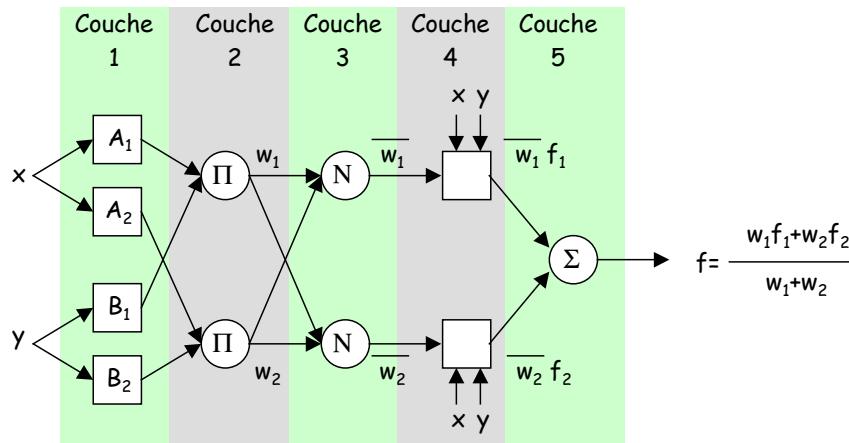


Figure 45 - Exemple d'un réseau ANFIS

Un réseau adaptatif est un réseau multicouches possédant deux type de nœuds : adaptatifs et fixes. Les nœuds adaptatifs réalisent des fonctions entre des entrées et une sortie et les paramètres de cette fonction peuvent être optimisés par une procédure d'apprentissage. Les nœuds fixes réalisent une fonction prédéfinie entre des entrées et une sortie.

La représentation d'un SIF en utilisant le formalisme ANFIS est donnée sur la figure 45. Les carrés représentent les nœuds adaptatifs et les cercles des nœuds fixes.

- La couche 1 réalise le calcul des degrés d'appartenance. Classiquement les fonctions d'appartenance utilisées sont des Gaussiennes.

$$A_j(x) = \exp\left[-\left(\frac{x-c_j}{a_j}\right)^2\right]$$

- La couche 2 réalise le ET flou entre les éléments des prémisses des règles (opérateur multiplication)

$$w_j = A_j(x) \times B_j(y)$$

- La couche 3 réalise le calcul de l'importance relative d'une règle par rapport aux autres.

$$\bar{w}_j = \frac{w_j}{\sum_{i=1}^{\text{Nombre de Règles}} w_i}$$

- La couche 4 réalise le calcul de la sortie pondérée d'une règle (combinaison linéaire des entrées)

$$\bar{w}_j \times z_j = \bar{w}_j \times (p_j x + q_j y + r_j)$$

- La couche 5 réalise le calcul de la sortie du SIF.

$$z = \sum_{j=1}^{\text{Nombre de Règles}} \bar{w}_j \times z_j$$

L'auteur de la méthode propose une procédure hybride pour le réglage des paramètres du modèle. Pour les paramètres linéaires une méthode classique des moindres carrés est utilisée. Pour les paramètres non linéaires, une technique d'optimisation basée sur la rétro propagation du gradient de l'erreur est préconisée.



1.1.4. Les règles d'associations

Attirés par les résultats qu'elles promettent et les enjeux auxquels elles sont consacrées, de nombreux algorithmes de fouille de données sont destinés à la recherche de règles d'association. Nous porterons, dans cette section, un regard couvrant l'ensemble des méthodes issues des travaux effectués, au travers des différents courants qui ont émergé. Les différentes méthodes présentées ont toutes un intérêt que chaque application peut exploiter en fonction de ses caractéristiques.

Problématique

Formellement, le problème des règles d'associations est présenté dans [Agrawal *et al.*, 1993] de la façon suivante : soit $I = \{i_1, i_2, \dots, i_m\}$ un ensemble d'items. Soit DB un ensemble de transactions²⁶, tel que chaque transaction T , dotée d'un identifiant unique noté (TID), soit constituée d'un ensemble d'items, appelés *itemset* vérifiant $T \subseteq DB$. T contient X , un ensemble d'items de I , $X \subseteq T$. Une règle d'association est une implication de la forme $X \rightarrow Y$, où $X \subseteq I$, $Y \subseteq I$ et $X \cap Y = \emptyset$.

Une règle d'association $X \rightarrow Y$ traduit le fait que si les objets X sont présents dans une transaction, alors les objets Y le sont avec une certaine probabilité. La force d'une règle est mesurée par deux indicateurs :

- Une règle est plus ou moins vraie pour un certain pourcentage constituant le support de la règle. Le support de la règle est donc une mesure indiquant le pourcentage de transactions de DB qui vérifient une règle d'association. En désignant par $|X|$ le nombre de transactions comportant l'ensemble X et par $|DB|$ le nombre total de transactions dans la base nous avons :

$$\text{Support}(X \rightarrow Y) = |X \& Y| / |DB| = \% \text{ de transactions vérifiant la règle.}$$

Il faut remarquer que le support de $X \rightarrow Y$ est le même que celui de $Y \rightarrow X$. Il est donc possible de simplement le noter $\text{support}(XY)$ car il s'agit du support de l'ensemble des produits $X Y$ ($X \cup Y$).

- Pour mesurer la validité d'une règle, la notion de confiance a été introduite et correspond à la probabilité conditionnelle $P(Y/X)$. Elle indique donc le pourcentage de transactions qui vérifient la conclusion d'une règle d'association parmi celles qui vérifient la prémissse. En utilisant la notation introduite ci-dessus, nous avons :

$$\text{Confiance}(X \rightarrow Y) = |XY| / |X| = \% \text{ de transactions qui vérifient l'implication.}$$

La confiance peut donc se calculer par rapport au support de la manière suivante :

$$\text{Confiance}(X \rightarrow Y) = \text{Support}(XY) / \text{Support}(X).$$

Ainsi, à partir d'une base de transactions DB , la problématique de la recherche de règles d'association consiste à extraire toutes les règles d'association dont le support est supérieur au support spécifié par l'utilisateur (*minSupp*) et dont la confiance est supérieure à la confiance spécifiée par l'utilisateur (*minConf*). La recherche de règles d'association s'effectue donc par calcul des supports de tous les ensembles d'objets de la base et par conservation des

²⁶ Nous considérons ici le terme de transaction dans le sens d'une transaction financière et non pas celui d'une transaction dans une base de données.



ensembles dont le support est supérieur au minimum demandé (minSupp). Un ensemble ayant un support supérieur à minSupp est qualifié de *fréquent*.

Tous les algorithmes de recherche de règles d'association utilisent les propriétés suivantes de manière à optimiser la recherche de tous les ensembles fréquents.

Propriété 1 (Support pour les sous-ensembles) :

Si $X \subseteq Y$ pour les itemsets X et Y alors $\text{Support}(X) \geq \text{Support}(Y)$ car toutes les transactions de DB qui supportent X supportent nécessairement Y.

Propriété 2 (Les sur-ensembles d'ensembles non fréquents ne sont pas fréquents)

Si l'itemset X ne vérifie pas le support minimal dans DB, i.e. $\text{Support}(X) < \text{minSupp}$, alors tous les sur-ensembles Y de A ne seront pas fréquents car $\text{Support}(Y) \leq \text{Support}(X) < \text{minSupp}$ (Cf. Propriété 1).

Propriété 3 (Les sous-ensembles d'ensembles fréquents sont fréquents)

Si l'itemset Y est fréquent dans DB, i.e $\text{Support}(Y) \geq \text{minSupp}$, alors tout sous-ensemble X de Y est fréquent dans DB car $\text{Support}(X) \geq \text{Support}(Y) \geq \text{minSupp}$ (Cf. Propriété 1). En particulier si $A = \{i_1, i_2, \dots, i_k\}$ est fréquent, tous les $(k-1)$ -sous-ensembles sont fréquents. L'inverse n'est pas vrai.

En résumé, la problématique de la recherche de règles d'association se divise en deux étapes distinctes :

1. Rechercher dans un premier temps tous les itemsets fréquents, i.e tous les itemsets dont le nombre d'apparition dans la base est supérieur à minSupp . Si la base possède m items, 2^m items sont alors potentiellement fréquents, d'où le besoin de méthodes efficaces pour limiter cette recherche exponentielle.
2. Générer, à partir de ces itemsets fréquents, des règles dont la confiance est supérieure à minConf . Cette étape est relativement simple dans la mesure où il s'agit de conserver les règles du type $X \rightarrow Y = \text{Support}(XY)/\text{Support}(X)$, ayant un taux de confiance suffisant.

La deuxième problématique étant triviale, nous nous intéressons par la suite à la recherche des itemsets fréquents dans une base de données. La plupart des algorithmes de recherche utilisent les mêmes principes généraux (Générer-Elaguer), que nous présentons ici. Extraire toutes les règles pertinentes d'une base signifie examiner toutes les combinaisons d'items possibles et vérifier leur fréquence et confiance par rapport aux données effectivement stockées. L'examen de toutes les combinaisons d'items revient à explorer, comme espace de recherche, l'ensemble des parties de l'ensemble des itemsets I. Les algorithmes par niveaux s'avèrent très performants pour procéder à une telle exploitation. L'idée générale est d'opérer de manière itérative en traitant, à chaque pas de l'algorithme, des itemsets de longueur croissante.

Fonction algoGenerique

Entrée :Un support minimum minSupp et une base de données DB.

Sortie : L'ensemble L des itemsets ayant une fréquence d'apparitions supérieure à minSupp .

$k=1; L = \emptyset;$

$C_1=\{\{i\} / i \in I\};$

Pour chaque $d \in D$ **compterSupport**(C_1 , minSupp , d);

$L_1=\{c \in C_1 / \text{Support}(c) > \text{minSupp}\};$



```
Tant que Lk ≠ ∅
    Ck = genererCandidats(T, k);
    Pour chaque d ∈ D compterSupport(Ck, minSupp, d);
    Lk = {c ∈ Ck / Support(c) > minSupp};
    L ← L ∪ Lk;
    k = k + 1;
Retourner L ;
```

Ainsi la première étape concerne des itemsets simplement réduits à un singleton, la deuxième manipule des couples d'items et, de manière générale, la $i^{\text{ème}}$ étape opère sur des itemsets contenant i éléments (ils sont appelés i -itemsets, ou itemsets de longueur i). Supposons qu'au début de la $i^{\text{ème}}$ étape, l'ensemble de toutes les combinaisons de i items à examiner soit fourni, il s'agit de l'ensemble des candidats. Il suffit alors, pour chaque candidat, de vérifier la présence effective dans la base, par une simple opération de comptage (*compterSupport*). Cette opération sous-entend le parcours complet de la base. Si cette dernière satisfait le support minimum, i.e. si le nombre d'occurrences du candidat considéré est supérieur au support minimum, celui-ci est considéré comme fréquent. Dans ce cas, il se peut qu'un sur-ensemble de cette combinaison se révèle également fréquent lors de l'étape ultérieure. Le candidat examiné est donc conservé pour être traité à l'étape suivante. A l'opposé, si l'opération de comptage conclut à la non fréquence d'un candidat, il devient inutile de calculer ses sur-ensembles (Propriété 2). Une telle combinaison est donc éliminée de l'espace de recherche préparé pour l'étape suivante (phase d'élagage). Ce dernier est donc réduit à l'ensemble de tous les candidats de longueur i qui sont validés, i.e. fréquents, après le parcours de la base.

Il s'agit à présent, à partir de cet ensemble, de construire toutes les combinaisons de $(i+1)$ -items qu'il conviendra d'examiner au pas suivant. Cette procédure d'extension, appelée phase de génération de candidats (*genererCandidat*), diverge selon les approches, mais la plupart du temps opère de manière optimale en réutilisant la connaissance acquise au cours des pas précédents. Le $i^{\text{ème}}$ pas s'achève donc en ayant exhibé les itemsets fréquents de longueur i qui serviront, à l'itération suivante, à la génération de nouveaux candidats de longueur $(i+1)$. Nous rejoignons donc l'hypothèse posée comme point de départ de chaque pas : un espace de recherche sur-ensemble de l'espace des solutions. La démarche exposée permet de plus de restreindre au plus tôt cet espace de recherche, en éliminant dès que possible toute combinaison non pertinente de l'espace des solutions.

Etat de l'art

Pionnières

Le premier algorithme, AIS, destiné à résoudre le problème de la recherche de règles d'associations, selon la définition donnée dans cette section, est présenté dans [Agrawal *et al.*, 1993]. Dans cet algorithme, des candidats (un candidat est un itemset susceptible d'être fréquent dans la base de données) sont générés et testés à la volée, pendant une passe sur la base de données. Après avoir lu une transaction, l'algorithme détermine quels itemsets fréquents de la passe précédente sont contenus dans cette transaction. De nouveaux candidats sont alors générés par extension de ces itemsets fréquents avec des items de la transaction. Il s'agit là des premières pierres sur lesquelles va se construire la méthode Générer-Elaguer. Les candidats générés à partir d'une transaction sont ajoutés à l'ensemble des candidats construits pendant cette passe (ou leurs compteurs sont incrémentés s'ils appartiennent déjà à cet ensemble).

Dans le but, également, de construire des règles d'associations, mais motivés par les possibilités d'utiliser SQL pour calculer les itemsets fréquents, [Houtsama *et al.*, 1995]



propose l'algorithme SETM. Basé sur le principe d'AIS, cet algorithme utilise SQL pour calculer les itemsets candidats, grâce à une opération de jointure standard de SQL entre les fréquents. Cet approche a révélé des inconvénients dus à la taille excessive des candidats générés.

A partir de raisonnements plus fins concernant les inclusions de fréquents, [Agrawal *et al.*, 1994] propose un nouvel algorithme qui reste encore une référence : Apriori. L'idée fondatrice de cet algorithme est alors la méthode Générer-Elaguer qui peut-être présentée, de manière informelle, comme ceci : puisque les fréquents ont entre eux des propriétés particulières, relatives à l'inclusion et puisque l'on sait qu'il existe une taille maximale de fréquents (le fréquent de plus grande taille dans la base), il faut procéder par étapes (passes sur la base de données). Pour cela, déterminer les fréquents de taille 1 (les items fréquents), construire les candidats de taille 2 (à partir des fréquents de taille 1), déterminer les fréquents de taille 2 par une phase qui détermine pour chaque transaction dans D, quels sont les 2-candidats contenus dans cette transaction et réitérer (2-candidats $\rightarrow_{\text{prune}}$ 2-fréquents \rightarrow_{gen} 3-candidats $\rightarrow_{\text{prune}}$ 3-fréquents \rightarrow_{gen} 4-candidats ...) jusqu'à ce que l'on ne trouve plus aucun fréquent.

Minimisant le nombre de passes

Pour tenter d'accélérer la production des résultats, de nouvelles méthodes sont alors élaborées à partir de l'observation suivante : le nombre de passes sur la base de données est à l'origine d'un trop grand nombre d'accès disque, dont la lenteur n'est plus à discuter.

Pour envisager une solution à ce problème, la méthode Partition [Savasere *et al.*, 1995] propose de diviser la base de données afin de déterminer les fréquents en seulement deux passes. Pour cela l'algorithme détermine un nombre de partitions de tailles identiques et met en œuvre une méthode similaire à Apriori pour calculer en mémoire les fréquents locaux à chaque partition. Notons L^i l'ensemble des fréquents de la partition i . Du fait de la différence entre les données stockées dans chaque partition, les fréquents qu'elles contiennent sont également différents. Pour déterminer les fréquents de manière globale, une autre passe sur la base (*counting phase*) est nécessaire. Pour créer les candidats qu'il faut tester dans cette phase, l'ensemble des candidats - globaux $C^G = \cup_i L^i$ est construit comme l'union de tous les fréquents locaux. Les travaux de [Gardarin *et al.*, 1998] étendent l'algorithme de Partition, en ajoutant une structure de recherche plus efficace sur les ensembles fréquents.

[Brin *et al.*, 1997] propose une solution destinée également à déterminer les k -fréquents en moins de k passes. Pour cela, l'algorithme DIC proposé, divise lui aussi la base en plusieurs parties de tailles égales. Supposons que la base soit divisée en trois parties et que les fréquents aient une taille maximale de trois items. DIC construit sur une première partie de la base tous les fréquents de longueur 1 rencontrés puis continue en créant les fréquents de longueur 2 à partir des 1-fréquents déjà extraits. Pendant qu'il construit les 2-fréquents, DIC continue de créer les 1-fréquents et commence à créer les 3-fréquents. Une fois la première passe terminée l'algorithme a obtenu les items fréquents, mais aussi les 2-fréquents aux environs des deux tiers de leur totalité et le premier tiers des 3-fréquents. La deuxième passe n'a pas besoin d'être achevée car le premier tiers de la base doit être vérifié pour les 2-fréquents (jusqu'ici seul les deux derniers tiers ont servi à les construire) et les deux premiers tiers suffisent àachever la construction des 3-fréquents (le troisième est déjà exploré pour les 3-fréquents).

Il existe également des méthodes stochastiques incomplètes qui peuvent être classées parmi les méthodes économies en accès disque. Citons par exemple [Gunopulos *et al.*, 1997] qui propose une méthode stochastique basée sur une généralisation du concept d'hypergraphe transversal. [Toivonen, 1996] utilise une méthode d'échantillonnage, Sampling, dont la première version se trouve dans [Mannila *et al.*, 1994]. Cette méthode permet d'obtenir la



plupart des fréquents (les auteurs affichent une très faible probabilité pour qu'un fréquent n'apparaisse pas).

Malgré les temps de réponse affichés pour ces algorithmes, il faut considérer les inconvénients engendrés par la compensation des accès disques par des calculs qui peuvent affecter leurs performances.

Adaptées à l'organisation des données

Les approches précédentes ont en commun le fait qu'aucune d'entre elles ne tient compte du fait que les données peuvent être corrélées. En effet dans [Pasquier *et al.*, 1999], l'algorithme Close proposé offre des temps de réponse particulièrement performants sur des bases de données telles que les données de recensement. Si cet algorithme, basé sur l'utilisation de treillis de Galois, n'accélère pas les temps de réponse sur les benchmarks proposés par [Agrawal *et al.*, 1994], il a l'avantage de garder les meilleures performances sur un type de données particulier : les données corrélées.

Rapides mais limitées aux bases de données de petite taille

[Wang, 1997] propose une méthode d'extraction de règles d'association, basée sur l'utilisation de suffix tree. Si l'algorithme exposé semble afficher des temps de réponse performants, sa complexité en mémoire reste un point obscur. En effet, il est certain qu'une méthode basée sur des suffix tree nécessite un espace en $O(n)$ (la seule information disponible quand à la complexité en espace de ce travail est qu'il connaît une complexité en mémoire linéaire) avec n la taille de la base de données.

Sans génération de candidats

Les auteurs de [Han *et al.*, 2000] considèrent que le goulot d'étranglement des processus d'extraction de type A-priori, se situe au niveau des candidats. Leur étude montre, en effet, que pour 10^4 items fréquents découverts, l'algorithme Apriori devra tester 10^7 candidats de taille 2 sur la base de données. De plus pour découvrir des itemsets fréquents de taille 100, cet algorithme devra tester et donc générer jusqu'à 10^{30} candidats au total. L'approche proposée par [Han *et al.*, 2000] consiste donc à contourner le problème de la génération des candidats, en supprimant cette notion grâce à une structure de FP-Tree. Cette structure, pour être mise en place, demande tout d'abord, une passe sur la base de données, afin d'en collecter les items fréquents. En effet, la structure de FP-Tree, ne considère que les items fréquents, afin d'extraire les itemsets fréquents. Une fois cette passe effectuée, la majeure partie de l'algorithme consiste à construire l'arbre, qui est une transformation de la base d'origine, limitée aux items fréquents. L'étape finale consiste à ajouter un index des sommets représentant chaque item fréquent. Pour extraire les itemsets fréquents, à partir de la structure FP-tree, les auteurs proposent d'exploiter les connaissances mises en place dans la structure et dans l'index. La méthode consiste donc à trouver les itemsets fréquents qui contiennent chacun des items de l'index.

1.1.5. Les motifs séquentiels

Proche de la problématique de la recherche de règles d'association, le problème de la recherche de séquences dans une base de données de transactions étend le problème en prenant en compte le fait que les objets de la base sont estampillés.

Problématique

Dans le contexte des motifs séquentiels, nous définissons les transactions manipulées de la manière suivante [Agrawal *et al.*, 1995] :



Une *transaction* constitue, pour un client C, l'ensemble des items achetés par C à une même date. Dans une base de données client, une transaction s'écrit sous la forme d'un ensemble : {identification-client, identification-date, itemset} où un itemset est défini de la même manière que dans le cas des règles d'association. Une *séquence* est une liste ordonnée, non vide, d'itemsets notée $\langle s_1 \ s_2 \dots \ s_n \rangle$ où s_j est un itemset (une séquence est donc une suite de transactions qui apporte une relation d'ordre entre les transactions). Soit $T_1, T_2 \dots T_n$ les transactions d'un client C, ordonnées par dates d'achat croissantes et soit $\text{itemset}(T_i)$ l'ensemble des items correspondants à T_i , alors la séquence de données de C est $\langle \text{itemset}(T_1) \ \text{itemset}(T_2) \ \dots \text{itemset}(T_n) \rangle$.

En guise d'illustration, considérons l'exemple suivant. Soit C un client et $S_1 = \langle (3) \ (4 \ 5) \ (8) \rangle$, la séquence de données représentant les achats de ce client. S_1 peut être interprétée par "C a acheté l'item 3, puis en même temps les items 4 et 5 et enfin l'item 8". Cette séquence est incluse dans la séquence $S_2 = \langle (7) \ (3 \ 8) \ (9) \ (4 \ 5 \ 6) \ (8) \rangle$ car $(3) \subseteq (3 \ 8)$, $(4 \ 5) \subseteq (4 \ 5 \ 6)$ et $(8) \subseteq (8)$. En revanche, $\langle (3) \ (5) \rangle \not\subseteq \langle (3 \ 5) \rangle$ (et vice versa).

Un client *supporte* une séquence S (fait partie du support pour S) si S est incluse dans la séquence de données de ce client. Le *support* d'une séquence S est calculé comme étant le pourcentage des clients qui supportent S. Une séquence qui vérifie le support minimum spécifié par l'utilisateur est une *séquence fréquente*.

Soit une base de données DB, l'ensemble E des *séquences fréquentes maximales* (également notées *motifs séquentiels*) est constitué de toutes les séquences fréquentes telles que pour chaque S dans E, S n'est incluse dans aucune autre séquence de E. Le problème de la recherche de motifs séquentiels consiste donc à trouver l'ensemble E.

Le problème de la recherche de séquences, tel qu'il est défini, présente les limites suivantes qui ont conduit au problème de la recherche de séquences généralisées [Srikant *et al.*, 1996] :

- Absence de contraintes de temps. Souvent, l'application peut avoir besoin de séparer les transactions pour ne pas prendre en compte des achats trop éloignés ou trop rapprochés (selon le but, on peut vouloir réaliser par exemple une étude statistique à court terme ou bien à long terme).
- Rigidité de la définition des transactions. Nombre d'applications peuvent avoir besoin de ne pas tenir compte d'une très faible durée séparant deux transactions et considérer qu'elles ont eu lieu, non pas de manière séquentielle, mais simultanée, comme une seule et même transaction.

Etat de l'art

Elaborés en premiers lieux, les algorithmes de recherche d'associations connaissent de grandes difficultés d'adaptation aux problèmes d'extraction de motifs séquentiels. En effet, si le problème de la recherche de règles d'associations est proche de celui des motifs séquentiels (il est à son origine), les études dans ce sens montrent que, lorsque l'adaptation est possible, c'est au prix de temps de réponse inacceptables [Agrawal *et al.*, 1995]. Plusieurs approches, destinées à présenter une solution correcte au problème de l'extraction de motifs séquentiels, ont été proposées depuis la définition du problème dans [Agrawal *et al.*, 1995]. Il est important de constater que l'intérêt de la recherche de motifs séquentiels par rapport aux règles d'associations se situe à deux niveaux :



- Les motifs séquentiels sont un moyen d'obtenir une catégorie particulière de règles d'associations (intégrant une relation d'ordre entre les itemsets), par une phase de réécriture des séquences en fin de processus. Le problème de la recherche de règles d'association est donc un sous problème de celui de l'extraction de motifs séquentiels.
- Les motifs séquentiels posent un problème plus complet et donc plus intéressant. En effet, pour définir le problème de l'extraction de motifs séquentiels, il faut prendre en compte de nouvelles contraintes apportées par la datation

Utilisant Générer-Elaguer

L'algorithme GSP [Srikant *et al.*, 1996] est un algorithme basé sur la méthode Générer-Elaguer mise en place depuis Apriori et destiné à effectuer un nombre de passes raisonnable sur la base de données. Pour évaluer le support de chaque candidat en fonction d'une séquence de données, GSP utilise une structure d'arbre de hachage destinée à effectuer un tri sommaire des candidats. Les candidats sont stockés en fonction de leur préfixe, mais sans tenir compte des contraintes de temps. Pour ajouter un candidat dans l'arbre des séquences candidates, GSP parcourt ce candidat et effectue la descente correspondante dans l'arbre. En arrivant sur une feuille, GSP ajoute ce candidat à la feuille et si la taille de la feuille dépasse la taille maximale alors elle est scindée en deux nouvelles feuilles dans lesquelles les candidats sont répartis. Pour trouver quelles séquences candidates sont incluses dans une séquence de données, GSP parcourt l'arbre en appliquant une fonction de hachage sur chaque item de la séquence. Quand une feuille est atteinte, elle contient des candidats potentiels. Cet ensemble de séquences candidates est constitué de candidats inclus dans la séquences de données et de candidats "parasites".

Enfin, pour déterminer quelles sont les séquences réellement incluses, GSP utilise une nouvelle structure pour rapidement trouver les séquences.

Dans [Masseglio *et al.*, 1999], les auteurs proposent un algorithme appelé PSP qui est également basé sur la méthode Générer-Elaguer mais avec une nouvelle structure d'arbre préfixée pour gérer efficacement les séquences fréquentes et les séquences candidates.

Minimiser le nombre de candidats

La méthode PrefixSpan, présentée dans [Pei *et al.*, 2001], se base sur une étude attentive du nombre de candidats qu'un algorithme de recherche de motifs séquentiels peut avoir à produire afin de déterminer les séquences fréquentes. En effet, selon les auteurs, pour envisager d'utiliser un algorithme comme GSP il faut s'attendre à devoir générer, uniquement pour la seconde passe, pas moins de $n^2 + n.(n-1)/2$ candidats de taille 2 à partir des n items trouvés fréquents dans la première passe. L'objectif des auteurs est alors de réduire le nombre de candidats qui est l'un des facteurs les plus pénalisants des algorithmes de recherche de motifs séquentiels.

Pour parvenir à cet objectif, PrefixSpan propose d'analyser les préfixes communs que présentent les séquences de données de la base à traiter. A partir de cette analyse, l'algorithme construit des bases de données intermédiaires qui sont des projections de la base d'origine déduites à partir des préfixes relevés. Ensuite, dans chaque base obtenue, PrefixSpan cherche à faire croître la taille des motifs séquentiels découverts, en appliquant la même méthode de manière récursive.

Deux sortes de projections sont alors mise en place pour réaliser cette méthode : la projection dite "niveau par niveau" et la "bi-projection". Au final, les auteurs proposent une méthode



d'indexation permettant de considérer plusieurs bases virtuelles à partir d'une seule, dans le cas où les bases générées ne pourraient être maintenues en mémoire en raison de leur taille.

Spade, présenté dans [Zaki, 1999], se classe dans la catégorie des algorithmes qui, à l'instar de PrefixSpan, cherchent à réduire l'espace des solutions en regroupant les motifs séquentiels par catégorie. Pour Spade, les motifs fréquents présentent des préfixes communs qui permettent de décomposer le problème en sous-problèmes qui seront traités en mémoire.

Le calcul de F2 (les fréquents de taille 2) par Spade, passe par une inversion de la base qui la transforme d'un format vertical vers un format horizontal. Il gère les candidats et les séquences fréquentes à l'aide de classes d'équivalence comme suit : deux k-séquences appartiennent à la même classe si elles présentent un suffixe commun de taille (k-1). Plus formellement, soit $P_{k-1}(\alpha)$ la séquence de taille k-1 qui préfixe la séquence α . Comme α est fréquente, $P_{k-1}(\alpha) \in F_{k-1}$ les fréquents de taille k-1. Une classe d'équivalence est définie de la manière suivante : $[\rho \in F_{k-1}] = \{\alpha \in F_k | P_{k-1}(\alpha)=\rho\}$. Chacune de ces classes d'équivalence contient alors deux types d'éléments : $[\rho.l_1] = <\rho(x)>$ ou bien $[\rho.l_2] = <\rho x>$ selon que l'item x appartient ou pas à la même transaction que le dernier item de ρ .

Les candidats sont ensuite générés selon trois critères : Autojointure ($[\rho.l_1] \times [\rho.l_1]$), Autojointure ($[\rho.l_2] \times [\rho.l_2]$) et jointure ($[\rho.l_1] \times [\rho.l_2]$).

Le reste de l'algorithme, à savoir le comptage du support pour les candidats générés, repose sur la ré-écriture préalable de la base de données. En effet, la transformation consiste à associer à chaque k-séquence l'ensemble des couples (client, itemset) qui lui correspondent dans la base.

Similarités avec d'autres domaines

Le problème de l'extraction de motifs séquentiels est proche de celui défini par [Wang *et al.*, 1994], situé dans la découverte de similarités dans les bases de données de séquences génétiques pour lesquelles les séquences sont des caractères consécutifs séparés par un nombre variable de caractères parasites. Dans notre définition du problème, une séquence est une liste ordonnée d'ensemble de caractères et non une liste de caractères. Si l'approche proposée par [Agrawal *et al.*, 1995] se révèle très différente, c'est également en raison de la taille des instances traitées. En effet, les travaux de [Wang *et al.*, 1994] sont des algorithmes basés sur des suffix-tree qui travaillent principalement en mémoire et révèlent une complexité en mémoire en $O(n)$, avec n la somme des tailles de toutes les séquences de la base. Il est impossible, dans le cadre des motifs séquentiels, d'accepter une complexité en mémoire qui dépend de la taille de la base car nous travaillons sur des instances de très grande taille et la mémoire volatile n'est jamais aussi importante que la mémoire permanente (nous ne pouvons pas non plus nous permettre de dupliquer la base ou de la réécrire sous une autre forme).

Dans [Mannila *et al.*, 1995], nous sommes confrontés au problème de la découverte d'épisodes fréquents dans une séquence d'événements. S'ils utilisent la notion de durée et de time-window, leur approche est encore différente de celle définie par la recherche de motifs séquentiels généralisés. En effet leur but est de trouver les répétitions à l'intérieur d'une même séquence, en faisant se déplacer une fenêtre de temps sur cette séquence. Le problème qui nous concerne consiste, pour sa part, à rechercher des répétitions entre différentes séquences et en fonction d'une fréquence d'apparition minimum (le support).



De nombreux algorithmes performants ont été définis pour la recherche de motifs et de sous mots dans une phrase, ainsi que la recherche de sous mots communs. Nous pouvons citer par exemple les algorithmes de Boyer-Moore ou Knuth-Morris-Pratt [Cormen *et al.*, 1994], mais aussi plus récemment les travaux de [Califano *et al.*, 1993]. Cependant ces techniques sont destinées à la localisation d'un pattern dans une phrase. Dans le problème de la recherche de motifs séquentiels, il s'agit d'exhiber des patterns pour plusieurs phrases et pour cela, de trouver quel est le nombre de phrases qui contiennent ces patterns.

1.1.6. Exemples d'applications

Dans cette section, nous décrivons quelques domaines d'applications mettant en œuvre les techniques décrites précédemment.

Le Web Mining

Avec la popularité du Web, de très grandes quantités de données comme l'adresse des utilisateurs ou les URLs demandées sont automatiquement récupérées par les serveurs Web et stockées dans des fichiers "Access Log". L'analyse de tels fichiers peut offrir des informations très utiles pour, par exemple, améliorer les performances, restructurer un site ou même cibler le comportement des clients dans le cadre du commerce électronique.

La découverte d'informations à partir du Web est généralement appelée Web Mining et peut recouvrir deux aspects : le Web Content Mining et le Web Usage Mining [Han *et al.*, 2001]. La première approche concerne la découverte et l'organisation d'informations extraites du Web. Par exemple, des approches basées sur les agents sont utilisées pour découvrir et organiser, de manière autonome, les informations extraites à partir du Web et des approches Bases de Données s'intéressent aux techniques d'intégration, d'organisation et d'interrogation de données hétérogènes et semi-structurées sur le Web. Le Web Usage Mining s'intéresse par contre au problème de la recherche de motifs comportementaux des utilisateurs à partir d'un ou de plusieurs serveurs Web afin d'extraire des relations entre les données stockées.

Le Web Content Mining

Dans le cadre des approches Bases de Données, les techniques de Bases de Données fédérées permettent de collecter des informations depuis des sources variées dans des entrepôts de données. Cependant avant de pouvoir rediffuser cette information il est souvent nécessaire de l'intégrer et de l'enrichir. Dans ce cadre, les techniques de fouille de données sont utilisables. La recherche de connaissances dans des données structurées a fait l'objet de nombreux travaux de recherche ces dernières années. La plupart des approches proposées s'intéressent à des structures plates ou fortement structurées. Avec la popularité du Web, le nombre de documents semi-structurés disponibles augmente très rapidement. Contrairement aux applications de bases de données traditionnelles où l'on décrit d'abord la structure des données, i.e. le type ou le schéma, où l'on crée ensuite les instances de ces types, dans les données semi-structurées, les données n'ont pas de schéma prédéfini et chaque objet contient sa propre structure. La majorité des documents "en ligne", tels que les fichiers HTML/XML, Latex, Bibtex, ou SGML sont semi-structurés. Par conséquent, la structure des objets est irrégulière et il est judicieux de penser qu'une requête sur la structure des documents est aussi importante qu'une requête sur les données [Wang *et al.*, 1999]. Cependant malgré cette irrégularité structurelle, il peut exister des similitudes structurelles parmi les objets semi-structurés. Il est même assez fréquent de constater que des objets semi-structurés qui décrivent le même type d'information ont des structures similaires. L'analyse de telles



structures implicites dans des données semi-structurées peut alors fournir des informations importantes pour : optimiser les évaluations de requêtes, obtenir des informations générales sur le contenu, faciliter l'intégration de données issues de différentes sources d'information, améliorer le stockage, faciliter la mise en place d'index ou de vues et aider à la classification de documents semi-structurés. Il est alors possible d'utiliser des méthodes de classification pour regrouper les fragments de documents les plus proches. La recherche de règles d'association ou de motifs séquentiels est également très adaptée [Gardarin, 1999].

Les approches basées sur les agents dans le cadre du Web Content Mining nécessitent le développement de techniques d'apprentissage pour découvrir et organiser les informations contenues sur le Web. Généralement, trois catégories de systèmes à agents sont considérés :

- Les agents de recherche intelligents : différents systèmes d'agents ont été développés pour rechercher l'information pertinente en utilisant les caractéristiques d'un domaine particulier (ou même le profil de l'utilisateur) afin d'organiser et d'interpréter les informations découvertes.
- Les agents chargés de filtrer et de catégoriser l'information : ces agents utilisent différentes techniques de recherche d'information. Par exemple, dans HyPursuit [Weiss et al., 1996], le système utilise l'information sémantique des liens ainsi que le contenu des documents pour créer des clusters hiérarchiques de documents hypertexte et le système BO (Bookmark Organizer) [Maarek et al., 1996] combine des techniques de groupement hiérarchique avec l'interaction de l'utilisateur pour organiser une collection de documents Web.
- Les agents personnalisés : cette catégorie d'agents utilise les préférences de l'utilisateur et l'évaluation des documents retournés pour optimiser la recherche d'informations. Par exemple, le système Syskill & Webert [Pazzani et al., 1996] utilise le profil de l'utilisateur et utilise un classifieur Bayésien pour atteindre les pages les plus opportunes.

Le Web Usage Mining

Dans le contexte du Web Usage Mining [Masand *et al.*, 2000], les données brutes sont collectées dans des fichiers Access Log des serveurs Web. Une entrée dans le fichier log est automatiquement ajoutée chaque fois qu'une requête pour une ressource atteint le serveur Web. Spécifiée par le CERN et la NCSA, une entrée contient des enregistrements formés de 7 champs séparés par des espaces²⁷ :

Host user authuser [date : time] " request " status bytes
Bien que des outils d'analyse existent pour indiquer, par exemple, le nombre d'accès à des URL ou la liste des URL les plus demandées, les rapports existant entre les ressources demandées et les accès des clients ne sont pas analysés par de tels outils dont les performances sont assez limitées.

Dans notre contexte, en analysant les informations stockées sur le serveur Web, un exemple de règles d'association peut être : "50% des clients qui ont visité les URL `plaquette/info-f.html` et `labo/infos.html` ont également visité `situation.html` ou bien 86% des clients qui ont acheté des produits via `/e-comm/books` et `/e-comm/computer` ont également commandé via l'URL `/e-comm/games`".

²⁷ Les informations contenues dans les fichiers logs peuvent varier selon les systèmes utilisés mais comportent globalement les mêmes types d'information.



L'extraction de motifs séquentiels permet, d'autre part, de mettre en évidence le type de relation suivantes : " 60% des clients qui ont acheté /e-comm/books/datamining et /e-comm/books/ computerscience ont également acheté dans les 30 jours suivants le livre /e-comm/books/algorithmic ou bien que 34% des clients ont visité l'URL /relnotes/deprecatedlist.html entre le 20 septembre et le 30 novembre.

Outre les règles d'association et les motifs séquentiels il est tout à fait possible également de classer les différents catégories de clients en utilisant des techniques de classification. Dans ce cadre un système OLAP (On Line Analytical Processing) est utilisé pour extraire des informations significatives. Les données sont tout d'abord filtrée pour éliminer les informations non pertinentes puis stockées dans une base de données relationnelle. Ensuite, une structure de tableau multidimensionnelle appelée Web Log Data Cube est construite et chaque dimension représente un champ avec toutes les valeurs possibles décrites par les attributs. Les système OLAP est alors utilisé pour appliquer des opérations de drill-down, slice et dice sur le data cube et offre ainsi la possibilité d'examiner les données sous différents angles. Enfin des techniques de fouille de données, comme la caractérisation, la recherche de règles, la prédition ou la classification peuvent être utilisées sur le Web Log Data Cube [Dyreson, 1997] [Han *et al.*, 2001].

Un panorama des systèmes et approches de Web Usage Mining ainsi que de leur utilisation est proposé dans [Cooley, 2000].

Le Text Mining

Le Text Mining consiste à utiliser des techniques linguistiques et de fouille de données pour analyser et synthétiser de grand volumes de textes. L'un des objectifs du Text Mining est d'extraire des informations qu'il aurait été difficile de trouver sans une analyse automatique et systématique de gros volumes de données. En appliquant les techniques précédentes, de nombreuses analyses peuvent être effectuées.

Dans ce cadre du processus d'extraction de connaissances à partir de données textuelles, un prétraitement des données est d'abord effectué en utilisant généralement des techniques issues du traitement linguistique. Ces traitements initiaux ont pour but d'optimiser la représentation du texte manipulé. Il peuvent être plus ou moins complexes en fonction des résultats attendus : élimination des mots vides de sens (articles, prépositions, ...), lemmatisation, gestion des synonymes et des hyperonymes, polysémie, etc. Une fois cette phase de traitement linguistique effectuée, une étape de lexicométrie mesure la fréquence d'apparition des mots, des expressions ou encore des cooccurrences de certains mots au sein d'un même texte.

A partir de ces données textuelles pré-traitées, différentes techniques de fouille de données peuvent être appliquées.

L'utilisation d'algorithme de recherche de règles d'association offre la possibilité de rechercher des corrélations entre des mots clés du texte et permet d'optimiser l'étape de lexicométrie [Feldman *et al.*, 1996] [Rajman *et al.*, 1998]. Ainsi, la problématique revient à considérer chaque document comme un couple <identification du document, {ensemble de mots clés}> et à rechercher tous les ensembles de mots clés W tels que $|W| \geq minSupp$ (ensemble fréquents de mots clés). A partir de ces fréquents découverts, il est donc possible de générer des règles d'association. Par exemple, en recherchant toutes les associations incluant embouteillage et ville, nous pouvons trouver la règle suivante : (Embouteillage,



Péphérique) → Paris (80%, 55%). Dans le cas où nous préférons analyser les documents non plus sous une forme résumé de mots clés, il est indispensable d'appliquer des techniques de traitements linguistiques plus complexes pour extraire non plus des mots clés mais des composants linguistiques de documents (séquence de mots par exemple) [Rajman *et al.*, 1998].

