
Avec le soutien de l'ARC et du CNET
Sous le patronage du comité Systémique et Cognition de l'AFCEP

Intelligence Collective

*Journées de Rochebrune
24 - 30 Janvier 1993*

rap e

Sommaire

- Danièle DUBOIS**, LEPC & LCPE - CNRS / Ecole Pratique des Hautes Etudes : "*Intelligence collective : codes, langages et intelligibilité*"
- Guy THERAULAZ, Jacques GERVET**, CNRS - UPR 38 Ethologie Marseille : "*Du Super-Organisme à l'Intelligence en Essaim : Modèles et Représentation du fonctionnement de sociétés d'insectes*"
- Anne BRYGOO**, DRI - Université P. & M. Curie Paris, **Sylvie DESPRES**, EHEI - Université R.Descartes Paris : "*Vers une spécification de la notion de système coopératif*"
- Laure BAZZOLI, Emmanuelle CONESA, Jean-Pierre MICAELLI**, ECT - Université Lumière Lyon 2 : "*L'intelligence collective entre le naturel, l'artificiel et le social*"
- Jean-Louis DESSALLES**, Télécom Paris : "*Détection Collective*"
- Evelyne ANDREEWSKY**, INSERM - Salpêtrière : "*Dynamique cognitive et interactions sociales : quelques phénomènes de résonance*"
- Herminia PERAITA**, Facultad de Psicología - Uned. Madrid : "*Intelligence collective : émergence des concepts sociaux*"
- Wolf-Dieter EBERWEIN**, WZB-FIB Berlin : "*Stabilité politique : une fonction des décisions individuelles et collectives ?*"
- Marie-Noëlle SARGET**, CEMS - Ecole des Hautes Etudes en Sciences Sociales : "*Les conditions d'émergence de l'intelligence collective dans les systèmes politiques*"
- Pierre-Yves RACCAH**, CNRS : "*Le lexique : Cristallisateur des connaissances collectives*"
- Marie-Pierre GLEIZES, Pierre GLIZE**, IRIT - Université Paul Sabatier Toulouse : "*Comportements collectifs en IA distribuée*"
- Equipe MIRIAD**, LAFORIA : "*Approcher la notion de collectif*"
- Anne NICOLLE-ADAM**, LAIAC - Université de Caen : "*Auto-organisation et intelligence collective*"
- Manuel ZACKLAD**, Dépt. de Génie Informatique - UTC de Compiègne, **Jean ERCEAU**, GIA - ONERA : "*Coordination via l'environnement vs. coordination via une organisation*"
- Joëlle BIONDI**, Université de Nice-Sophia Antipolis : "*Coopération entre agents économiques, physiques ou cognitifs : l'algorithme 'bucket brigade'*"
- A.M.ALQUIER-BLANC**, Université Toulouse I : "*Reflexion méthodologique sur les systèmes d'information coopératifs dans les organisations*"
- Pascal BOLDINI**, ENST Bretagne : "*De l'internalisation à l'autonomie, et à l'action collective : un modèle mathématique*"
- Eric BONABEAU**, CNET Lannion B : "*Transitions de phase dans les systèmes collectifs*"
- Paul BOURGINE**, CEMAGREF : "*Phénomènes critiques auto-organisés*"

- Hugues BERSINI**, Iridia - Université Libre de Bruxelles : "*Intelligence collective : Adaptation par renforcement et recrutement*" [Texte non parvenu]
- Bruno CORBARA**, LAFORIA : "*Sociétés animales*" [Texte non parvenu]
- Jean-Louis DENEUBOURG**, Institut de Physique et de Chimie, Bruxelles : "*Colonies d'insectes*" [Texte non parvenu]
- John STEWART**, Institut Pasteur : "*Une forme non neuronale de cognition : le système immunitaire*"
- Yves-Marie VISETTI**, LIMSI : "*Une présentation du principe de LINSKER et de ses applications*" [Texte non parvenu]
- Guillaume DEFFUANT**, **Etienne MONNERET**, CEMAGREF : "*Adaptation et extraction de régularités : Une approche constructiviste*"
- Renaud CAZOULAT**, LAIAC - Université de Caen : "*Simulation de systèmes auto-organisés*"

page 2

Intelligence collective: codes, langages et intelligibilité

Danièle Dubois

LEPC
Laboratoire
d'Ergonomie
physiologique
& cognitive

LCPE
Langages,
Cognitions,
Pratiques
& Ergonomie

C.N.R.S / E.P.H.E..

41 rue Gay Lussac 75005 Paris

Tel (33) 1. 46.33.63.23 Télécopie (33) 1. 43. 26. 88.16

Ma contribution revêt en fait deux aspects :

- l'un réactif d'abord puis réflexif et épistémologique dans un second temps, relativement à la notion même d'intelligence collective à une époque où l'intelligence est régulièrement attribuée aux objets du développement technologique (la voiture, la route, la machine à café etc ...) à travers les progrès de l'intelligence artificielle.
- l'autre aspect concerne directement le thème quant à son contenu et revoie à la question (détaillée en sous questions) de l'exercice de l'intelligence par une collection d'individus avec ou sans communication, avec ou sans langage.