L'analyse de tendance dans une série de documents est une application possible des algorithmes de recherche de motifs séquentiels [Lent *et al.*, 1997]. Dans ce cadre, nous recherchons les changement de fréquence d'apparition des termes (mot/phrase) dans une période de temps données (par exemple, "augmentation pollution"). Pour résoudre cette problématique, une première étape consiste à identifier les champs de textes fréquents à l'aide d'algorithme de recherche de motifs et d'en générer l'historique. En interrogeant ce dernier, il est donc possible d'évaluer différentes tendances (tendance récente à monter, transition récente, résurgence d'usage, ...). Par exemple, l'analyse de documents journalistiques entre 1999 et 2001 montre une tendance récente à monter de l'intérêt des utilisateurs pour les transports en communs.

Le Data Mining Spatial

La fouille de données spatiales est un domaine de recherche en pleine expansion [Aufaure *et al.*, 2000]. Elle s'attache à effectuer une analyse dirigée par les données sur des bases de données spatiales [Longley *et al.*, 1999]. Elle offre de nouvelles perspectives pour beaucoup d'applications à caractère décisionnel comme le géo-marketing ou l'analyse de risques d'accidents. Elle possède sa propre spécificité par rapport à la fouille de données traditionnelle car elle doit prendre en compte les relations spatiales entre objets. Cependant, les tâches dans un processus de Fouille de Données Spatiales (FDS) sont définies comme une extension par les critères spatiaux de celles utilisées dans la fouille de données. Ces différentes tâches permettent de résumer les données, de trouver des règles de classification, de former des clusters d'objets similaires, de trouver des associations ou des dépendances pour caractériser les données et d'extraire les tendances et les déviations. Ainsi, par exemple, dans le domaine des bases de données, des algorithmes de fouille de données spatiales ont été proposés et des prototypes ont été développés. GeoMiner [Han *et al.*, 1997] est une extension de DBMiner [Han *et al.*, 2001] qui intègre les techniques OLAP (On-Line Analytical Process) et fouille de données.

La transposition au domaine spatial des méthodes de clustering s'appuie sur une similarité d'objets localisés suivant leur distance métrique. Néanmoins, la finalité du clustering dans ce contexte correspond plus à la détection de concentrations anormales (par exemple, la détection de zones à risque en accidentologie). Les travaux sur le clustering spatial sont surtout axés sur l'optimisation des algorithmes existants de manière à prendre en compte la masse de données disponibles. Par exemple, [Ester *et al.*, 1998] propose une extension de l'algorithme DBScan en utilisant l'index spatial R*Tree ainsi que des extensions incrémentales des algorithmes DBClass et DBLearn.

L'extension de la découverte de règles d'association aux bases de données spatiales permet de générer des règles du type :

$X \rightarrow Y (s, c)$ où s correspond au support et c à la confiance

telles que X et Y sont des ensembles de prédicts spatiaux et non spatiaux. En d'autre termes, ceci revient à chercher des associations entre des propriétés des objets et celles de leur



voisinage. Ainsi, la recherche d'associations impliquant les aéroports et les autres entités géographiques génère par exemple la règle suivante :

est-un (x, "aéroport ") → proche_de(x, "zone pavillonnaire ") (61%, 70%)

L'extension au domaine spatial des algorithmes de classification a été définie par l'extension aux propriétés de leurs voisins jusqu'à un ordre N de voisinage [Ester *et al.*, 1997]. Ainsi, il est possible de trouver une règle du type :

Si population élevée et type de voisin = route et voisin de voisin = aéroport **Alors** puissance économique élevée (95%).

1.2. La recherche opérationnelle

La recherche opérationnelle apparaît en 1940 en Angleterre puis aux États-Unis à des fins de recherche militaire : il s'agissait pour le Royaume Uni d'utiliser au mieux ses moyens militaires, à l'époque insuffisants (avions, forces antiaériennes (D.C.A.), moyens maritimes. L'idée fondamentale était de mettre autant de soin dans l'emploi des moyens qu'on en avait mis pour les concevoir et les construire.

Après la guerre, la recherche opérationnelle s'introduit dans le monde des affaires, l'objectif étant d'organiser, produire, stocker et vendre de façon optimale. L'arrivée de l'ordinateur allait accélérer l'essor de cette nouvelle branche des mathématiques qu'est la théorie des graphes, support théorique de la recherche opérationnelle, laquelle utilise également trois autres branches fondamentales : algèbre linéaire et programmation linéaire (résolution de systèmes d'équations et d'inéquations linéaires; méthode du simplexe), la statistique, usant du calcul des probabilités, notamment la théorie des files d'attente (queueing theory), la théorie des graphes qui s'est surtout développée au 20^{ème} siècle, dans les années 50, avec Claude Berge, en France, et D. König en Allemagne. Elle s'appuie sur la théorie des ensembles et les relations binaires, dont le principal initiateur, après Euler et les premières bases de la logique mathématique jetées par Boole et De Morgan, fut Cantor. L'essor cohérent des réseaux de chemin de fer (SNCF), des lignes aériennes, du réseau téléphonique doit beaucoup à la théorie des graphes.

Dans le cadre des nouvelles branches mathématiques du 20^{ème} siècle, il faut noter l'apparition de la théorie des jeux, imaginée par Borel mais principalement développée aux États-Unis par Tucker, Dantzig, von Neumann et Morgenstern à l'issue de la seconde guerre mondiale et issue de la recherche opérationnelle appliquée un système concurrentiel économique ou militaire. Par jeu, on entend stratégie à mettre en œuvre (en jeu) afin d'optimiser une production, un bénéfice, face à une concurrence, de s'opposer (militairement) efficacement à un adversaire, de maîtriser, dans la mesure du possible, les phénomènes naturels (météorologie, séismes, etc.).

1.2.1. Programmation linéaire

La programmation linéaire a pour objet l'étude et la résolution des « programmes linéaires » c'est-à-dire des problèmes d'optimisation dans lesquels la fonction objectif aussi bien que les contraintes sont exprimées de manière linéaire.

Les modèles de programmation linéaire permettent d'aborder un grand nombre de problèmes d'optimisation en apparence très différents, dans des contextes très divers. La programmation



linéaire relève des mathématiques de la Recherche Opérationnelle et possède un vaste domaine d'application : en gestion ainsi qu'en économie, en statistique, en physique, etc..

Les applications classiques de la programmation linéaire sont les problèmes de maximisation du profit sous contraintes de disponibilité de ressources, de minimisation des coûts sous contraintes de satisfaction de la demande, de transport, de mélanges, etc. La méthode du simplexe (qui est un est un algorithme de résolution proposé par George Dantzig; voir [Dantzig, 1963][Chvátal, 1983] est une méthode simple, robuste et efficace, qui permet d'attaquer avec succès des problèmes comportant plusieurs dizaines de milliers de variables et de contraintes. Elle resta en fait sans concurrence pendant près de 40 ans, jusqu'à la découverte d'une famille d'algorithmes connus sous le nom de « méthodes de points intérieurs » [Taha, 1992].

A ce jour, plusieurs méthodes de résolution coexistent souvent dans les logiciels commerciaux, mais la méthode du simplexe reste une des meilleures alternatives disponibles pour la solution de nombreux problèmes de programmation linéaire.

Il est à noter que la programmation linéaire s'avère être un outil mathématique très riche, qui donne une vue pénétrante des méthodes d'optimisation continue et qui constitue une technique fondamentale de l'optimisation combinatoire.

Modèle de programmation linéaire et terminologie élémentaire.

Sous sa forme la plus générale, le modèle de programmation linéaire (P.L.) est le modèle d'optimisation

$$\text{minimiser (ou maximiser)} \quad z(x) = \sum_{j=1}^n c_j x_j \quad (\text{fonction objectif})$$

sous les contraintes (s.c.)

$$\sum_{j=1}^n a_{ij} x_j \begin{cases} \leq \\ = \\ \geq \end{cases} b_i \quad (i=1, \dots, m).$$

Les nombres c_j , a_{ij} , b_i sont les paramètres du modèle; ce sont des nombres connus avant la résolution (coût unitaire, capacité disponible, coefficients techniques,...).

Pour $i = 1, \dots, m$ et $j = 1, \dots, n$, on appelle :

c_j = coefficients de la fonction objectif $z(x)$ (appelée également fonction économique)

a_{ij} = coefficients des contraintes

b_i = membres de droite (*M.d.D.*) des contraintes.

Les variables (de décision) x_j sont indéterminées a priori : la résolution du modèle consiste précisément à déterminer les valeurs des variables qui optimisent la fonction objectif $z(x)$ tout en respectant les contraintes formulées. L'appellation donnée au modèle est justifiée par le fait que la fonction objectif et les membres de gauche des contraintes sont des fonctions linéaires des variables.



En termes généraux, il est utile d'interpréter $\{1, \dots, n\}$ comme un ensemble d'activités économiques qu'une firme peut entreprendre à différents niveaux. La valeur de la variable x_j indiquera le niveau choisi (« programmé ») pour l'activité j ($j = 1, \dots, n$). On peut alors voir c_j comme le profit ou le coût unitaire associé à l'activité j . Si $\{1, \dots, m\}$ est un ensemble de ressources consommées par les activités, alors b_i est la quantité disponible de la ressource i et a_{ij} est la quantité de la ressource i consommée par unité de l'activité j .

La terminologie suivante est utilisée :

Solution (point) réalisable : c'est un point x qui satisfait à toutes les contraintes, c'est-à-dire un point $x \in R^n$ tel que $\sum_{j=1}^n a_{ij} x_j \leq b_i$ pour tout $i = 1, \dots, m$.

Région, espace réalisable - Polyèdre des contraintes : c'est l'ensemble P des points réalisables, c'est-à-dire :

$$P = \left\{ x \in R^n : \sum_{j=1}^n a_{ij} x_j \leq b_i \text{ pour } i = 1, \dots, m \right\}.$$

Géométriquement, P est un polyèdre convexe, c'est-à-dire une région définie comme l'intersection de m demi-espaces.

Solution optimale : c'est un point x^* réalisable et qui optimise $z(x)$ sur P , c'est-à-dire, pour un problème de maximisation, un point x^* tel que

$$x^* \in P \text{ et } z(x^*) \geq z(x) \text{ pour tout } x \in P.$$

Valeur optimale : c'est la valeur $z(x^*)$ atteinte par toute solution optimale x^* .

Forme standard des problèmes de programmation linéaire.

Pour appliquer la méthode du simplexe, nous supposerons que le P.L. à résoudre est sous *forme standard (FS)*, c'est-à-dire qu'il comporte une contrainte de non-négativité pour chaque variable et que les autres contraintes sont mises sous forme d'égalités. Un P.L. sous forme standard s'écrit donc :

$$\begin{aligned} \max / \min \quad & z = \sum_{j=1}^n c_j x_j \\ \text{s.c.} \quad & \sum_{j=1}^n a_{ij} x_j = b_i \quad (i = 1, \dots, m) \\ & x_j \geq 0 \quad (j = 1, \dots, n) \end{aligned}$$

Cette hypothèse n'est absolument pas restrictive: en effet, tout P.L. peut être mis sous forme standard au prix d'une augmentation (parfois considérable) du nombre de contraintes et de variables. Il suffit pour cela de procéder aux trois types de remplacements décrits ci-dessous.

1) Toute contrainte d'inégalité

$$\sum_{j=1}^n a_j x_j \leq b \tag{1}$$



peut être remplacée, en introduisant une *variable d'écart (slack)* s , par un système de contraintes « équivalent »:

$$\begin{cases} \sum_{j=1}^n a_j x_j + s = b \\ s \geq 0 \end{cases} \quad (2)$$
$$(3)$$

L'interprétation d'une variable d'écart en termes économiques est simple : si b est le nombre d'unités disponibles d'une certaine ressource, et $\sum_{j=1}^n a_j x_j$ le nombre d'unités de cette ressource utilisée par la solution x , alors la quantité $s = b - \sum_{j=1}^n a_j x_j$ représente le nombre d'unités résiduelles, non utilisées par la solution x .

2) Toute contrainte d'inégalité

$$\sum_{j=1}^m a_j x_j \geq b$$

est « équivalente » au système

$$\begin{cases} \sum_{j=1}^n a_j x_j - s = b \\ s \geq 0 \end{cases} \quad (4)$$
$$(5)$$

où s est une variable d'écart, qui peut être considérée dans ce cas comme une variable de surplus.

3) Si la variable x_j n'est pas contrainte à être non négative, on peut la remplacer dans la fonction objectif et dans toutes les contraintes par l'expression

$$x_j' - x_j''$$

(où x_j' , x_j'' sont deux nouvelles variables) et introduire les contraintes additionnelles :

$$x_j' \geq 0, \quad x_j'' \geq 0.$$

On peut conclure qu'en combinant les transformations 1), 2), 3) décrites ci-dessus, n'importe quel programme linéaire peut être réécrit sous forme standard. L'avantage essentiel de cette transformation est qu'elle va nous permettre d'utiliser notre connaissance des systèmes d'équations linéaires. La méthode du simplexe va en effet consister essentiellement en la résolution successive de systèmes linéaires, et les contraintes de non négativité seront traitées implicitement.

La méthode de Gauss-Jordan.

Un système d'équations linéaires de la forme

$$\sum_{j=1}^n a_{ij} x_j = b_i \quad (i = 1, \dots, m) \quad (L)$$

peut être résolu par la méthode de Gauss-Jordan, ou méthode d'éliminations successives.



Supposons que les m équations soient linéairement indépendantes (le système est de rang m ; ceci implique en particulier que $n \geq m$). Il est alors possible de trouver m colonnes linéairement indépendantes parmi les colonnes de la matrice des coefficients (a_{ij}). Supposons par exemple que les colonnes 1 à m soient linéairement indépendantes. Appelons les variables correspondantes (x_1, \dots, x_m) *variables en base* et les variables restantes (x_{m+1}, \dots, x_n) *variables hors base* (remarquons qu'en général, il existe de nombreux choix possibles pour les variables en base: il suffit que les colonnes associées soient linéairement indépendantes).

On peut ensuite transformer le système par éliminations successives jusqu'à obtenir un système sous la forme:

$$\begin{array}{rcl} x_1 & + & \bar{a}_{1,m+1} x_{m+1} + \dots + \bar{a}_{1,n} x_n = \bar{b}_1 \\ x_2 & + & \bar{a}_{2,m+1} x_{m+1} + \dots + \bar{a}_{2,n} x_n = \bar{b}_2 \\ \dots & + & \dots \\ x_m & + & \bar{a}_{m,m+1} x_{m+1} + \dots + \bar{a}_{m,n} x_n = \bar{b}_m \\ | \underline{\hspace{1cm}} & | \underline{\hspace{1cm}} & | \underline{\hspace{1cm}} \\ \text{variables en base} & & \text{variables hors base} \end{array} \quad (\text{Dico})$$

Ce système (Dico) est appelé le *dictionnaire du système* (L) associé aux *variables de base* x_1, \dots, x_m : il est équivalent à (L) (c'est-à-dire qu'il possède les mêmes solutions que (L)) et peut être interprété comme définissant les variables en base en terme des variables hors base (comme on le voit en faisant passer x_{m+1}, \dots, x_n dans le membre de droite de (Dico)).

Une solution particulière (*solution de base*) associée à (Dico) est obtenue en fixant x_{m+1}, \dots, x_n (variables hors base) à 0 et en posant $x_i = \bar{b}_i$ pour $i = 1, \dots, m$.

La méthode du simplexe repose exclusivement sur la considération de solutions de base. Les principales caractéristiques d'un problème de programmation linéaire, qui le différencient clairement d'un système d'équations linéaires sont les suivantes :

- il comporte une fonction objectif à optimiser, ce qui implique que l'on s'intéresse à une solution bien particulière parmi les solutions réalisables plutôt qu'à une solution quelconque comme dans le cas des systèmes d'équations ;
- il comporte également des contraintes de non négativité; il est facile de se convaincre que, sans elles, le problème (FS) tomberait dans une des catégories suivantes : (i) pas de solution réalisable, ou (ii) une solution réalisable unique (et donc nécessairement optimale), ou (iii) problème non borné ; un tel problème pourrait être résolu par simple résolution du système de contraintes.

La méthode du simplexe.

La méthode du simplexe procède par *itérations* successives. Au début de chaque itération, le problème à résoudre est sous forme standard « modifiée ». Plus précisément, le système est écrit sous la forme générique:

$$\max/\min z \quad (6)$$



$$\text{s.c. } z - \sum_{j=1}^n \bar{c}_j x_j = v \quad (7)$$

$$\sum_{j=1}^n \bar{a}_{ij} x_j = \bar{b}_i \quad (i = 1, \dots, m) \quad (8)$$

$$x_j \geq 0 \quad (j = 1, \dots, n) \quad (9)$$

La valeur des coefficients \bar{a}_{ij} , \bar{c}_j , \bar{b}_i et v varie d'itération en itération. Le système (7)–(8) est un dictionnaire associé à un ensemble de variables en base ($x_j : j \in B$) (où l'ensemble B varie en cours d'exécution).

A chaque itération de la méthode du simplexe, le dictionnaire courant est transformé en un nouveau dictionnaire. Cette transformation repose sur la sélection d'une *variable entrante*, disons x_e , telle que $\bar{c}_e > 0$ pour un problème de maximisation (et $\bar{c}_e < 0$ pour un problème de minimisation). Les propriétés fondamentales de la méthode du simplexe peuvent être formulées comme suit:

- i) *Si la solution de base associée au dictionnaire initial est réalisable, alors chacune des solutions de bases produites par la méthode du simplexe est réalisable; en d'autres termes, à chaque itération, $\bar{b}_i \geq 0$ pour $i = 1, \dots, m$.*
- ii) *Pour un problème de maximisation, si $\bar{c}_j \leq 0$ pour chaque $j = 1, \dots, n$, alors la solution de base courante est optimale.*
- iii) *Pour un problème de minimisation, si $\bar{c}_j \geq 0$ pour chaque $j = 1, \dots, n$, alors la solution de base courante est optimale.*

Remarquons que, avant de pouvoir démarrer l'exécution de la méthode du simplexe, il est nécessaire de disposer d'une solution de base initiale *réalisable*, c'est-à-dire d'une solution de base $x \in P$ (où P désigne la région réalisable). Or, il est bien évident que certains problèmes de programmation linéaire n'admettent *aucune solution réalisable*.

Par ailleurs, même si le problème considéré possède des solutions réalisables, *trouver*, c'est-à-dire *calculer*, une telle solution peut constituer en soi un problème difficile (en fait, on peut établir rigoureusement que trouver une solution réalisable est *exactement* aussi difficile que résoudre un problème de programmation linéaire). Il existe des techniques générales permettant de construire une solution de base initiale réalisable ou de décider que le problème considéré est irréalisable : *méthode du grand M*, *méthode des deux phases*. Cette dernière méthode utilise la programmation linéaire appliquée à un programme appelé programme auxiliaire, qui possède évidemment une solution réalisable de départ et dont la solution optimale donne soit la solution de départ du problème initial à traiter, soit l'indication que le problème à traiter n'est pas réalisable.

Enfin, tout problème de programmation linéaire réalisable n'admet pas nécessairement de solution optimale. De tels problèmes sont dits *non bornés* : quelle que soit la valeur Z , il



existe une solution réalisable $x \in P$ telle que la valeur de la fonction objectif en x est meilleure que Z .

En fait, on peut démontrer rigoureusement le résultat suivant:

Pour un problème de programmation linéaire de la forme

optimiser z

s.c. $x \in P$,

seuls quatre cas peuvent se présenter:

- (i) $P = \emptyset$: le problème n'est pas réalisable;
- (ii) $P \neq \emptyset$ et l'optimum de z n'existe pas: le problème est non borné;
- (iii) $P \neq \emptyset$ et le problème possède une solution optimale unique qui coïncide nécessairement avec un sommet, ou point extrême de P ;
- (iv) $P \neq \emptyset$ et le problème possède une infinité de solutions optimales qui forment un côté, ou une face, de P .

En termes géométriques, les solutions de base correspondent aux sommets (points extrêmes) du polyèdre des contraintes. Les itérations successives de la méthode du simplexe définissent un chemin parcourant ces sommets par ordre de valeur croissante.

Description matricielle des dictionnaires.

Il est commode de disposer de notations matricielles qui permettent de décrire de façon compacte les dictionnaires du simplexe. Considérons donc à nouveau le programme linéaire sous la forme standard familière:

$$\begin{aligned} \max z &= \sum_{j=1}^n c_j x_j \\ \text{s.c. } &\sum_{j=1}^n a_{ij} x_j = b_i \quad (i=1, \dots, m) \quad (\text{FS}) \\ x_j &\geq 0 \quad (j=1, \dots, n). \end{aligned}$$

Notons: $c = [c_1, \dots, c_n]$, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$,

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{m1} & \dots & a_{mn} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}.$$

En utilisant ces notations matricielles, (FS) peut alors s'écrire de façon compacte:

$$\begin{aligned} \max z &= c x \\ \text{s.c. } &Ax = b \\ x &\geq 0 \end{aligned} \quad (\text{FS})$$

où cx et Ax sont des produits matriciels (vecteur \times vecteur et matrice \times vecteur). Nous allons à présent formuler l'hypothèse que le système $Ax = b$ admet au moins une solution et que $\text{rang}(A) = m$. (Ces hypothèses ne sont pas très restrictives, puisqu'une application de la



méthode de Gauss-Jordan permet de vérifier la première hypothèse et de supprimer, le cas échéant, les lignes linéairement dépendantes du système.)

Soit maintenant $x = (x_B, x_N)$ une solution de base de (FS), où x_B représente le sous-vecteur des variables en base et x_N celui des variables hors base. Les variables en base et hors base induisent également une partition des colonnes de A de la forme $A = [B | N]$ et l'on peut donc réécrire le système des contraintes sous la forme:

$$Ax = Bx_B + Nx_N = b. \quad (10)$$

Remarquons que le nombre de variables en base, c'est-à-dire le nombre de colonnes de B, est exactement égal au nombre de contraintes. Donc, B est une matrice carrée de dimension $m \times m$. Supposons à présent que B est également *inversible* et prémultiplions le système d'équations (10) par B^{-1} . On obtient ainsi le système équivalent:

$$x_B + B^{-1} N x_N = B^{-1} b. \quad (11)$$

Sous cette forme, chaque variable de base apparaît dans une et une seule contrainte, comme requis dans un dictionnaire. Par ailleurs, la fonction objectif peut s'écrire

$$z = cx = c_B x_B + c_N x_N \quad (12)$$

où c_B (respectivement c_N) est le sous-vecteur des coefficients de la fonction objectif associés aux variables en base (respectivement hors base). Pour éliminer les variables en base de cette expression, on peut substituer (11) dans (12) et exprimer ainsi la fonction objectif sous une forme équivalente, mais dans laquelle interviennent seulement les variables hors base:

$$\begin{aligned} z &= c_B (B^{-1} b - B^{-1} N x_N) + c_N x_N \\ &= c_B B^{-1} b + (c_N - c_B B^{-1} N) x_N. \end{aligned} \quad (13)$$

En annulant les variables hors base, c'est-à-dire en posant $x_N = 0$ dans (11) et (13), on obtient alors une solution de base pour laquelle

$$x_B = B^{-1} b, \quad x_N = 0 \quad (14)$$

et

$$z = cx = c_B B^{-1} b. \quad (15)$$

On appelle *base* de (FS) une sous-matrice carrée inversible de A. Si B est une base, alors le dictionnaire associé à B est le système:

$$\left\{ \begin{array}{l} z + (c_B B^{-1} N - c_N) x_N = c_B B^{-1} b \\ x_B + B^{-1} N x_N = B^{-1} b. \end{array} \right. \quad (16)$$

La *solution de base* associée à B est la solution $x = (x_B, x_N) = (B^{-1} b, 0)$. La base B est dite *réalisable* si la solution de base associée à B est réalisable, c'est-à-dire si toutes les composantes du vecteur $B^{-1} b$ sont non négatives.

Cette discussion montre que, lorsqu'un ensemble de variables en base est connu, il n'est pas nécessaire d'effectuer des itérations de la méthode du simplexe pour déterminer le dictionnaire associé à ces variables en base. Il suffit d'identifier les vecteurs (c_B, c_N) et les matrices (B,N) associés au choix des variables en base, d'inverser B, et de calculer le système (16)-(17).



Bien sûr, ceci ne constitue pas un fait très surprenant: en effet, si l'on connaît l'ensemble des variables en base, nous savons déjà que la méthode d'élimination de Gauss-Jordan permet d'obtenir rapidement le dictionnaire associé à ces variables. Au delà de cette observation élémentaire, le système (16)-(17) est surtout intéressant en ceci qu'il nous fournit une représentation compacte (analytique) du dictionnaire considéré *en terme des données originales du problème* (B , N , b , etc). Notons enfin, pour terminer, que les critères de réalisabilité et d'optimalité de la solution de base courante sont aisément exprimés pour le dictionnaire (16)-(17) :

- la solution de base associée à B est réalisable si et seulement si $B^{-1}b \geq 0$;
- pour un problème de maximisation, la solution de base associée à B est optimale si $c_B B^{-1}N - c_N \geq 0$;
- pour un problème de minimisation, la solution de base associée à B est optimale si $c_B B^{-1}N - c_N \leq 0$.

Dualité

La notion de dualité est un concept fondamental en programmation linéaire et conduit à un résultat de grande portée théorique et pratique : le théorème de dualité.

A tout problème de maximisation sous forme standard

$$\begin{aligned} \max \quad & z = c x \\ \text{s.c.} \quad & Ax = b \\ & x \geq 0, \end{aligned} \tag{P}$$

on associe un problème de minimisation, dit *problème dual*:

$$\begin{aligned} \min \quad & w = b y \\ \text{s.c.} \quad & y A \geq c \\ & y \text{ non restreint en signe.} \end{aligned} \tag{D}$$

De même, au problème de minimisation sous forme standard

$$\begin{aligned} \min \quad & z = c x \\ \text{s.c.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{P}$$

on associe le problème de maximisation *dual*

$$\begin{aligned} \max \quad & w = b y \\ \text{s.c.} \quad & y A \leq c \\ & y \text{ non restreint en signe.} \end{aligned} \tag{D}$$

Dans ce contexte, (P) est appelé *problème primal* de (D). Remarquons que, si (P) comporte n variables et m contraintes, alors (D) comporte m variables et n contraintes (une variable par contrainte de (P), et une contrainte par variable de (P)).

Au prix d'une mise sous forme standard, tout PL a un problème dual. En particulier, le dual du dual (c'est-à-dire, le problème dual de (D)) existe. On peut montrer que le dual du dual est un problème de programmation linéaire équivalent au problème primal .



Existe-t-il une relation entre les valeurs optimales de (P) et de (D) ? Le théorème suivant apporte une première réponse à cette question.

Si x est une solution réalisable de (P) et y est une solution réalisable de (D), alors $cx \leq by$ lorsque (P) est du type 'maximisation' et $cx \geq by$ lorsque (P) est du type 'minimisation'.

Le corollaire suivant en découle :

Si (P) et (D) admettent des solutions optimales (disons x^ et y^* , respectivement), alors*
 $cx^* \leq by^*$ lorsque (P) est du type 'maximisation'
et $cx^* \geq by^*$ lorsque (P) est du type 'minimisation'.

Plus généralement, pour montrer qu'une solution réalisable x^* de (P) est également optimale pour (P), il suffit d'exhiber une solution réalisable y^* de (D) telle que $cx^* = by^*$.

Le théorème de dualité suivant montre que si x^* est optimale, alors ceci est toujours possible:

Si (P) admet une solution optimale x^ , alors (D) admet une solution optimale y^* et $cx^* = by^*$. Plus précisément, si x^* est une solution de base optimale de (P) associée à la base B, alors $y^* = c_B B^{-1}$ est optimal pour (D) et $cx^* = by^*$.*

Enfin le théorème suivant, appelé théorème de complémentarité :

Deux solutions optimale (x^, y^*) du primal et du dual respectivement sont optimales si et seulement si $(y^* \cdot A_j - c_j) \cdot x_j^* = 0$, pour tout j variant de 1 à n .*

permet d'obtenir la solution optimale d'un des programmes, à partir de la connaissance de la solution optimale de son problème dual.

Programmation linéaire avec variables entières.

Lorsque les variables d'un programme mathématique représentent des décisions stratégiques, ou simplement lorsque certaines variables ne peuvent prendre des valeurs fractionnaires, il faut intégrer dans le modèle des contraintes d'intégralité. Ceci modifie profondément la nature des programmes linéaires. Le programme mathématique suivant est un programme linéaire en nombres entiers (PLNE) :

$$\begin{aligned} \text{max } z &= \sum_{j=1}^n c_j x_j \\ \text{s.c. } \sum_{j=1}^n a_{ij} x_j &\leq b_i \quad (i = 1, \dots, m) \\ x_j &\text{ entier non négatif} \end{aligned}$$

En général ce problème n'a de solutions que si les éléments a_{ij} et b_i sont aussi entiers. Si les variables valent seulement 0 ou 1 (variables dites booléennes, binaires ou de décision), on dit que l'on a un programme linéaire en « 0-1 ». Enfin, comme dans l'exemple suivant, on peut avoir seulement certaines variables astreintes à être entières.



Exemple : Une entreprise envisage plusieurs sites de construction pour des usines qui serviront à approvisionner ses clients. A chaque site potentiel i correspond un coût de construction a_i , une capacité de production u_i , un coût de production unitaire b_i et des coûts de transport c_{ij} des usines vers les clients.

Soit y_i une variable binaire prenant la valeur 1 si un entrepôt est construit sur le site i , d_j la demande de l'usine j et x_{ij} la quantité produite à l'usine i et destinée au marché j (fot de i à j). Un plan de construction optimal peut être obtenu en résolvant le programme :

$$\min z = \sum_i a_i y_i + \sum_i b_i \sum_j x_{ij} + \sum_i \sum_j c_{ij} x_{ij}$$

$$\sum_i x_{ij} = d_j$$

$$\sum_j x_{ij} \leq u_i y_i$$

$$x_{ij} \geq 0 \text{ et } y_i = 0 \text{ ou } 1$$

Cette formulation contient deux éléments intéressants : un coût fixe (construction) modélisé par une variable binaire y_i ainsi qu'une contrainte logique forçant les flots provenant d'un site à être nuls si aucune usine n'est construite en ce site. Notons aussi que certaines variables sont entières alors que d'autres (flots) sont réelles. Un tel programme linéaire est appelé mixte.

Comment résoudre un PLNE ?

Une première stratégie de résolution : l'énumération implicite.

Considérons un programme mathématique mixte et laissons de côté, dans un premier temps, les contraintes d'intégralité, et contentons nous de résoudre un programme linéaire (relaxation linéaire, notée RL, du programme en variables entières). Si la solution optimale de ce programme satisfait aux contraintes d'intégralité, alors cette solution est aussi solution optimale du programme avec variables entières.

Sinon, il doit exister au moins une variable x_j dont la valeur α est fractionnaire. On sépare alors le problème en deux : un sous-problème contiendra la contrainte $x_j \leq \text{Int}(\alpha)$, Int désignant la partie entière, et le second la contrainte $x_j \geq \text{Int}(\alpha) + 1$. Il est clair que ceci crée une partition du problème relaxé. On répète le processus pour chacun des sous-problèmes. Cette procédure est habituellement représentée sous forme d'un arbre binaire où, à chaque niveau, une partition du sommet père s'effectue suivant la règle décrite précédemment. Il s'agit alors de parcourir cet arbre d'énumération afin d'y trouver la solution optimale.

L'exploration d'un chemin de l'arbre peut prendre fin pour trois raisons :

- la solution devient entière ;
- le domaine admissible d'un sous-problème devient vide ;
- la valeur de l'objectif correspondant à la solution optimale du problème relaxé est inférieure (moins bonne) à celle d'une solution admissible connue, possiblement obtenue à un autre sommet de l'arbre.

Dans chacun de ces trois cas on dit que le sommet est sondé, et il est inutile de pousser plus loin dans cette direction. L'algorithme s'arrête lorsque tous les sommets sont sondés. La meilleure solution obtenue au cours du déroulement de l'algorithme est alors l'optimum *global* de notre problème.



Cette technique d' énumération partielle est connue sous le vocable de Branch-and-Bound.

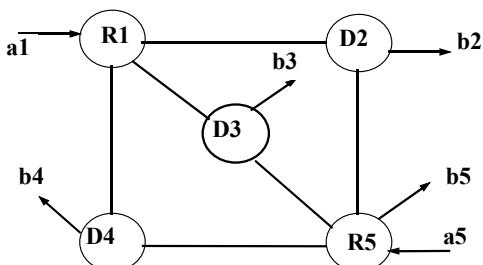
Une seconde stratégie : Méthodes de coupe .

La motivation des méthodes de coupe ou de troncature est de tenter de trouver l'enveloppe convexes des solutions entières, c'est-à-dire le plus petit polyèdre contenant toutes les solutions entières du PLNE.

Le premier à développer des algorithmes spécialisés pour la programmation en nombre entiers a été R.Gomory. La méthode qu'il a proposée consiste essentiellement à tronquer graduellement l'ensemble convexe des solutions réalisable du programme linéaire continu correspondant (RL), par l'ajout successif de contraintes linéaires supplémentaires, de sorte que les points extrêmes des nouveaux ensembles polyédriques ainsi obtenus soient discrets dans la région de la solution optimale. Lorsqu'on résout les nouveaux programme linéaires ainsi engendrées à l'aide de la méthode du simplexe, si la solution trouvée est entière, on conclut qu'elle correspond à la solution optimale du problème du PLNE. Si , en revanche, la solution calculée n'est pas entière, on conclut qu'il existe dans le voisinage de cette solution une région qui ne contient pas de points à coordonnées discrètes et par conséquent on l'élimine de l'ensemble convexe précédent par une troncature convenable. On procède ainsi jusqu'à ce que une solution discrète soit obtenue. De cette stratégie de résolution il existe plusieurs variantes, qui exploitent toutes le mécanisme élaboré par Gomory pour obtenir les contraintes de coupe. L'inconvénient de ces méthodes est qu'elles fournissent une solution réalisable uniquement à la fin, contrairement aux méthodes arborescentes qui livrent en cours de recherches une suite de bonnes solutions. Enfin les méthodes de coupe sont complexes mais sont cependant utiles comme technique d'appoint dans les méthodes arborescentes.

Quelques applications de la programmation linéaire.

Les problèmes de distribution, correspondent à une classe de problèmes susceptibles d'être résolus par les méthodes de la programmation linéaire que leur structure rend aisée à résoudre : ceux où chaque colonne des équations de contrainte contient au plus deux éléments non nuls et qui sont de modules égaux à 1.



Prenons l'exemple d'une compagnie pétrolière avec 2 raffineries (R1 et R5) qui approvisionnent 3 centres de distributions (D2, D3 et D4) par des conducteurs de capacité finie.

Comment doit-elle router son pétrole pour minimiser ses coûts de transport tout en satisfaisant aux demandes. En posant :

ai : production au point i (nulle ou non)

bi : consommation au point i (nulle ou non).

xij : quantité transportée de i à j.

kij : capacité de la route (i,j)

cij : coût unitaire sur la route (i,j).

La conservation du flot nous indique qu'en en chaque ville (i) :

total reçu + production = total expédiée + consommation.



Qui peut se traduire mathématiquement par :

$$\sum_{j \neq i} x_{ji} + a_i = \sum_{j \neq i} x_{ij} + b_i \text{ pour tout } i=1, \dots, n$$

De plus , on doit satisfaire les contraintes de capacité :

$$0 \leq x_{ij} \leq k_{ij} \text{ sur chaque arc } (i,j).$$

Le problème est donc :

$$\text{Min } \sum_{i,j} c_{ij} x_{ij}$$

avec

$$\sum_{j \neq i} x_{ij} - \sum_{j \neq i} x_{ji} = a_i - b_i \text{ pour tout } i=1, \dots, n$$

$$j \neq i$$

$$0 \leq x_{ij} \leq k_{ij}$$

(en général $a_i \cdot b_i = 0$, un nœud est soit producteur, soit consommateur)

En laissant de côté les contraintes de capacité, on a bien un problème de distribution.

Le problème de transport , ou le problème d'affectation, fait partie de cette classe de problème de distribution, où on ne peut faire transiter de matière par un nœud, mais où chaque nœud est soit source soit destination de matière, chaque source étant reliée directement à toutes les destinations.

En n'imposant pas de contraintes de capacités (on suppose alors que la capacité de chaque arc est ∞) le problème est alors :

$$\text{Min } \sum_{i,j} c_{ij} x_{ij}$$

avec

$$j=n$$

$$\sum_{j=1}^n x_{ij} = a_i \quad (i=1, \dots, n) \quad \text{contraintes de production}$$

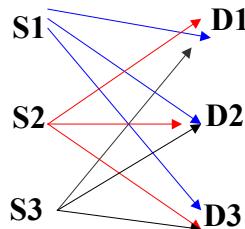
$$j=1$$

$$i=m$$

$$\sum_{i=1}^m x_{ij} = b_j \quad (j=1, \dots, n) \quad \text{contraintes de consommation}$$

$$i=1$$

$$0 \leq x_{ij} \quad (i=1, \dots, m; j=1, \dots, n) \quad \text{contraintes de positivité}$$



De manière évidente, un problème de transport admet une solution si et seulement si les productions équilibrivent les demandes. De plus cette solution est bornée si a_i et b_j sont bornées.

Du fait que toute base d'un problème de transport est totalement unimodulaire (i.e le déterminant de toute sous-matrice carrée extraite = +1 ou -1 ou 0, démonstration faite par récurrence sur l'ordre des sous-matrices) on démontre facilement que si toutes les productions et demandes sont entières, les solutions de base sont entières.



Pour résoudre ce type de problème, on a à notre disposition plusieurs méthodes pratiques de recherche d'une solution de base réalisable (méthodes du coin Nord-Ouest, du coût minimum, de Vogel). La recherche de l'optimum ensuite peut se baser sur les propriétés du problème dual associé :

$$\begin{aligned} [\max] \quad & \sum_i a_i u_i + \sum_j b_j v_j \\ & u_i + v_j \leq c_{ij} \quad i=1, \dots, m ; j=1, \dots, n \\ & u_i, v_j \text{ de signe quelconque.} \end{aligned}$$

Les propriétés du dual indiquent en particulier que :

- d'une part $C_{ij} = u_i + v_j$ pour les variables de bases .
- d'autre part, si la quantité $C_{ij} - u_i - v_j$ pour les variables hors base est négative, on peut faire décroître la fonction objectif en sélectionnant la variable correspondante comme variable entrante dans la base.

Ce problème dispose par ailleurs d'algorithmes rapides travaillant directement sur le graphe associé à ce problème de flot de coût minimal.

1.2.2. Programmation non linéaire.

Afin de décrire la place de l'optimisation dans les applications nous prenons un exemple concernant l'organisation d'un réseau de transports en commun. A travers cet exemple on mettra en évidence des problèmes qu'on peut modéliser et résoudre par des techniques mathématiques d'optimisation non linéaire et de la théorie des graphes et des réseaux.

Formulation de la demande

Dans un département fortement rural et avec des fréquentations des transports en commun à caractère saisonnier, il est nécessaire d'organiser la prise à domicile et le rassemblement des usagers sur des points de concentration (Problème 1), départs ou arrêts d'une ligne régulière de transport en commun (Problème 2). Tout le trafic de rassemblement doit être organisé autour de la grille horaires des transports en commun et doit s'adapter à sa modularité saisonnière ou conjoncturelle. Il est nécessaire de dimensionner la capacité des véhicules de rabattement (Problème 3). Il faut étudier les conséquences du couplage entre le système de rassemblement et la ligne régulière afin de réduire les points d'arrêt compressant d'autant la durée de chaque liaison (Problème 4). L'objectif est de rendre le réseau de transport public plus compétitif au regard de son principal concurrent : la voiture particulière. Ceci améliorerait l'utilisation des subventions, elles ne serviraient plus à faire circuler des véhicules vides mais à amener des passagers vers la ligne régulière qui nécessite moins de soutien.

Parmi les objectifs visés par la mise en place du réseau de rassemblement, on peut citer :

- les voitures individuelles restent au domicile en diminuant ainsi les risques de pollution (Problème 5) ;
- l'investissement doit être bien moins onéreux qu'un parking ;
- le réseau améliore sa rentabilité et ses taux de remplissage,
- des emplois nouveaux et viables économiquement sont créés (Problème 6) ;



Les solutions éventuelles doivent tenir compte :

- des capacités des véhicules ;
- des caractéristiques de chaque client ;
- des tolérances de temps ou d'exactitude souhaitée par le service, des possibilités de groupage ou de circuit, etc...

Une base de données comporte toutes les rues des villes, tous les hameaux du département, ainsi que les principales voies de circulation. Sur cette base, la planification de trajet (Problème 7) pour chaque véhicule de rassemblement doit être réalisée. En équipant les véhicules de balises GPS (Problème 8) on peut indiquer en temps réel les trajets à suivre répondant ainsi au mieux aux besoins des usagers.

Commentaires préliminaires.

Nous avons essayé de décomposer la demande ci-dessus en un certain nombre de sous demandes (appelées Problème 1, ..., Problème 8) afin de faciliter la modélisation et la recherche de solutions. Cette *décomposition* est nécessaire, car le problème ne peut pas être traité dans sa totalité par une seule méthode ou technique (de même qu'un seul corps de métier n'est pas en mesure de construire toute une maison). Il nous apparaît clairement qu'on ne peut pas traiter le Problème 6 (création de nouveaux emplois) avec des méthodes d'optimisation. En fait, un tel problème fait intervenir des facteurs qu'on ne peut pas (ou alors difficilement) quantifier. Les données quantitatives sont la condition nécessaire pour réaliser l'optimisation dans le sens expliqué dans les paragraphes suivants. De même, le Problème 8 (équipement des véhicules avec des balises GPS) est hors du domaine de l'optimisation.

A partir de la base de données et des données cartographiques, on pourrait modéliser (voir les paragraphes suivants) le Problème 1 (organisation des rassemblements) comme problème de couverture, la structure minimum de la ligne régulière qui concerne le Problème 2 comme problème de recherche d'un arbre dans un graphe et le Problème 7 (planification de trajet) comme problème d'ordonnancement. Ces problèmes peuvent être modélisés mathématiquement, généralement en faisant une représentation *simplifiée* du problème réel.

Quant à l'impact des modifications souhaitées sur la diminution des risques de pollution (Problème 5) on pourrait envisager des modèles de prédiction faisant ainsi appel aux experts en acquisition de connaissances à partir de données (provenant de capteurs par exemple). Dans ce cas aussi, on n'est pas concerné directement par le domaine de l'optimisation.

La simplification des modèles nécessite une phase de *validation* de même que la décomposition du problème nécessite une phase d'*intégration* des résultats des sous modèles. Ces deux phases ne font pas partie du domaine de l'optimisation proprement dit, elles nécessitent une collaboration obligatoire et impliquent l'interaction entre les acteurs intervenants.

Les remarques ci-dessus délimitent le champ d'action des méthodes d'optimisation. On verra aussi que des raisons intrinsèques aux techniques de solution font que l'optimisation non linéaire utilisée pour résoudre le Problème 1 n'est d'aucune utilité pour le Problème 7. Ceci justifie la cohabitation au sein des méthodes d'optimisation de plusieurs techniques.



Modélisation du Problème 1.

Pour simplifier l'exposé, on supposera que le département étudié possède un réseau de transport en commun présenté par des traits gras sur la figure 46. On représentera par un cercle la surface couverte par un véhicule de rassemblement. Une étude préliminaire permettra de définir le domaine de variation du centre du cercle et aussi de son rayon. Il semble envisageable de situer le centre du cercle dans tel ou tel pâté de maisons. Le rayon doit varier de sorte que le cercle touche le réseau principal en au moins un point. La qualité de couverture sera exprimée par la fonction objectif et le fait que les centres et les rayons des cercles ne peuvent pas prendre de valeurs quelconques sera exprimé par les contraintes.

Pour écrire formellement le modèle nous définissons les variables suivantes :

- Les coordonnées (x_k, y_k) des centres C_k pour $k = 1, \dots, n$;
- Les limites $x_{min}(k), x_{max}(k), y_{min}(k)$ et $y_{max}(k)$ définissant les rectangles contenant les centres C_k ;
- Les rayons z_k des cercles pour $k = 1, \dots, n$;
- Les limites $z_{min}(k), z_{max}(k)$ définissant les limites des rayons des cercles ;

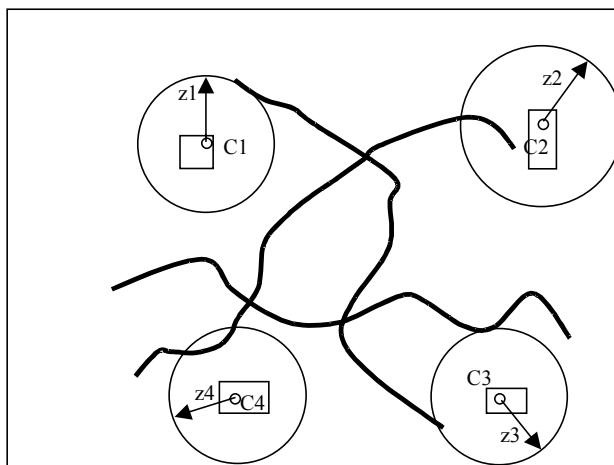


Figure 46 : Présentation schématique d'un réseau de transport en commun et du réseau de rassemblement.

Fonction objectif à maximiser

$$f(x_1, \dots, x_n, y_1, \dots, y_n, z_1, \dots, z_n) = \sum_{k=1}^n \pi \cdot z_k^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n d(C_i, C_j) \quad (18)$$

Le premier terme de cette fonction exprime la superficie totale couverte par les cercles. Le deuxième terme exprime l'éloignement des cercles les uns par rapport aux autres. La distance entre les centres des cercles est $d(C_i, C_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$.

Afin de couvrir au mieux la zone desservie, on cherchera à maximiser la fonction objectif. Remarquons toute de suite que ce n'est pas une fonction linéaire.

Contraintes

1. Le centre C_k doit appartenir au rectangle respectif :

$$x_{min}(k) \leq x_k \leq x_{max}(k),$$



$$\begin{aligned} y_{min}(k) \leq y_k \leq y_{max}(k) \\ \text{pour } k = 1, \dots, n. \end{aligned} \tag{19}$$

2. Les rayons doivent prendre des valeurs dans les limites prescrites :

$$\begin{aligned} z_{min}(k) \leq z_k \leq z_{max}(k), \\ \text{pour } k = 1, \dots, n. \end{aligned} \tag{20}$$

3. Les cercles ne doivent pas se couvrir.

$$\begin{aligned} d(C_i, C_j) \geq z_i + z_j \\ \text{pour } i = 1, \dots, n - 1 \text{ et } j = i + 1, \dots, n. \end{aligned} \tag{21}$$

Forme générale d'un problème d'optimisation non linéaire

Ainsi le Problème 1 peut se formaliser comme suit :

Maximiser la fonction objectif (18) en satisfaisant les contraintes (19), (20) et (21).

On considérera que les fonctions participant à la formulation du problème sont des fonctions de vecteurs $\mathbf{x} = [x_1, x_2, \dots, x_n]$ appartenant à l'espace \mathbf{R}^n ou à un sous ensemble \mathbf{D} de \mathbf{R}^n .

Formulation du problème

$$\begin{aligned} & \text{Minimiser } f(\mathbf{x}) \\ & \text{pour } \mathbf{x} \text{ vérifiant les contraintes} \\ & \quad \mathbf{x} \in \mathbf{D} \subseteq \mathbf{R}^n \\ & \quad \mathbf{D} = \{ \mathbf{x} \in \mathbf{R}^n \mid h_1(\mathbf{x}) = 0, \dots, h_m(\mathbf{x}) = 0, g_1(\mathbf{x}) \leq 0, \dots, g_p(\mathbf{x}) \leq 0 \} \end{aligned}$$

La fonction $f(\mathbf{x})$ est la *fonction objectif*. Les contraintes $h_i(\mathbf{x}) = 0, i = 1, \dots, m$ et $g_j(\mathbf{x}) \leq 0, j = 1, \dots, p$ sont les contraintes fonctionnelles. La plupart des développements théoriques nécessitent la continuité et la dérivabilité des fonctions ci-dessus jusqu'au deuxième ordre.

Habituellement, on traite séparément le problème d'optimisation sans contraintes fonctionnelles du problème avec ces contraintes. Nous suivrons cette ligne classique de présentation en commençant par l'optimisation non linéaire sans contraintes et en traitant par la suite le problème avec contraintes.

Optimisation non linéaire sans contraintes

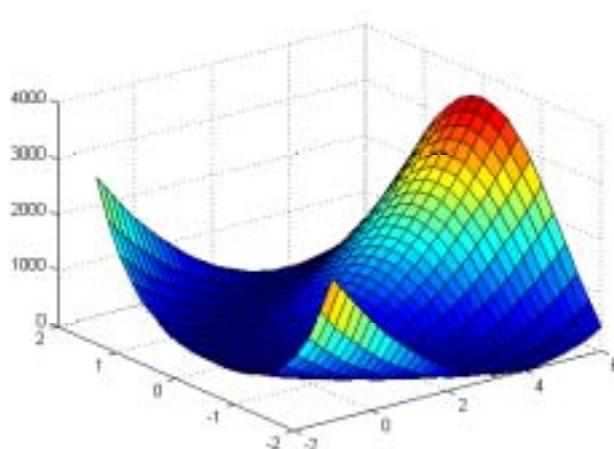
Soit $f(\mathbf{x})$, $\mathbf{x} = [x_1, \dots, x_n] \in \mathbf{R}^n$, une fonction non linéaire de n variables réelles. Ce paragraphe traite le problème suivant :

Trouver $\mathbf{x}^* \in \mathbf{R}^n$ tel que $f(\mathbf{x}^*) \leq f(\mathbf{x})$ pour tout $\mathbf{x} \in \mathbf{R}^n$.

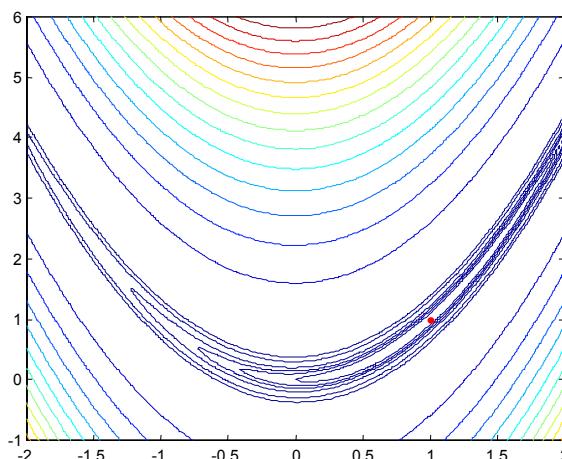
Il s'agit donc, de minimiser $f(\mathbf{x})$ dans l'espace \mathbf{R}^n .

Nota : tous les vecteurs apparaissant sont des vecteurs colonnes. Les lignes sont des vecteurs transposés \mathbf{x}^T

Une fonction utilisée largement comme fonction test pour les différentes méthodes numériques d'optimisation est la fonction suivante (appelée "banana function") :

Figure 47 : graphique de la "banana fonction" dans $[-2, 2] \times [-1, 6]$.

Cette fonction est non négative pour toutes les valeurs de x_1 et x_2 et s'annule pour $x_1=1$ et $x_2=1$ qui est donc son optimum. Malgré la simplicité du calcul analytique de son optimum, cette fonction est difficile à optimiser par des méthodes numériques à cause de sa forme en vallée courbée. Une autre difficulté vient du fait que cette fonction est relativement "plate" (voir les lignes de niveau ci-dessous) autour de l'optimum ce qui rend la détection de celui-ci difficile.

Figure 48 : les lignes de niveau de la "banana fonction" dans $[-2, 2] \times [-1, 6]$. L'optimum $[1, 1]$ est marqué par un point.

Conditions de minimum

Avant de présenter quelques algorithmes d'optimisation il est nécessaire de rappeler le développement limite d'une fonction de n variables, à proximité d'un point donné \mathbf{x}^0 :

$$f(\mathbf{x}) = f(\mathbf{x}^0) + \sum_{i=1}^n \frac{\partial f(\mathbf{x}^0)}{\partial x_i} (x_i - x_i^0) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f(\mathbf{x}^0)}{\partial x_i \partial x_j} (x_i - x_i^0) \cdot (x_j - x_j^0) + \dots$$

Pratiquement, toutes les méthodes d'optimisation travaillent avec une approximation jusqu'à l'ordre 2 de $f(\mathbf{x})$. Autrement dit, au lieu d'optimiser $f(\mathbf{x})$ autour d'un point donné \mathbf{x}^0 , on optimise son approximation $g(\mathbf{x})$:



$$g(\mathbf{x}) = f(\mathbf{x}^0) - \mathbf{b}^T \cdot (\mathbf{x} - \mathbf{x}^0) + \frac{1}{2} \cdot (\mathbf{x} - \mathbf{x}^0)^T \cdot H \cdot (\mathbf{x} - \mathbf{x}^0) \quad (5)$$

Le vecteur $\mathbf{b} = \nabla f(\mathbf{x}^0)$ contient les n dérivées partielles de $f(\mathbf{x})$ calculées au point \mathbf{x}^0 . Ce vecteur s'appelle le *gradient* de $f(\mathbf{x})$ au point \mathbf{x}^0 .

La matrice carrée H d'ordre n , est le *Hessien* de $f(\mathbf{x})$ au point \mathbf{x}^0 . Cette matrice contient les dérivées partielles doubles de $f(\mathbf{x})$ au point \mathbf{x}^0 , $H(i, j) = \frac{\partial^2 f(\mathbf{x}^0)}{\partial x_i \partial x_j}$.

Condition de minimum local :

Le point \mathbf{x}^* est un minimum local pour une fonction $f(\mathbf{x})$ si et seulement si le gradient de $f(\mathbf{x})$ s'annule pour $\mathbf{x} = \mathbf{x}^*$ et le Hessien est défini positif.

Ces conditions sont connues depuis le 17^e siècle grâce à Fermat. Elles ont été formalisées au 18^e siècle par Lagrange [Lagrange, 1797] et utilisé largement depuis les années 1960. C'est pendant cette période qu'on a vu les premiers algorithmes d'optimisation et les premiers ouvrages sérieux traitant le sujet. On peut mentionner [Luenberger, 1973], [Fletcher, 1980 ; 1981], [Dantzig *et al.*, 1960], [Wolfe, 1961]. Les conditions théoriques d'optimum qui utilisent l'information des dérivées de premier et deuxième ordre sont connues sous le nom commun des méthodes de Newton. Le calcul des dérivées seconde est assez consommateur en temps de calcul et avant la mise au point de la différentiation automatique ceci était considéré comme un sévère inconvénient. Ainsi pendant une vingtaine d'années plusieurs méthodes d'optimisation n'utilisant pas les dérivées seconde ont été mises au point formant comme la classe des méthodes quasi-Newton. Curieusement, l'utilisation de certaines méthodes bien analysées de cette classe donne souvent de meilleurs résultats par rapport aux méthodes utilisant l'information des dérivées seconde. Une des raisons est le fait que ces méthodes se "souviennent" d'une certaine façon du comportement global de $f(\mathbf{x})$ alors que les méthodes de Newton utilisent seulement l'information locale. On peut affirmer actuellement que les méthodes quasi-Newton peuvent résoudre sans difficultés des problèmes d'optimisation sans contraintes allant jusqu'à quelques centaines de variables (problèmes d'échelle moyenne) même dans le cas où le Hessien est mal conditionné. Malheureusement, aucune méthode de cette classe n'est capable de traiter correctement le cas où le Hessien est une matrice creuse. Beaucoup d'effort de recherche concerne récemment les problèmes de grande échelle (quelques milliers de variables) [Buckley *et al.*, 1983], [Liu *et al.*, 89]. Il existe actuellement des programmes informatiques utilisant ces idées récentes pouvant résoudre des problèmes allant jusqu'à 10000 variables.

En dimension élevée l'effort principal concerne la détermination d'une direction de descente \mathbf{v} . Ayant \mathbf{v} , la majorité des algorithmes d'optimisation non-linéaire suivent la recherche sur une ligne dont la trame est ci-dessous :

Algorithme itératif d'optimisation

Trouver un point initial $\mathbf{x}(0)$;

Tant que critère d'arrêt non satisfait **faire**

Trouver une direction de descente \mathbf{v} ;

Trouver $\mathbf{x}(1) = \mathbf{x}(0) + \sigma \mathbf{v}$ tel que $f(\mathbf{x}(1)) < f(\mathbf{x}(0))$;

Remplacer $\mathbf{x}(0)$ par $\mathbf{x}(1)$;

Fin tant que.

Le choix du paramètre σ est fait pour réaliser $f(\mathbf{x}(1)) < f(\mathbf{x}(0))$. Ces méthodes présentent quelques difficultés :

- comment trouver un point initial $\mathbf{x}(0)$ qui ne soit pas "trop" éloigné de l'optimum ?



- comment définir un critère d'arrêt qui soit suffisamment fin pour détecter un minimum même lorsqu'on se trouve sur un "plateau" ?
- quelle est la meilleure direction de descente v et le meilleur pas de descente Δ ?

Les difficultés mentionnées ci-dessus ont pour conséquence principale de converger vers des minima locaux. Ainsi, les algorithmes de descente, peuvent se bloquer dans un minimum local tout en étant loin de l'optimum global. Une stratégie simple pour éviter les minima locaux est de recommencer la descente à partir de différents points initiaux.

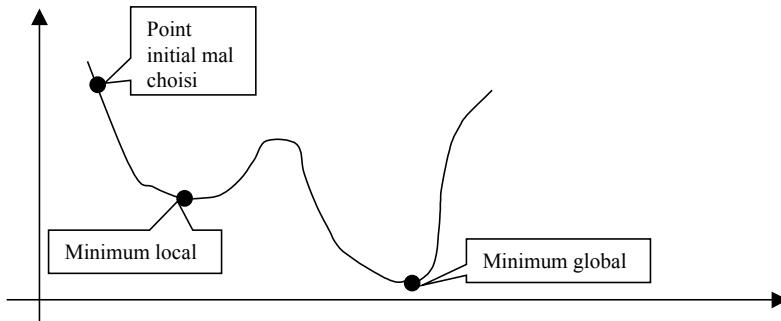


Figure 49 : Point initial, minimum local et minimum global.

Revenons maintenant à la minimisation d'une forme quadratique. C'est un des cas les plus rencontrés car il traduit le principe fondamental en mécanique : le mouvement est orienté de sorte à minimiser l'énergie. Souvent l'énergie s'exprime par une fonction quadratique. Son minimum se trouve au point où sa dérivée s'annule et la dérivée d'une fonction quadratique est une fonction linéaire.

L'importance des formes quadratiques vient aussi du fait qu'on peut approcher suffisamment bien une fonction quelconque par les trois premiers termes de son développement limité (voir formule (5)). Soit donc la forme quadratique suivante :

$$g(\mathbf{x}) = a - \mathbf{b}^T \cdot \mathbf{x} + \frac{1}{2} \cdot \mathbf{x} \cdot H \cdot \mathbf{x}$$

où, a est une constante. On peut montrer le résultat suivant :

Si H est définie positive alors $g(\mathbf{x})$ est minimisée au point $\mathbf{x} = H^{-1} \cdot \mathbf{b}$.

Ainsi, la solution d'un problème d'optimisation quadratique se transforme en solution d'un système algébrique d'équations linéaires, sujet qui est bien maîtrisé en théorie comme en pratique. Nous utiliserons l'optimisation de la fonction quadratique dans l'algorithme de Newton décrit par la suite.

Algorithmes Descente à pas, Newton et quasi-Newton

L'algorithme Descente à pas (en anglais "steepest descent"), dont l'idée remonte à Cauchy, utilise le gradient. Il entre dans la classe des algorithmes itératifs. L'idée de base repose sur le résultat suivant :

Soit $g(\mathbf{x}) = a - \mathbf{b}^T \cdot \mathbf{x} + \frac{1}{2} \cdot \mathbf{x}^T \cdot H \cdot \mathbf{x}$, l'approximation quadratique

de $f(\mathbf{x})$. La fonction $f(\mathbf{x})$ est minimisée localement dans la direction décrite par le vecteur $-\mathbf{b}$.



Le fait d'utiliser le gradient en plus des valeurs de $f(\mathbf{x})$ peut accélérer la convergence vers le minimum. Ceci est vrai pour un bon nombre de fonctions. Pourtant, le nombre d'itérations peut être assez élevé pour des fonctions présentant des "vallées étroites" comme c'est le cas de la "banana function".

Algorithme "Descente à pas"

Trouver un point initial $\mathbf{x}(0)$;

Tant que critère d'arrêt non satisfait faire

 Calculer le gradient \mathbf{b} de la fonction $f(\mathbf{x})$ au point $\mathbf{x}(0)$;

 Trouver le scalaire a^* minimisant la fonction $F(a) = f(\mathbf{x}(0)) - a \mathbf{b}$;

 Remplacer $\mathbf{x}(0)$ par $\mathbf{x}(1) = \mathbf{x}(0) - a^* \mathbf{b}$;

Fin tant que.

Les critères d'arrêt peuvent changer selon la nature du problème qu'on traite. Un critère d'arrêt peut être un seuil sur le nombre d'estimations de $f(\mathbf{x})$. Un autre critère peut être la comparaison de l'erreur relative avec une tolérance définie d'avance.

Le calcul du gradient \mathbf{b} demande en général n estimations fonctionnelles. La minimisation unidimensionnelle pour trouver a^* nécessite plusieurs estimations. Ces deux points sont coûteux en temps de calcul. Généralement, ceci est compensé par le fait que le gradient pointe dans la "bonne direction".

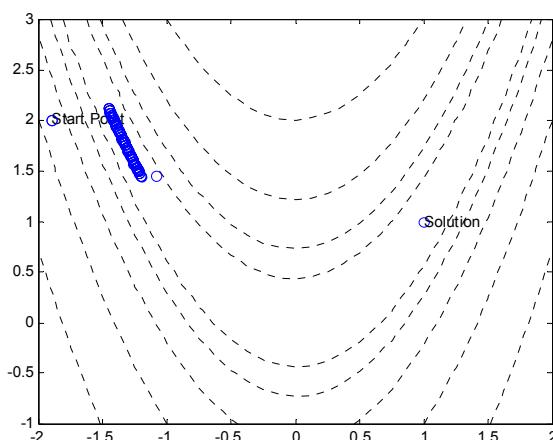


Figure 50 : La méthode "descente à pas" après 302 itérations appliquée à "banana function". On voit que la méthode "piétine assez loin du point initial [-1.9, 2].

Les méthodes Newton et quasi-Newton sont les plus répandues parmi celles qui utilisent les dérivées partielles de $f(\mathbf{x})$. En fait, elles utilisent le gradient et le Hessien de $f(\mathbf{x})$ pour donner une approximation de sa courbure.

Soit $g(\mathbf{x}) = a - \mathbf{b}^T \cdot \mathbf{x} + \frac{1}{2} \mathbf{x} \cdot H \cdot \mathbf{x}$ l'approximation quadratique de $f(\mathbf{x})$. Nous savons que $g(\mathbf{x})$

est minimisée au point $\mathbf{x} = H^{-1} \cdot \mathbf{b}$. Ci-dessous est donné la méthode de Newton pour marquer, dans l'encadré qui suit, la différence avec les méthodes quasi-Newton.

Algorithme Newton

Trouver un point initial $\mathbf{x}(0)$;

Tant que critère d'arrêt non satisfait faire



Calculer le gradient \mathbf{b} de la fonction $f(\mathbf{x})$ au point $\mathbf{x}(0)$;
Calculer le Hessien H de la fonction $f(\mathbf{x})$ au point $\mathbf{x}(0)$;
Remplacer $\mathbf{x}(0)$ par $\mathbf{x}(1) = \mathbf{x}(0) - H^{-1} \cdot \mathbf{b}$;
Fin tant que

Le problème avec l'algorithme Newton est le *calcul* de H à chaque itération. En fait, en dimension n , H nécessite le calcul de n^2 dérivées partielles secondes, ce qui est très coûteux en temps de calcul. Pour éviter ces calculs les méthodes quasi-Newton se contentent de faire une *estimation* du Hessien H au lieu de le calculer :

Algorithme quasi-Newton

Trouver un point initial $\mathbf{x}(0)$;

Tant que critère d'arrêt non satisfait **faire**

Calculer le gradient \mathbf{b} de la fonction $f(\mathbf{x})$ au point $\mathbf{x}(0)$;
Estimer le Hessien H de la fonction $f(\mathbf{x})$ au point $\mathbf{x}(0)$;
Remplacer $\mathbf{x}(0)$ par $\mathbf{x}(1) = \mathbf{x}(0) - H^{-1} \cdot \mathbf{b}$;
Fin tant que

Il existe un grand nombre d'estimations du Hessien. La plus répandue est la formule BFGS (par Broyden, Fletcher, Goldfarb et Shanno [Broyden, 1967]), qui apparaît être la plus effective pour une grande classe de fonctions :

$$H_{k+1} = H_k + \frac{\mathbf{q}(k) \cdot \mathbf{q}(k)^T}{\mathbf{q}(k)^T \cdot \mathbf{s}(k)} - \frac{H_k^T \cdot \mathbf{s}(k)^T \cdot \mathbf{s}(k) \cdot H_k}{\mathbf{s}(k)^T \cdot H_k \cdot \mathbf{s}(k)} \quad \text{où}$$

$$\mathbf{s}(k) = \mathbf{x}(k+1) - \mathbf{x}(k) \text{ et } \mathbf{q}(k) = \mathbf{b}(k+1) - \mathbf{b}(k).$$

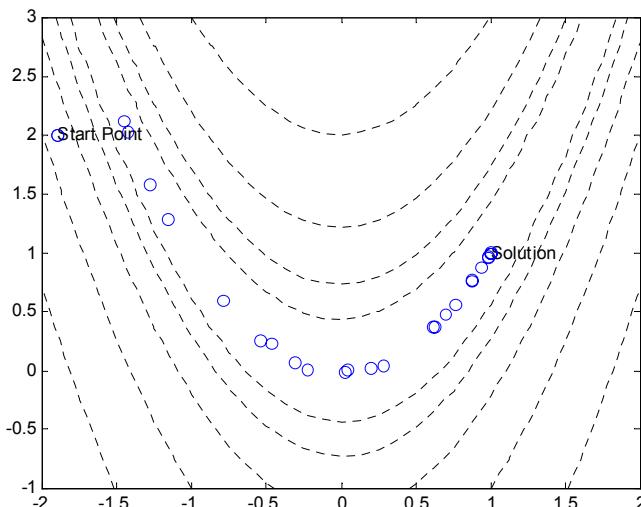


Figure 51 : La méthode BFGS pour la "banana function". La convergence est très rapide

La valeur de la fonction trouvée après 105 itérations est $9 \cdot 10^{-9}$ ce qui peut être considéré très proche de 0.

Ainsi, la formule ci-dessus construit une approximation du Hessien à l'itération $k + 1$, en n'utilisant que les gradients à l'itération k et $k + 1$. Ceci nécessite seulement le calcul de n dérivées partielles au lieu de n^2 pour le calcul direct du Hessien.



Optimisation non linéaire avec contraintes

Le traitement de problèmes avec des contraintes quelconques n'a pas connu le même succès que le problème avec contraintes linéaires même si quelques approches réussies existent. La difficulté principale vient du fait que les fonctions non linéaires présentent des courbures et variations très difficiles à détecter et tracer numériquement. Malgré cela, la solution des problèmes de petite à moyenne échelle est actuellement une routine et des librairies mathématiques de solvers existent en version commerciale. On peut mentionner la librairie NAG et IMSL. Un guide sur les codes existants en optimisation est donné dans [Fourer, 2001]. Il existe aussi des programmes dans le domaine public qui présentent une bonne performance [Mittelmaier, 2000].

Toutes les méthodes d'optimisation avec contraintes utilisent le principe de Lagrange formulé ci-dessous:

Principe de Lagrange 1788 :

Un problème de minimisation d'une fonction $f(x)$ avec p contraintes $f_k(x) \leq 0$, peut être transformé en recherche d'un point stationnaire du *lagrangien* :

$$L(x, \lambda) = f(x) + \sum_{k=1}^p \lambda_k f_k(x)$$

où les λ_k sont des constantes positives appelés coefficients de Lagrange.

Il est intéressant de noter qu'une contribution majeure dans la théorie est survenue seulement 173 ans après Lagrange ! Cette contribution apportée par Kuhn et Tucker est formulée comme suit :

Principe de Kuhn et Tucker 1951 :

Lorsque les fonctions $f(x)$ (à minimiser) et les contraintes $f_k(x) \leq 0$ sont convexes, on peut chercher un point selle (x_0, λ^*) du lagrangien qui vérifie :

$$\min_x L(x, \lambda^*) = L(x_0, \lambda^*) = \max_{\lambda \geq 0} L(x_0, \lambda)$$

L'importance du cas convexe en optimisation peut être résumée par la phrase de Rockafellar [Rockafellar, 1993]

"La vraie ligne de démarcation en optimisation n'est pas entre linéaire et non linéaire mais entre convexe et non convexe".

Les principes ci-dessus donnent les grandes lignes d'approche pour l'optimisation sous contraintes. Pourtant les problèmes de grande échelle restent difficiles à résoudre et ils font l'objet actuellement d'une recherche intensive. Sans entrer dans les détails très techniques on peut mentionner cinq axes principaux d'approches pour les problèmes de grande échelle :

1. Modification de méthodes bien maîtrisées en introduisant dans la partie algébrique des techniques de traitement des matrices creuses.
2. Transformation des contraintes et de la fonction objectif pour obtenir un problème équivalent avec seulement des contraintes de borne, c'est-à-dire des contraintes de



type $a_k < x_k < b_k$. Ainsi, on simplifie les contraintes en compliquant la fonction objectif.

3. Linéarisation de la fonction objectif et des contraintes. Ceci peut se faire en prenant des approximations linéaires de ces fonctions et résoudre un problème linéaire localement.
4. Utilisation de nouvelles versions de la Programmation Quadratique Séquentielle. On fait une approximation quadratique de la fonction objectif et une approximation linéaire des contraintes.
5. On adopte des méthodes du point intérieur lesquelles, contrairement à la méthode de simplexe en programmation linéaire qui parcourt les bords du domaine faisable, traversent ce domaine.

Actuellement, on ne connaît pas une approche clairement gagnante par rapport aux autres. Ainsi, devant un problème d'optimisation non linéaire avec contraintes on teste plusieurs méthodes pour choisir la mieux adaptée.

1.2.3. La théorie des graphes et réseaux

Le premier article sur les graphes remonte à L. Euler en 1736 [Euler, 1736]. Il avait comme point de départ une curiosité mathématique : comment parcourir une fois et une seule les sept ponts de la ville de Koenigsberg ? Pendant longtemps les travaux sur les graphes ne déborderont guère le cadre des récréations mathématiques. C'est seulement à partir de 1946 que la théorie des graphes connaît un grand développement sous l'impulsion de nombreux spécialistes de la recherche opérationnelle motivés par des problèmes concrets [Kuhn, 1955], [Ford *et al.*, 1956]. On assiste alors à une explosion des recherches et des applications. Ceci a fait sentir le besoin d'unification et d'établissement de bases théoriques qui sont traités dans le livre fondamental de Claude Berge [Berge, 1958] réédité à plusieurs reprises. Depuis, plusieurs revues spécialisées publient des centaines d'articles par an sur des classes de graphes ou sur des algorithmes de résolution de tel ou tel problème.

Le succès des graphes vient en grande partie de la facilité qu'ils ont pour représenter simplement la structure d'un grand nombre de situations. L'exemple le plus classique est la représentation d'un réseau de communication : réseau de routes, de chemins de fer, de téléphone, d'informations,

Un graphe $\mathbf{G} = (\mathbf{X}, \mathbf{E})$ est défini par un ensemble de *sommets* \mathbf{X} et par un ensemble d'*arêtes* \mathbf{E} qui représentent des liens entre les sommets.

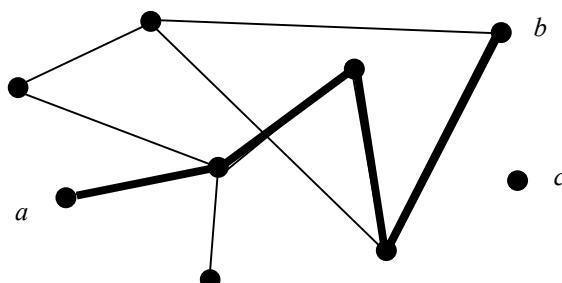


Figure 52 : Un graphe avec 9 sommets et 8 arêtes.



D'autres informations numériques peuvent venir s'ajouter à la structure d'un graphe. Un exemple courant est la capacité des arêtes qui peut représenter le nombre de véhicules en une unité de temps sur un tronçon de route ou bien le débit dans un réseau informatique.

Afin de donner un aperçu des problèmes traités par les techniques et méthodes de la théorie des graphes nous avons sélectionné quatre problèmes typiques et leur méthode de résolution.

Problème de connexité et d'arbre minimum

La première question à résoudre lors de la construction d'un réseau est sa *connexité*. Autrement dit, si a et b sont deux points quelconques du réseau il doit exister au moins un chemin les reliant. La formulation de ce problème en terme de graphes est immédiate. Un graphe $\mathbf{G} = (\mathbf{X}, \mathbf{E})$ est connexe si pour tout $a, b \in \mathbf{X}$, il existe au moins un chemin $a = a_1b_1, \dots, a_pb_p = b$ tel que $a_ib_i \in \mathbf{E}$ et $a_{i+1} = b_i \in \mathbf{X}$.

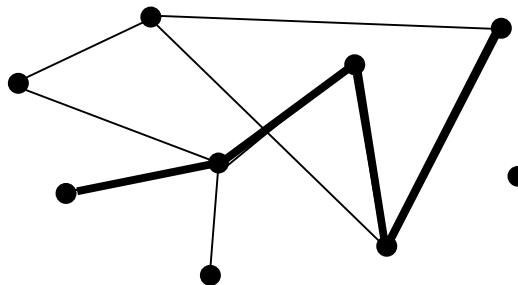


Figure 53 : Chemin entre a et b (en gras).

Le graphe montré sur la figure ci-dessus n'est pas connexe car le sommet c ne peut être relié à aucun autre.

Il est clair qu'un réseau est conçu initialement pour assurer la connexité. Or on a vu que pendant les inondations ou autres catastrophes naturelles, la configuration du réseau est modifiée et on souhaite savoir en temps réel s'il assure la connexité et sinon connaître les zones isolées. L'algorithme de détection de la connexité est donné ci-dessous. La cardinalité ou le nombre d'éléments d'un ensemble \mathbf{X} est noté par $|\mathbf{X}|$.

Algorithme : Connexité

Entrée : Graphe $\mathbf{G} = (\mathbf{X}, \mathbf{E})$ avec $|\mathbf{X}| = n$ et $|\mathbf{E}| = m$.
Sortie : Réponse = 1 si \mathbf{G} est connexe et 0 sinon.

```
A = {a} ; B = X \ {a} ;
Tant qu'il existe xy ∈ E avec x ∈ A et y ∈ B faire
    A = A ∪ {y} ; B = B \ {y} ;
Fin tant que
Si |A| = n alors
    Réponse = 1 ;
sinon
    Réponse = 0 ;
Fin si
```

L'idée de cet algorithme est de commencer avec un ensemble \mathbf{A} contenant un sommet quelconque a et de l'augmenter tant qu'on peut avec des sommets atteignables en dehors de \mathbf{A} . Le temps de calcul est proportionnel avec le nombre de sommets n . En langage informatique



on dit que la complexité de l'algorithme est polynomiale $O(n)$. Ceci demande une fraction de seconde pour tester la connexité d'un graphe avec plusieurs milliers de sommets. En fait, la connexité fait partie de la classe de problèmes combinatoires pour lesquels il existe un algorithme polynomial de résolution. On verra par la suite d'autres problèmes dont le temps de calcul est exponentiel, ce qui en pratique veut dire infini si la taille du problème est grande.

Supposons maintenant qu'on cherche le réseau minimal qui relie tous les sommets du graphe. Le terme *minimal* est utilisé ici dans le sens où le réseau n'est plus connexe si on enlève au moins une arête. Ce réseau minimal serait une sorte de squelette reliant tous les sommets sans redondance, autrement dit il y aura un seul chemin reliant n'importe quel couple de sommets. Ainsi, étant donné un graphe connexe $\mathbf{G} = (\mathbf{X}, \mathbf{E})$ on cherche un sous-graphe $\mathbf{T} = (\mathbf{X}, \mathbf{E}_T)$ de \mathbf{G} tel que \mathbf{T} est connexe et $\mathbf{T} \setminus e$ n'est pas connexe pour n'importe quelle arête $e \in \mathbf{E}_T$. Un tel graphe \mathbf{T} s'appelle un *arbre*. Remarquons que \mathbf{T} a le même ensemble de sommets que \mathbf{G} et généralement $\mathbf{E}_T \subset \mathbf{E}$.

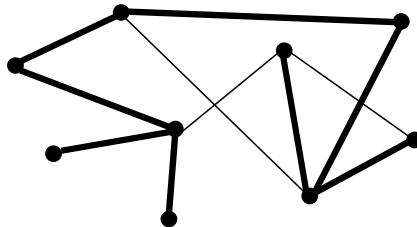


Figure 54 : Un graphe connexe et un arbre (en gras).

On peut remarquer dans la figure ci-dessus que si on enlève les arêtes en trait fin on ne pourra enlever aucune arête en gras sans déconnecter le graphe.

Nous nous intéressons maintenant à la recherche d'un arbre dans un graphe donné. Si on connaît les longueurs $l(e)$ des arêtes on peut même chercher l'arbre \mathbf{T} de longueur totale $L(\mathbf{T}) = \sum_{e \in E_T} l(e)$ minimum. La "longueur" d'une arête peut être le temps moyen de traversée du tronçon de réseau représenté, ce qui est une information pertinente dans un réseau urbain où la longueur et le temps de traversée ne sont pas forcément proportionnels.

Les algorithmes de recherche d'un arbre minimum procèdent pas à pas à partir d'une arête du graphe \mathbf{G} . A chaque itération on ajoute une arête dans l'arbre partiel ainsi créé seulement si elle ne forme pas de cycle avec les arêtes choisies précédemment.

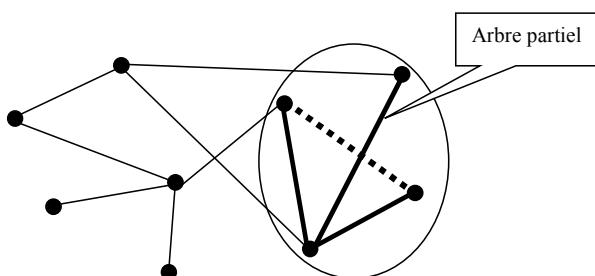


Figure 55 : L'arête en pointillé ne peut pas être ajouté à l'arbre partiel (en gras) car elle crée un cycle.



Algorithme : Arbre de poids minimum [Kuhn et al., 1956]

Entrée : Graphe connexe $\mathbf{G} = (\mathbf{X}, \mathbf{E})$ avec $|\mathbf{X}| = n$ et $|\mathbf{E}| = m$.

Sortie : Arbre $\mathbf{T} = (\mathbf{X}, \mathbf{E}_T)$ de poids minimum.

Ordonner les arêtes de \mathbf{G} dans l'ordre de longueurs croissantes ;

$$l(e_1) \leq l(e_2) \leq \dots \leq l(e_m)$$

$\mathbf{E}_T = \{e_1\}$; $k = 1$;

Tant que $k < m$

$k = k + 1$;

Si e_k ne forme pas de cycle avec les arêtes de \mathbf{E}_T **alors**

$$\mathbf{E}_T = \mathbf{E}_T \cup \{e_k\}$$

sinon

$k = k + 1$; // On passe à l'arête suivante

Fin si

Fin tantque

Il existe plusieurs algorithmes de recherche d'un arbre minimum dans un graphe. Ils sont plus ou moins efficaces selon que le graphe est plus ou moins dense. Ils sont tous polynomiaux et ne dépassent pas un temps de calcul de l'ordre $O(n^2)$. Plusieurs bons algorithmes dans les graphes, dont celui de recherche d'un arbre minimum, et leur analyse sont donnés dans [Tarjan, 1977].

Problème du cycle eulérien et hamiltonien

Etant donné un graphe $\mathbf{G} = (\mathbf{X}, \mathbf{E})$ et les longueurs des arêtes $l(e)$ pour tout e dans \mathbf{E} on peut poser les problèmes suivants :

Problème du cycle Eulerien.

Peut-on traverser toutes les *arêtes* de \mathbf{G} une et une seule fois et revenir au point de départ ? Si la réponse est non, trouver les arêtes à retraverser dont la somme des longueurs est minimale.

Typiquement, ce problème décrit la tournée d'un camion de ramassage de poubelles qui part d'un point donné, traverse toutes les routes du réseau et revient au point de départ.

Problème du cycle Hamiltonien.

Peut-on parcourir tous les *sommets* de \mathbf{G} une et une seule fois et revenir au point de départ ? Si la réponse est oui, trouver le circuit de longueur minimale.