Question UN : sur la définition du thème des journées

L'intelligence peut-elle être autre que collective ?

L'intelligence (de chaque individu) et l'identité même de chaque individu, au moins humain, étant par essence sociale (collective), se seraient **les situations de mise en oeuvre de cette (ces) intelligence(s) individualisée(s) qui seraient individuelles ou collectives.**

Je trouve en effet un peu bizarre de formuler le problème de cette manière surtout dans la mesure où, en première analyse celà me semble lié :

- 1) à l'hypertrophie de l'individualisme et précisément du statut de la notion d'individu dans notre culture et tradition intellectuelles. De ce point de vue, j'aurai souhaité, étant donné mon incompétence en ce domaine, que des interrogations sur ce thème soient évoquées par des psychosociologues, sociologues, médecins ou psychiatres, anthropologues afin de nous aider à préciser les relations entre individualité, personnalité, identité de soi, groupes et ... intelligence.
- 2) au développement actuel des modélisations et formalisations issues des développements autonomes des technologies de traitement de l'information, qui interviennent à des niveaux d'abstraction tels qu'ils escamotent la prise en compte

des natures diverses des données empiriques, en particulier la restriction de l'intelligence à de la computation ou de la résolution de problèmes.

3) à l'effacement du rôle des systèmes symboliques dans l'analyse contemporaine de l'intelligence, devenue majoritairement artificielle au sein des sciences cognitives

C'est sur ce dernier point que je me permets de proposer à discussion quelques questions, que je pense très directement liées au thème :

Question DEUX : Intelligence ou intelligences ?

Dans quelle mesure peut-on considérer comme équivalent et donc faire entrer sous la même dénomination d'intelligence la diversité des exemples que nous aurons à traiter dans le séminaire.?

Il s'agit donc de tenter de préciser d'abord si des "matériaux biologiques", des cellules et bêtes (à 6 ou 8 pattes de préférence), à l'homme, il est néanmoins possible de considérer qu'il s'agit de phénomènes sémantiquement identiques. Ce qui revient à tenter d'abord d'explicitier **quelques critères de définition de l'intelligence ou d'identification de réponses ou conduites intelligentes** (ce qui, j'insiste, n'est pas nécessairement la même chose et ne conduira pas aux mêmes réponses) et de comparer des définitions bestiales, humaines et artificielles de ce concept. Je pense que déjà quelques perplexités peuvent nous saisir: je vois en particulier poindre le débat sur intelligence et rationalité(s ?) et une opposition relative au couple adaptation/vérité, déjà abordée dans le groupe.

Question TROIS : mécanismes et genèse de l'intelligence collective en psychologie

Quels sont ou plus exactement quelles hypothèses a-t-on (en psychologie en ce qui me concerne, en référence aux "grands" comme Piaget ou Wallon) concernant les mécanismes régulateurs de l'intelligence collective ?

à confronter aux hypothèses issues d'autres domaines disciplinaires (biologie, éthologie, économie ?)

Question QUATRE : Réalisation de tâches collectives et communications

Peut-on se passer dans le cadre d'une analyse de "l'intelligence collective" enfin identifiée (!) de la question d'une (de la) **transmission d'information(s) entre individus qu'elle soit centralisée et contrôlée, ou distribuée ?** En d'autres termes, l'intelligence collective est-elle assimilable à la résolution d'un problème ou plus généralement à l'exécution d'une tâche sous forme coopérative, ou compétitive ?

Quelques exemples de situations de travail ou de tâches collectives à travers les communications animales et humaines seront donnés (à confronter à d'autres en

particulier communications homme/machine et au niveau infra-organisme qui, je souhaite, seront présentés dans les journées)

Question CINQ : Communications et transmission d'information

A partir de cette dernière question peut-on considérer l'intelligence collective **humaine** sous la dépendance de mécanismes de transmission d'informations identiques à ceux des autres organismes ?
ou en d'autres termes, les **langues humaines** sont-elles assimilables à des codes ?

J'argumenterai (théoriquement et empiriquement) en faveur d'une réponse non à cette question, en particulier à partir d'arguments liés à l'histoire des langues, à une critique de la conception sémantique "classique" (qui présuppose une adéquation (mapping) de la chose, du concept et du nom).

Question SIX : Vous avez dit : Intelligence collective ?