Typiquement, ce problème décrit la tournée d'un bus de ramassage scolaire qui doit visiter tous les points d'arrêt des élèves et revenir au point de départ. On peut considérer qu'un réseau routier qui n'a pas cette possibilité est sûrement à revoir. Mais, l'existence du cycle Hamiltonien n'est pas évidente à tester. En plus, la recherche d'un circuit de longueur minimale (connu dans la littérature comme le problème du voyageur de commerce) est un problème difficile dans le sens où les seuls algorithmes connus qui fourniraient une solution *exacte* nécessitent un temps de calcul exponentiel vis-à-vis de la taille du problème. Rechercher la solution exacte de ce problème pour un graphe avec 30 sommets demanderait plusieurs dizaines d'années aux ordinateurs actuels. Ainsi, on peut se contenter de solutions approchées qu'on trouve avec des méthodes heuristiques.



Revenons au problème du cycle eulerien. Même si ces deux problèmes se ressemblent, les conditions d'existence et la recherche d'un cycle eulerien sont faciles (toujours en termes d'efficacité algorithmique). En fait Euler à démontré le résultat suivant :

Condition d'Euler.

Le graphe \mathbf{G} a un cycle eulerien si et seulement si les degrés de tous les sommets sont pairs.

Le *degré* d'un sommet est le nombre d'arêtes adjacentes.

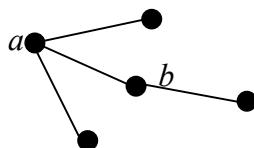


Figure 56 : le sommet a est de degré 3 et le sommet b de degré 2. Il est évident que le graphe ci-dessous n'a pas de cycle eulerien.

L'algorithme de test et de recherche d'un cycle eulerien est simple à mettre en œuvre :

Algorithme : Cycle Eulerien.

Entrée : Graphe connexe $\mathbf{G} = (\mathbf{X}, \mathbf{E})$ avec $|\mathbf{X}| = n$ et $|\mathbf{E}| = m$.

Sortie : Un cycle $\mathbf{C} = e_{i1}, e_{i2}, \dots, e_{ik}, \dots e_{im}$ parcourant toutes les arêtes de \mathbf{G} .

Soit \mathbf{C} le cycle contenant un sommet quelconque a ;

Tant que il existe une arête $e = uv$ avec $u \in \mathbf{C}$ et $v \in \mathbf{X} \setminus \mathbf{C}$

 Calculer un Chemin \mathbf{CC} partant de $e = uv$;

Si \mathbf{CC} est un cycle

 Insérer \mathbf{CC} dans \mathbf{C} au sommet u ;

Sinon

$\mathbf{C} = \text{cycle_vide}$;

Finsi

Fin tant que

La complexité de cet algorithme est $O(m)$, où m est le nombre d'arêtes de \mathbf{G} . Lorsque le cycle eulerien n'existe pas, on peut demander à minimiser la somme des longueurs des arêtes à retraverser pour balayer toutes les arêtes de \mathbf{G} . Ce problème est plus complexe mais si on met la contrainte de ne traverser les arêtes de \mathbf{G} pas plus de 2 fois, le problème se ramène à la recherche d'un cycle maximal et d'un couplage minimum. L'explication détaillé sort du cadre de cet exposé mais on peut dire que le problème reste polynomial (voir [Xuong, 1992]).

Problème du plus court chemin

Etant donné un graphe $\mathbf{G} = (\mathbf{X}, \mathbf{E})$ et deux sommets quelconques $a, b \in \mathbf{X}$ on cherche le plus court chemin $\mathbf{P}(a, b)$ reliant a à b . Pour mieux répondre à la réalité on peut chercher les chemins $\mathbf{P}_1(a, b), \mathbf{P}_2(a, b), \dots, \mathbf{P}_k(a, b)$, ordonné dans l'ordre de longueurs croissantes. La longueur d'un chemin \mathbf{P} n'est d'autre que la somme des longueurs des arêtes de \mathbf{P} . D'autres variantes de ce problème sont :

Chercher tous les plus courts chemins $\mathbf{P}(a, x)$, x sommet quelconque différent de a .

Chercher les plus courts chemins $\mathbf{P}(x, y)$, entre tous les couples x et y de \mathbf{X} .



Ces variantes peuvent être résolues facilement si on sait trouver le plus court chemin pour un couple de sommets donné. Par conséquent, nous ne traitons que ce dernier problème.

L'algorithme le plus efficace est fourni par Dijkstra [Dijkstra, 1959]. Nous donnons une description relativement détaillée de cet algorithme typique pour donner un aperçu des techniques utilisées en théorie des graphes.

Cet algorithme met en oeuvre une technique d'étiquetage, largement utilisée dans plusieurs algorithmes des graphes. Sans perdre en généralité on cherchera le plus court chemin du sommet 1 à tous les autres $\{2, \dots, n\}$. On notera par $l(i, j)$ la longueur de l'arête $ij \in E$. Définissons par $\mu^*(i)$ comme la longueur minimum des chemins de 1 à i ; en particulier $\mu^*(1) = 0$. L'algorithme procédera en $n - 1$ itérations. Au début de chacune des itérations l'ensemble des sommets est partitionné en deux sous-ensembles, S contenant tous les sommets i d'étiquette $\mu^*(i)$ et $X \setminus S$ contenant les sommets j munis d'une étiquette temporelle $\mu(j) = \min(\mu^*(i) + l(i, j))$, où le minimum est pris sur toutes les arêtes ij avec $i \in S$ et $j \in X \setminus S$. On peut prouver que si $\mu(j) = \min_{k \in X \setminus S} (\mu(k))$ alors $\mu^*(j) = \mu(j)$. Avec ces remarques, l'algorithme de recherche du plus court chemin s'énonce comme suit :

Algorithme : Plus court chemin d'un sommet aux autres.

Entrées : Un graphe G et les longueurs des arêtes.

Sortie : Les longueurs des plus courts chemins entre le sommet 1 et tous les autres.

Initialisation

$S = \{1\}$; $X \setminus S = \{2, \dots, n\}$;
 $\mu(1) = 0$; $\mu(i) = l(1, i)$ si $1i \in E$ et ∞ sinon.

Itération

Tant que $X \setminus S$ est non vide **faire**

 Sélectionner $j \in X \setminus S$ tel que $\mu(j) = \min_{k \in X \setminus S} (\mu(k))$;

$S = S \cup \{j\}$; $X \setminus S = (X \setminus S) \setminus \{j\}$;

Pour tous les voisins k de j dans $X \setminus S$ **faire**

$\mu(k) = \min(\mu(k), \mu(j) + l(j, k))$;

Fin pour tous

Fin tant que

Une analyse de la complexité de cet algorithme montre que le temps maximal requis sera proportionnel avec $O(m \log n)$. Ainsi, cet algorithme et toutes les variantes sont très efficaces en temps de calcul.

Problème du flot maximum

Etant donné un réseau de transport on veut savoir quel est le flot maximum qu'on peut transporter d'un point a à un point b . Il est clair que ce flot dépend de la capacité du réseau. Ainsi, trouver le flot maximum a un lien directe avec le sous réseau de capacité minimum qui est représenté en pratique par des "goulots" qui se forment presque toujours aux mêmes endroits. Rappelons qu'en général, l'information que l'on a concerne les capacités des tronçons unitaires. A partir de cette information on cherche à trouver le flot maximal en même temps que les coupes de capacité minimum par rapport au réseau tout entier.

En terme de graphes les problèmes énoncés ci-dessus peuvent se formuler comme suit :

Soit un graphe connexe orienté $G = (V, A)$ tel que pour tout arc $ij \in A$ (l'arc commence au sommet i et finit au sommet j) on connaît sa capacité $c(ij) > 0$.



Un *flot* dans \mathbf{G} est une affectation de nombres $f(ij)$ à chaque arc ij tels que

- *le flot dans l'arc ij ne peut pas dépasser la capacité*
$$0 \leq f(ij) \leq c(ij)$$
- *loi de Kirchoff, le flot entrant dans un sommet est égal au flot sortant*
$$\sum_{j \in \text{prec}(i)} f(ji) = \sum_{k \in \text{succ}(i)} f(ik)$$

Une *coupe* dans \mathbf{G} est un ensemble d'arcs $\mathbf{K} = \{i_1j_1, \dots, i_kj_k\}$ tel que le graphe devient non connexe si on enlève les arcs de \mathbf{K} . Une coupe de capacité minimum est une coupe \mathbf{K} telle que :

$$c(\mathbf{K}) = \sum_{ij \in \mathbf{K}} c(ij) \quad \text{soit minimale.}$$

Un des résultats principaux de la théorie des flots dans les réseaux montre la dualité (dans le sens défini en programmation linéaire) entre le problème de recherche de *flot maximal* et la recherche de la *coupe minimum*. Ce résultat est la source d'une multitude d'utilisations des algorithmes de flot et coupes tant dans les applications économiques et industrielles que dans la déduction de résultats en mathématiques combinatoires.

La présentation de l'algorithme de recherche de flot maximum et de coupe minimum est relativement compliquée. La première version de cet algorithme est apparue en 1956 dans [Ford *et al.*, 1956]. Pendant au moins 30 ans de recherches intensives, montrant l'intérêt du problème, des améliorations significatives ont été apportées tant théoriques que algorithmiques [Hu, 1982]. Grâce aux travaux fondamentaux de Edmonds et Karp [Edmonds *et al.*, 1972] on connaît actuellement un algorithme de complexité polynomiale $O(m n^2)$ (m - nombre d'arcs et n - nombre de sommets de \mathbf{G}).

Parmi les applications économiques et industrielles des flots et coupes, citons : l'extension d'un réseau de transport au moindre coût, le problème de placement d'entrepôts, l'ordonnancement de tâches, la conception de réseaux urbains, ferroviaires, aériens, etc.

1.3. Les choix aléatoires

1.3.1. Caractère des problèmes aléatoires

Dans le domaine des choix aléatoires, il n'est plus possible d'évaluer avec certitude les éléments nécessaires au calcul des conséquences des décisions possibles. Ces éléments oscillent autour d'une valeur moyenne. C'est par le jeu des répétitions statistiques qu'on peut étudier la façon dont se répartissent ces fluctuations.

La difficulté tient au fait que les conséquences des décisions ne s'expriment plus sous la forme d'un *nombre*, mais d'une *distribution de fréquences*. On doit alors comparer différentes distributions pour choisir la meilleure décision. Ces distributions peuvent heureusement se résumer par leurs paramètres ; en général, la moyenne et la variance suffisent. Ces renseignements seraient de bien piètre utilité si les mathématiciens n'avaient pas donné aux praticiens l'outil du calcul des probabilités. Mais un pont reste à franchir entre l'observation des répétitions d'un phénomène, qui est un souci de la statistique descriptive, et la théorie des probabilités, riche en lois, en formules, en tables et en abaques susceptibles de fournir les solutions à de nombreux problèmes, à la seule condition qu'on se soit assuré au préalable qu'il n'est pas déraisonnable de supposer que les répétitions observées sont bien le reflet de telle loi de la théorie des probabilités : c'est l'objet de la statistique mathématique et des tests



d'hypothèses. Rationnel et raisonnable à la fois, l'analyste est alors à même de donner une solution acceptable aux problèmes aléatoires de l'entreprise. De nombreux problèmes de gestion se classent dans cette catégorie ; le hasard intervenant dans le contexte d'un choix, il s'agit en général de trouver le meilleur compromis entre deux phénomènes antagonistes.

1.3.2. Les phénomènes d'attente

Les goulets d'étranglement proviennent de l'impossibilité d'exiger la rentabilité d'un service individuel. Il faut alors se résigner à partager ce service avec d'autres clients et s'armer bien souvent de patience dans les files d'attente. Dans une file d'attente, les arrivées des clients à des stations où ils viennent demander un service s'opèrent d'une façon aléatoire ; la durée du service est elle-même en partie gouvernée par le hasard. On cherche à établir un compromis entre le temps perdu par les clients du fait de l'attente et la multiplication des stations de service qui ne fonctionnent pas gratuitement.

L'analyste, observant d'une part la répétition des entrées et d'autre part celle des services, en déduit le nombre moyen d'entrées ainsi que le nombre moyen de services par unité de temps. Sachant que bien souvent un phénomène tel que celui des entrées de la clientèle est régi par une loi de Poisson et celui des services par une loi exponentielle, il s'assure de la vraisemblance de ces hypothèses en soumettant le résultat de ses observations à un test statistique (test de χ^2 ou de Pearson). Si le résultat de ce test n'est pas défavorable, un modèle de file d'attente est rigoureusement établi à l'aide du calcul des probabilités, et l'analyste n'a plus qu'à en utiliser les résultats.

Il peut, dans le cas d'ouvriers et de magasiniers par exemple, minimiser le coût supporté du fait de l'attente des clients dans la file et de l'inoccupation de certaines stations pour un meilleur profit de l'entreprise.

1.3.3. Entretien préventif et renouvellement des équipements

La fiabilité, qui est l'étude de la résistance des matériels aux pannes, est une partie importante de la recherche opérationnelle. L'entretien préventif (c'est-à-dire, par exemple, le changement d'une pièce avant que cette pièce soit hors d'usage) permet de réduire les conséquences d'une panne pour des matériels d'usure. L'usure, présentant un caractère aléatoire, oblige à prévoir le renouvellement des équipements si on veut les maintenir en service permanent. On recherche dans ce cas le meilleur taux de réapprovisionnement. Ce taux dépend de la loi d'usure qu'on s'efforcera de définir statistiquement par l'observation des équipements gérés. Un taux de service trop élevé limite les risques d'arrêt intempestif, mais il augmente les frais d'entretien et de stockage. Or, ce taux n'est pas constant, ce qui est une nouvelle cause de souci pour le gestionnaire qui cherchera, par un entretien préventif, le meilleur compromis entre le sacrifice qu'entraîne le remplacement d'un matériel incomplètement usé et les conséquences d'une panne.

Pour un équipement comportant une pièce fragile, la fonction de survie de la pièce donne la probabilité de son fonctionnement à l'époque t . Si l'on attend la défaillance pour procéder au remplacement de la pièce usée, on subit avec certitude toutes les conséquences d'une mise hors de service de l'équipement. Si, au contraire, on décide de remplacer systématiquement la pièce fragile à l'occasion des heures d'entretien de l'équipement dès que la pièce a atteint un âge j , âge qui est justement à déterminer, on ne subira le coût de la panne qu'avec une probabilité réduite, la panne n'étant pas certaine. Évaluant les coûts moyens des deux décisions possibles, on calcule en fonction de j et à l'aide de la fonction de survie le minimum du coût relatif à l'entretien préventif. Si ce coût est inférieur au coût de la première décision,



la valeur de j obtenue donne l'âge à partir duquel on a tout intérêt à effectuer le remplacement.

1.3.4. Gestion scientifique des stocks

Dans les problèmes de stocks, les demandes de la clientèle ne pouvant être exactement connues ni en date, ni en quantité, et le réapprovisionnement n'étant pas instantané, on s'efforce, devant ces phénomènes aléatoires, de se prémunir contre deux risques : celui d'une *rupture de stock*, cause d'un manque à la vente pénalisé par un coût de pénurie qu'entraîne obligatoirement une commande trop faible ; et celui d'un *surstock*, cause d'une immobilisation d'un capital et d'un coût de stockage sans utilité, conséquence d'une commande trop élevée. L'observation du phénomène permet d'estimer au mieux la demande moyenne de la clientèle (détermination d'un modèle par l'étude des séries chronologiques) ainsi que son aléa, c'est-à-dire l'ampleur des fluctuations (étude statistique de la distribution des fréquences de la demande). On détermine ensuite, en limitant les risques selon les possibilités ou le désir du gestionnaire (fixation du taux de service, c'est-à-dire du pourcentage des demandes qui seront en moyenne honorées), la commande destinée à satisfaire non seulement la demande moyenne estimée, mais encore une partie du supplément éventuel dû aux fluctuations et susceptible de se manifester (stock de sécurité), ainsi que les règles de déclenchement des commandes successives.

1.4. Stratégies en situation de concurrence, théorie des jeux

1.4.1. La caractéristique des problèmes de concurrence

Dans l'univers de la concurrence, l'aléatoire et le combinatoire se rejoignent ; de plus, les éléments nécessaires au calcul des conséquences des décisions possibles ne peuvent plus être évalués ni par un nombre certain, ni par une distribution statistique.

Il faut savoir abandonner la prétention d'une détermination strictement objective des facteurs entrant dans l'évaluation du choix. La théorie des jeux est, dans ce domaine, d'un grand secours. On peut sommairement classer ces problèmes en deux catégories : les jeux contre la nature et les jeux à deux personnes. L'espérance mathématique y est largement mise à contribution.

1.4.2. Les jeux contre la nature

Imaginons un compétiteur se trouvant en présence de *situations naturelles* et hésitant entre plusieurs décisions à prendre. Supposons, par exemple, trois décisions envisagées, d_1 , d_2 et d_3 , contre trois éventualités possibles, e_1 , e_2 et e_3 . Pour guider son choix, celui qui décide doit s'efforcer d'apprécier le gain, ou éventuellement la perte, r_{ij} , que lui vaut la simultanéité de la décision d_i devant l'éventualité e_j ; il ne peut pour cela se fier qu'à son jugement fondé sur la connaissance des faits ; il détermine ainsi la matrice des utilités $M = (r_{ij})$. Il lui faut ensuite définir un critère de décision, et cela va dépendre de ses possibilités du moment ; il peut être optimiste, et il accepte le risque, ou prudent, car il a peut-être des raisons de se méfier de la nature : c'est le critère de Wald et von Neumann, lequel, supposant que la nature nous oppose toujours l'éventualité qui nous est la plus défavorable, incite le joueur à la prudence. Dans le premier cas, le joueur prendra la décision correspondant au maximum des maximums de ses décisions et, dans le second cas, celle relative au maximum des minimums ; il limite alors les risques de perte.



Selon Laplace, on peut supposer que, pour des phénomènes naturels sur lesquels on a peu de renseignements, sinon aucune information, il est raisonnable de considérer les éventualités comme équiprobables et de retenir la décision à laquelle est attachée l'espérance mathématique maximale. Par exemple, si la matrice des utilités est:

$$\begin{array}{c} e_1 \quad e_2 \quad e_3 \\ d_1 \left(\begin{array}{ccc} 2 & 0 & 5 \\ 2 & 1 & 1 \\ 3 & 4 & 3 \end{array} \right) \\ d_2 \\ d_3 \end{array}$$

les espérances mathématiques sont respectivement :

$$E(d_1) = -2.1/3 + 0.1/3 + 5.1/3 = 1$$

$$E(d_2) = 2.1/3 + 1.1/3 + (-1).1/3 = 2/3$$

$$E(d_3) = 3.1/3 + 4.1/3 + (-3).1/3 = 4/3,$$

l'espérance mathématique la plus forte correspond à la décision d_3 .

Dans certains cas, une appréciation du phénomène étant déjà un peu moins floue, on peut se croire autorisé à attribuer aux éventualités une gamme de *coefficients de vraisemblance*, c'est-à-dire une distribution subjective de probabilités : p_1 , p_2 et p_3 , par exemple, pour e_1 , e_2 et e_3 . On retiendra à nouveau la décision à laquelle correspond l'espérance mathématique la plus élevée.

D'autres critères de décision sont utilisables, celui de Savage en particulier, qui s'efforce de tenir compte du regret attaché à l'abandon volontaire de décisions, certes susceptibles d'entraîner des pertes, mais aussi des possibilités de gain important. Ces différents critères, à fondement souvent subjectif, peuvent conduire à des choix totalement différents ; c'est ce qui distingue ce dernier volet de la recherche opérationnelle des précédents dans lesquels la décision s'impose en toute objectivité.

1.4.3. Les jeux à deux personnes

Supposons maintenant que l'incertitude ne provienne plus de la nature des choses, mais de l'existence d'une volonté antagoniste.

Remplaçons la nature par un concurrent N disposant librement des décisions e_1 , e_2 et e_3 de telle sorte que les résultats qu'il peut escompter sont justement les valeurs opposées des gains du premier joueur (jeu à somme nulle). Par hypothèse, nous supposerons les joueurs prudents (utilisation du critère de Wald, ce qui est bien normal quand on a à se défendre contre un adversaire) et intelligents ; cela signifie que le raisonnement d'un des joueurs peut aussi être tenu par l'autre. Pour un jeu en un coup, la tactique à adopter par chaque joueur est la suivante : pour D, jouer le *maximin* (maximum des minimums) et pour N, jouer le *minimax* (minimum des maximums), puisqu'il se trouve dans la situation opposée à celle de D. Par exemple, si l'on reprend pour matrice des gains de D la matrice :

$$\begin{array}{c} N \\ e_1 \quad e_2 \quad e_3 \\ D \quad d_1 \left(\begin{array}{ccc} 2 & 0 & 5 \\ 2 & 1 & 1 \\ 3 & 4 & 3 \end{array} \right) \\ d_2 \\ d_3 \end{array}$$



pour D jouer le maximin, ici -1, correspond à la décision d_2 et pour N jouer le minimax, ici 3, correspond à la décision e_1 .

Ces choix sont peut-être décevants, mais ils ne sont guère susceptibles d'amélioration. Par exemple, N, qui est intelligent et qui sait donc que D va choisir d_2 peut être tenté de prendre e_3 à la place de e_1 pour espérer gagner 1 au lieu de perdre 3 ; il est, certes, intelligent, mais aussi imprudent ; car D, procédant de la même finesse d'esprit, peut alors prendre la décision d_1 à la place de la décision d_2 et ainsi s'assurer un gain de 5 au lieu d'une perte de 1. Il se peut que le minimax et le maximin coïncident, auquel cas, assurément, les joueurs n'ont aucun intérêt à s'écartez de cette valeur, de ce point d'équilibre du jeu.

Émile Borel puis John von Neumann et Oskar Morgenstern se sont efforcés de généraliser le jeu précédent par la recherche d'une *stratégie mixte* qui consiste à calculer la fréquence avec laquelle on doit opposer chaque décision, ou *stratégie pure*, à l'adversaire dans un jeu à plusieurs coups. Ce problème conduit à la résolution d'un programme linéaire.

C'est par des considérations de ce genre qu'un chef d'entreprise s'installant sur un marché peut étudier, face à la concurrence, les résultats des variations des prix de vente d'un produit et définir une politique d'action.

Pour préciser ces techniques dans ce contexte d'application, considérons la situation suivante : face à un devenir indéterminé, le décideur, lorsqu'il dispose d'une marge de manœuvre peut envisager plusieurs solutions {a, b, c}. Par exemple, le chef d'entreprise qui a des difficultés financières peut décider de réduire ses effectifs (a), de lancer une campagne de promotion (b) ou de se diversifier dans de nouveaux produits (c)... La méthode des arbres de décision permet d'envisager les divers scénarios possibles, de leur affecter un budget, une probabilité de réussite, de simuler les effets attendus. Parmi ces scénarios, certains sont coûteux, mais peuvent rapporter beaucoup, d'autres sont moins coûteux mais risquent d'entraîner de maigres bénéfices. Laquelle des options choisir ? Pour trancher entre les diverses formules, il faut un critère de choix qui corresponde à une attitude face au risque.

Le critère de Laplace consiste à adopter, parmi les choix possibles, celui qui rapporte un résultat moyen. Il est calculé à partir de la moyenne arithmétique entre des alternatives jugées également probables.

Le critère de Bernoulli, dit aussi critère de l'espérance mathématique, consiste à affecter à chacune des solutions a, b, c, une plus ou moins grande probabilité de réalisation. Cette probabilité est définie subjectivement, puis traduite en termes mathématiques. Cette espérance mathématique permet donc de pondérer les résultats escomptés.

Le meilleur choix est celui qui obtient le meilleur rapport résultats/chances de réussite.

Le critère de Wald ou critère Minimax est un critère pessimiste, qui ne consiste à envisager pour chaque alternative a, b et c que leurs plus mauvaises conditions de réalisation. On choisit alors parmi eux, celui qui coûterait le moins...

Le critère de Savage ou la matrice des regrets consiste à construire une matrice où l'on affecte à chaque alternative (représentée par une case), un regret, c'est-à-dire le montant du manque à gagner si on avait connu d'avantage les résultats. On applique alors au regret le critère Minimax. Pour ce qui concerne ces approches, nous proposons la lecture de l'annexe E de la thèse de Fabiani que nous joignons intégralement en annexe [Fabiani, 1996 (disponible sur internet)].



Illustration

Une personne joue à pile ou face. Elle dispose de deux options de réponse {Pile, Face}. Deux états de la nature peuvent advenir : Pile ou Face. La combinaison des options et des états définit les issues possibles : Pile-Pile, Pile-Face, Face-Pile, Face-Face.

La théorie des probabilités dit que la probabilité objective de chaque état est ici de 0.5. Posons que des gains ou des pertes sont attachés à chaque issue : gain de 120F si Pile-Pile, gain de 200F si Face-Face, perte de 10F si Pile-Face, perte de 100F si Face-Pile.

Sous un principe de maximisation des gains, l'option de réponse doit être choisie en fonction de la valeur espérée VE associée à chaque option (soit la somme des valeurs des issues correspondantes, pondérées par la probabilité des états correspondants).

Ainsi, dans notre exemple, la personne a intérêt à parier Pile puisque :

$$VE(Pile) = .5.(120) + .5(-10) = 55F$$

$$VE(Face) = .5.(200) + .5.(-100) = 50F$$

Par rapport à ce modèle, dit de la valeur espérée, un autre modèle, dit de l'utilité subjective espérée (USE) introduit les concepts de probabilité subjective et d'utilité.

En fonction d'informations dont elle dispose à un moment donné la personne peut considérer que son incertitude est réduite : autrement dit, à la suite de l'occurrence de plusieurs Pile successifs, la probabilité subjective de Face augmente. Posons qu'à un moment du jeu, la personne considère que la probabilité d'occurrence de Face est .8.

En fonction de la valeur subjective (utilité) qu'elle attache aux valeurs monétaires, la personne peut considérer qu'il est 2 fois plus intéressant de gagner 200F que 100F et qu'il est indifférent de perdre 10F.

La personne considérera alors qu'elle a intérêt à parier Face puisque :

$$USE(Pile) = .2.(100) + .8.(0) = .2$$

$$USE(Face) = .8.(200) + .2.(-100) = 140$$

1.4.4. Responsable et analyste dans la prise de décision

Si la recherche scientifique débouche sur une notion de loi, la recherche opérationnelle, abordant des phénomènes très complexes dans lesquels il faut étudier l'interaction de nombreuses grandeurs, a recours à un modèle pour simplifier et schématiser la réalité. L'approche expérimentale est difficile; il n'est pas possible de faire des expériences en vraie grandeur, et l'homme intervient par sa décision. L'analyste dispose, pour ces modèles de situations complexes, des *méthodes de simulation* lui permettant de juger de l'opportunité ou de la validité de ses hypothèses quand un modèle analytique ne peut pas être élaboré. Ces méthodes «consistent essentiellement à réaliser, par des moyens artificiels, des expériences sur les phénomènes économiques». Grâce à elles, le responsable peut à son tour mettre en évidence, à partir de données comptables et techniques, les résultats de politiques nouvelles.

Les cadres de l'entreprise peuvent aussi être initiés ou confirmés dans de nouvelles fonctions ou retrouver plus facilement leur place dans une nouvelle organisation.

Un problème n'ayant de sens en recherche opérationnelle que si l'on connaît la fonction à optimiser, le choix du critère dans l'évaluation d'une décision est primordial. La définition de cette fonction incombe par conséquent au responsable.

L'analyste peut toujours proposer, à celui qui fait appel à son concours, une éventuelle paramétrisation mettant en évidence les conséquences de certaines solutions possibles. Il doit insister sur le danger que peut présenter l'application de la recherche opérationnelle à des aspects fragmentaires de l'entreprise qui risque, si l'on néglige les interactions d'autres compartiments, de perturber l'équilibre de l'ensemble, mais il n'a pas à se soucier de la conduite du responsable qui, maître de ses décisions, doit savoir en accepter les conséquences.



1.5. Les méthodes multicritères

1.5.1. Justification des méthodes multicritères

Un premier cas typique d'application des méthodes multicritères est le choix du tracé de grandes voies de circulation. Une autoroute qui, de par la planification de sa trajectoire, n'exproprie quasiment personne va rencontrer l'adhésion de beaucoup d'utilisateurs.

Cependant, elle mécontentera certainement les conservateurs de la nature, si elle engendre la disparition d'une zone de protection spéciale. De la même façon, un tracé plus ou moins rectiligne procurera davantage de sécurité et permettra de gagner du temps, par rapport à un tracé plus sinueux mais qui ne nécessite aucun ouvrage d'art important pour éviter les obstacles naturels. On constate donc déjà que le choix du tracé implique de nombreux critères qui ne sont pas toujours aisément traduisibles en unité monétaire et qui peuvent parfois être contradictoires, il faudra donc trouver un compromis.

Cet exemple donne, déjà dans son énoncé, une idée de la difficulté qui peut surgir si l'on tente d'utiliser les méthodes classiques de *recherche opérationnelle*. En effet, une simple *analyse « coûts-bénéfices »* montre vite ses limites et son inadéquation à traiter les problèmes complexes auxquels se confronte un décideur : comment ramener à une unité commune, de surcroît monétaire, des impacts aussi hétéroclites que le bruit, la pollution des eaux, et la dégradation du paysage [Maystre *et al.*, 1994] ?

Même si toutes ces notions peuvent être quantifiées par un moyen ou un autre, elles impliquent une subjectivité qui ne permet pas de traduire les unités de l'une dans les unités de l'autre. En tout cas, cela semble impossible sans une certaine perte d'information due à une quantité d'estimations qui est parfois invraisemblable.

Avant l'apparition des méthodes multicritères, les problèmes de décision se ramenaient le plus souvent à l'optimisation d'une fonction économique, constituant l'unique critère de sélection. Cette approche monocritère avait le mérite de déboucher sur des problèmes mathématiques bien posés mais qui n'étaient pas toujours représentatifs de la réalité car :

- la comparaison de plusieurs actions possibles se fait rarement suivant un seul critère ;
- les préférences sur un critère sont, dans bien des cas, difficilement modélisables par une fonction ; et lorsqu'il y a plusieurs objectifs, il est impossible de les atteindre tous à la fois.

Ainsi, on peut dire que le domaine de réussite de la recherche opérationnelle est constitué de tous les problèmes qu'il est possible d'isoler du processus de gestion du système (comme, par exemple, le choix du mélange optimal de rations alimentaires destinées au bétail). Par complément, le domaine d'échec de la recherche opérationnelle comprend toutes les décisions de gestion qu'on ne peut isoler de leur contexte (comme, par exemple, le tracé d'une autoroute). Ainsi, la recherche opérationnelle n'a pas donné complète satisfaction car on lui avait fixé un objectif (trop) ambitieux : désigner, en toutes circonstances, la meilleure décision (un optimum), même quand cette notion pouvait être vide de sens [Schärlig, 1985]. En effet, choisir d'optimiser, c'est implicitement se situer dans une approche à critère unique. Cela se démontre aisément par l'absurde : dès que l'on prend plusieurs points de vue pour juger des conséquences de plusieurs actions, on risque de désigner comme optimale une action différente pour chaque point de vue et, en fin de compte, de ne dégager aucun optimum des calculs. Or, toute la réalité humaine est « à points de vue multiples » ou encore multicritère.



Reprenons notre exemple: si, pour choisir un tracé d'autoroute, le ministre de l'équipement et des transports ne considérait que l'aspect financier, il limiterait au strict minimum les travaux (terrassements, viaducs, tunnels, etc.) afférents à ce genre de projet.

C'est évidemment loin d'être le cas. Et cela ne concerne qu'un individu à la fois. Mais tout le monde sait que, dans un service public ou dans la gestion d'un Etat, les intervenants sont multiples. Multiplicité des critères, multiplicité des intervenants : les deux phénomènes ne se superposent pas uniquement, ils se multiplient. Cela devrait déjà suffire pour envisager de nouvelles méthodes par rapport à l'optimisation.

On peut encore ajouter deux autres faits : le côté non commensurable de certains critères et le fait qu'ils puissent être contradictoires. Reprenons l'exemple de l'autoroute. L'ensemble des futurs usagers désire que l'autoroute soit sûre, assure la fluidité du trafic, permette un gain de temps considérable, etc. Ces différents aspects ne sont pas (ou très difficilement) traduisibles en coûts car ils n'ont pas de cours ni dans l'économie ni dans la tête de l'individu.

Il faut donc utiliser des méthodes qui sachent tenir compte de plusieurs critères sans les réduire à un seul.

Si on rassemble tous les critères énumérés dans le choix d'un tracé d'autoroute, à savoir le coût des travaux, la sécurité, la fluidité de trafic, le gain de temps, l'accroissement de l'économie régionale et qu'on y ajoute les nuisances pour les éventuels riverains, les expropriations nécessaires ou encore les forêts détruites, il n'est pas difficile de constater que toutes ces notions sont assez contradictoires.

1.5.2. Concepts fondamentaux

Même si, en dernier ressort, la responsabilité d'une décision incombe à un individu clairement identifié, celle-ci est souvent la résultante d'interactions entre de multiples *acteurs* au cours du processus de décision. Parmi les acteurs, on appelle *intervenants* ceux qui, de par leur intervention, conditionnent directement la décision en fonction du système de valeurs dont ils sont porteurs. A leur côté, figurent tous ceux qui, de façon normalement passive, subissent les conséquences de la décision, laquelle est seulement censée tenir compte de leurs préférences.

Ainsi, l'aide à la décision est l'activité de celui qui, se basant sur certains modèles, aide à obtenir des éléments de réponses aux questions posées par un intervenant dans un processus de décision. Ces éléments doivent permettre d'accroître la cohérence entre l'évolution du processus, d'une part, et les objectifs et le système de valeurs de l'intervenant, d'autre part. Il est important à ce stade, de noter que l'aide à la décision ne relève donc que très partiellement de la recherche d'une vérité [Roy et al., 1993].

Dans cette section, nous présentons la terminologie et les conditions minimales nécessaires pour définir un problème d'aide à la décision multicritère que nous présentons sous la forme suivante :

S : ensemble de solutions ou alternatives (projets, actions, options...) parmi lesquelles le décideur doit choisir.

C : ensemble des critères permettant d'évaluer les solutions. $X=[0,1]$: l'ensemble des degrés de satisfaction (ou évaluations partielles).

Nous considérons l'ensemble fini des n critères $C = \{c_1, c_2, \dots, c_n\}$, $P(C)$ étant l'ensemble des parties de C , $x = (x_1, x_2, \dots, x_n)$ représente les valeurs numériques de satisfaction des critères, de sorte que, x_i est le degré de satisfaction du critère c_i . On considère aussi l'ensemble des alternatives $S = \{s^1, s^2, \dots, s^p\}$. A chaque solution $s^k \in S$ est associé un profil $s^k = (x_1^k, x_2^k, \dots, x_n^k)$



avec x_i^k représente l'évaluation partielle de s^k selon le critère c_i . Pour des raisons de simplicité , les valeurs des scores sont exprimées dans l'intervalle [0,1], 1 correspond à la satisfaction complète d'un critère, 0 exprime la non satisfaction complète. $|A|$ désigne le cardinal d'un ensemble A, A/B désigne la différence au sens des ensembles et \wedge, \vee , le minimum et le maximum respectivement.

1.5.3. Approches d'agrégation multicritère

Si les méthodes d'agrégation sont nombreuses, c'est parce qu'il est dans la nature des choses qu'aucune méthode ne respecte la totalité des exigences qu'un utilisateur pourrait trouver «normales» dans l'idée du multicritère [Schärlig, 1985]. Il faut donc décider sur quelle exigence on va céder. Face à ce choix, on peut distinguer deux approches opérationnelles [Roy *et al.*, 1993]. La première attitude serait d'inclure tous les jugements dans ce qu'on appellera en mathématique une fonction d'utilité ou d'agrégation [Roy, 1985], en leur attribuant d'éventuels poids. Cela suppose que tous les jugements sont commensurables. Il est à noter que cette technique suppose en outre que les jugements sont transitifs, d'où l'appellation « *agrégation complète transitive* ». Roy [Roy, 1985] ainsi que Maystre *et al.* [Maystre *et al.*, 1994] l'appellent encore « *approche du critère unique de synthèse évacuant toute incomparabilité* ». Ce critère unique risque évidemment d'effacer toutes les nuances. Toutefois, les méthodes d'agrégation complète peuvent s'avérer intéressantes ou tout simplement les seules utilisables [Schärlig, 1985]. Une seconde attitude est de respecter l'incomparabilité et l'intransitivité au prix de la clarté des résultats ! En effet, cette technique ne permet pas d'obtenir un résultat indiscutable comme on s'y attend en général, pour tout ce qui est à base de mathématique [Schärlig, 1985]. Ici, on cède donc sur la clarté. Comme on se contente dans ce cas d'appréhender partiellement les conséquences des divers jugements, cette attitude est dite *d'agrégation partielle*, ou encore l'appellent encore « *approche du surclassement de synthèse acceptant l'incomparabilité* » ou « *méthodes de surclassement* » [Roy *et al.*, 1993 ; Maystre *et al.*, 1994 ; Vincke, 1989].

Dans cette approche, la technique consiste à comparer les alternatives deux à deux et à vérifier si, selon certaines conditions préétablies, l'une des deux actions surclasse l'autre et à quel degré. A partir de toutes ces comparaisons, on tente ensuite de réaliser une synthèse.

Approche cardinale

L'un des problèmes d'aide à la décision multicritère est d'aboutir à un classement des solutions de la meilleure à la moins bonne, au vu de l'ensemble critères, ou tout simplement à trouver la meilleure solution. Dans l'approche cardinale, l'objectif est de construire une fonction : $H : [0,1]^n \rightarrow [0,1]$ telle que, pour chaque solution s^k : on a $H(s^k) = H(x_1^k, \dots, x_n^k)$, $H(s^k)$ représente l'évaluation globale de la solution s^k relativement à tous les critères, H est l'opérateur d'agrégation à déterminer.

Les conditions nécessaires sur l'opérateur h sont les suivantes :

- h est continu ;
- $h(0,0,\dots,0) = 0$ et $h(1,1,\dots,1) = 1$;
- $\forall (u_i, v_i) \in [0,1]^2$, si $u_i \geq v_i$ alors $h(u_1, \dots, u_p) \geq h(v_1, \dots, v_p)$.

Pour une opération exprimant la satisfaction simultanée des objectifs un axiome naturel est :



$\forall (u_1, \dots, u_p), h(u_1, \dots, u_p) \leq \min(u_1, \dots, u_p)$, c'est-à-dire que l'évaluation globale d'une action ne peut être meilleure que la plus mauvaise des évaluations partielles. On appellera ces opérateurs des conjonctions. Les principales conjonctions associatives sont :

$$\min(u_1, \dots, u_p), \prod_{i=1}^p u_i \text{ et } \max\left(0, \sum_{i=1}^p u_i - p + 1\right).$$

Un opérateur exprimant la redondance de deux objectifs satisfera un axiome dual du précédent :

$\forall (u_1, \dots, u_p), \max(u_1, \dots, u_p) \leq h(u_1, \dots, u_p)$ c'est-à-dire que c'est la meilleure des évaluations partielles qui conditionne l'évaluation globale. Ce sont les disjonctions dont une sous-classe importante sont les disjonctions associatives. Les plus utilisées sont :

$$\max(u_1, \dots, u_p), 1 - \prod_{i=1}^p (1 - u_i), \min\left(1, \sum_{i=1}^p u_i\right).$$

Une opération exprimant le compromis satisfera l'axiome suivant, complémentaire des deux précédents : $\forall (u_1, \dots, u_p), \min(u_1, \dots, u_p) \leq h(u_1, \dots, u_p) \leq \max(u_1, \dots, u_p)$.

L'identification de l'opérateur h peut être décomposée en deux étapes. On peut cerner le comportement du décideur en lui demandant de fournir une appréciation linguistique du niveau de compatibilité entre conséquence et objectif sur un ensemble réduit d'actions. Il est alors possible de se fixer rapidement sur le type d'agrégation à employer : conjonction, compromis ou disjonction. Une fois déterminée l'une des trois attitudes fondamentales possibles du décideur, on peut utiliser des familles de fonctions paramétrées pour affiner la détermination de l'opérateur d'agrégation : le choix de h est ramené à un problème d'identification paramétrique.

Les opérateurs conjonctifs sont couverts, par exemple, par l'opération de Yager qui définit une famille paramétrée d'opérateurs :

$$Y_q(u_i) = 1 - \min\left(1, \left(\sum_{i=1}^p (1 - u_i)^q\right)^{1/q}\right) \text{ pour } q \geq 0$$

Par exemple : $\lim_{q \rightarrow \infty} Y_q(u_i) = \min(u_i)$ et $Y_1(u_i) = \max\left(0, \sum_{i=1}^p u_i - p + 1\right)$.

La famille des conormes associées permet de couvrir l'ensemble des disjonctions :

$$CY_q(u_i) = \min\left(1, \left(\sum_{i=1}^p u_i^q\right)^{1/q}\right) \text{ pour } q > 0$$

$$\text{Les compromis sont couverts par : } Ym_q(u_i) = \left(\frac{\sum_{i=1}^p u_i^q}{p}\right)^{1/q}$$



Des valeurs remarquables de q donnent la moyenne arithmétique $Y_{m-1}(u_i)$ et la moyenne harmonique $Y_{m-1}(u_i)$; lorsque $q \rightarrow 0$, le compromis est la moyenne géométrique.

Les caractéristiques symétriques de h n'impliquent pas que l'agrégation soit symétrique. En effet, l'agrégation repose sur les ensembles flous décrivant les objectifs partiels, et des dissymétries peuvent facilement être introduites en manipulant les fonctions d'appartenance. Les fonctions d'agrégation ci-dessus peuvent être également généralisées avec des coefficients de pondération ou poids pour introduire les dissymétries de façon naturelle [Dubois, 1983 ; Dubois et al., 1985].

Des techniques d'optimisation à partir de ces trois familles d'opérateurs permettent de construire des outils d'identification de l'opérateur h lorsque l'on dispose d'une base d'échantillons de tests suffisamment riche (en nombre et en diversité des appréciations partielles). L'identification très précise d'une telle opération est parfois impossible tant l'incertitude sur le paramètre q est grande. Les résultats d'identification doivent alors être accompagnés d'une sensibilité et d'un pouvoir de discrimination afin que l'on soit à même d'accorder un degré de validité au résultat lors d'une utilisation en simulation (recherche d'extrêma « très plats »). Avoir recours à ces tests de sensibilité n'est pas sans implication sur l'interprétation sémantique des coefficients des fonctions d'agrégation. Cela n'invalide pas l'approche pour autant, en effet :

- une identification très précise peut être jugée inutile ; seul un certain comportement de l'opérateur h peut être souhaité. Cela est à rapprocher du caractère arbitraire des valeurs d'appartenance, et de la traduction numérique des jugements émis par le décideur. Il vaut mieux dans ce cas là, se donner un catalogue restreint mais bien dispersé d'opérations simples et identifier h de façon *qualitative* en demandant au décideur de formuler à l'aide d'une échelle linguistique de jugement (totalement compatible, plutôt compatible,..., incompatible) un jugement global sur chaque action.
- si on trouve que plusieurs opérations sont des modèles possibles, on pourra malgré tout tenter de classer les éléments. On pourra introduire une incertitude sur le résultat de l'agrégation. Par exemple si on n'a pas pu (ou voulu) discriminer entre les opérations h_1, h_2, \dots, h_t on peut poser :

$$\forall (u_1, \dots, u_p), h(u_1, \dots, u_p) \in \left[\min_{i=1,t} h_i(u_1, \dots, u_p), \max_{i=1,t} h_i(u_1, \dots, u_p) \right]$$

La notion d'importance relative des critères (le sens accordé à la notion d'importance restant par ailleurs *vague* dans la littérature) conduit naturellement à introduire des distributions de poids sur les critères dans le cadre de la théorie précédente en généralisant si possible les opérateurs définis par :

$$\mu_D(s) = \sum_{i=1}^P p_i [\mu_i(s)]$$

La notion de poids peut être transposée sur de nombreux opérateurs d'agrégation comme évoqués ci-dessus (en particulier normes et conormes) [Dubois et al., 1984] dès lors que l'opération présente une structure additive sous-jacente. Ainsi, lorsque h est de la forme $\Psi^{-1}(\Psi(.) + \Psi(.))$, une extension de l'équation précédente peut être proposée :



$$\bar{h}(u_1, \dots, u_n) = \Psi^{-1}\left(p \sum_{i=1}^p p_i \Psi(u_i)\right) \text{ où } \sum_{i=1}^p p_i = 1$$

A titre d'exemple, pour le produit, en posant $\Psi = \log$, on obtient : $\bar{h}(u_1, \dots, u_p) = \prod_{i=1}^p [\mu_i(s)]^{p_i}$

AHP (Analytic Hierarchy Process)

La résolution de problèmes décisionnels selon la méthode AHP (Analytic Hierarchy Process) de Saaty [Saaty, 1980] est sans doute l'AMC (Analyse MultiCritère) la plus répandue justement parce qu'elle accompagne le décideur dans la méthodologie de formulation de son problème et en particulier parce qu'elle propose une Méthode d'Evaluation des Paramètres d'Importance (MEPI). Le processus d'analyse hiérarchique repose sur la construction d'une hiérarchie qui s'opère selon une démarche descendante. La structuration de priorités consiste à définir des priorités entre critères en comparant paire par paire tous les couples de critères de même niveau hiérarchique. L'IRC ainsi obtenue est agrégée selon une démarche ascendante pour aboutir à un critère unique de synthèse en racine de l'arborescence. Ces principes garantissent une cohérence logique dans l'homogénéité et la pertinence des groupements de critères, ainsi que dans le rapport de proposition entre les paramètres d'importance.

La MEPI repose sur un processus de comparaison par paire. Les critères de même niveau sont comparés deux à deux relativement au critère de niveau supérieur g_A . Les comparaisons de paires de critères (g_i, g_j) sont effectuées à l'aide d'une échelle sémantique à laquelle est associée une échelle numérique (voire tableau ci-dessous). L'échelle reflète en nature et intensité l'expression de préférence partielle entre chaque critère.

Valeur numérique	Échelle sémantique	Description
1	importance égale des deux critères	Les deux critères contribuent autant au critère père
3	faible importance d'un critère par rapport à un autre	L'expérience et l'appréciation personnelles favorisent légèrement un élément par rapport à un autre
5	importance forte ou déterminante d'un critère par rapport à un autre	L'expérience et l'appréciation personnelles favorisent fortement un critère par rapport à un autre
7	importance attestée d'un critère par rapport à un autre	Un critère est fortement favorisé et sa dominance est attestée dans la pratique
9	importance absolue d'un critère par rapport à un autre	Les preuves favorisant un critère par rapport à un autre sont aussi convaincantes que possible
2,4,6,8	valeurs intermédiaires entre deux appréciations voisines	Un compromis est nécessaire entre deux appréciations

Tableau 14 : échelle de mesure de la méthode AHP

Les comparaisons par paire sont présentées dans une matrice carrée, réciproque, de dimension n , notée $M = (m_{i,j})$ où $m_{i,j}$ représente l'importance de g_i sur g_j relativement à g_A et est telle que :

$$m_{i,j} > 0 \text{ et } m_{i,i} = 1 \quad i, j = 1, \dots, n.$$

Les positions de transposition de la matrice sont calculées selon l'axiome de réciprocité :

$$m_{i,j} = \frac{1}{m_{j,i}} \text{ ce qui réduit le nombre de comparaisons à } \frac{n(n-1)}{2}.$$



L'étape de synthèse des appréciations suivante permet d'évaluer l'IRC à partir des appréciations formulées au cours du processus de comparaison par paires. L'IRC prend alors la forme d'un vecteur $W = (w_1, w_2, \dots, w_n)$ où w_i représente l'importance relative du critère g_i par rapport à sa famille d'appartenance. Le vecteur W peut être obtenu en calculant la moyenne de chaque colonne de la matrice M normalisée (chaque élément est divisé par la somme des éléments de même colonne).

Soit M' la matrice de comparaison normalisée issue de M et telle que :

$$\mathbf{m}'_{i,j} = \left(\sum_{k=1}^n \mathbf{m}_{k,j} \right)^{-1} \cdot \mathbf{m}_{i,j} \quad i,j,k = 1..n.$$

Alors le vecteur d'importances relatives W est tel que :

$$w_i = \frac{1}{n} \cdot \sum_{j=1}^n \mathbf{m}'_{i,j} \quad i,j = 1..n$$

Des appréciations sont dites cohérentes si elles satisfont la propriété de transitivité suivante :

$$\mathbf{m}_{i,j} = \mathbf{m}_{i,k} \cdot \mathbf{m}_{k,j} \quad i, j, k = 1..n$$

Les situations de cohérence parfaite sont rares. Aussi, la méthode AHP propose de calculer un Indice de Cohérence (IC) qui mesure l'écart entre les réponses du ou des décideurs et des réponses jugées cohérentes.

La dernière phase de la méthode AHP est la synthèse des appréciations sous la forme d'un critère unique. Appliquons cette méthode à une hiérarchie simple de deux niveaux :

Soit H cette hiérarchie avec :

$F = (g_1, g_2, \dots, g_n)$ pour famille de critères de niveau 1 ;

$D = (d_1, d_2, \dots, d_n)$ pour ensemble des données descriptives de niveau 2 telles que : d_i est en relation avec le critère g_i de niveau supérieur et $d_i = \{d_{i,1}, d_{i,2}, \dots, d_{i,k_i}\}$;

$A = (a_1, a_2, \dots, a_m)$ pour ensemble d'actions potentielles avec $a = \{d_1(a), d_2(a), \dots, d_n(a)\}$.

Soit W, le vecteur d'importances relatives associé à F tel que $w_i, i = 1..n$ représente l'importance du critère g_i .

Soit W^i , le vecteur d'importances relatives associé à d_i tel que $w_{i,j}, j = 1..k_i$ représente l'importance relative de $d_{i,j}$.

Soit $a \in A$ une action quelconque. L'expression de préférence globale de a par rapport à la famille de critères F est : $g(a) = \sum_{i=1}^n w^i(d_i(a)) \cdot w_i$.

Les désavantages de cette méthode sont :

- la comparaison par paire de AHP prête un sens spécifique à la notion de poids (IRC) lié à la fois à l'agrégation arithmétique utilisée et au mode de détermination des poids. L'interprétation donnée par le concepteur de la méthode à ces poids n'est pas nécessairement celle perçue par le décideur qui se prête à la MEPI ;
- l'utilisation de la moyenne arithmétique peut conduire à des renversements de préférences lorsqu'une action supplémentaire est ajoutée ;



- les liens entre les critères sont des liens de subordination mais la notion de coordination entre ceux-ci ne peut être modélisée à travers une somme pondérée (voire intégrale sur les intégrales floues).

Approche relationnelle: méthodes de sur-classement

L'idée de cette approche, est que les alternatives sont comparées deux à deux, et que l'on exprime par un nombre le degré de préférence de l'une des alternatives par rapport à une autre vis à vis d'un critère donné. Toutes ces relations de préférence sont alors agrégées pour tenir compte de tous les critères. Cette approche a été développée essentiellement par B.Roy depuis plusieurs années [Roy *et al.*, 1993].

La comparaison de deux solutions fait apparaître diverses situations de préférence qui peuvent être modélisées par des relations binaires:

- la relation de dominance Δ est définie par : $\forall a, b \in S, a \Delta b \Leftrightarrow x_j^a \geq x_j^b$.
- $a I b$: a est indifférent à b ; I est réflexive et symétrique.
- $a P b$: a est préféré à b ; P est irréflexive et asymétrique
- $a Q b$: a est préféré ou est indifférent à b ; Q est irréflexive et asymétrique.
- $a R b$: a est incomparable à b , irréflexive et symétrique.
- S (relation de surclassement) : $S = P \cup I$.

L'idée de base de l'approche relationnelle, est d'enrichir la relation de dominance (Δ) de manière à progresser dans la résolution d'un problème décisionnel . Il ne s'agit pas cependant, contrairement aux méthodes cardinales de chercher à bâtir directement un opérateur d'agrégation unique qui agrège les n critères en un seul. Il s'agit plutôt d'une approche alternative qui passe par une agrégation intermédiaire et partielle des n critères en relation de surclassement, une méthode de surclassement intègre deux phases bien distinctes : la construction de la relation de sur-classement, et l'exploitation de cette dernière à des fins d'aide à la décision comme nous allons voir dans la description de la méthode ELECTRE I.

La méthode ELECTRE

ELECTRE I [Roy *et al.*, 1993] suppose résolue la problématique de construction d'un jeu de critères non réductible en un seul, c'est une famille cohérente de critères. Mais très souvent, le décideur souhaite voir tel ou tel critère avoir beaucoup d'importance (poids) dans la procédure d'agrégation. La traduction de ces jugements qualitatifs en un jeu de coefficients d'importance ayant une signification économico-mathématique précise n'est pas triviale. Afin de caractériser l'importance de l'ensembles des critères concordant avec l'affirmation $a S b$, on attribue à chaque critère un indice d'importance noté w_i positif d'autant plus grand que le critère c_i est important. Dès lors, on peut obtenir l'indice de concordance $P(a,b)$ donnant une idée de l'importance de la « coalition » des critères appuyant la proposition $a S b$ comme suit :
$$P(a,b) = \sum_{c_i \in S(a,b)} w_i / \sum_1^n w_i$$

A partir d'un seuil de concordance choisi $\hat{c} \in]0.5,1]$ par l'utilisateur, on définit la relation de concordance Co par : $a Co b \Leftrightarrow P(a,b) \geq \hat{c}$. Soulignons que cette relation exclut toute



possibilité de compensation entre les critères, puisque la conclusion $a \text{ Co } b$ ne dépend pas des évaluations de a et b sur les critères c_i tels que $b \not\leq_i a$. Pour atténuer les effets parfois excessifs de cette méthode non compensatoire, dans la version originale d'ELECTRE I, l'indice de discordance est définie par:

$$d(a,b) = \begin{cases} 0, & x_i^a \geq x_i^b, \forall i \\ \frac{1}{\delta} \max[x_i^b - x_i^a], & \text{sinon} \end{cases}$$

où $\delta = \max_{a,b,i} [x_i^b - x_i^a]$, $d(a, b)$ est un indice (compris entre 0 et 1) d'autant plus grand que la préférence de b sur a est marquée sur au moins un critère.

Respectivement, s'il existe des critères qualitatifs, une autre façon de prendre en compte le concept de discordance est de définir la relation de discordance (D). Il s'agit de donner sur chaque critère, un ensemble D_j de couples d'évaluations (x_j, y_j) tels que, si $x_j^a = x_j$ et $x_j^b = y_j$ on refuse le sur-classement $a \not\leq b$, car $x_j^b - x_j^a$ est estimée si importante qu'elle n'autorise pas l'affirmation $a \leq b$, puis, on fixe le seuil de veto V_j sur les c_j , qui représente la plus petite différence des performances $x_j^b - x_j^a$ incompatible avec le sur-classement $a \not\leq b$. Plus la valeur V_j est faible, plus le pouvoir de veto augmente : $(x_j^a - x_j^b) \geq V_j \Leftrightarrow (x_j^a, x_j^b) \in D_j$. Dès lors on obtient la relation de discordance D , $a D b \Leftrightarrow \exists j \in \{1,2,\dots,n\}$ tel que $(x_j^a, x_j^b) \in D_j$.

Ayant défini le seuil de concordance \hat{c} (relativement grand) et, un seuil de discordance \hat{d} (relativement petit) [respectivement la relation de discordance D], on définit la relation de sur-classement $\$$: $a \$ b \Leftrightarrow P(a,b) \geq \hat{c}$ et $d(a, b) \leq \hat{d}$ [Resp. $a \$ b \Leftrightarrow a \text{ Co } b$ et non [$a D b$]].

La relation de surclassement $\$$ peut être représentée sous la forme d'un graphe dont les sommets sont les solutions. On cherche donc un sous ensemble N de solutions tel que toute solution n'appartenant pas à N , est surclassée par au moins une solution de N et les solutions de N sont incomparables entre elles (cette deuxième condition permet de rendre N minimal pour l'inclusion). Dans la théorie des graphes, un tel ensemble porte le nom de noyau du graphe et des algorithmes existent pour le déterminer. Pour évoluer vers le meilleur compromis, il reste alors à analyser de façon plus fine les solutions du noyau. D'autres variantes de cette méthode ont été proposées, suivant les problématiques étudiées [Roy *et al.*, 1993 ; Vincke, 1989].

Intégrales floues

Nous avons traité les relations de subordination entre les critères dans le paragraphe *Approche cardinale*, il est nécessaire dans la pratique de considérer les relations de coordination entre critères d'évaluation. En effet, dans le cas général, les critères sont liés, en plus des liens de subordination, par des relations transversales [Mousseau, 1993] : ces relations transversales expriment des liens de coordination qui assurent la cohérence de scores d'un même niveau hiérarchique. Pour bénéficier de cette distinction sémantique des relations entre critères, il est nécessaire d'introduire de nouvelles fonctions d'agrégation comme l'intégrale de Choquet



[Grabisch *et al.*, 2000] [Berrah *et al.*, 2000]. Cet opérateur est une bonne illustration de l'agrégation de performances liées par des dépendances de subordination et de coordination : elle permet en particulier de modéliser l'importance objective d'un critère et son interaction avec les autres critères d'évaluation.

Plusieurs situations peuvent être distinguées :

- les objectifs déclinés en critères d'évaluation sont parfaitement non-interactifs ou indépendants, c'était l'hypothèse du paragraphe précédent ;
- les objectifs sont interactifs ; un type d'interaction concerne la notion de conflit. D'un autre point de vue, ces objectifs sont partiellement *complémentaires*, c'est-à-dire que leur satisfaction simultanée affecte la performance agrégée plus significativement que des satisfactions prises séparément (il y a synergie) ;
- un autre type d'interaction d'objectif est la redondance, c'est-à-dire des objectifs qui sont en quelque sorte interchangeables parce que les performances respectives auxquelles ils se rapportent évoluent dans le même sens.

Comment en pratique prend-on en compte ce type de caractéristiques dans une évaluation ? Il est clair que si l'on que si l'on est capable de construire une hiérarchie de critères où l'on a pris soin de n'utiliser que des objectifs indépendants et additifs, cette problématique n'a pas lieu d'être. Malheureusement, en pratique, l'indépendance suppose en particulier qu'il n'y ait pas de redondance entre les critères et la classique moyenne pondérée n'est pas capable de rendre compte de cet aspect. Plutôt que d'essayer de construire des structures hiérarchiques indépendantes et additives, on peut modéliser l'interaction d'objectifs comme le permet l'intégrale de Choquet.

L'idée initiale est de construire une fonction d'ensemble sur C (l'ensemble des critères), non additive, permettant de définir un poids non seulement sur chaque critère, mais aussi sur chaque sous ensemble de critères, ce problème a été partiellement résolu depuis l'introduction des mesures et intégrales floues [Grabisch *et al.*, 1996].

Ces intégrales, telles que l'intégrale de Choquet [Grabisch et al., 1996], sont définies sur la base du concept de mesure floue, qui est une généralisation des mesures de Lebesgue.

Définition. La mesure floue sur l'ensemble des critères C est la fonction d'ensemble $\mu: P(C) \rightarrow [0,1]$ satisfaisant les axiomes :

- i) $\mu(\emptyset) = 0, \mu(C) = 1;$
- ii) $I \subset J \subset C \Rightarrow \mu(I) \leq \mu(J)$. (monotonocité)

Dans ce contexte, $\mu(I)$ représente le coefficient d'importance de l'ensemble des critères I .

Une mesure floue est dite additive si $\mu(I \cup J) = \mu(I) + \mu(J)$ lorsque $I \cap J = \emptyset$. Dans ce cas, il suffit d'identifier les n coefficients (poids) $\mu(c_1), \mu(c_2), \dots, \mu(c_n)$ pour identifier la mesure floue entièrement. Mais, de façon générale, on a besoin d'identifier 2^n coefficients correspondant à la dimension de $P(C)$.

Définition. Soit μ une mesure floue sur C . L'intégrale de Choquet d'un profil $x = (x_1, x_2, \dots, x_n)$ suivant la mesure μ est défini par :

$$C_\mu(x_1, \dots, x_n) = \sum_{i=1}^n (x_{(i)} - x_{(i-1)}) \cdot \mu(A_{(i)}) \quad \text{avec } 0 \leq x_{(1)} \leq \dots \leq x_{(n)} \leq 1 \text{ et } A_{(i)} = \{c_{(i)}, \dots, c_{(n)}\}.$$



L'intégrale de Choquet est idempotente, continue et monotone croissante, ce qui implique que, les intégrales floues sont toujours comprises entre le minimum et le maximum de leurs arguments. Les opérateurs connus : minimum, maximum, médiane, moyennes, les OWA peuvent être vus comme des cas particuliers des intégrales de Choquet.

On peut montrer que l'intégrale de Choquet dans le cas 2-additif prend la forme :

$$CI_g(u_1, u_2, \dots, u_p) = \sum_{\substack{I_{ij} > 0 \\ i > j}} \min(u_i, u_j) I_{ij} + \sum_{\substack{I_{ij} < 0 \\ i > j}} \max(u_i, u_j) |I_{ij}| + \sum_{i=1}^p u_i (\nu_i - \frac{1}{2} \sum_{j \neq i} |I_{ij}|)$$

avec $\nu_i - \frac{1}{2} \sum_{j \neq i} |I_{ij}| \geq 0$ et où :

- les ν_i sont les indices de Shapley et représentent l'importance globale de chaque objectif relativement à tous les autres ($\sum_{i=1}^p \nu_i = 1$).
- les I_{ij} représentent les interactions entre les paires d'objectifs (C_i, C_j) qui prennent leur valeur dans l'intervalle $[-1;1]$; une valeur de 1 signifie qu'il y a un effet synergique positif entre les deux objectifs, une valeur de -1 qu'il y a une synergie négative, et une valeur nulle que les objectifs sont indépendants.

Ainsi, peut-on distinguer décomposer l'intégrale de Choquet en trois parties, une conjonctive, une disjonctive et une additive :

- un I_{ij} positif implique que la satisfaction simultanée des critères C_i et C_j a un effet significatif sur la performance agrégée et qu'une satisfaction unilatérale n'a pas contre aucun effet ;
- un I_{ij} négatif implique que la satisfaction de C_i ou C_j est suffisante pour avoir un effet significatif sur la performance agrégée ;
- un I_{ij} nul implique qu'il n'existe pas d'interaction entre les deux critères considérés ; ainsi les ν_i jouent le rôle de poids des moyennes pondérées classiques.

Les propriétés mathématiques de l'intégrale de Choquet ont été largement traitées dans les travaux de Grabisch [Grabisch *et al.*, 1995 ; Grabisch, 1996 ; Grabisch *et al.*, 1998 ; Grabisch *et al.*, 2000].

2. LES APPROCHES COGNITIVES DE LA DECISION

Pour aborder la décision sous l'angle de la cognition, il est utile de s'attarder déjà sur les différents sens que l'on prête communément au terme de décision : acte ou processus.

Dans une décision finale se mêlent raisonnements et intuitions, calculs et appréciations globales, volonté et doutes, hypothèses et certitudes, expériences et rêveries, débats et compromis. Comment qualifier cette connaissance action dont l'usage est si courant ? Nous proposons dans cette partie quelques éléments concourant à éclairer l'éénigme de la décision sous l'angle des sciences humaines...

2.1. Le processus de la décision

La décision est au cœur d'une des grandes énigmes de la pensée et de l'action. Elle semble partout présente dans l'Histoire et dans les petites histoires de la vie de tous les jours.



L'examen de la réalité des décisions montre que décider ne correspond pas à une phase précise, clairement identifiable où tombe le couperet. La langue française emploie l'expression « prendre une décision » un peu comme si la décision était un objet identifiable. Le point d'arrivée est confondu avec le processus. La langue anglaise, qui emploie l'expression de « decision making process » (la fabrication de la décision) rend mieux compte de ce fait. La décision est un cheminement : elle se construit, se négocie, suit des voies sinuées au cours du temps. Elle pourrait sembler insaisissable. Depuis longtemps, les sciences de l'homme tentent de percer les mystères de la décision, d'en mettre en évidence les logiques, d'y déceler des régularités. Les sciences de l'homme et de la société apportent un éclairage à la compréhension de situations de décisions.

2.2. Le kaléidoscope de la décision et l'apport des sciences humaines

Il n'existe clairement pas de science de la décision. Il existe simplement plusieurs entrées, plusieurs points de vue, plusieurs lectures. Enfin, il est légitime de distinguer quelques grands domaines qui constituent des centres d'intérêt particulier pour des études sur la décision : le domaine politique [Allison, 1971], celui des entreprises et des organisations, celui de la vie quotidienne.



TABLEAU COMPARATIF CAMESCOPES NUMÉRIQUES										
MODÈLE	Caract. MV300									
SORBIER	01/01/00	01/01/00	01/01/00	01/01/00	01/01/00	01/01/00	01/01/00	01/01/00	01/01/00	01/01/00
Prix	Entre 10000 et 15000 euros									
200004	Entre 10000 et 15000 euros									
Capacité	Entre 10000 et 15000 euros									
Risque	Entre 10000 et 15000 euros									
Sensibilité	Entre 10000 et 15000 euros									
Affichage	Entre 10000 et 15000 euros									

Figure 57 : Stratégie politique ou choix du consommateur, la décision semble partout présente dans l'Histoire et dans les petites histoires de la vie de tous les jours

Les théories de la décision pourraient être déclinées en cinq grands types d'approches qui privilégient selon le cas : la rationalité de l'action, la psychologie du décideur, les contraintes organisationnelles, le poids des compromis, le contexte de l'action.

Selon le premier modèle rationnel, la décision prise doit être le résultat d'un choix comparatif entre les diverses solutions possibles. Le décideur et ses conseillers doivent mesurer avec soin les risques et les issues probables de chaque formule, peser leurs avantages et inconvénients pour retenir finalement celle qui représente le meilleur rapport « coût/efficacité ». Cette analyse, en terme de calcul rationnel, postule l'existence d'un acteur unique qui agirait en vertu de préférences hiérarchisées en fonction de la meilleure utilité.

Ce séduisant modèle théorique fait abstraction des aspects organisationnels souvent implicites dans un processus décisionnel. En effet, la panoplie des solutions théoriquement possibles est





plus large que celles réellement envisagées par une approche purement rationnelle. Les propositions émanent de scénarios préétablis par des membres bien identifiés de l'organisation. De plus les informations disponibles sont généralement incomplètes, imprécises ou contradictoires !

L'idée de cette approche organisationnelle est que le décideur ne possède pas en fait une connaissance totale de la situation, d'où le terme de « rationalité limitée » cher à H.A. Simon. Les limitations dans la connaissance des faits et hypothèses proviennent des contraintes de l'organisation qui sélectionne ou favorise tel ou tel scénario en fonction de ses intérêts.

Une approche dite politique de la décision consiste à mener une analyse qui privilégie le jeu des acteurs, leur capacité à manœuvrer, à produire des coalitions. Le choix fait l'objet de jeux de pouvoirs entre les parties concernées. Une solution de compromis entre les positions radicales est généralement adoptée.

Les approches psychologiques s'intéressent selon l'angle de vue adopté, aux processus mentaux mis à l'œuvre lors d'une décision individuelle ou bien à la personnalité des décideurs.

La psychologie cognitive envisage la décision comme un cas particulier des stratégies de résolution de problèmes. Elle explore des rationalités subjectives employées par les individus en situation. Beaucoup d'expériences concernent les biais cognitifs c'est-à-dire les raisonnements en apparence corrects qui conduisent à de fausses conclusions. Heureusement, les heuristiques (c'est-à-dire les méthodes de résolution de problèmes approximatives) employées dans la vie courante sont souvent fiables.

L'autre voie explorée par la psychologie de la décision concerne la personnalité des décideurs. Les spécialistes américains de management ont décrit toute une panoplie de profils de personnalité ou styles décisionnels qui correspondent à autant de styles de leadership : les intuitifs, les rationnels, les ingénieux, les rigides, les autoritaires, les innovants, etc. Alors que l'intuitif aime suivre son inspiration, le rationnel s'informe, calcule, soupèse, s'arme de prudence. Le premier est créatif, le second plus conformiste.

On peut encore ranger les études de psychosociologie parmi ces analyses psychologiques. On souligne alors comment les interactions entre participants vont contribuer à polariser la décision soit dans le sens d'une radicalisation, soit dans le sens du conformisme.





Les modèles dits composites désignent une panoplie d'autres approches dont le seul trait commun est de combiner plusieurs aspects des modèles précédents. Derrière l'image trompeuse d'une décision consciente et unifiée il y a, en fait, une multiplicité de rationalités différentes qui s'imbriquent, se superposent, se confrontent.



Le nombre d'approches de la décision reflète bien sûr la complexité du phénomène mais aussi la diversité des domaines d'application. On ne décide pas de la même façon dans un ministère ou dans une PME, dans une assemblée générale de grévistes ou dans une firme multinationale. Selon le contexte d'action, la dimension politique, psychologique, organisationnelle, le poids de l'environnement,... seront plus ou moins prépondérants [Stratégor, 1988].

<i>Types d'approches</i>	<i>Dimensions mises en lumière</i>
<i>Approche rationnelle</i>	<ul style="list-style-type: none">• définition précise du problème• liste systématique des choix possibles• calcul de la meilleure utilité
<i>Approche organisationnelle</i>	<ul style="list-style-type: none">• rationalité limitée du décideur• contraintes de l'organisation• choix d'une solution satisfaisante
<i>Approche politique</i>	<ul style="list-style-type: none">• importance des négociations et compromis entre les groupes concernés
<i>Approche psychologique</i>	<ul style="list-style-type: none">• stratégie mentale de résolution de problèmes• profil de personnalité des décideurs et styles décisionnels
<i>Approche composite</i>	<ul style="list-style-type: none">• intègre plusieurs dimensions des modèles précédents• étude des phases, des niveaux hiérarchiques et des contextes de la décision

Quelles caractéristiques du processus de décision nous enseignent ces quelques lignes ?

D'abord dans le domaine du politique au sens large du terme, c'est la distribution du pouvoir entre les hommes politiques, technocrates et société civile qui constitue la trame des acquis de la science politique.

Le décideur n'est pas monolithique : les actions ou mesures résultantes sont élaborées à travers un processus long et enchevêtré mobilisant de nombreux acteurs ; il n'y a pas de décision, un décideur, mais une série de stratégies et de compromis entre les points de vue, entre des groupes qui ne partagent pas la même solution. Dans sa théorie du surcode [Sfez, 1992], L. Sfez montre que derrière l'image trompeuse d'une décision consciente et unifiée, il y a en fait une multiplicité de

rationalités différentes qui s'imbriquent, se superposent se confrontent.





Le décideur in fine sert de détonateur à l'élaboration de nouvelles politiques au sens large du terme. Même si c'est lui qui signe et tranche en dernier ressort, il s'appuie sur des solutions concrètes dont la formulation détaillée, techniquement constituée, n'est pas son œuvre. Des groupes tiers fournissent ces solutions, élaborent les options et assurent leur légitimité opérationnelle. Les mesures prises en définitive ne sont pas nécessairement des réponses à des exigences générales de la société, de l'entreprise mais le produit d'activités intermédiaires d'experts et de conseillers. A la limite, on peut raisonnablement se demander si le décideur formel décide vraiment, et même si décision il y a, qui soit repérable et prise à un moment donné [Thoenig, 1987].

L'argumentaire est un incontournable exercice de rhétorique du monde politique. A ce propos, Henri Nallet, ex-ministre de l'agriculture puis de la justice, déclarait lors de sa première rencontre avec les représentants de l'institution judiciaire : « Je ne demanderai au premier ministre une « rallonge » budgétaire que si nous sommes capables de justifier tout franc supplémentaire par une amélioration correspondante dans le fonctionnement du service. Si on ne peut pas justifier, je ne demanderai rien.».

Sur le plan de la psychologie sociale, un premier constat est que lors d'une prise de décision, les groupes peuvent adopter des choix plus risqués que les individus ; le groupe est alors supposé favoriser la dilution des responsabilités, la réduction de l'engagement des individus. On peut penser au contraire, que dans certaines circonstances, les individus s'engagent plus au sein d'un groupe : il a été en effet observé que la décision en groupe provoque un renforcement des options initiales quelles qu'elles soient, on parle de processus de polarisation.

Un autre point de la décision de groupe concerne la nécessité de parvenir à un consensus [Doise, 1992] : l'engagement des individus s'en trouve augmenté car, s'il n'y a pas nécessité d'aboutir à une décision commune, les membres d'un groupe débattent de leur opinions personnelles, mais sans chercher à retravailler cognitivement les opinions exprimées par les autres.



L'objectif d'un consensus ne doit pas être de supprimer les conflits, mais de les tolérer. Plus les options individuelles sont variées, plus la confrontation est importante et plus la reformulation du problème peut devenir explicite. La recherche d'un compromis par le groupe incite les individus à reconsidérer leurs positions sur le problème. Lorsque les membres du groupe se sont mis d'accord sur une conclusion, on constate qu'ils la maintiennent ensuite lorsqu'on les interroge individuellement. La possibilité de laisser s'exprimer les opinions divergentes favorise donc la cohésion du groupe. Par ailleurs, il faut noter que l'innovation vient souvent des points de vue minoritaires dissidents : ainsi, pour être véritablement efficaces, les mécanismes de décision doivent favoriser la contradiction.

Qu'en est-il de la décision stratégique d'entreprise ? [Koenig, 1990 ; STRATEGOR, 1988 ; March, 1991]. Compte tenu des remarques précédentes, on peut affirmer qu'aujourd'hui l'action d'un dirigeant d'entreprise ne correspond pas à l'image mythique du grand patron, clairvoyant et omniprésent. Sa connaissance de l'environnement est limitée, il se heurte à une organisation souvent rétive à ses choix. L'élaboration d'une stratégie émergente doit tenir compte de ces contraintes.





Le dirigeant a bien souvent été représenté comme le grand stratège, à la tête d'un arsenal de prévisions, scénarios, calculs, méthodes, plans, programmes, projets, qu'une organisation disciplinée et efficace élabore sous sa direction, et dont il conduit fermement mais sereinement la réalisation.

Représentation mythique ou modèle normatif ? L'important est que se révèle une trame commune, une sorte de modèle de base, implicite, de la décision stratégique, qui comporte trois temps successifs :

- l'anticipation : les décisions procèdent de l'état futur de l'environnement ;
- le choix : le décideur est le dirigeant qui exprime sa volonté, fruit de son analyse ;
- la mise en œuvre : le choix arrêté par le dirigeant est réalisé par l'entreprise conformément à sa volonté.

Malheureusement, de nombreuses études empiriques ont montré que les processus de décision stratégiques dans les entreprises s'écartent sensiblement de ce modèle rationnel et linéaire. Tout d'abord l'environnement décourage les anticipations. Ensuite, l'entreprise est bien moins docile qu'on ne le croit, pour exécuter les décisions du dirigeant. Enfin, le dirigeant lui-même ne ressemble guère à ce décideur aux capacités exceptionnelles.

C'est d'un environnement incertain que naissent et disparaissent les menaces qui peuvent affecter la stratégie de l'entreprise. La stratégie implique donc de mettre l'entreprise en adéquation avec son environnement, et tout d'abord avec le ou les marchés visés. Dans ce domaine, la fiabilité des prévisions est plus que douteuse et certaines entreprises estiment que des prévisions fausses sont plus nuisibles que l'absence de prévisions, car elles produisent de fausses certitudes, qui limitent les capacités d'intelligence et de réaction des managers. L'environnement stratégique d'une entreprise est aussi un univers d'acteurs—concurrents, fournisseurs, distributeurs—parties prenantes de toutes sortes. Les alliances de plus en plus nombreuses que les firmes passent entre elles sont souvent décrites comme des moyens de réduire l'incertitude générée par la multiplicité des acteurs, en stabilisant leurs relations.

Pour ce qui est de l'organisation rétive, les moyens humains et organisationnels qui composent l'entreprise sont censés servir les desseins stratégiques élaborés par le dirigeant. La formulation de la stratégie serait l'apanage du dirigeant, et sa mise en œuvre serait la mission de l'ensemble de l'entreprise. Cette distinction si naturelle occulte le rôle considérable de l'organisation dans la réalisation de la stratégie. L'organisation se montre bien souvent rétive et cela de plusieurs manières.

Les grandes entreprises sont dotées de structures et de systèmes de gestion qui déterminent largement leur fonctionnement. Ces dispositifs constituent souvent une force d'inertie non négligeable. Pour nombre de grands groupes, il a fallu des années pour cesser de mesurer la performance de l'entreprise en quantité produite et en coût par unité produite et centrer leur comportement sur d'autres critères, tels que l'adaptation des produits aux demandes du marché, la qualité des produits et du service, la diversification de l'offre, etc.

Les structures produisent leur propre dynamique, mais les acteurs, individus et groupes, qui peuplent ces structures, s'en servent également pour satisfaire leurs propres intérêts. Les décisions stratégiques, par essence, désignent des enjeux importants. Notamment, elles



répartissent les ressources humaines et financières au sein de l'organisation. Les actions stratégiques de la direction générale sont décodées en ces termes par les acteurs, qui de plus nourrissent souvent des projets spécifiques. Des processus politiques (luttes, conflits, négociations, alliances, ruses, etc.) s'engagent autour de la formulation, de la sélection, et de la réalisation de ces actions et de ces projets. Il n'est pas toujours facile de distinguer clairement les intérêts particuliers et l'intérêt de l'entreprise.

A titre d'exemple, la conception d'une automobile dans une grande firme implique une myriade de micro-décisions prises au niveau des services techniques. Les choix sont fondés sur le compromis, l'autorégulation, l'ajustement spontané. Les conflits potentiels sont tranchés par le sommet à des dates clés qui scandent les phases de la réalisation.

Revenons sur les limites de la pensée stratégique du dirigeant. Le décideur ne domine ni l'environnement auquel il est confronté ni l'organisation qu'il dirige. Il n'est pas non plus, ainsi que le mythe le dépeint, idéalement lucide et rationnel. La liste est longue des biais cognitifs auxquels les dirigeants sont exposés. Il faut convenir que les problèmes stratégiques, par leur importance, leurs ambiguïtés, leur singularité, favorisent les erreurs et les désillusions.

Le dirigeant est confronté à un dilemme : en privilégiant les approches méthodiques, traitant des données objectives et autant que possible quantifiées, il perd les informations qualitatives, impalpables mais essentielles, car un problème stratégique est toujours une configuration complexe de faits et d'interprétations ; mais en faisant confiance à l'expérience, aux analogies, à sa perception, à l'intuition, il risque d'être victime de sa subjectivité. Le dirigeant n'aborde pas les situations d'un œil vierge. Il a notamment des représentations de ce qui est important ou non, de ce qu'il convient de faire ou non. Il s'appuie sur des règles et principes qui sont pour lui des évidences, mais qui pour l'observateur apparaissent comme des croyances, c'est-à-dire des manières de penser dont la validité n'est pas démontrée.

On peut ainsi s'interroger sur la capacité des dirigeants à contrôler la stratégie ? Entre les incertitudes générées par l'environnement, les dérives organisationnelles, les croyances des dirigeants, y a-t-il donc une chance de voir se former des décisions stratégiques sensées et cohérentes ? Il est vrai que ces tensions peuvent empêcher la formation d'une stratégie véritable mais dans beaucoup de cas, elles peuvent être sinon résolues, du moins contrôlées.

Souligner que les processus qui fabriquent la stratégie s'écartent souvent de la rationalité mise en scène par les mythes du dirigeant décideur ne signifie pas que les dirigeants soient impuissants. Ils peuvent développer des moyens indirects de contrôle et d'influence. Ainsi, par la mise en place de structures et de systèmes de gestion, ils définissent des règles du jeu qui peuvent induire une dynamique allant dans le sens souhaité. Le dirigeant ne manque pas d'atouts du fait de sa position privilégiée à l'interface entre les unités de l'entreprise et en contact avec l'environnement. Il peut se poser en arbitre régulateur des conflits internes et favoriser les projets et les acteurs qui sont compatibles avec ses propres préférences et ses propres ambitions.





La véritable question semble pourtant être ailleurs. La configuration des structures et la répartition du pouvoir dans l'entreprise sont généralement des parties d'un système plus ou moins cohérent qui comprend également un ensemble de croyances portant sur la stratégie, largement partagé par le groupe des dirigeants, et souvent aussi par les autres niveaux du personnel. On parle alors de paradigme stratégique, par analogie avec les paradigmes scientifiques. Le paradigme décrit l'environnement, l'organisation dans l'environnement, et ce qui marche et ce qui ne marche pas pour survivre et prospérer dans cet environnement.

Le paradigme stratégique constitue une logique dominante qui détermine largement les résultats de la plupart des décisions. La question centrale est la capacité des dirigeants à en prendre conscience et à en jouer de manière à profiter de l'élan, tout en évitant les écueils et les dérives. Dans cette perspective, les événements stratégiques—faillite d'un gros client, entrée d'un nouveau concurrent puissant, apparition d'une nouvelle technologie de substitution—constituent des mises à l'épreuve du paradigme stratégique, des tests de sa capacité à maintenir l'entreprise sur une trajectoire favorable. Ils sont aussi pour les dirigeants l'occasion de faire évoluer le paradigme : redéfinir ce qui doit être tenu pour vrai et ce qui doit être tenu pour bon, et réorienter les forces qui façonnent la stratégie de l'entreprise.

On voit que le rôle du dirigeant, s'il n'a plus l'aura du grand décideur, est cependant fondamental, déterminant et... difficile ! Sa situation est assez inconfortable. On attend de lui une lecture claire d'un environnement incertain et souvent ambigu. Tenu d'incarner une volonté, il est pourtant dépendant, pour l'action concrète, d'une foule de relais—hommes, unités, procédures, systèmes. Entre les choix énoncés et les résultats promis et attendus, des années entières peuvent s'écouler, pendant lesquelles les orientations prises ne manqueront pas d'être contestées ou détournées, à la faveur d'événements extérieurs qui sembleront remettre en question son projet. La réalité des décisions stratégiques est donc bien éloignée de l'image mythique. Les réussites stratégiques se construisent rarement à l'aide d'intuitions fulgurantes ou de plans rigoureux. La capacité à produire une stratégie émergente à partir d'un flux de décisions et d'événements apparemment disjoints est une des premières conditions de survie pour les entreprises où le dirigeant est devenu un arbitre plus qu'un décideur mythique.

Toujours dans les stratégies d'entreprise, on peut signaler un modèle particulier, dit *modèle de la poubelle* (garbage can model [March, 1988]). Deux américains définissent à la fin des années 60, deux nouvelles notions. La première est celle d'anarchies organisées dont les universités sont selon, eux, un parfait exemple. Cette expression désigne les organisations :

- sans objectifs cohérents et partagés par tous,
- où le processus de production relève d'une technologie complexe, peu matérialisable
- dont les membres participent de façon active aux prises de décision.

La seconde est le modèle de la poubelle qui décrit le style de décision propre aux anarchies organisées. Ce modèle remet en cause les théories où les décisions résultent d'une confrontation entre des objectifs identifiés, des solutions disponibles et leurs conséquences (modèle rationnel) et les théories où les décisions sont le résultat d'une



négociation entre des groupes aux intérêts divergents (modèle politique). Dans les anarchies organisées, des choix sont à la recherche de problèmes, des questions cherchent des opportunités pour décider, des solutions cherchent des questions auxquelles elles pourraient être une réponse et des décideurs cherchent du travail...

Des décisions se produisent quand les flux de problèmes, de solutions, de participants et d'opportunité de choix se rencontrent. Toute prise de décision est ainsi assimilable à une poubelle où des types de problèmes et de solutions sont déchargés par les participants dès qu'ils sont générés et qui, se rencontrant, font émerger un choix.

La simulation informatique de ce modèle, a priori chaotique, ne fait pourtant apparaître que trois styles de choix possibles : les décisions par inattention, par déplacement des problèmes et par résolution des problèmes. Les deux premiers étant plus fréquents que le dernier. Cette émergence d'ordre dans les processus de décision anarchique a inspiré des travaux sur la capacité d'apprentissage des anarchies organisées.



Les apports du modèle de la poubelle sont nombreux. Premièrement, il invite à ne pas surestimer la rationalité des acteurs, à ne pas surévaluer leur capacité à appréhender le processus où ils interviennent. De même, il incite les décideurs à rester modestes et à ne pas recourir à des outils managériaux trop ambitieux (planification, rationalisation...). Mieux vaut travailler à la marge et utiliser une technologie de l'absurde : traiter les objectifs comme des hypothèses, considérer que sa perception intuitive correspond à la réalité, que son expérience a valeur de théorie...

Ce modèle permet également de mieux prendre en compte les aléas : le timing est un élément décisif de l'analyse, tout comme les événements qui se produisent au cours d'une décision. Ainsi, la présence (ou l'absence) d'une personne à une réunion peut modifier le sens de la décision.

Enfin, il permet de raisonner en dissociant les intentions des actions, les causes des effets. On peut ainsi trouver des problèmes parce qu'on détient une solution. On peut prendre une décision, non parce qu'on a un problème, mais parce qu'on a l'opportunité de faire un choix. On peut opter pour une solution même si elle ne répond pas au problème soulevé.

Stimulant et ludique, ce modèle donne un éclairage moins rationnel et linéaire et, par conséquent, plus contingent et aléatoire. Mais il comporte une philosophie du décideur souvent critiquée : inconstant dans ses préférences, infidèle à ses intentions, il est aussi incapable d'action rationnelle. A tel point qu'il disparaît de la scène et n'est pas central dans l'analyse du processus. Par crainte de sur-rationaliser, ce modèle tend aussi à sous-estimer tous modes de régulation, toutes règles du jeu formelles ou informelles pouvant circonscrire le comportement des acteurs [Musselin, 1990].



2.3. Les pièges de la raison

2.3.1. La logique des erreurs

L'objectif de la théorie mathématique de la décision était initialement d'expliquer comment un homme rationnel doit prendre des décisions dans l'incertitude. Le décideur construit un arbre de décision pour projeter les conséquences des différents choix possibles quant à la probabilité et à l'utilité des événements correspondants ; il décide entre les choix selon une règle de maximisation de la valeur espérée ou de l'utilité attendue. Cette première orientation s'est révélée incapable de prédire les décisions effectivement prises par des individus et donc a fortiori de décrire les processus mentaux qu'ils mettent en œuvre pour décider. On a expliqué cette incapacité par le fait que d'une part le système humain de traitement de l'information est limité, d'autre part l'individu n'estime pas et ne traite pas les probabilités selon les principes des théories probabilistes.

Le choix rationnel suppose la connaissance, l'examen et l'évaluation de toutes les options possibles, de tous les états de la nature actuels et futurs, des différentes issues possibles. Une fois tous ces aspects estimés sur plusieurs dimensions, il reste à faire la sélection présentant l'utilité maximale. Un tel processus requiert à la fois une grande disponibilité d'informations et une grande capacité de calcul. Kahneman, Slovic et Tversky ont mis quant à eux en évidence les propriétés de prise en compte et de traitement des probabilités par les individus [Kahneman *et al.*, 1982].

Illustration

Un taxi est impliqué, la nuit, dans un accident de la circulation. Deux compagnies de taxi exercent dans la ville, la verte et la bleue. 85% des taxis de la ville sont verts et 15% sont bleus. Un témoin a identifié le taxi de l'accident comme étant bleu. Le tribunal a testé la capacité du témoin à identifier les taxis la nuit. Lorsqu'on lui a présenté un échantillon équiperenti de taxis, le témoin a correctement identifié les taxis dans 80% des cas. Quelle est la probabilité que le taxi impliqué dans l'accident soit bleu ?

Une très forte majorité des personnes soumises à ce problème répond que la probabilité que le taxi soit bleu est de 80%. La réponse exacte, selon la théorie bayésienne des probabilités est 41%. En général cette dernière estimation surprend et ne paraît pas intuitivement plausible. Les réponses observées traduisent le fait que l'individu néglige les probabilités a priori (ici 85% et 15%), qui donnent le cadre général de l'estimation, pour ne considérer que l'information spécifique (ici la fiabilité du témoin).

Théorie bayésienne

B1 : le taxi est bleu

B2 : le taxi est vert

A : le témoin a vu un taxi bleu

$P(A/B_k)$ est la probabilité que le témoin voit un taxi bleu la nuit quand celui-ci est B_k .

$$P(A/B1) = .8$$

$$P(A/B2) = .2$$

$$P(B1) = .15$$

$$P(B2) = .85$$

La probabilité $p(A)$ est la probabilité de voir un taxi bleu :

$$P(A) = p(A/B1).p(B1) + p(A/B2).p(B2) = 0.29$$

Enfin, on applique la règle de Bayes pour calculer la probabilité que le taxi soit bleu quand le témoin affirme qu'il est bleu :

$$P(B1/A) = \frac{p(A/B_1).p(B_1)}{p(A)} = \frac{0.12}{0.29} = 0.41$$



Dans la mesure où le type de phénomène d'écart à une norme illustré dans la table précédente est très général, l'idée s'est imposée qu'il traduit une propriété de l'activité mentale (cognitive). Il a été désigné sous le terme de biais cognitif. Les principales propriétés de l'activité mentale dans la décision ont été inférées à partir d'un inventaire des principaux biais cognitifs [Caverni *et al.*, 1990].

Les biais que subit le décideur sont de quatre types : l'acquisition de l'information, le traitement de l'information, l'expression de la réponse, l'information qu'il reçoit en retour.

L'acquisition d'informations est biaisée dans plusieurs circonstances : lors de l'évocation de données en mémoire, lors de la sélection de données dans l'environnement, et en fonction du mode sous lequel les données se présentent.

Ainsi la fréquence des événements qui font l'objet d'une certaine publicité est généralement surestimée par rapport à celle des événements dont on parle peu. La fréquence absolue des événements est préférée à leur fréquence relative. Les informations concrètes sont plus prégnantes en mémoire que les informations abstraites. L'ordre dans lequel les informations sont présentées peut produire des effets dits de primauté ou de récence. Les premières ou les dernières informations induisent la décision finale.

L'individu apprend-il à décider ? Cela supposerait qu'il puisse reconnaître des relations stables entre événements dans l'information qu'il reçoit en retour de ces décisions. Or plusieurs biais interfèrent avec la possibilité d'apprentissage. Le fait que certains événements ne se produisent jamais est un obstacle pour estimer les conséquences de certaines décisions. D'autre part certaines décisions sont d'une fréquence trop faible pour qu'un individu puisse tirer profit d'éventuelles conséquences qu'il aurait observées. Enfin, lorsque des événements consécutifs à une décision sont accessibles, une tendance est de les attribuer systématiquement à l'action du décideur (illusion du contrôle).

Les biais mis en évidence par la psychologie cognitive [Hogart, 1987] attestent que le décideur ne fonctionne pas selon les principes des théories formelles. Toutefois les explications fournies par la psychologie cognitive sur les processus en œuvre dans la décision sont encore limitées : perception sélective, traitement séquentiel, capacités de calcul et de mémorisation limitées, sensibilité aux caractéristiques de la tâche, action sur l'environnement... Au lieu d'être algorithmique l'activité mentale du décideur serait heuristique : les procédures mises en œuvre seraient habituellement efficaces mais n'auraient pas de justifications rigoureuses.

Nous retiendrons que les processus psychologiques de la décision individuelle ne correspondent qu'imparfaitement au modèle d'un calculateur froid. Même lorsque l'individu se veut rationnel, de nombreuses erreurs logiques viennent piéger son raisonnement.

2.3.2. Décider dans l'urgence

Ajoutons que la décision doit parfois être prise dans l'urgence. En situation de crise, la mécanique de la décision se gripe parce que :

- la crise peut provoquer des dommages importants, et dans certains cas menacer la survie de l'organisation ;
- la réponse à apporter au problème n'est pas planifiée ;
- une attention immédiate est nécessaire ;
- l'événement est tel qu'il exige la mise en œuvre de toutes les ressources de l'organisation ;



- l'ensemble des caractéristiques de la crise va déclencher un stress important chez le décideur.

Les dommages que la crise peut provoquer, la complexité de la situation, et la rapidité avec laquelle il faut réagir créent une tension qui va peser sur la prise de décision. L'entrée en crise se matérialise souvent par un choc qui peut paralyser toute réaction du décideur. Le problème est tellement démesuré, on semble avoir tellement peu de prise sur les événements que l'on ne sait que faire. Passé ce cap, c'est en réfléchissant aux façons de répondre à la crise que l'on réalise la complexité de la situation. La résolution du problème ne va-t-elle pas entraîner l'aggravation d'un autre (la crise peut se révéler à la conjonction de défaillances techniques, humaines et organisationnelles).

La décision dans les situations de crise donne la limite des processus classiques. La pression empêche une évaluation sérieuse des possibilités de réponse et de leurs conséquences. Le stress initial provoqué par l'étendue des dommages est renforcé par la pression temporelle, l'incertitude et la complexité de la situation. Ainsi, ce qui peut aider le décideur, ce n'est pas tant la préparation de scénarios ou de procédures, que la prise de conscience qu'une crise peut survenir dans son entreprise, et donc de la préparation matérielle et psychologique à l'affronter [Gilbert, 1992][Lagadec, 1988].

2.3.3. Entre stratégie consciente et force aveugle

La décision peut être vue sous l'angle du déterminisme social ou de l'autonomie individuelle, des pulsions inconscientes ou du raisonnement lucide. Et si la diversité des regards tenaient moins aux conflits théoriques qu'à l'échelle d'observation adoptée ?

La décision en situation collective porte sur l'étude des processus collectifs, c'est-à-dire sur les interactions et les transactions sociales qui concourent à la production d'une décision. Les acteurs ont de « bonnes » raisons de décider ou de faire pression en faveur de telle ou telle solution. La décision est donc la résultante d'une somme d'interactions dont la logique peut parfois échapper aux acteurs eux-mêmes. Une clarification du débat sur la décision se trouve dans l'existence d'une différence d'échelles d'observation, l'obligation de découper dans la réalité, l'intégration d'un effet de situation dans le découpage de la réalité, la pertinence de la méthode des itinéraires pour reconstituer un processus de décision [Desjeux, 1988].

La première échelle est l'échelle macro-sociale. C'est celle de la société. Elle permet de montrer comment les individus ont incorporé les modèles culturels, les codes, les styles de vie de leur groupe d'appartenance. Dans ce type d'approches macro-sociales, la décision individuelle disparaît au profit des régularités sociales. A cette échelle, la tendance est plutôt de sous-valoriser les calculs de l'acteur.

La deuxième échelle est interactionniste. L'observation porte sur les liens concrets entre les acteurs. Elle montre comment les décisions individuelles sont contingentes, fruit de négociations sociales ; comment le discours, les enjeux sociaux, les intérêts, le sens, l'émotionnel organisent le champ décisionnel. A cette échelle, la raison et/ou l'émotion de l'acteur ont tendance à être survalorisées.

La troisième échelle est micro-individuelle. Elle permet de comprendre les arbitrages par lesquels un individu raisonne ses choix. La méthode consiste alors à faire reconstruire par les décideurs les qualités qu'ils cherchent, puis à leur faire élucider les signes par lesquels ils reconnaissent cette qualité.

Ainsi, il est illusoire de croire que même le consommateur qui achète son poulet au supermarché procède à une décision individuelle, c'est la situation qui est individuelle. La



décision, elle, est tout autant qu'en organisation la résultante d'un processus dans le temps. L'acteur subit des influences sociales liées à son origine sociale, aux médias ou interactions familiales.

2.3.4. Décision raisonnée ou raisonnable

B.Jarrosson explique dans [Jarrosson, 1994] les raisons pour lesquelles la décision ne peut être ramenée à la norme du calcul :

- le calcul néglige la complexité²⁸, c'est-à-dire l'existence de critères d'évaluations multiples et non commensurables ;
- le calcul néglige le sujet connaissant qui interprète l'information, qui se construit en construisant ses croyances²⁹ ;
- le calcul néglige que le décideur existe en tant qu'être autonome.

Ce n'est pas l'information brute qui sert à décider mais plutôt le sens qu'a cette information³⁰. Une information ne dit rien de ce qu'il faut faire, seule son interprétation l'indique. Son interprétation, autrement dit le sens qu'elle prend ou pas pour une personne. Le sens de l'information n'est jamais contenu dans l'information elle-même.

Le schéma causal de la décision se décline selon une vision linéaire de celle-ci [Sfez, 1992]. La décision s'inscrit dans un schéma à cinq temps.

- Formulation d'un désir et conception d'un projet y répondant
- Prise d'information
- Délibération
- Décision proprement dite
- Exécution.

L'absence d'une ou plusieurs de ces phases correspond à une déviation que l'on peut étiqueter :

- si l'il manque l'exécution, on parlera de velléité ;
- si l'on passe directement de la conception à la décision on parlera d'impulsivité ou d'instinct ;
- si l'on ne saute que la prise d'informations, on parlera de paresse, de manque de rigueur ;
- si l'on saute la délibération on parlera de légèreté ou de précipitation ;
- si l'on s'arrête à la délibération, il s'agira plutôt d'intellectualisme.

Ce schéma linéaire rencontre les limites de toute linéarité : l'impossibilité de considérer des causalités enchevêtrées. L'exécution, par exemple, ne va-t-elle pas changer la conception même du projet, le but recherché ? D'une manière générale, chacune des phases modifie l'objectif initial. La délibération conduit à rechercher de nouvelles informations, etc. Le schéma est simple mais il ne correspond pas à la réalité de causalités enchevêtrées. Le simple

²⁸ Les sciences humaines portent souvent un regard dur et un peu obsolète sur les mathématiques de la décision. Ceci dit, la réciproque est fréquemment et tristement vérifiée aussi.

²⁹ Voir plus loin la notion d'apprentissage collectif.

³⁰ Se référer à la notion de connaissances actionnables.



et le compliqué se comprennent analytiquement, mais le réel n'est ni simple, ni compliqué puisqu'il est complexe.

Il convient de distinguer l'évidence et la pertinence. L'évidence relève d'une cohérence interne, d'une compatibilité avec une structure logique ; la pertinence relève d'une cohérence externe, d'une compatibilité avec une situation extérieure. Il y a les solutions logiques (évidence) et les solutions efficaces (pertinence). Dans la mesure où l'on peut regarder le monde avec différentes logiques, les solutions logiques et les solutions efficaces ne sont pas toujours les mêmes. La décision est un processus d'interactions entre le décideur et le monde. Il faut donc qu'elle soit en accord avec le monde plutôt qu'avec sa logique interne. Elle doit être pertinente plutôt qu'évidente.

Dans l'idée de pertinence apparaît la notion de contexte. Une action est pertinente par rapport à un contexte, qu'elle soit logique ou pas, évidente ou pas. La décision évidente est raisonnée, la décision pertinente est raisonnable. Le schéma causal linéaire est évident mais il n'est pas pertinent, raisonné mais pas raisonnable. Certaines causes sont obscures et ambiguës, soit qu'elles existent de façon souterraine soit qu'elles ne possèdent pas l'intensité que l'évidence leur attribue.

Jarrosson explique encore à propos du pouvoir qu'il existe deux façons d'être : s'opposer ou devenir. La première façon est réaliste et enfantine, tout au moins dans la vision du pouvoir qu'elle met en œuvre, la deuxième est idéaliste et adulte. Ces deux façons d'être relèvent de la décision. Pour s'opposer comme pour mettre ses projets en œuvre, il faut décider. L'être est au cœur de la décision. En premier lieu, la décision a pour fonction de permettre à l'acteur d'agir. Le système agit à travers ses acteurs. Il faut donc que les acteurs aient l'impression (vraie ou fausse) d'être libres, de prendre des décisions. Ces décisions étant prises dans la perspective de projets, la décision est une façon d'être un devenant. En second lieu, la décision a pour fonction de permettre à l'acteur de supporter le monde. Si l'acteur n'opposait pas à chaque instant sa liberté de décider à la pesanteur des déterminismes, la vie ne lui serait pas agréable. La décision est une façon d'être en s'opposant au monde. On supporte plus facilement ce monde qui n'est pas tel qu'on le voudrait si on peut à chaque instant projeter sa liberté sur l'imperfection du monde, le redressant tout en distinguant ses scories.

2.4. Le modèle S.T.I³¹ de H.A. Simon

2.4.1. Rationalité substantive versus rationalité procédurale

Pour H.A.Simon, l'émergence de la décision comme domaine d'étude scientifique remonte aux années 1943 et 1948, moment où se créent trois courants de recherche parallèle.

- Les théories mathématiques de la décision centrées autour de la théorie des jeux, des théories normatives avec J.von Neumann, O.Morgenstern et Savage. Cette approche est surtout développée en économie mathématique [von Neumann *et al.* ; 1944][Savage, 1972][Yager, 1979] ;
- La cybernétique comme science de la communication et de la commande dans les systèmes naturels et artificiels créée par N.Wiener en 1948 et vite associée aux problématiques de la recherche opérationnelle ;
- H.A.Simon à travers sa thèse publiée en 1947 "A study of the decision making process in administrative organization".

³¹ Système de Traitement de l'Information



H.A.Simon remet alors en cause le modèle formel et mathématique de la décision encore adopté par la quasi-totalité des économistes classiques, des chercheurs opérationnels et des spécialistes de gestion de production. Cette vision part du postulat selon lequel choisir de façon rationnelle, c'est adopter la démarche déductive et analytique de la logique formelle des propositions. Face à un problème donné, il faut inventorier une liste complète des actions possibles, sélectionner par des procédures mathématiques la meilleure solution, c'est-à-dire celle qui satisfasse au mieux un critère d'utilité et enfin agir en appliquant cet optimum calculé. La référence idéale sera celle de l'abeille qui construit une cellule de cire selon une forme géométrique invariable : un dodécaèdre rhomboédrique dont l'angle de la base est de 136.4 degrés. Les calculs ne montrent-ils pas que cet angle correspond à la forme optimale qui minimise la quantité de cire nécessaire pour stocker un volume donné de miel ! La nature parvient donc à programmer en son sein une procédure totalement rationnelle, ou optimisatrice.

Toute notre cellule cartésienne de la décision est fondée sur ce principe "faire aussi bien que la nature" laquelle serait logique, rationnelle et déductive. Les recherches opérationnelles, la théorie des jeux, les calculs statistiques procèdent tous de cette démarche idéale. Cependant dans la réalité quotidienne, on utilise rarement de telles procédures.

La réflexion de Simon a consisté à remettre en cause cette démarche en proposant une conception alternative des processus de décision dans les organisations.

D'abord il faut remettre en cause l'idée selon laquelle la décision est une réponse précise à un problème donné, prédéfini. La décision est un processus où problème et réponse se construisent en même temps. Simon ne parle pas de la décision mais de "decision making process". La première phase de la décision consiste à identifier la nature de la question à traiter. Parfois la décision conduira à changer les données du problème.

Le second apport de Simon concerne la conception et l'évaluation des solutions alternatives possibles. L'exemple qu'il cite est celui des joueurs d'échecs. Aux échecs il est impossible à un cerveau humain, ni même à un ordinateur d'envisager toutes les combinaisons possibles. Les joueurs utilisent donc des heuristiques, stratégies habiles, raisonnées, et non pas une procédure algorithmique qui consisterait à passer en revue la liste complète des solutions.



Pour la plupart des problèmes de la vie quotidienne, nous mettons en œuvre de telles heuristiques, c'est-à-dire des raisonnements plausibles mais non certains (des inférences plutôt que des déductions). Parfois même, il arrive qu'une démarche de la pensée viole les principes de la logique déductive formelle et soit pourtant efficace. L'argumentation de fond de Simon est donc d'abandonner ce concept de stratégie optimale. Premièrement, il est très rare que nous soyons dans une telle situation avec un seul critère. Et à supposer que l'on soit dans une telle situation, on a rarement la capacité cognitive de traiter les milliards d'informations et de solutions possibles.

Simon a longtemps médité sur le concept de rationalité avant de proposer un diagnostic. Il a pendant très longtemps utilisé le concept de "Bound rationality" traduit habituellement par rationalité limitée et qui renvoie à l'idée d'une connaissance imparfaite, ou bornée que le sujet a de son environnement.

Depuis les années 70, Simon préfère opposer le concept de rationalité substantive, qui est le raisonnement formel, analytique et déductif à la rationalité procédurale qui correspond à la



façon dont l'être humain conduit fort correctement sa raison en reliant sans cesse ses intentions et ses perceptions du contexte dans lequel il raisonne.

Il y a une sorte de renversement. Ce n'est pas la rationalité procédurale qui est limitée mais plutôt la rationalité déductive. La rationalité déductive ou substantive ne correspond qu'à une petite partie des formes possibles du raisonnement humain.

Il s'agit d'élargir l'éventail des formes de raisonnement possibles sans se limiter au seul raisonnement déductif. L'étude de la rationalité procédurale ouvre un important champ d'études. Il est expliqué aujourd'hui par les théories de l'argumentation, la nouvelle rhétorique, ou encore certains courants de la psychologie cognitive. On n'est plus dans le cadre de la déduction formelle mais dans celui de la capacité de l'esprit à produire des solutions rusées, malicieuses pour résoudre les problèmes.

Avec l'essor actuel des sciences de la cognition se développent aujourd'hui de nouvelles réflexions épistémologiques des sciences de la complexité, jetant ainsi les bases des "Nouvelles sciences de la décision" [Simon, 1991].

2.4.2. Acquisition et traitement de l'information

Pour Simon, les modèles mathématiques utilisés en particulier en sciences économiques présentent une conception qui se caractérise par quatre traits :

- ignorance des phases d'intelligence et de conception, dont le résultat est supposé faire partie des croyances du décideur dès le départ (ensemble des possibles, forme du problème) ; la phase de sélection est jugée suffisante pour modéliser la décision ;
- ignorance de la phase de bilan, même pour un simple bouclage sur la phase de sélection, puisque celle-ci est seule modélisée, et «linéarité», en conséquence, du raisonnement ;
- recours à une modalité de rationalité forte puisque optimisant le critère (unique) retenu («utilité» pour le consommateur, profit pour le producteur) ;
- réduction de la complexité du système de pilotage par la sorte d'hypostase du décideur.

Une telle conception de la décision, même en se plaçant dans l'hypothèse d'avenir certain, semble aujourd'hui difficilement justifiable sauf dans des cas bien spécifiques. En effet, on peut faire valoir que ce modèle «traditionnel» de la décision individuelle peut être utilisé à des fins prescriptives dans les domaines purement techniques, et en particulier pour la gestion à court terme d'un parc d'équipements ou d'un chantier. Les phases d'intelligence et de conception sont alors souvent triviales, et les quatre hypothèses implicites relevées plus haut peuvent être admises au moins en première approximation. C'est sur ce créneau spécifique de la gestion de la production et de la gestion financière que fleurissent, surtout depuis les années cinquante, les nombreux modèles de la recherche opérationnelle et du calcul économique dont les apports ne sauraient être discutés.

Il en va tout autrement dans le cadre de la décision organisationnelle où le vocable «système de traitement de l'information» (S.T.I) permet de désigner commodément la lignée des modèles issus de la pensée de H.A.Simon. Cet ensemble de propositions se fonde sur quatre remarques critiques vis-à-vis des applications un peu hâtives du modèle de conception précédent.

- Les décisions ne sont pas toutes de même nature ni de même niveau. Certaines sont telles qu'un automate, correctement informé et programmé, est en mesure de les prendre aussi bien – voire de façon plus fine – que l'homme lui-même, dont la capacité purement combinatoire est singulièrement limitée. On les appellera décisions «programmables».



Mais, dès lors notamment que des ensembles humains ou des interfaces hommes-machines sont concernés, les décisions sont de l'ordre du «non-programmable». Gommer les phases d'intelligence et de conception du processus de décision est alors une réduction inacceptable.

- De façon plus générale encore, une conclusion identique s'impose lorsqu'on remarque que les ensembles de possibles ne sont jamais faciles à explorer. Or le meilleur général n'est sans doute pas celui qui réalise le dosage optimal entre aile droite et aile gauche sur un terrain donné, mais celui qui sait imaginer une possibilité stratégique de plus, permettant d'accroître sa palette de tactiques. Il en va de même du chef d'entreprise, du préfet, ou du couple élevant ses enfants. Les phases d'intelligence et de conception sont véritablement les phases clés du processus de décision.
- On pourrait se donner pour programme de recherche de tenter de formaliser la démarche décisionnelle à travers toutes les phases évoquées. Mais on buterait très vite sur des difficultés de calcul inextricables. Une autre voie consiste à essayer plutôt de représenter – et d'améliorer – la façon dont les hommes utilisent leurs capacités de raisonnement et de traitement des informations. Au lieu de chercher à désigner une décision ambitionnant d'être optimale, il est plus modeste dans les objectifs, mais peut-être plus efficace pour le résultat, de chercher à user d'une procédure de traitement de l'information et de raisonnement plus satisfaisante. La rationalité «limitée» ou «procédurale» vient ainsi se substituer à la rationalité optimisante et «substantive».
- Enfin, le fait qu'on se place dans le cadre d'une organisation comportant évidemment des services, des départements, etc., non totalement intégrés introduit une contradiction avec la nécessité, pour optimiser de façon significative, de traiter simultanément dans le temps et partout dans l'organisation l'ensemble des informations disponibles. Les problèmes de délégation et de coordination se heurtent, là encore, à d'inextricables difficultés. Aussi, dans le modèle S.T.I., les différentes phases de la décision ne se présentent-elles pas de façon linéaire, mais en boucles. De nombreuses itérations sont nécessaires, au vu de la faible capacité cognitive des hommes et de la complexité des problèmes de décision, avant qu'un terme puisse être apporté au processus de décision. Davantage encore, chacune des phases engendre des sous-problèmes qui, à leur tour, appellent des phases d'intelligence, de conception, de sélection et de bilan. Les phases sont ainsi des «engrenages d'engrenages»³².

Par ailleurs, le processus de décision est mis en œuvre, dans une organisation, par un système de décision, lui-même complexe et dont l'hypostase du décideur peut donner une représentation mutilante et être une source d'erreurs de décision si elle conduit à ignorer les difficultés de communication et de coordination.

Le modèle S.T.I insiste en définitive sur les aspects cognitifs de la décision, l'acquisition et le traitement de l'information apparaissant comme plus importants pour prendre une «bonne» décision que la recherche fine illusoire d'une décision en apparence «la meilleure». La notion d'organisation est centrale dans la justification de ce modèle.

Il y a encore une dizaine d'années, la mise en œuvre du modèle S.T.I de Simon aurait pu être résumée comme suit. Les techniques de mise en œuvre du modèle S.T.I sont principalement celles de l'informatique, et notamment des Systèmes Interactifs d'Aide à la Décision (SIAD). Ceux-ci constituent une interface entre un ordinateur et une ou plusieurs personnes (SIAD individuel, ou SIADI ; et SIAD de groupe ou SIADG). Sur la «machine», on implante soit un

³² Voir le schéma que nous proposons plus en aval pour une instantiation du modèle S.T.I de Simon aux systèmes de gestion dynamique de des connaissances (SGDC).



algorithme, correspondant à la phase de sélection, soit un système heuristique pouvant aider à la créativité (aide à la phase de conception) ou, de façon plus ambitieuse, utilisant une base de connaissances (système expert). Dans tous les cas de figure, le système implanté en machine est considéré comme n'«apportant» pas «la» solution, mais comme jouant le rôle d'amplificateur cognitif, mettant l'accent sur une ou plusieurs phases du processus de décision.

Dans ce qui suit nous montrons que les avancées techniques et technologiques des systèmes d'information de ces dernières années, couplées à un véritable accompagnement mathématique et informatique des processus cognitifs collectifs semblent définitivement donner raison à H.A.Simon et nous proposons une mise en œuvre pratique du modèle S.T.I pour la réalisation d'un SIADG basé sur le couplage d'un système de gestion dynamique des connaissances (**SGDC**) et d'un système d'évaluation multicritère.

2.5. Le couplage d'un SIADG avec un système de gestion des connaissances

Cette section propose une mise en œuvre pratique du modèle précédent avec les outils mathématiques et technologiques d'aujourd'hui.

Face à la complexité d'un monde incertain, le développement de systèmes techniques, organisationnels ou stratégiques maîtrisés dépend de plus en plus d'une meilleure utilisation par chaque acteur de ses capacités cognitives (savoir, savoir-faire et savoir-agir). Ceci implique que chacun d'eux, mais aussi les organisations dont ils sont membres, perfectionnent leur propre système d'observation et d'appréhension de la réalité.

La maîtrise de la complexité est affaire de bonne organisation collective. Elle devient ainsi l'émanation de savoirs multiples dans les domaines techniques, sociaux, culturels, organisationnels. Résultante complexe d'actions et de décisions individuelles et de construits collectifs, elle se développe au travers des multiples structures de communication reliant les hommes.

Par exemple, l'analyse de la production des faits scientifiques amène à s'intéresser aux controverses scientifiques et technologiques, qui fournissent une prise empirique particulièrement féconde pour étudier la mise en place des mécanismes de clôture du débat et les négociations socio-techniques à l'origine des faits scientifiques. Depuis quelques années, la prolifération des controverses publiques liées au développement scientifique et technologique a conduit à tourner l'analyse dans une autre direction. Par symétrie, il s'agit de comprendre les mécanismes qui favorisent la "déconstruction" de faits scientifiques pourtant considérés comme stables et indiscutables, notamment dans les domaines marqués par l'émergence de la notion de risque. D'où le développement d'études sur la mobilisation des non-spécialistes sur les questions scientifiques et l'apport des profanes dans les controverses technologiques...

Le résultat de telles réflexions permet de reposer la question classique du débat public et de la démocratie technique, conçue sous la forme trop limitée de la participation du collectif ou du public aux décisions stratégiques techniques. On est alors conduit à s'intéresser aux nouvelles procédures, structures, institutions manageriales ou politiques qui permettent d'organiser la confrontation entre spécialistes et profanes, en mêlant des questions de programmes de recherche et de démocratie procédurale. Plutôt que d'envisager ces procédures de mise en débat public comme de simples instruments de légitimation politique ou manageriale, les recherches doivent en montrer la «performativité», et tenter de déterminer en quoi elles sont ou non capables de modifier la dynamique de production des connaissances scientifiques et



techniques. Ces situations controversées peuvent être analysées à travers des « forums hybrides » qu’elles instituent. Il s’agit d’espaces de discussion et de confrontation dans lesquels s’expriment des scientifiques et des experts, mais également des groupes estimant avoir des intérêts en jeu et s’efforçant de participer aux débats, de peser sur les décisions, d’opposer des arguments à d’autres arguments, de monter des alliances, d’en appeler à l’opinion publique ou aux médias. Comment sont gérés de tels espaces ? Comment se déroulent les controverses qui y prennent place ? Quelles procédures sont privilégiées dans la gestion de ces controverses ?

A ce titre, l’impact des Technologies de l’Information et de la Communication devient un facteur essentiel du développement de nos modes d’organisation, voire même de nos sociétés. La complexité des systèmes quelle qu’en soit la nature dans lesquels l’homme est impliqué aujourd’hui conduit à l’émergence de systèmes dynamiques de traitement de l’information de plus en plus complexes et incontournables. Les thèmes de travail à cet égard tant dans les Sciences de l’Information que dans les Sciences Humaines doivent donc à la fois relever de l’étude des systèmes complexes et des risques associés à l’usage de ces technologies de l’information. L’objectif est double : élaborer des modèles dont la qualité principale doit être leur compatibilité avec les modes cognitifs utilisés par l’homme confronté à une situation complexe, et instrumenter la relation que l’homme ou plutôt un collectif organisé entretient avec son système d’information dans l’action et la décision.

2.5.1. Connaissances actionnables

Nous avons abordé dans la section *Intelligence collective humaine* la problématique de la gestion des connaissances scientifiques et techniques relatives aux organisations. Nous avons proposé des solutions quant au déploiement de solutions efficaces pour les SGDCs et introduit la notion de connaissance actionnable CA (connaissance utile à l’action).

Nous rappelons ici les éléments essentiels sur les CAs quant au processus de décision. Un corpus de CAs recouvre un domaine de connaissances qui est aussi un domaine d’action autrement dit de décision.

Une CA doit peut donc être définie tout à la fois comme :

- une donnée informative jugée utile,
- un savoir élémentaire interprété,
- une trace de raisonnement,
- une unité de sens partageable et réutilisable.

La fonction de doter le collectif d’une mémoire commune est assurée par le SGDC qui permet de gérer et de contrôler l’évolution du corpus de connaissances produit par le collectif sur le problème à résoudre. On construit à cette fin des structures dynamiques de CAs, qui prennent souvent la forme de graphes (réseaux ou arbres...). Les notions de mémoire partagée et de processus cognitif d’apprentissage collectif acquièrent alors une dimension dynamique essentielle au suivi d’un corpus de connaissances en expansion : le graphe de CAs favorise une connaissance évolutive et réactive en offrant d’adoindre à une CA un complément, une contradiction, une opposition ou encore une simple mise à jour sous la forme d’une nouvelle CA. L’échange de points de vue « métier », de systèmes de valeurs, d’intérêts ou de cultures, trouve ainsi une structure dynamique qui favorise la réactivité, la délibération et



l'argumentation et qui in fine peut modifier la dynamique du processus d'apprentissage collectif.

L'apprentissage d'un savoir et d'une mémoire collectifs est le premier des processus cognitifs pour lequel le SGDC apporte des éléments de solution. Dans ce qui suit, il est montré en quoi il constitue également un support naturel pour les deux autres processus cognitifs relevant d'une intelligence collective que sont la décision et l'argumentation.

Le mythe du dirigeant omniscient est clairement obsolète dans cette approche de la décision en organisation. Néanmoins, le décideur peut se poser en arbitre régulateur et utiliser cette approche pour lui permettre de mieux gérer les conflits internes et d'argumenter sa position lorsqu'il décide de favoriser certaines solutions ou certains candidats en concurrence. Le risque qu'il ne favorise uniquement les solutions ou les candidats qui sont compatibles avec ses propres préférences et ses propres ambitions ou que la décision soit purement technocratique est réduit.

Contrairement aux objectifs de la Recherche Opérationnelle et en accord avec le modèle de H.A.Simon, cette approche cognitive consiste à essayer plutôt de représenter—et d'améliorer—la façon dont les hommes utilisent leurs capacités de raisonnement et de traitement des informations. On ne cherche pas une solution nécessairement optimale mais argumentable, plutôt qu'un compromis, on cherche un consensus. L'utilisation des STIC donne une certaine transparence à la décision : en proposant un cadre et une instrumentation explicites pour les processus cognitifs d'apprentissage puis d'argumentation, de délibération, de concertation, de sélection, elle permet (tant dans l'aspect gestion des connaissances que dans la décision elle-même) de favoriser l'échange hiérarchique et transversal de données, d'informations, de connaissances et de décisions à travers une modélisation systémique et mathématique des liens de subordination et de coordination qui lient les acteurs d'un processus de décision organisationnel.

2.5.2. Evaluation multicritère

Sur la base de la distribution des connaissances du projet et du partage d'une représentation partagée [Teulier-Bourgine, 1997] des objectifs de celui-ci, le collectif peut alors être doté d'un SIADG avec des fonctionnalités propres au couplage avec le SGDC.

En effet, les principes de partage des CAs et de réactivité aux opinions exprimées par des acteurs de métier, de culture et de fonction très différentes orientent le SIADG vers des outils de décision multicritère (voir § 1.5.3) [Roy, 1985 ; Roy *et al.*, 1993 ; Pomerol *et al.*, 1993]. L'orientation cognitive de ce genre d'approches signifie que l'on doit porter autant d'attention à la sémantique des opérateurs mathématiques d'agrégation ou de fusion multicritère qu'à leurs propriétés mathématiques [Koenig, 1990 ; Dubois, 1983 ; Dubois et Prades, 1985]. En effet, lorsque l'évaluation globale d'un objectif est complexe, multipoints de vue, il est nécessaire de décomposer l'objectif à atteindre en structurant l'ensemble des critères d'évaluation des solutions candidates envisagées par des relations hiérarchiques. Les relations hiérarchiques expriment des liens de subordination c'est-à-dire qu'à partir de scores ou évaluations partielles d'un niveau donné, on est capable d'évaluer le score de niveau supérieur. Il peut également exister des liens de dépendance entre les critères d'un même niveau.

Les CAs en tant que données informatives jugées utiles pour le projet et savoirs élémentaires interprétés sont alors vues comme des jugements de valeur sur l'information produite dans le SGDC dans le cadre finalisé du projet, c'est-à-dire de l'action. Ces jugements de valeurs sont



rapportés à l'évaluation partielle d'une solution pour un critère donné lorsque la CA est cartographiée dans la grille d'évaluation solutions-critères (Figure 58).

L'agrégation multicritère consiste donc à évaluer globalement les différents candidats ou solutions étudiées dans le projet à partir de la fusion des appréciations partielles portées par les CAs [Montmain *et al.*, 2002a]. Pour une stratégie d'agrégation définie, cela signifie que la solution retenue sera le résultat des évaluations partielles et négociées des différents acteurs impliqués dans la décision et que ce choix repose uniquement sur leur production de CAs : toute décision est prise sur la base des connaissances partagées par le collectif projet à un instant donné et gérées par le SGDC.

Une CA est donc un jugement de valeur ou une intensité de préférence exprimé sous la forme d'un vote par son rédacteur vis-à-vis d'une solution candidate et d'un critère d'évaluation donné. L'évaluation multicritère des solutions nécessite ensuite plusieurs niveaux d'agrégation :

- Agrégation partielle des CAs attribuées à une case donnée de la grille d'évaluation pour évaluer la solution vis-à-vis d'un critère donné ;
- Agrégation globale sur l'ensemble des scores partiels sur une colonne de la grille au regard d'une stratégie d'évaluation définie au préalable à travers l'opérateur de fusion sur les critères ;
- Agrégation implicite des jugements de valeur portés par les CAs dans le temps en respectant la dynamique du processus décisionnel.

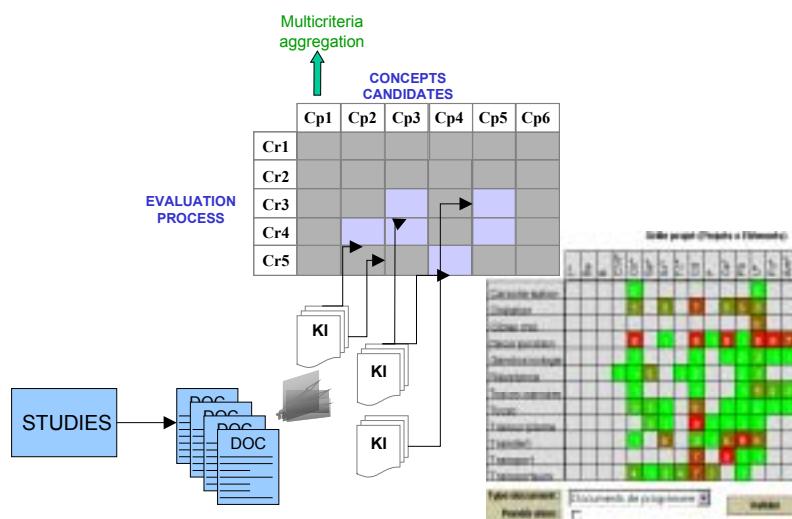


Figure 58 : Evaluation multicritère par les CAs

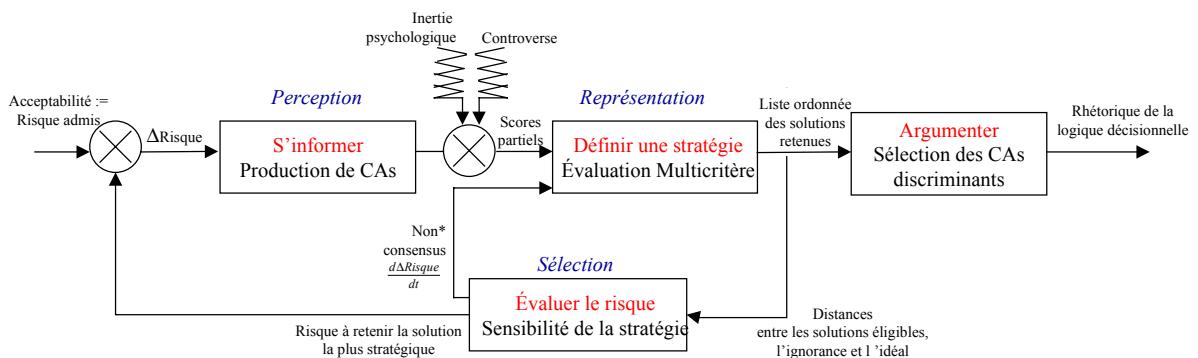
2.5.3. Acceptabilité d'une décision collective : argumentation et gestion du risque

Nous avons souligné en introduction de ce document que si la recherche opérationnelle repose sur une hypothèse d'ordre, l'approche cognitive de la décision en organisation est régie selon une hypothèse d'équivocité : à une situation, il est possible d'associer un grand nombre d'interprétations et de stratégies décisionnelles satisfaisantes, la notion de solution optimale n'a pas nécessairement de sens. Un point important de la décision de groupe concerne donc la nécessité de parvenir à un consensus, aussi est-il fondamental de doter un SIAD collectif de fonctionnalités d'argumentation des choix proposés et d'évaluation des risques associés pour favoriser l'acceptabilité des décisions.



Ainsi, la capacité du système à identifier les critères les plus déterminants et/ou les plus discriminants dans la phase de sélection des solutions est d'autant plus pertinente si l'on en revient à la fonctionnalité initiale d'un SGDC. Les SGDCs développés dans le cadre du management de projet sont en premier lieu des outils de gestion des connaissances : ils sont conçus à la fois sur la base de la capitalisation des savoirs et du savoir faire et d'autre part sur le partage dynamique des connaissances actionnables [Penalva *et al.*, 2002 a et b]. Ainsi, si dans les processus d'agrégation multicritère résumés précédemment, les CAs sont assimilées aux scores partiels qu'ils portent, il ne faut pas oublier qu'une CA est d'abord l'expression en langage naturel de l'interprétation finalisée par son rédacteur d'une information dans le cadre de son projet, c'est-à-dire un élément d'argumentation ou de rhétorique. Identifier les critères les plus déterminants ou les plus discriminants dans le choix d'une solution revient alors à dire que le système est capable d'extraire automatiquement de la base de connaissance du SGDC les éléments décisionnels (c'est-à-dire la partie « rhétorique » des CAs) à faire valoir pour une argumentation automatisée de la décision. Le couplage du système d'aide à la décision à un système de gestion des connaissances permet de s'orienter vers la génération automatique de rapports d'argumentation pour le décideur final [Akharraz *et al.*, 2002 a et b]. L'argumentation est le premier point à faire valoir pour faire accepter une décision consensuelle par un collectif.

Le second élément que le décideur final doit gérer est donc le risque associé à la décision. Pour introduire cette notion de risque dans le modèle S.T.I de la décision que nous proposons, il faut revenir sur l'idée que, dans les approches cognitives de la décision, toute décision raisonnée doit reposer sur une analyse de la situation qui permet d'expliciter et de caractériser les phases de perception, de représentation et de sélection du processus décisionnel, ces trois phases de la décision ne se présentant pas de façon linéaire, mais en boucles (loops). Pour exprimer cette idée de boucle d'itération dans le processus de décision, nous allons introduire la notion de pilotage ou contrôle de projet par gestion des risques associés. Nous ne détaillerons pas l'aspect mathématique de cette partie, nous souhaitons juste illustrer qu'avec les outils technologiques et mathématiques décrits dans ce document, le modèle S.T.I du processus décisionnel et ses « loops » proposés par H.A.Simon peut être instrumenté de façon pratique pour une décision d'organisation. Le schéma (Figure 59) que nous proposons ci-dessous est une instantiation possible des boucles du processus décisionnel vu par les sciences cognitives dans notre système.



* Si le risque associé à la décision à l'instant t n'évolue plus et reste trop fort, alors on ne peut obtenir de consensus et il faut revoir la stratégie.

Figure 59 : Pilotage par le risque

Pour un corpus de CAs donné et une stratégie décisionnelle modélisée par un opérateur d'agrégation fixée, on obtient une liste ordonnée des sélections potentielles pour laquelle on peut calculer le risque d'élire la solution retenue en fonction de la distance définie par la stratégie appliquée entre le candidat retenu et les solutions concurrentes. Si le risque calculé



n'évolue plus et reste en deçà de la consigne fixée (le seuil d'acceptabilité), indépendamment de nouveaux flux d'informations, la stratégie d'agrégation peut être révisée (les paramètres la déterminant peuvent être réajustés, les critères d'évaluation réexaminés) mais en premier lieu, il s'agit de s'assurer qu'aucune information supplémentaire ne pourra modifier la sélection (boucle de « régulation » principale). En fait, s'il est possible de jouer sur la monotonie de l'entropie de la base de connaissances par l'adduction de CAs, un processus de régulation (les perturbations correspondant à l'inertie psychologique ou à la controverse) est mis en jeu : on s'informe plus précisément sur tel ou tel point du projet (lancer ou accélérer des études dans le cadre d'un projet) pour décider avec un risque inférieur au seuil d'acceptabilité. Si l'on est en dessous de ce seuil, la situation est décidable, aucune étude supplémentaire n'est requise pour proposer un choix qui ne pourra être « stratégiquement » désavoué (arrêt ou ralentissement des études devenues inutiles compte tenu de l'état des connaissances suffisant).

Ce schéma simpliste qui fait intervenir les trois phases d'un processus décisionnel au sens cognitif du terme est uniquement destiné à dresser le parallèle entre le contrôle de processus technique et la gestion d'une décision collective (ou le pilotage d'un projet) par le contrôle du risque encouru. Le risque a dans ce modèle une définition particulière, il n'est pas la classique fonction d'une gravité et d'une fréquence d'occurrence, mais est associé à la distance séparant deux candidats concurrents : lorsque l'élu distance de beaucoup ses «adversaires», cela laisse entendre qu'il faudrait une très forte (donc peu probable) variation d'entropie de la base de connaissances du SGDC pour que sa « supériorité » soit remise en cause ; par contre, s'il devance de peu ses adversaires, l'introduction de nouvelles CAs sur les bons items de la grille d'évaluation, risque d'affecter le classement.

Il est clair que la mathématique mise en jeu dans ce processus n'a rien à voir avec la modélisation par équations différentielles de l'automatique classique pour le contrôle de processus de production industriels : les termes régulation, perturbation et contrôle ne sont employés que par analogie avec le contrôle des systèmes techniques de l'automatique. Le but de cette représentation par analogie est de mettre en évidence la nécessité d'un « rebouclage » sur les phases d'apprentissage et de décision collectives, et de montrer la pertinence du couplage d'un SDGC et d'un SIAD pour résoudre ce problème.

2.6. Les modèles graphiques probabilistiques pour la représentation des croyances et pour la décision stratégique

2.6.1. Introduction : visualisation, intuition et raisonnement

Les modèles graphiques probabilistiques permettent d'exprimer un problème de croyance ou de décision stratégique de manière graphique, claire et intuitive afin de suggérer des schémas de décision à des décideurs en fonction des connaissances dont on dispose sur l'état de certaines variables.

De plus, s'appuyant sur des modèles probabilistiques éprouvés, ils permettent d'inférer de manière rigoureuse des diagnostics, des décisions optimales et leurs effets. La rigueur des inférences ne masque pas cependant le caractère subjectif et intuitif de la démarche qui autorise le dialogue et la mise au point.

Enfin, la capacité de raisonner des causes vers les effets et des effets vers les causes que permettent les modèles graphiques probabilistiques autorisent une mise à jour des croyances et des recommandations au vu d'observations ou de faits révélés.



2.6.2. Les modèles graphiques probabilistes de diagnostic et décision

Les réseaux bayésiens.

Un réseau Bayésien est un graphe paramétré dont les nœuds sont des variables qui représentent différentes connaissances ou croyances et des arcs qui représentent des relations de dépendances entre ces nœuds, le plus souvent des relations de causalité. Les nœuds en amont d'un nœud sont appelés ses ‘nœuds parents’. Les variables peuvent prendre leurs valeurs dans un ensemble discret ou continu. Le graphe est acyclique et orienté (DAG : Directed Acyclic Graph). L’absence d’un arc entre deux nœuds signifie l’indépendance entre les deux variables (les deux variables n’ont pas de relation causale entre elles).

Des poids sont donnés aux liaisons. Ils représentent les probabilités conditionnelles entre les variables, c'est-à-dire la probabilité d'observer l'effet dans ses différents états considérant la cause (respectivement les causes) dans ses différents états (respectivement combinaisons d'états).

Exemple d'un réseau Bayésien : le lecteur trouvera un exemple très simple d'un réseau Bayésien dans la figure 60. Il s'appuie sur le problème ‘rail-route’ présenté dans l'exposé des besoins par METL.

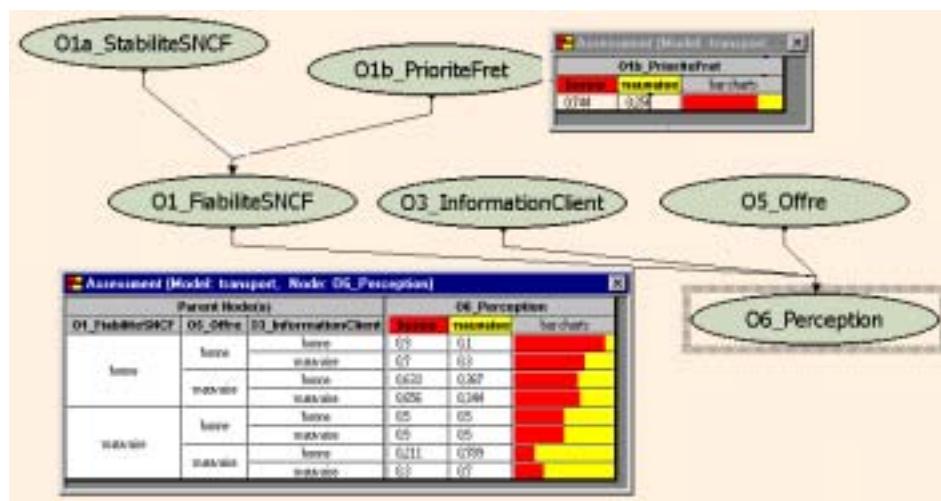


Figure 60 : Exemple simple d'un réseau bayésien (implémenté sur l'outil MSBNx de Microsoft)

Chaque nœud du réseau représente l'une des variables possibles qui influencent la décision de l'affréteur (nous avons omis les variables O2 et O4 pour ne pas alourdir la présentation). En effet, nous avons analysé le problème dans un premier temps comme un problème de perception (croyance) de la qualité de service du combiné rail-route pour un décideur qui a un problème de transport. Nous avons considéré chaque variable comme pouvant prendre l'un des deux états ‘mauvais’ (le décideur a une mauvaise image du -ou une mauvaise information sur le- combiné rail-route), ou ‘bon’. Chaque flèche représente l'influence d'une variable sur son enfant. Chaque nœud est ensuite nanti soit d'une probabilité (ou dans notre cas d'une croyance assimilée à une probabilité) a priori (AP ou ‘priors’), ou d'une table de probabilités conditionnelles (PC) qui conditionne la probabilité des états de la variable par la combinaison des états des variables parentes dans le réseau. Nous n'avons visualisé que deux tableaux pour ne pas alourdir le schéma. Dans l'exemple, nous avons mis des valeurs de probabilité selon l'interprétation que nous avons faite du texte en tenant compte de la remarque “ordre de gravité décroissante”. La simple construction du réseau et des tables de probabilités fait



émerger des interrogations sur le modèle qui peuvent conduire à des vérifications (études statistiques) et ensuite à la construction de diagrammes d'influence (décisions qui peuvent influer sur le modèle).

Discussion sur l'exemple :

La modélisation de cet exemple oblige à expliciter des relations et à réviser le modèle de croyance initial. Par exemple nous avons mis 0.5 pour les valeurs ‘bonne’ et ‘mauvaise’ des Probabilités Conditionnelles (PC) de la perception de l'affréteur dans le cas où toutes les autres variables sauf la fiabilité sont perçues bonnes. Par contre nous avons mis à 0.05 la probabilité a priori de l'instabilité sociale de la SNCF. Autrement dit, nous considérons qu'un affréteur perçoit de manière beaucoup plus sévère que les faits la stabilité de la SNCF. Alors on pourrait considérer qu'une action de communication sur la stabilité pourrait corriger cette impression abusive.

Les Diagrammes d'Influence (DI)

Les DI sont des RB qui présentent des nœuds particuliers pour introduire la possibilité de prendre des décisions sur certaines variables et d'en observer les conséquences sur d'autres par propagation. De plus certains nœuds représentent l'utilité que le décideur peut attacher à la valeur d'une variable (gain économique, facteur de satisfaction, ...).

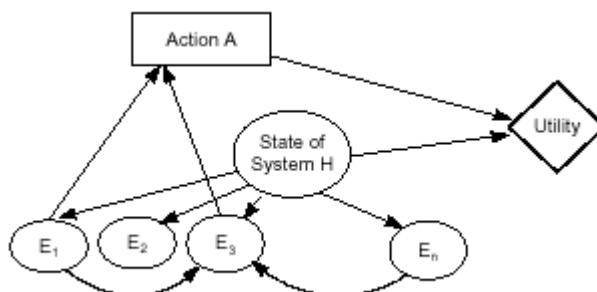


Figure 61 : Exemple d'un diagramme d'influence tiré de [Horvitz *et al.*, 1997]

Le lecteur comprendra intuitivement l'intérêt des diagrammes d'influence pour éclairer la prise de décision au travers de ses conséquences soit en terme de coût, soit en terme de gain, quelle qu'en soit sa nature. On pourra se reporter à la bibliographie pour considérer les champs d'applications.

Réseaux bayésiens (RB) et Diagrammes d'Influence dynamiques

L'exemple de la figure 61 présente aussi la particularité d'être un diagramme d'influence dynamique. Une action (ici l'actionA) a pour conséquence de modifier l'état d'un système, sachant que cet état peut aussi influencer l'action. Cette approche permet de suivre l'évolution d'un système au fil des décisions qui sont prises, ou de l'évolution de différentes variables.

L'intérêt des réseaux bayésiens (RB) et des Diagrammes d'Influence(DI)

La modélisation d'une croyance ou d'un espace de facteurs influant une décision selon un modèle RB ou DI présente de multiples intérêts.



Intérêts des RB ou des DI

Visualisation et intuition : Un réseau bayésien ou un diagramme d'influence est un modèle graphique qui offre des représentations visuelles intuitives des différents paramètres entrant dans une décision.

Modularité : Expression d'un problème de relation 'cause-effets' par parties. Ceci permet l'expression modulaire du problème.

Inférence (prognostic, diagnostic, inférence mixte) : L'application de théorèmes bien connus de la théorie des probabilités (théorèmes de Bayes...) permet d'inférer des 'croyances' ou des recommandations à partir de la connaissance dont on dispose.

Modèle dynamique : Actualisation de la connaissance sur les nœuds au fur et à mesure que des 'évidences' surviennent, c'est-à-dire de la connaissance certaine sur le réel. Les croyances peuvent donc être révisées au fur et à mesure de l'arrivée de nouvelles informations. L'actualisation peut se faire dans le temps.

Décision stratégique : Des fonctions d'utilité peuvent être associées à certains nœuds pour suggérer des actions en situation d'information incomplète.

Hétérogénéité des connaissances : les variables d'un même réseau peuvent être de nature très diverses, certaines continues d'autres discontinues.

Prise en compte du subjectif et de l'objectif : les probabilités a priori et les probabilités conditionnelles peuvent faire l'objet de connaissances objectives (statistiques), ou à défaut de connaissances subjectives (avis d'experts).

2.6.3. Applications et bibliographie des RB et des DI

Applications

Les Réseaux Bayisiens et les Diagrammes d'Influence ont été largement utilisés et connaissent un intérêt croissant étant donné les avancées tant théoriques (modélisation, apprentissage, ...) que pratiques (outils en ligne) récemment observées. Ils sont largement utilisés dans des domaines où prévaut la décision automatique ou la décision assistée dans un environnement incertain et éventuellement évolutif :

- La décision stratégique : affectation de ressources, déclenchement d'opérations, etc. (par exemple : [Noh *et al.*, 1997], [Das, 1999])
- La modélisation d'un utilisateur : d'un élève dans l'enseignement, d'un abonné de e-news, d'un receveur de courriels, ou de programmes de e-télévision (par exemple : [Sahami *et al.* 1998])
- La prise d'initiative d'une machine dans un environnement incertain [Horvitz *et al.*, 97] [Horvitz *et al.*, 1998] [Horvitz, 1999]
- Le diagnostic : médical, maintenance, etc. (par exemple : [Stangl *et al.*, 2000], [Pearl 2001])
- L'interprétation du langage naturel (par exemple : [Eizirik *et al.*, 1993])
- L'apprentissage : découverte d'un réseau d'interactions génétiques à partir de transcriptomes (expressions de gènes sur des puces ADN), (par exemple : [Segal *et al.* 2001])



- Etc.

Nous renvoyons le lecteur au paragraphe suivant pour une brève analyse de certains types d'application.

Bibliographie

Ouvrages ou articles introductifs :

[Pearl ; 1988] : cet ouvrage de Judea Pearl (Université de Californie-Los Angeles) est notamment considéré comme l'ouvrage fondateur du domaine des réseaux Bayésiens, même si les fondements théoriques sont antérieurs. Ceux-ci renvoient en effet au fameux théorème du Révérend Thomas Bayes (1701-1761) : $P(A|B,c) = P(B|A, c) * P(A|c) / P(B,c)$, où $P(B|A, c)$ signifie la probabilité d'occurrence de l'événement B connaissant l'événement A et le contexte c. L'ouvrage de Pearl donne un modèle d'inférence dans les réseaux Bayésiens de type arbres, des théorèmes et des algorithmes de transformation de réseaux en arbres, et des méthodes d'apprentissage de structures et de valeurs de probabilités de réseaux à partir de données statistiques.

Malgré ses qualités pédagogiques, l'ouvrage (en anglais) reste difficile d'accès pour les non initiés. Mais il pose rigoureusement les fondements du domaine.

[Becker et Naïm, 1999] : Cet ouvrage a le double mérite d'être en Français et de faire un point bien renseigné du domaine. Par contre, il est de construction hétérogène, les chapitres passant de la vulgarisation à la formalisation poussée avec des raccourcis de démonstration parfois gênants. Les chapitres de vulgarisation sont une bonne introduction en Français au domaine.

[Gonzales et Wuillemin, 1998] : Bien qu'il ne fasse pas référence, nous aimons cet article qui résume la matière en quelques pages et en français. De plus, les RB sont appliqués ici à la modélisation d'utilisateurs, domaine d'un grand intérêt. Les outils cités sont pour certains déjà dépassés ou inaccessibles.

[Murphy, 2001a] Ce tutoriel en anglais en ligne est d'excellente facture, mais d'un abord pas toujours facile.

[Charniak, 1991] Cet article disponible en ligne est une bonne vulgarisation, écrit par l'un des pères fondateurs de l'intelligence artificielle.

[Pearl, 2001] Plusieurs articles ou ouvrages récents de Pearl reviennent aux sources de la relation de causalité. Dans quels cas l'observation d'une corrélation entre deux ou plusieurs variables peut-elle être assimilée à une relation de causalité ? L'interrogation est fondamentale puisqu'elle questionne la légitimité d'une action sur une variable pour contrôler les autres. Le domaine médical est privilégié ici. Le développement de la génomique est aussi fortement concerné par cette interrogation avec en particulier l'interprétation des puces à ADN ou à protéines. Les modèles de décision ont tout à gagner avec un tel éclairage.

Autres articles ou ouvrages (en relation avec l'étude):

Ils portent principalement sur les méthodes d'apprentissage de RB, et les applications des RB et des DI dans différents domaines. Nous reprenons ici quelques domaines d'application qui



nous semblent rapprochés du sujet de notre étude, à savoir la décision dans les systèmes de transport.

C3I : [Das, 1999] Détaille les applications possibles des RB et des DI dans les problèmes de décision stratégique de type C3I (Command, Control, Communications and Intelligence). Ce rapport technique présente l'avantage non seulement de situer les apports des RB et des DI dans un environnement de décision incertain, mais encore détaille un exemple précis en utilisant l'outil Netica (voir ci-dessous les outils). Certains systèmes de transport peuvent être considérés comme des C3I, plus particulièrement en situation de crise. Nous avons eu l'occasion de travailler sur des situations de crise comme la marée blanche (région bloquée par une soudaine chute de neige) pour laquelle il convient d'affecter des unités d'intervention en fonction des accès disponibles et des urgences. Il nous apparaît qu'une approche de type RB et DI peut alors s'envisager.

Initiative de la machine : [Horvitz et Seiver, 1997] [Horvitz et al., 1998] [Horvitz, 1999] l'équipe d'Horvitz au groupe Adaptive Systems & Interaction de Microsoft Research a proposé plusieurs solutions pour la prise de décision d'une machine en environnement incertain à l'aide de RB. Ces projets ont fait l'objet d'applications réellement opérationnelles. Il est à noter que l'application Office Assistant intégrée dans les outils Microsoft (le trombone farceur de Word ou de Excel) est en fait une mauvaise mise en œuvre opérationnelle des modèles d'Horvitz puisque les ingénieurs de Microsoft, emportés par l'enthousiasme de l'idée, ont retiré toute la partie de décision probabilistique basée sur les RB, à savoir le cœur du modèle. En conséquence, l'outil est plus souvent perçu comme une gène que comme une aide par les utilisateurs en raison de son caractère intrusif.

Ce cas d'utilisation des réseaux bayésiens est très intéressant dans le contexte d'analyse des nouveaux systèmes d'IHM tant dans les véhicules (voitures), que pour les superviseurs de trafic [Pauzie, 2001]. Il est en effet possible d'envisager des systèmes à initiatives d'information pour gérer les situations où l'opérateur (le conducteur, la personne en charge d'une supervision de trafic), ne dispose ni du temps, ni de la compétence pour à la fois traiter l'information et prendre des décisions. Nous renvoyons sur ce thème le lecteur à nos travaux sur les interfaces adaptatives [Vaudry et Crampes, 2001a,b,c].

Représentation d'un acteur : Le modèle d'Horvitz ci-dessus, et plus particulièrement les modèles présentés dans les articles précédents de cet auteur, intègrent un modèle de représentation d'un état mental d'un individu (dans son cas, un utilisateur). Ce type d'application des RB ou des DI est très prometteur et est abondamment illustré par des applications comme par exemple dans [Murray, 1999]. Nous avons utilisé une telle approche dans notre exemple de la figure 60. Nous voyons l'utilisation des RB et des DI très pertinente pour modéliser ainsi l'état mental de différents acteurs qui interviennent dans une prise de décision. *Les RB et les DI éclairent ici de manière très pertinente une approche probabilistique de la théorie des jeux*

Décision politique : [Cain, 2001] présente une application liée à la modélisation de la décision en matière d'environnement. Cette étude est directement une source d'inspiration pour le domaine des transports.



Documentation et Outils

Plusieurs outils sont maintenant disponibles et leur documentation a valeur de tutoriels et d'exemples d'applications. Nous citerons en particulier la documentation qui accompagne MSBNx de Microsoft [Kadie *et al.*, 2001] et celle qui accompagne le BN Toolkit de Matlab [Murphy, 2001b].

Plusieurs documents ou références à des documents sont accessibles à partir du site de Netica
<http://www.norsys.com/resources.htm>.

[<http://www.hugin.com/cases/>] l'autre outil commercial largement distribué propose nombre de cas d'études.

2.6.4. Analyse critique des représentations graphiques probabilistes

Nous présentons les différentes critiques et les réponses qu'il est possible de leur apporter. Nous proposons certaines de ces réponses en rapport avec nos propres travaux sur la visualisation dynamique de graphes de connaissances.

- 1) Problème : Expression graphique. Un RB ou un DI est reconnu très intéressant par la représentation graphique qu'il autorise. Or celle-ci est fastidieuse à effectuer et demande des révisions constantes.

Réponse : des outils de dessin de RB et de DI sont apparus récemment (voir par exemple MSBNx de Microsoft). La critique n'est plus de mise sauf sur deux points : l'organisation du réseau, et la gestion de sa complexité. Nous apportons des réponses complémentaires sur ces deux points ci dessous avec nos travaux dans le domaine de la représentation visuelle de la connaissance auto-adaptative.

- 2) Problème : Complexité du calcul des inférences. Les inférences, au vu d'évidence ou de décision sont longues à calculer

Réponse : cet obstacle est levé avec les outils déjà cités. Les inférences sont calculées automatiquement.

- 3) Problème : Difficultés à évaluer les probabilités a priori, les probabilités conditionnelles, et les utilités (pour les DI)

Réponse 1: *Ceci est effectivement le point délicat.* Cependant il en est ainsi de tout modèle à dimension quantitative. Les RB et les DI présentent cependant l'avantage ici d'accepter soit des données objectives (études statistiques), soit subjectives (voir à ce sujet l'argumentation dans [Pearl, 1988]).

Avec des valeurs subjectives, le modèle est représentatif d'un raisonnement humain à partir de croyances, ce qui est le but recherché.

Réponse 2 : Dans les situations où l'on dispose de données statistiques, il est possible de mettre en œuvre des technologies d'apprentissage. Cette approche nous semble peut applicable dans le cas de la décision dans le domaine du transport. Nous voyons ici plutôt des modèles subjectifs.

- 4) Problème : Taille du réseau. Le réseau peut être très complexe

Réponse 1: plusieurs théorèmes liés à la séparation entre variables (d-separation) permettent de limiter la taille du réseau.

Réponse 2 : Nos travaux portent sur une représentation partielle du réseau selon la pertinence des influences (seuil d'influence pertinent, seuil d'utilité pertinent).



- 5) Problème : Un décideur peut difficilement lire un RB ou un DI
Réponse : nous influençons un réseau par un profil d'utilisateur. Ceci nous permet des représentations multi points de vue et personnalisées.
- 6) Problème : Interprétation du réseau et de ses inférences : il n'est pas facile de comprendre ce que représente un réseau et ses inférences (voir l'exemple ci-dessus)
Réponse : nous proposons des rattachements sémantiques de documents et d'illustrations aux noeuds du réseau selon un principe de 'cartes topiques'. Nous avons pu observer comment le rattachement de certaines images ou symboles à certains noeuds rend le réseau tout à fait parlant.
- 7) Problème : Un réseau présente le raisonnement d'un seul agent
Réponse : en fait, ce type de modèle est justement très intéressant pour représenter le raisonnement de plusieurs agents et de leur décision dans le temps (voir par exemple [Noh et Gmytrasiewicz, 1997])

2.6.5. Facilitations de l'usage des modèles graphiques probabilistiques par interfaces adaptatives

Nous avons présenté dans la section précédente divers problèmes qui peuvent être des obstacles à l'utilisation des RB. En l'état, les RB sont des outils graphiques de spécialistes avec lesquels les décideurs ne peuvent pas dialoguer.

On peut reprocher à ces outils leur manque d'expressivité au niveau des noeuds (absence de symbolique représentative des spécificités des variables), la nécessité de les organiser à la main, leur non variation en fonction de l'évolution du contexte (disposition, présence/absence, taille,...), l'impossibilité de les décorer avec des informations complémentaires (hyperliens, annotations diverses, selon le principe des 'topic maps'), et la non possibilité de représenter divers points de vue.

Nos travaux de recherche sur les interfaces adaptatives s'intéressent dans ce domaine à simplifier, et à rendre plus visibles et manipulables différents modèles de graphes. Les outils que nous développons sont particulièrement adaptés à la visualisation des RB. En particulier, un décideur doit pouvoir explorer une décision, en explorer divers points de vue et en voir (en particulier par des effets visuels) tout de suite les effets possibles selon leur probabilité.

Par exemple, un graphe peut ne présenter que les variables pertinentes à un instant présent vu leur effet sur les variables aval. Le contexte peut amener une révision de la pertinence des variables étant donné l'observation d'évidences ou la prise de certaines décisions (cas des diagrammes d'influence). A la révision de la pertinence correspond une évolution du graphe.

Il est à noter que l'interfaçage des outils RB comme MSBNx ou Matlab peut s'effectuer sans difficulté avec des outils d'interfaçage adaptatif comme ceux que nous développons via des fichiers au format XML. Aussi est-il possible de composer une outillage de manière flexible, la vue (interface adaptative) étant décorrélée de son exploitation par le modèle Bayésien.

2.6.6. Conclusion

Les avantages

- 1) Les Réseaux Bayésiens et les Diagrammes d'Influences sont à l'heure présente sans doute une des réponses les plus pertinentes en matière de représentation graphique de la connaissance et de modèle de prise de décision en environnement incertain.



- 2) Les outils : La complexité des calculs qui avait jusqu'alors freiné la diffusion des RB et des DI est maintenant prise en charge par des outils puissants et au champ d'application étendu
- 3) Les applications : De nombreuses applications ont été réalisées, en particulier aux USA, dans des domaines très divers. Certaines touchent à la décision stratégique (C3I) ou à la décision politique (environnement), domaines proches de celui du transport.

Les freins

- 4) Les freins de la complexité des modèles : Le concept de RB ou de DI et les modes de calcul d'inférence restent inabordables au néophyte. La manipulation des RB relève donc de l'expert. Par contre, la capacité de visualisation permet un véritable dialogue intuitif avec le néophyte, pour peu que le mode de communication, c'est-à-dire l'interface, le permette.
- 5) Les freins de l'interface : La rigidité des interfaces des outils disponibles est un frein réel à la communication et à la manipulation aisée des outils RB ou DI. Nous disposons de solutions adaptées à la prise en compte de ce problème au travers de nos travaux sur les interfaces adaptatives et de la visualisation de cartes topiques.
- 6) Les freins de la modélisation : la modélisation d'une situation n'est pas chose aisée. Trois approches sont possibles : (i) on dispose d'une bonne connaissance du domaine (statistiques, connaissances probabilistiques des relations causes-effets), (ii) on dispose de beaucoup de données (statistiques d'expérience) mais pas de connaissance de relations causales, (iii) on ne dispose que de l'intuition. Selon les cas, on pourra construire des réseaux du plus objectif au plus subjectif. Cependant, il nous semble que dans un domaine tel que la prise de décision, la troisième solution est la plus opérante et finalement la moins coûteuse, les décisions supposées rationnelles se prenant souvent par consensus d'experts, éclairées simplement par des données de terrain. La perspective d'utiliser des techniques d'apprentissage de RB à partir de données statistiques comme en génomique nous apparaît donc peu opérante.

3. DISCUSSION

La Recherche Opérationnelle se fonde sur le principe de formalisation des choix et le calcul d'optimisation. Pour choisir dans l'incertain, trancher entre diverses formules, il faut un critère de choix qui corresponde à une attitude face au risque.

Pour illustrer ces propos, considérons la situation suivante : face à un devenir indéterminé, le décideur (entrepreneur, stratège, exploitant, joueur d'échecs), lorsqu'il dispose d'une marge de manœuvre peut envisager plusieurs solutions {a, b, c}. Par exemple, le chef d'entreprise qui a des difficultés financières peut décider de réduire ses effectifs (a), de lancer une campagne de promotion (b) ou de se diversifier dans de nouveaux produits (c)... Parmi ces scénarios, certains sont coûteux, mais peuvent rapporter beaucoup, d'autres sont moins coûteux mais risquent d'entraîner de maigres bénéfices. Laquelle des options choisir ? Quelle attitude adopter face au risque ?

- Le critère de Laplace consiste à adopter, parmi les choix possibles, celui qui rapporte un résultat moyen. Il est calculé à partir de la moyenne arithmétique entre des alternatives jugées également probables.
- Le critère de Bernoulli, dit aussi critère de l'espérance mathématique, consiste à affecter à chacune des solutions a, b, c, une plus ou moins grande probabilité de réalisation. Cette



probabilité est définie subjectivement, puis traduite en termes mathématiques. Cette espérance mathématique permet donc de pondérer les résultats escomptés. Le meilleur choix est celui qui obtient le meilleur rapport résultats/chances de réussite.

- Le critère de Wald ou critère Minimax est un critère pessimiste, qui ne consiste à envisager pour chaque alternative a, b et c que leurs plus mauvaises conditions de réalisation. On choisit alors parmi eux, celui qui coûterait le moins...
- Le critère de Savage ou la matrice des regrets consiste à construire une matrice où l'on affecte à chaque alternative (représentée par une case), un regret, c'est-à-dire le montant du manque à gagner si on avait connu d'avantage les résultats. On applique alors au regret le critère Minimax.

Qu'il s'agisse de maximiser des gains, de minimiser des regrets ou des pertes, ces objectifs sont à rapprocher des différentes activités concernant le management des risques : maîtrise des risques, gestion du danger, gestion de crises et gestion des situations critiques. Par ailleurs, la décision reste dans cette perspective un processus de planification où le principe directeur est la raison (la déduction) « *La logique est une valeur universelle, un outil de compréhension du monde autant qu'un moyen de le transformer* ».

S'il va de soi que toute stratégie repose nécessairement sur un ensemble d'informations sur la situation décisionnelle (« les ressources nécessaires au décideur »), l'obtention, la disponibilité, la gestion, l'organisation, le partage, l'échange, le traitement, la fiabilité, la pertinence, la complétude, l'interprétation de cette information sont autant de dimensions qui, dans nombre de « situations complexes » remettent largement en cause les théoriques modèles « linéaires » du processus de la décision de la ROP. Le schéma causal de la décision qui y est imposé, formulation d'un désir et conception d'un projet y répondant, prise d'information, délibération, décision proprement dite et exécution, rencontrent les limites de toute linéarité : l'impossibilité de considérer des causalités enchevêtrées. L'exécution, par exemple, ne va-t-elle pas changer la conception même du projet, le but recherché ? D'une manière générale, chacune des phases modifie l'objectif initial. La délibération conduit à rechercher de nouvelles informations, etc.

L'étude du processus de la décision doit donc inclure un véritable processus de traitement de l'information lorsque la situation n'est pas complètement « mathématisable », ce qui est généralement le cas lorsque la décision relève d'une organisation ou d'un quelconque collectif. H.A. Simon, qui a élaboré la théorie économique de la rationalité limitée présentée en 1947 dans *Administration et processus de décision*, précise à cet effet que la difficulté consiste à traiter l'information entre autre parce qu'elle est trop abondante. Pour cela, il met en avant les outils informatiques, qu'il appelle des « prothèses de l'homme » au sens où ils aident ce dernier à poser plus rationnellement les problèmes, filtrer les informations et simuler et planifier l'action qui devra suivre. Il édicte alors quelques principes pour la conception de ces outils et précise que l'essentiel est de comprendre la manière dont les décisions sont prises dans l'organisation, soulever les questions auxquelles l'information va répondre, adopter une approche arborescente et modulaire des problèmes. Dans le cadre de la décision organisationnelle, le vocable «système de traitement de l'information» (S.T.I) permet de désigner commodément la lignée des modèles issus de la pensée de H.A.Simon.

Ces dix dernières années ont été témoins du passage de l'information à la connaissance. Dans les approches mathématiques, objectifs et contraintes ont généralement été modélisés par le biais de modèles logico-mathématiques et la recherche de la décision optimale est une



« simple » question de « problem solving » où l'aspect cognitif du processus n'a que bien peu de place. La modélisation mathématique des objectifs et autres contraintes rend l'objectivité du modèle implicite, puisque sa formulation formelle a été établie. Dans une approche plus cognitive de la décision, celle-ci repose sur une hypothèse d'équivocité : pour une situation décisionnelle donnée, plusieurs interprétations de cette même situation sont possibles selon le point de vue, la fonction, les intérêts de l'acteur en jeu... La délibération se trouve ainsi être une étape majeure de la décision. L'information interprétée devient connaissance. Les techniques de gestion de connaissances apportent alors de nouvelles perspectives d'évolution aux modèles S.T.I. La dualité action/décision et le recours aux connaissances se rejoignent dans le concept de connaissances utiles à l'action ou *connaissances actionnables* [Argyris et al., 1978]. Le couplage de systèmes de gestion des connaissances et de systèmes d'aide à la décision ouvre la voie des *elucidative systems* ou systèmes automatiques d'argumentation où l'acceptabilité de la décision devient l'enjeu prioritaire [Penalva et al., 2002a,b ; Montmain et al., 2002b].

Pour nuancer ces approches encore « trop mécanistes », J.G. March insiste sur les conditions pratiques de la prise de décision et de l'élaboration de systèmes d'information. Il montre qu'en particulier les problèmes, les solutions et les actions se présentent rarement de façon séquentielle comme H.A. Simon le laisse entendre. Les décisions se développent d'elles-mêmes au fur et à mesure que les actions prennent sens pour les personnes de l'entreprise. Les managers ne sont pas des décideurs tout puissants mais jouent un rôle pour aider leurs employés à donner du sens à leur expérience. L'information prend la forme d'histoires pour interpréter l'action. Enfin, L. Suchman s'est opposée à H. Simon sur le modèle de la planification (ou de la programmation de l'information) où l'action est supposée être résolue par le plan et ne servirait éventuellement qu'à le raffiner. Elle montre au contraire que l'action est située, c'est-à-dire inscrite dans le contexte de circonstances concrètes et particulières. Le plan n'est alors qu'une ressource pour l'action : il est un construit social fortement dépendant de l'histoire et de l'expérience des individus qui l'élaborent. Le but du plan est de préparer à l'action pour mobiliser au mieux son savoir incorporé [Vacher et al., 2000].

Enfin peut-on s'interroger sur le « bien fondé » d'une décision. Il est en effet nécessaire de distinguer l'évidence et la pertinence de celle-ci. L'évidence relève d'une cohérence interne, d'une compatibilité avec une structure logique ; la pertinence relève d'une cohérence externe, d'une compatibilité avec une situation extérieure. Il y a les solutions logiques (évidence) et les solutions efficaces (pertinence). Dans la mesure où l'on peut regarder le monde avec différentes logiques, les solutions logiques et les solutions efficaces ne sont pas toujours les mêmes. La décision est un processus d'interactions entre le décideur et le monde. Il faut donc qu'elle soit en accord avec le monde plutôt qu'avec sa logique interne. Elle doit être pertinente plutôt qu'évidente. Dans l'idée de pertinence apparaît la notion de contexte. Une action est pertinente par rapport à un contexte, qu'elle soit logique ou pas, évidente ou pas. La décision évidente est raisonnée, la décision pertinente est raisonnable.

Ces réflexions nous laissent penser que si les modèles de la décision donnés par la ROP relèvent de mathématiques compliquées et ont eu d'incontestables résultats en particulier en gestion ou en économie, il n'en reste pas moins que le niveau de formalisation du problème qu'ils exigent n'est généralement pas celui de la modélisation de processus décisionnels d'une organisation engagée dans une dynamique d'action en environnement ouvert, où incertitude, imprécision et incomplétude des données décisionnelles s'amalgament dans l'évaluation des risques... En d'autres termes, la complexité des mathématiques n'est pas toujours la solution utile ou suffisante pour résoudre une situation décisionnelle complexe, l'apport des sciences humaines ouvre d'autres horizons sur l'immuable mystère de la décision.



BIBLIOGRAPHIE

- Adriaans P. and D. Zantinge (1996). Data Mining. *Syllogic*.
- Agrawal, R., T. Imielinski and A. Swami. (1993). Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of the International Conference on Management of Data (SIGMOD '93)*, Washington D.C, USA, pp. 207-216.
- Agrawal, R. and R. Srikant. (1994). Fast Algorithms for Mining Association Rules. In *Proceedings of the International Conference on Very Large Databases (VLDB '94)*, Santiago, Chile, pp. 487-499.
- Agrawal, R. and R. Srikant. (1995). Mining Sequential Patterns. In *Proceedings of the 11th International Conference on Data Engineering (ICDE '95)*, Tapei, Taiwan.
- Akharraz, A., Montmain, J., Mauris, G (2002a). Fonctionnalités explicatives d'un système de fusion de connaissances collectives par intégrale de Choquet 2-additive. *LFA'2002, Rencontres Francophones sur la Logique Floue et ses Application*, Montpellier, France.
- Akharraz, A. , Montmain, J., Mauris, G (2002b). A project decision support system based on an elucidative fusion system. *Fusion 2002, 5th International Conference on Information Fusion*, Annapolis, Maryland, USA.
- Allison, D. (1971). The essence of Decision. Explaining the Cuban Missile Crisis. *Boston, Little Brown*.
- Andreewski, E. (1998). Langage et construction de la cognition collective. *Grand Atelier MCX au Futuroscope*, Poitiers (F).
- Argyris, C., Schön, D. (1978). Organizational learning : a theory of action perspective. *Addison Wesley*.
- Aufaure, M-A., L. Yeh, K. Zeitouni. (2000). Fouille de données spatiales. chapitre dans "Le temps, L'espace et l'éolutif en sciences du traitement de l'information - Tome 2", Cépadès éditions, pp. 319-328.
- Avenier, M.J. (1992). Recherche-action et épistémologies constructivistes, modélisations systémique et organisations socio-économiques complexes : quelques « boucles étranges » fécondes. *Revue internationale de systémique*, Vol. 6, n°4, 1992
- Bachelard, G. (1934). Le nouvel esprit scientifique. *Quadrige, PUF, Paris (17ème édition, 1987)*
- Bartoli, J.A., Le moigne, J-L. (1996). Organisation intelligente et système d'information stratégique, *Ed. Economica, Gestion*.
- Becker, A., Naïm, P. (1999). Les réseaux Bayésiens. *Eyrolles*.
- Berge, C. (1958). La théorie des graphes et ses applications. *Paris, Dunod*.
- Berger, P., Luckman, T. (1986). La construction sociale de la réalité. *Méridiens-Klincksieck*.
- Berger, P., Luckman, T. (1996) : La construction sociale de la réalité ; *Armand-Colin*.
- Berrah, L., G. Mauris, L. Foulloy and A. Haurat (2000). The Choquet integral as a tool for industrial performance aggregation, *IPMU'2000*.
- Bertalanffy, L., von (1973). Théorie générale des systèmes, physique, biologie, psychologie, sociologie, philosophie, trad. *Française J.B. Chabrol, DUNOD*.
- Bezdek, J.C. (1981). Pattern Recognition with fuzzy Objective Function Algorithm. *New York, Plenum Press*
- Bono (de), E. (1985). Réfléchir mieux. *Les éditions d'organisation*.
- Bourrier, M., Laroche, H. (2001). Risque de défaillance : les approches organisationnelles. dans « *Risques, erreurs et défaillances. Approche interdisciplinaire* » sous la direction de René AMALBERTI, Catherine FUCHS, Claude GILBERT, CNRS-Maison des Sciences de l'Homme-Alpes, mai 2001, 362 p.
- Brin, S., R. Motwani, J. Ullman and S. Tsur. (1997). Dynamic Itemset Counting and Implication Rules for Market Basket Data. In *Proceedings of the International Conference on Management of Data (SIGMOD '97)*, Tucson, Arizona, pp. 255-264.



- Bonabeau, E., Theraulaz, G. (1994). Intelligence collective. *Collection Systèmes Complexes, 287p, Hermès, Paris.*
- Bougnoux, Daniel (1993). Science de l'information et de la communication. *Larousse, Paris, 808 p.*
- de Brabandere, L., Mikolajczak, A. (1998). Le plaisir des idées : libérer, gérer et entraîner la créativité au sein des organisations. *Editions Dunod.*
- Breton, P. (1987). Histoire de l'informatique. *Editions la Découverte, Paris, 239 p.*
Un livre clair sur les débuts de l'informatique jusqu'aux années 80. Pour ne pas oublier que la réalité technique n'échappe pas aux mythes.
- Broustail, J., Fréry, F. (1993). Le management stratégique de l'innovation. *Editions Dalloz.*
- Broyden, C.G. (1967). Quasi-Newton methods and their application to function minimization. *Mathematical computing, 21, pp. 368-391.*
- Buckley, A., Lenir, L. (1983). QN-Like variable storage conjugate gradients. *Mathematical programming, pp.155-175, 27.*
- Buzan, T. (1998). Une tête bien faite. Exploitez vos ressources intellectuelles. *Editions d'Organisation.*
- Cain, J. (2001). Planning improvements in natural resources management. *Guidelines for using Bayesian networks to manage development projects.*
- Califano, A. and I. Rigoutsos. (1993). FLASH: A Fast Look-up Algorithm for String Homology. In *Proceedings of the 1st International Conference on Intelligence Systems for Molecular Biology, Bethesda, pp. 353-359.*
- Caverni, J-P., J-M. Fabre et M. Gonzalez (1990). Cognitive biaises. *Amsterdam, North Holland.*
- Chapuy, P., R. Monti. (1998). La filière agricole et l'environnement – scénarios 2010 par la méthode Delphi – Abaque de Régnier. *LIPS Working Papers, Cahiers du LIPS N°9.*
- Charniak, E. (1991). Bayesian Networks Without Tears. *AI Magazine, 12.*
- Chiu, S. L. (1994). Fuzzy Model Identification Based on Cluster Estimation. *Journal of Intelligent and Fuzzy System, 2 :267-278.*
- Clergue, G. (1997). L'apprentissage de la complexité. *Hermès, Paris.*
- Cohen, M.D., J.G., March and J.P., Olsen. (1972). A garbage Can Model of Organizational choice. *Administrative Science Quarterly, 17 (1), pp.1-25.*
- Collerette, P. (1996). La méthode des études de cas. Dictionnaire des méthodes qualitatives en sciences humaines et sociales, *sous la direction d'Alex Mucchielli, pp.77-80, Armand Colin, Paris.*
- Cormen, T., C. Leiserson and R. Rivest. (1994). Introduction to Algorithms. *MIT Press, Cambridge, Massachussets.*
- Cooley, R.W. (2000). Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data. *In Phd Thesis, University of Minnesota.*
- Dantzig G. B., Wolge P. (1960). Decomposition principle for Linear Programs. *Operations Research, 8, pp. 101-111.*
- Das, B. (1999). Representing Uncertainties Using Bayesian Networks, *Information Technology Division Electronics and Surveillance Research Laboratory DSTO-TR-0918, www.dsto.defence.gov.au/corporate/reports/DSTO-TR-0918.pdf.*
- David, A. (1996). Structure et dynamique des innovations managériales. cahier n°12, *Ecole des Mines de Paris- Centre de Gestion Scientifique.*
- Denis, F. and R.Gilleron (2000). Apprentissage à partir d'exemples. *Notes de cours http://www.grappa.univ-lille3.fr/polys/apprentissage/*
- Desjeux, D., I.Orhant and Taponier (1988). L'édition en sciences humaines, la mise en scène des sciences de l'homme et de la société. *Ed. L'Harmattan.*
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik, 1, pp.269-271.*



- Djellal, F. (2001). Transport routier : la diversification des trajectoires d'innovation. *Revue Française de Gestion, Fondation nationale pour l'enseignement de la gestion des entreprises*. 133, pp.84.
- Doise, W., Moscovi, S. (1992). Dissensions et consensus. Ed. PUF.
- Dubois, D. (1983). Modèles mathématiques de l'imprécis et de l'incertain en vue d'applications aux techniques d'aide à la décision. *Thèse de Docteur es-sciences de l'Institut National Polytechnique de Grenoble*.
- Dubois, D. and H. Prade (1984). Criteria aggregation and ranking of alternatives in the framework of fuzzy set theory, in Fuzzy Sets and Decision Analysis (H-J. Zimmermann, L.A. Zadeh and B. Gaines, Eds.). *TIMS Studies in the Management Sciences*, **20**, 209-240.
- Dubois, D. and H. Prade (1985). A review of fuzzy set aggregation connectives. *Information Sciences*, **36**, 85-121.
- Durand, T. (2000). Alchimie de la compétence (l'). *Revue Française de Gestion, janvier-février*.
- Dyreson, C. (1997). Using an Incomplete Data Cube as Summary Data Sieve. *Bulletin of the IEEE Technical Committee on Data Engineering (March)*, pp. 19-26.
- Edmonds, J., Karp R.M. (1972). Theoretical improvements in algorithmic efficiency for network flow problem. *Journal of the A.C.M.*, **19**, 2, pp.248-264.,
- Eizirik, L.M.R., Barbosa, V.C., Mendes, S.B.T. (1993). A Bayesian-Network Approach to Lexical Disambiguation. *Cognitive Science*, **17**(2), pp. 257-283 .
- Ester, M., A. Frommelt, H.-P. Kriegel and J. Sander. (1997). Algorithms for Characterization and Trend Detection in Spatial Database. In *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD'97)*, New York, NY.
- Ester, M., H.-P. Kriegel, J. Sander and X. Xu. (1998). Clustering for Mining in Large Spatial Databases. *KI-Journal, Special Issue on Data Mining*, ScienTec Publishing, N°1.
- Euler, L. (1736). *Commentationes Arithmeticae collectae*, St. Petersburg.
- Fabiani, P. (1996). Représentation dynamique de l'incertain et stratégie de perception pour un système autonome en environnement évolutif. *Thèse de l'Ecole Nationale Supérieure de l'Aéronautique et de l'Espace*.
- Fayad, U.M., G. Piatetsky-Shapiro and P. Smyth. (1996). Advances in Knowledge Discovery and Data Mining. *AAAI Press*, Menlo Park, CA.
- Feldman, R., I. Dagan and W. Kloegsen. (1996). Efficient Algorithm for Mining and Manipulating Associations in Texts. In *Proceedings of the 13th European Meeting on Cybernetics and Research*
- Ferber, J. (1995). Les systèmes multi-agents : vers une intelligence collective. 522 p, InterEditions, Paris.
- Fernez-Walch, S., Romon, F. (2001). Peut-on innover dans les entreprises de BTP ? Revue Française de Gestion, Fondation nationale pour l'enseignement de la gestion des entreprises. 133, pp.94.
- Fletcher, R. (1980) Practical methods of optimisation 1: Unconstrained optimisation. John Wiley.
- Fletcher, R. (1981) Practical methods of optimisation 2: Constrained optimisation. John Wiley.
- Fodor, J.A. (1983) The Modularity of Mind. *M.I.T. Press*.
- Ford, L.R., Fulkerson, D.R. (1956). Maximal flow through a network. *Canadian Journal of Mathematics*.
- Fourer R. (2001) Linear/nonlinear programming faq's. <http://www.mcs.anl.gov/oct/guide/faq>.
- Fustier, M., Fustier, B. (2001). Exercices pratiques de créativité à l'usage du formateur. *Editions d'Organisation*.
- Gardarin, G., P. Pucheral and F. Wu. (1998). Bitmap Based Algorithms for Mining Association Rules. In *Proceedings of the 15th Workshop on Bases de Données Avancées*, Tunis.
- Gardarin, G. (1999). Internet/Intranet et Bases de Données : data web, data média, data warehouse, data mining. *Eyrolles*.
- Gavrilloff, I., Jarrosson, B. (2001). Une fourni de 18 mètres, ça n'existe pas... La créativité au service des organisations. *Dunod*.
- Gilbert, C. (1992). Le pouvoir en situation extrême. Catastrophes et politiques. Ed. *L'Harmattan*.



- Gilleron, R., M. Tommasi. (2000). Découverte de connaissances à partir de données - *notes de cours* <http://www.grappa.univ-lille3.fr/polys/fouille/>
- Glaserfeld von, E. (1981). Introduction à un constructivisme radical. In watzlawick P. (sous la direction de) : *l'invention de la réalité*, Seuil, Paris, trad. 1988
- Goffman, I. (1974). Les Rites d'interaction. Minuit, Paris.
- Guoguelin, P., Cuny, X. (1988). La prise de risque dans le travail. *Actes des journées d'étude et de réflexion organisées par les Chaires de psychologie du travail & sécurité du travail*, Paris, juin 1988, Editions O/E, Marseille.
- Gonzales, C., Wuillemin, P.H. (1998). Réseaux bayésiens en modélisation d'utilisateurs. *Sciences et Techniques Educatives*, 5 (2).
- Grabisch, M., H.T. Nguyen and E.A. Walker (1995). Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference (Chap. 8). Kluwer Academic.
- Grabisch, M., Roubens, M. (1996). The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, 89, pp.445-456.
- Grabisch, M., S.A. Orlowski and R.R. Yager (1998). Fuzzy aggregation of numerical preferences, in R. Slowinski, Ed., *Fuzzy Sets in Decision Analysis, Operations Research and Statistics*. Kluwer Academic.
- Grabisch, M. and M. Roubens (2000). Application of the Choquet Integral in Multicriteria Decision Making, in *Fuzzy Measures and Integrals : Theory and Applications* (M.Grabisch, T.Murofushi and M.Sugeno) , Physica-Verlag.
- Grosjean, M., Lacoste, M. (1999). Communication et intelligence collective. Le travail à l'hôpital. PUF, Paris
- Gunopulos, D., H. Mannila and S. Saluja. (1997). Discovering all Most Specific Sentences by Randomized Algorithms (Extended Abstract) . *Technical Report, Academy of Finland*.
- Guntern, G. (2001). Les 7 règles d'or de la créativité : l'imaginaire au cœur de l'entreprise. Village Mondial.
- Han, J., and M. Kamber. (2001). Data Mining: Concepts and Techniques. Morgan Kaufman Publishers.
- Han, J., K. Koperski and N. Stefanovic. (1997). GeoMiner: A System Prototype for Spatial Data Mining. In *Proceedings of the International Conference on Management of Data (SIGMOD'97)*, Tucson, Arizona.
- Han, J., J. Pei and Y. Yin. (2000). Mining Frequent Patterns without Candidate Generation. In *Proceedings of the International Conference on Management of Data (SIGMOD'00)*, Dallas, Texas.
- Hartigan, J.A. (1975). Clustering Algorithms, New York: John Wiley & Sons.
- Hofstadter, D. (1979). Gödel, Escher, Bach : an eternel golden braid. Basic book Inc., new york, 1979 (trad. J. Henry et R. French), Interéditions, Paris, 1985.
- Hogart, R. (1987). Judgement and choice : the psychology of decision. Chichester, Wiley.
- Höppner, F., F. Klawonn, R. Kruse, T. Runkler. (1999). Fuzzy Cluster Analysis. Wiley, Chichester.
- Horvitz, E. (1999). Principles of Mixed-Initiative User Interfaces. *Proceedings of CHI '99, ACM SIGCHI Conference on Human Factors in Computing Systems*, Pittsburgh, PA.
- Horvitz, E., Breese, J., Heckerman, D., Hovel, D., Rommelse, K. (1998). The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*.
- Horvitz, E., Seiver, A. (1997). Time-Critical Action: Representations and Application. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*.
- Hotier, H. (1997). L'induction ou l'emprise des sens. Communication et Organisation, *Textes préparatoires au colloque Induction et Communication du GRECO*, pp.115-149, Bordeaux, France.
- Houtsama, M. and A. Swami. (1995). Set-Oriented Mining of Association Rules. In *Proceedings of the International Conference on Data Engineering (ICDE'95)*.



- Hu, T.C. (1982). Combinatorial algorithms. *Addison-Wesley, Menlo Park, California.*
- Hutchins E. (1994). Comment le cockpit se souvient de ses vitesses. *Sociologie du Travail, avril, pp. 451-473.*
- Jain, A.K., M.N. Murty, P.J. Flynn. (1999). Data clustering : A review, *ACM Comput. Surv.* 31(3), pp.264 –323.
- Jang, J-S. R. (1993). ANFIS : Adaptive-Network-Based Fuzzy Inference System *IEEE Transactions on systems, Man, and Cybernetics*, 23, (3)
- Jarrosson, B. (1994). Décider ou ne pas décider ?. *Maxima, Laurent du Mesnil Editeur.*
- Journé, B. (1999). Les organisations à risques : gérer la sûreté par les ressources. *Thèse de doctorat, Ecole Polytechnique, janvier.*
- Kadie, C.M., Hovel, D., Horvitz, E. (2001). MSBNx: A Component-Centric Toolkit for Modeling and Inference with Bayesian Networks. *Microsoft Research Technical Report MSR-TR-2001-67.*
- Kahneman, D., P. Slovic and A. Tversky (1982). Judgement under uncertainty : heuristics and biases. *Cambridge University Press.*
- Koenig, G. (1990). Management stratégique : vision, manœuvres, tactiques. *Ed. Nathan*, chap.1 à 3.
- Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the travelling salesman problem, *Proc. Am. Math. Soc.*, 71, pp.48-50.,
- Kuhn, H.W. (1955). The hungarian method for the assignment problem. *Nav. Res. Log. Quart.*, 2, pp.83-97.
- Lagadec, P. (1988). Etats d'urgence. défaillances technologiques et déstabilisation sociale. *Ed. Seuil*
- Lagrange J.L. (1797). Théorie des fonctions analytiques. Impr. de la République, Paris.
- Langlois A., Philips M. (1997). Automates cellulaires, application à la simulation urbaine. *197 p, Hermes, Paris.*
- Latour, B. (1994). Sociologie des sciences, analyse des risques collectifs et des situations de crise. Point de vue de bruno Latour. *Séminaire du programme « Risques collectifs et situations de crise », Actes de la première séance, Ecole des Mines de paris, novembre 1994.*
- Laurgeau, C., Blosseville, J-M. (1999). Réflexions sur les systèmes de transports routiers futurs. *JNRR'99.*
- Lebart, L. (2002). Quelques méthodes de classification *Cycle Postgrade de Statistique, Université de Neuchâtel*, <http://www.unine.ch/statistics/postgrad/downloads.htm>
- Le Moigne, J-L. (1977). La théorie du Système général, théorie de la modélisation. *P.U.F., Paris, 1977, 3^{ème} édition mise à jour, 1990.*
- Le Moigne, J-L. (1986). Intelligence des mécanismes, mécanismes de l'intelligence. *Paris, Fayard, Fondation Diderot.*
- Le Moigne, J-L. (1990). La modélisation des systèmes complexes. *AFCET Systèmes, Dunod, Paris.*
- Le Moigne, J-L. (1998). Connaissance actionnable et action intelligente. *Grand Atelier MCX au Futuroscope, Poitiers (F)*
- Lent, B., R. Agrawal and R. Srikant. (1997). Discovering Trends in Text Databases. *In Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD'97).*
- Lévy, P. (1991) L'idéographie dynamique, vers une imagination artificielle. *Editions La Découverte, Paris.*
- Lévy, P. (1994). *L'intelligence et ses nouveaux outils. Dans l'empire des techniques*, pp. 154-166, *Editions du Seuil.*
- Lévy-Leblond, J.M. (1994). La physique, une science sans complexe. *In Les théories de la complexité. Autour de l'œuvre d'henri Atlan, sous la direction de F. Fogelmann-Soulié, Colloque de Cerisy, juin 1994, Editions du Seuil, Paris.*
- Liu, M. (1993). Etudes de situations socio-techniques complexes. *Actes du congrès AFCET, in Systémique et cognition*, pp 309-316, Versailles.
- Liu D.C., Nocedal J. (1989). On the limited memory BFGS methods for large-scale optimization. *Mathematical programming*, 45, pp.503-528.



- Longley, P.A., M.F. Goodchild, D.J. Maguire, D.W. Rhind. (1999). *Geographic Information Systems. Volume 1, Wiley.*
- Luenberger D. G. (1973). *Linear and non linear programming.* Addison Wesley.
- Maarek, Y.S. and I.Z. Ben Shaul. (1996). Automatically organizing bookmarks per content. In *Proceedings of the 5th International World Wide Web Conference.*
- Mann, D. (1998). 40 Inventive (Business) Principles With Examples. *TRIZ Journal,* <http://www.triz-journal.com/archives>.
- Mann, D. (2002). Hands-on systematic innovation. *Creak press*, pp.464.
- Mannila, H., H. Toivonen and A.I. Verkano. (1994). Improved Methods for Finding Association Rules. *Technical Report, University of Helsinki*, Finland.
- Mannila, H., H. Toivonen and A.I. Verkano. (1995). Discovering Frequent Episodes in Sequences. In *Proceedings of the 1st International Conference on Knolwdege Discovery in Databases and Data Mining (KDD'95)*, Montreal, Canada, pp. 210-215.
- March, J. (1988). *Decisions and Organization.* Blackwell, New-York
- March, J. G. (1989). Décisions et organisation. *Les Éditions d'Organisation, Paris.*
- March, J. (1991). Décisions et organisations. *Editions d'Organisation.*
- March J. (1999). Mythes, organisation et changement. *Revue française de gestion, n°120, octobre 1998, pp 15-21*
- Masand, B. and M. Spiliopoulou. (2000). *WebKDD-99: Workshop on Web Usage Analysis and User Profiling " SIGKDD Explorations, 1 (2).*
- Masseglia, F., P. Poncelet and R. Cicchetti. (1999). An Efficient Algorithm for Web Usage Mining. *Networking and Information Systems Journal (NIS), Hermès Editor, 2, (5-6)*, pp. 571-603.
- Mayère, A. (1997). Capitalisation des connaissances et nouveau modèle industriel. In *L'entreprise et l'outil informationnel, sous la direction de Monnoyer M.-C., l'Harmattan, Paris.*
- Maystre, L.Y., Pictet, J., Simos, J. Méthodes multicritères Electre. *Description, conseils pratiques et cas d'application à la gestion environnementale, 323 p, Presses polytechniques et universitaires romandes, Lausanne, Suisse.*
- Mélèse, J. (1972). L'analyse modulaire des systèmes de gestion. *Hommes et Techniques.*
- Mélèse, J. (1979). *Approche systémique des organisations.* Réédition de l'édition de 1979, *Editions d'Organisation, Paris, 1990.*
- Métais, E., Moingeon, B. (2001). Management de l'innovation : le "learning mix". *Revue Française de Gestion, Fondation nationale pour l'enseignement de la gestion des entreprises, 133*, pp.113.
- Miermont, J. (1995). Réalité et construction des connaissances. *Revue Internationale de systémique, vol. 9, n°2.*
- Mittelman H.D. (2000). Benchmarks for optimization software. <http://plato.la.asu.edu/bench.html>.
- Montmain, J., Gentil, S. (2000). Dynamical causal model diagnostic reasoning for online technical process supervision, *Automatica, 36*, 1137-1152.
- Montmain, J., Akharraz, A. and Mauris, G. (2002a). Knowledge management as a support for collective decision-making and argumentation processes. *IPMU'2002, 9th International Conference on Information processing and Management of uncertainty in Knowledge-Based Systems, Annecy, France.*
- Montmain, J., A. Akkharaz and G. Mauris (2002b). Decision acceptability in project management: argumentation and risk control. *IAR Intelligent Control and Diagnosis, Grenoble, France.*
- Morin, E. (1980). La méthode 2, la vie de la vie. *Editions du Seuil.*
- Morin, E. (1982). Science avec conscience. *Fayard.*
- Morin, E. (1991). La méthode 4, les idées, leur habitat, leur vie, leurs mœurs, leur organisation. *Editions du Seuil.*
- Morin, E. (1994). La Complexité humaine. *Flammarion, Paris.*
- Mousseau, V. (1993). Problèmes liés à l'évaluation de l'importance relative des critères en aide multicritère à la décision : réflexion théoriques, expérimentations et implémentations informatiques. *Thèse de doctorat, université de Paris-Dauphine.*



- Mucchielli, A. (1995). Les Sciences de l'Information et de la Communication. *Hachette*.
- Mucchielli, A. et al. (1996) Dictionnaire des Méthodes Qualitatives en Sciences Humaines et Sociales. *Armand Colin, Paris.*
- Mucchielli, A. et al. (1998). Théorie des processus de la communication. *Armand Colin, Paris.*
- Mucchielli, A. (1999). Théorie systémique des communications. *Armand-Colin, Paris.*
- Mucchielli, A. (2000). La nouvelle communication : épistémologie des sciences de l'information–communication. *Armand Colin, Paris.*
- Mugur-Schächter, M. (1995). Une méthode de conceptualisation relativisée : vers une épistémologie nouvelle apte à faire face aux complexités. *Revue Internationale de Systémique, Vol.9, n°2, 1995.*
- Murphy, K. (2001a). Brief Introduction to Graphical Models and Bayesian Networks, <http://www.cs.berkeley.edu/~murphyk/Bayes/bayes.html>
- Murphy, K. (2001b). The Bayes Net Toolbox for MATLAB
- Murray, W.R. (1999). An easily implemented, linear-time algorithm for bayesian student modeling in multi-level trees, *In actes de la conférence Artificial Intelligence in Education*, Le Mans, France.
- Musselin, C. (1990). Quelle stratégie de recherche pour les anarchies organisées. *Actes du séminaire CONDOR, Vol.2, CRD-ESCP, pp.151-168.*
- Noh, S., Gmytrasiewicz, P. J. (1997). Bayesian Belief Update in Antiair Defense. *In the Workshop on Machine Learning for User Modeling of the Sixth International Conference on User Modeling, Italy.*
- Nonaka, I., Takeuchi, H. (1995). Knowledge Creating Company (The). *Oxford University Press .*
- Norman, D. A. (1993). Les artefacts cognitifs. *In Les objets dans l'action. De la maison au laboratoire, Textes réunis par B. Conein, N. Dodier et L. Thévenot, Éditions de l'EHESS, Paris, pp. 15-34.*
- Oury, J-M. (1983). Économie politique de la vigilance. *Calmann-Lévy, Paris, 187 p.*
- Paillet, P. (1996). Problématique d'une recherche qualitative. Dictionnaire des méthodes qualitatives en sciences humaines et sociales, *sous la direction d'Alex Mucchielli, pp.165-167, Armand Colin, Paris.*
- Pasquier, N., Y. Bastide, R. Taouil and L. Lakhal. (1999). Efficient Mining of Association Rules Using Closed Itemset Lattices. *In Proceedings of the International Conference on Database Theory (ICDT'99), pp. 398-416, Israel.*
- Pauzie, A. (2001). Problématique générale des IHM (Interface Homme / Machine) pour tous systèmes d'information dans les véhicules : ergonomie et sécurité. *Actes de la journée technique du 18 décembre 2001 Interface homme / machine Services télématiques et nomades dans la conduite des véhicules : Risques et apports.*
- Pazzani, M., J. Muramatsu and D. Billsus. (1996). Syskill & Webert: Identifying Interesting Web Sites. *In Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access, Portland, Oregon.*
- Pei, J., J. Han, B. Mortazavi-Asl, H. Pnto, Q. Chen, U. Dayal and M.C. Hsu. (2001). PrefixSpan: Mining Sequential Patterns Efficiently by Prefix Projected Pattern Growth. *In Proceedings of the 17th International Conference on Data Engineering (ICDE '01).*
- Pearl, J. (2001). Causal Inference in the Health Sciences: A Conceptual Introduction. *Contributed to Health Services and Outcomes Research Methodology.*
- Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. *Morgan Kaufmann, San Mateo, CA.*
- Penalva, J-M. (1997). La modélisation par les systèmes en situations complexes. Thèse de l'Université de Paris-Sud, UFR d'Orsay
- Penalva, J-M. (2000). Connaissances actionnables et intelligence collective. Nimestic'2000, Ingénierie système et NTIC, Nîmes, France.



- Penalva, J-M. and Montmain, J. (2002). Travail collectif et intelligence collective : les référentiels de connaissances. *IPMU'2002, 9th International Conference on Information processing and Management of uncertainty in Knowledge-Based Systems*, Annecy, France.
- Perrin, J. (2001). Concevoir l'innovation industrielle-Méthodologies de conception de l'innovation. *CNRS Editions*.
- Peton, N. (1999). Méthode du groupement par soustraction pour l'identification de modèle flou : amélioration et application à la prévision de la pollution atmosphérique. *Thèse de doctorat de Montpellier II*.
- Piaget, J. (1970). L'épistémologie génétique. *Que sais-je ?, P.U.F., Paris, 1970 (quatrième édition 1988)*.
- Pitrat, J. (1991). An intelligent system must and can observe its own behavior. *Proceedings of Cognitiva 1991 (Madrid), Ed. AFCET, Paris*.
- Pomerol, J-C. and S.Barba-Romero (1993). Choix multicritère dans l'entreprise. *Ed. Hermès*.
- Proust, J. (1997). La mémoire entre réalité et apparence. *Centre de Recherche en Epistémologie Appliquée, UMR7656, Ecole polytechnique, Institut du Management d'EDF*.
- Quach, R. (2002). Identification d'un modèle flou appliquée à un problème de classification. *Thèse de doctorat de l'Université Jean Monnet de Saint-Étienne*.
- Rajman, M. and R. Besancon. (1998). Text Mining - Knowledge Extraction from Unstructured Textual Data. In *Proceedings of the 6th Conference of International Federation of Classification Societies (IFCS-98)*, Rome, Italy.
- Rea, K.C. (1999a). TRIZ and Software - 40 Principle Analogies. *TRIZ Journal*, <http://www.triz-journal.com/archives>.
- Rea, K.C. (1999b). Using TRIZ in Computer Science Concurrency. *TRIZ Journal*, <http://www.triz-journal.com/archives>
- Revillard J.P. (1998). Immunologie. 461p, *De Boeck Université*, 3^{ème} édition.
- Ricoeur, P., Michaud, B. (1995). Le paradoxe d'apprendre. *Rencontres philosophiques de l'UNESCO : qu'est-ce qu'on ne sait pas ?, Gallimard, pp 12-17, Paris*.
- Rockafellar R. T. (1993). Lagrange multipliers and optimality. *SIAM Rev.*, 35, pp.183-238.
- Rossel, P., Bassand, M., Roy, M.A. (1998). Au delà du laboratoire : les nouvelles technologies à l'épreuve de l'usage. *Presses Polytechniques et Universitaires Romandes*
- Roy, B. (1985). Méthodologie multicritère d'aide à la décision. *Economica, Paris*.
- Roy, B. (1992). Science la décision ou science de l'aide à la décision ? *Revue Internationale de systémique*, 6 (5), pp.497-529. Version anglaise in *EJOR Special Issue on Model Validation in Operations research*, 66, 1993.
- Roy, B., Bouyssou, D. (1993). Aide multicritère à la décision : méthodes et cas. *Economica, Paris*.
- Saaty, T.L. (1980). The analytical hierarchy process. *McDraw Hill, New-York*.
- Sahami, M., Dumais, S., Heckerman, D., Horvitz, E. (1998). A Bayesian approach to filtering junk e-mail. *AAAI Workshop on Learning for Text Categorization, Madison, Wisconsin*.
- Sakarovitch, M. (1984a). Optimisation combinatoire, méthodes mathématiques et algorithmiques- Graphes et programmation linéaire. *Hermann, enseignements des sciences*.
- Sakarovitch, M. (1984b). Optimisation combinatoire, méthodes mathématiques et algorithmiques- Programmation discrète. *Hermann, enseignements des sciences*.
- Savage, L.J. (1972). The foundation of statistics. *Dover, New-York*.
- Savasere, A., E. Omiecinski and S. Navathe. (1995). An Efficient Algorithm for Mining Association Rules in Large Databases. In *Proceedings of the 21st International Conference on Very Large Databases (VLDB '95)*, Zurich, Switzerland, pp. 432-444.
- Schärlig, A. (1985). Décider sur plusieurs critères, panorama de l'aide à la décision multicritère. *Presses polytechniques et universitaires romandes*, 304 p, *Lausanne, Suisse*.
- Schutz, A. (1987). Le Chercheur et le Quotidien. *Méridiens-Klincksieck*.



- Segal, E., Taskar, B., Gasch, A., Friedman, N., Koller, D. (2001). Rich probabilistic models for gene expression. *Bioinformatics*, 17 Suppl 1:S243-52
- Sfez, L. (1992). Critique de la décision. *Presses de la Fondation des sciences politiques*, 4^e ed. (1^{er}ed. 1973)
- Simon, H.A. (1969). Science des systèmes, science de l'artificiel. *Traduction française par J-L. Le Moigne, Bordas, 1991.*
- Simon, H. A. (1980). Le nouveau management. La décision par les ordinateurs. *Économica, Paris*, 159 p.
- Simon, H. A. (1983). Administration et processus de décision. (*trad. de Administrative Behavior, 1947*), *Économica, Paris*.
- Simon, H.A. (1991). Sciences des systèmes, sciences de l'artificiel. *Ed. Dunod, 1969 (réédition).*
- Simon, H.A. (1997). Models of bounded rationality. *MIT Press, Cambridge Massachusetts*. 3.
- Simon, H. A., March, J. J. (1964). Les organisations. *Dunod, Paris.*
- Srikant, R. and R. Agrawal. (1996). Mining Sequential Patterns: Generalizations and Performance Improvements. In *Proceedings of the 5th International Conference on Extending Database Technology (EDBT'96)*, Avignon, France, pp. 3-17.
- Stake, R.E. (1995). The Art of Case Study Research. *Sage Publications, Thousand Oaks, California.*
- Stangl, D.K., Berry D.A. (2000). Meta-analysis: past and present challenges. In: Stangl, Berry, editors. *Meta-Analysis in Medicine and Health Policy*. New York: Marcel Dekker; pp. 1-28.
- STRATEGOR, coll. (1988). Stratégie, structure, décision, identité, politique générale d'entreprise. *Interéditions*, chap.13 à 16.
- Suchman, L. (1987). Plans and Situated Actions. *Cambridge University Press.*
- Sutton, R.S., Barto, A.G. (1998). Reinforcement Learning. 322 p, *MIT Press, London.*
- Tarjan, R. E. (1977). Finding optimum branchings. *Networks*, 7, pp.25-35.
- Terninko, J. (1998). Systematic innovation : an introduction to TRIZ (Theory of inventive problem solving). CRC press.
- Terninko, J. (1999). 40 Inventive Principles with Social Examples. *TRIZ Journal*, <http://www.triz-journal.com/archives>.
- Teulier-Bourgine, R. (1997). Les représentations, médiations de l'action stratégique. La stratégie chemin faisant, M.J. Avenier, Ed. Economica.
- Thoenig, J-C. (1987). L'ère des technocrates. *Ed. L'Harmattan, Paris.*
- Tiberghien, G. (1991). Modèles de l'activité cognitive. *Dans psychologie cognitive : modèles et méthodes*, pp. 13-26, *Presses universitaires de grenoble*.
- Toivonen, H. (1996). Sampling Large Database for Association Rules. In *Proceedings of the 22nd International Conference on Very Large Databases (VLDB'96)*.
- Tumer K., Wolpert D. (2000). Collective Intelligence and Braess'Paradox. In *the Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pp. 104-109, Austin, TX.
- Vacher, B. (2000). « Formes d'organisation et modes de gestion à travers le siècle », « Grilles de lecture de l'organisation : cartographies pour la décision », « Utilisations, organisation et malentendus de l'information et de ses systèmes », chapitre 1/1 à 1/3 de la partie 7. *Techniques Documentaires*, Ed. Weka, mises à jour de mars et septembre.
- Vaudry, C., Crampes, M. (2001b). AMEBICA : une interface adaptative pour la supervision industrielle. *Actes de la Conférence JIM 2001, Interaction Homme Machine et Assistance*, Metz, France.
- Vaudry, C., Crampes, M. (2001b). Modèle de la composition multimédia adaptative dans l'interaction homme-machine. *Actes de MFI 2001, Modèles Formels de l'Interaction, Volume III*, B. Chaaib-Draa et P. Enjalbert editors, Toulouse, France.
- Vaudry, C., Crampes, M. (2001c). AMEBICA : An adaptive user interface for process control.. *Short paper presented, IHM-HCI '01. Actes de IHM-HCI 2001, Volume II*, pp.17-20, Lille, France.
- Vernadat, F. (1999). Techniques de modélisation en entreprises : application aux processus opérationnels. *Economica*.



- Verduyn, D. (2000). Integrating Innovation into Design for Six Sigma. *TRIZ Journal*, <http://www.triz-journal.com/archives>.
- Vincke, P. (1989). L'aide multicritère à la décision. *Ed. de l'Université de Bruxelles*, 179 p, Bruxelles.
- Von Neumann, J. and O.Morgensten (1944). Theory of games and economic behavior. *Princeton Univ. Press, Princeton N.J.*
- Walliser, B. (1992). Systèmes et modèles. *Seuil, Paris, 1977*.
- Wang, J.T., G-W Chirn, T. Marr, B. Shapiro, D. Shasha and K. Zhang. (1994). Combinatorial Pattern Discovery for Scientific Data : some preliminary results. In *Proceedings of the International Conference on Management of Data (SIGMOD'94)*, Minneapolis, USA.
- Wang, K. and H. Liu. (1999). Discovering Structural Association of Semi Structured Data. *IEEE Transactions on Knowledge and Data Engineering*.
- Wang, K. (1997). Discovering Patterns from Large and Dynamic Sequential Data. *Journal of Intelligent Information Systems*, pp. 8-33.
- Watzlawick *et al.* (1972). Une logique de la communication. *Paris, Le Seuil*.
- Watzlawick,P., Weakland, J., Fish, R. (1973). Change. Principles of problem formation and problem resolution (traduit de l'anglais par P. Furlan : changements, paradoxes et psychothérapies). *Editions du Seuil, 1975*.
- Watzlawick, P. (1981). L'invention de la réalité. Comment savons nous ce que nous croyons savoir ? Contribution au constructivisme. *Seuil, Paris, 1988 (traduction française de la version originale, 1981, 1985)*
- Weick, K. E., Roberts, K. H. (1993). Collective Mind in Organisations : Heedful Interrelating on Flight Decks. *ASQ, Vol 38, N°3, September, pp. 357-381*.
- Weisbuch G. (1989). Dynamique des systèmes complexes, Une introduction aux réseaux d'automates. *Savoirs Actuels, 212p, InterEditions/Editions du CNRS*.
- Weiss, R., B. Velez, M. A. Sheldon, C. Nampempre, P. Szilagyi, A. Duda, and D. K. Gifford. (1996). Hypursuit: a Hierarchical Network Search Engine that exploits Content-Link Hypertext Clustering. In *Proceedings of the 7th ACM Conference on Hypertext (Hypertext'96)*.
- Winkin, Y. (1981). La nouvelle communication. *Paris, Le Seuil*.
- Winkless, B. (2000). Food Product Development and the 40 Inventive Principles. *TRIZ Journal*, <http://www.triz-journal.com/archives>.
- Wolfe P. (1961). A duality theorem for nonlinear programming. *Quar. Appl. Math.*, 19, pp.239-244.
- Wolpert, D.H., Tumer K. (2000). An Introduction to Collective Intelligence, *NASA Ames Research Center, 96p, http://ic.arc.nasa.gov/ic/projects/coin_pubs.html*.
- Xuong, N. H. (1992). Mathématiques discrètes et informatique. *Masson*.
- Yager, R. Possibilistic decision-making (1979). *IEEE Trans. Systems Man Cybernetics*. 9, 177-200
- Yin, R.K. (1984). Case Study Research, Design and Methods. *Sage, London*.
- Zaki, M. (1999). Fast Mining of Sequential Patterns in Very Large Databases. *Technical Report, The University of Rochester*, New York, N° 14627.