Ce détour linguistique ou plus exactement sémiotique ramène au sujet dans la mesure où il repose la question des propriétés spécifiques de l'intelligence collective et individuelle sur les couples d'opposition tels que :

interprétation/résolution de problèmes

prototype/stéréotype

créativité/exactitude

évolutif/figé

non déterministe/déterministe

aveugle/finalisé

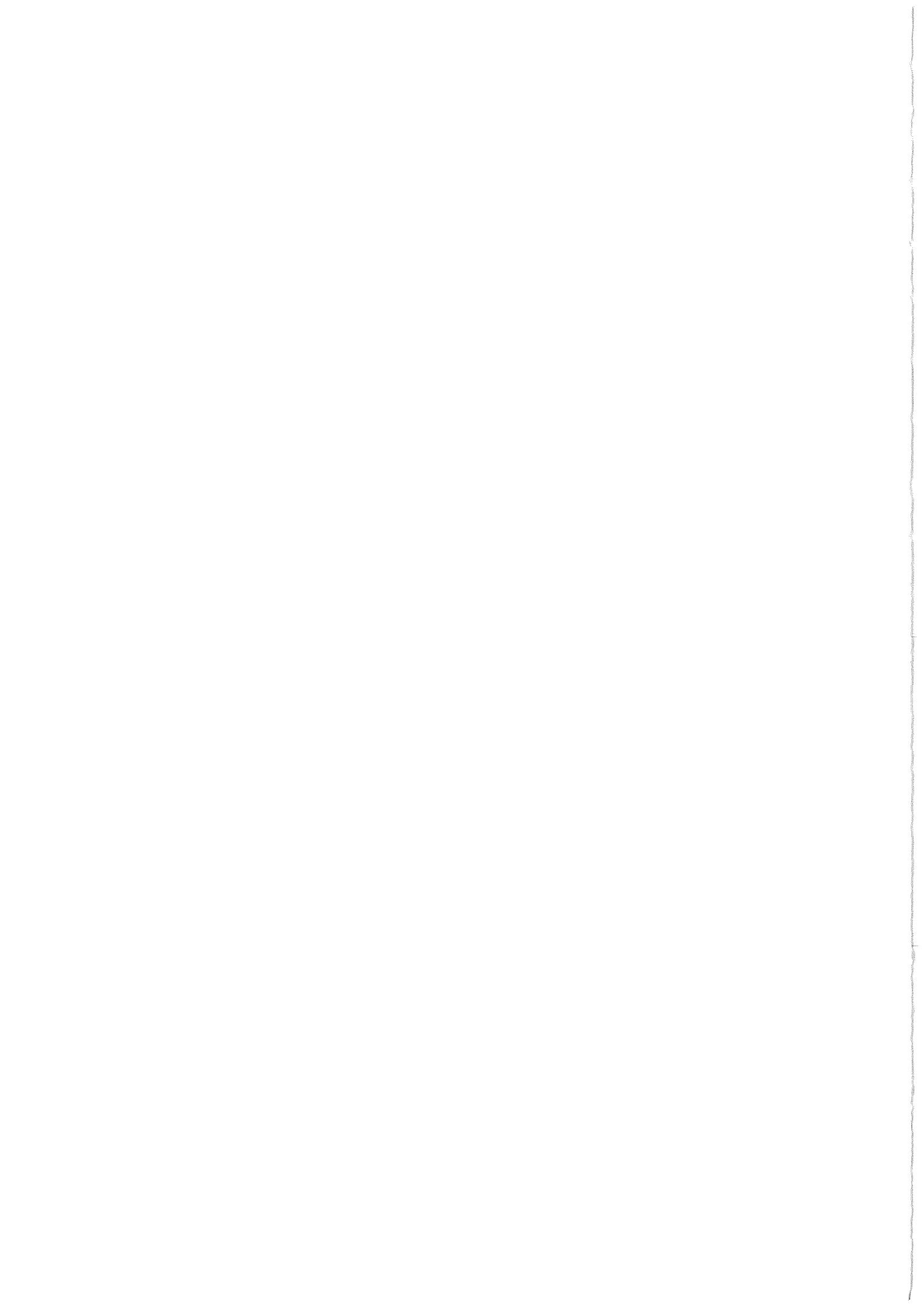
.....

Conclusion

Elle reprendra les arguments développés dans le parcours de ces questions, sur les deux thèses suivantes :

- **fondamental** : L'intelligence est par essence collective si on prend comme référence de la définition d'intelligence l'intelligence **humaine** (le problème est évidemment différent si les systèmes artificiels sont considérés comme plus intelligents et plus rationnels...). Elle est indissolublement dépendante de connaissances partagées, par la communication de ces connaissances et donc par l'existence des langues et langages et plus génériquement des systèmes symboliques porteurs de significations.

- **métascientifique** : est-il scientifiquement fondé de parler et d'utiliser ce vocable d'intelligence collective ? N'est pas précisément un **jeu de langage** qui détournerait le travail scientifique (sur les concepts) vers le **marketing publicitaire** au risque d'impasses théoriques ...mais de succès commerciaux ?



**Du Super-Organisme à l'Intelligence en Essaim :
Modèles et Représentation du fonctionnement de sociétés d'insectes.**

I. La colonie comme système adaptatif : SCHNEIRLA et la première cybernétique.

1. L'analyse des relations entre l'individu et la colonie dans les colonies de fourmis légionnaires *Eciton burchelli*.

2. Le modèle de l'organisation sociale des Dorylines.

II. Les développements de l'analyse systémique du fonctionnement collectif des colonies.

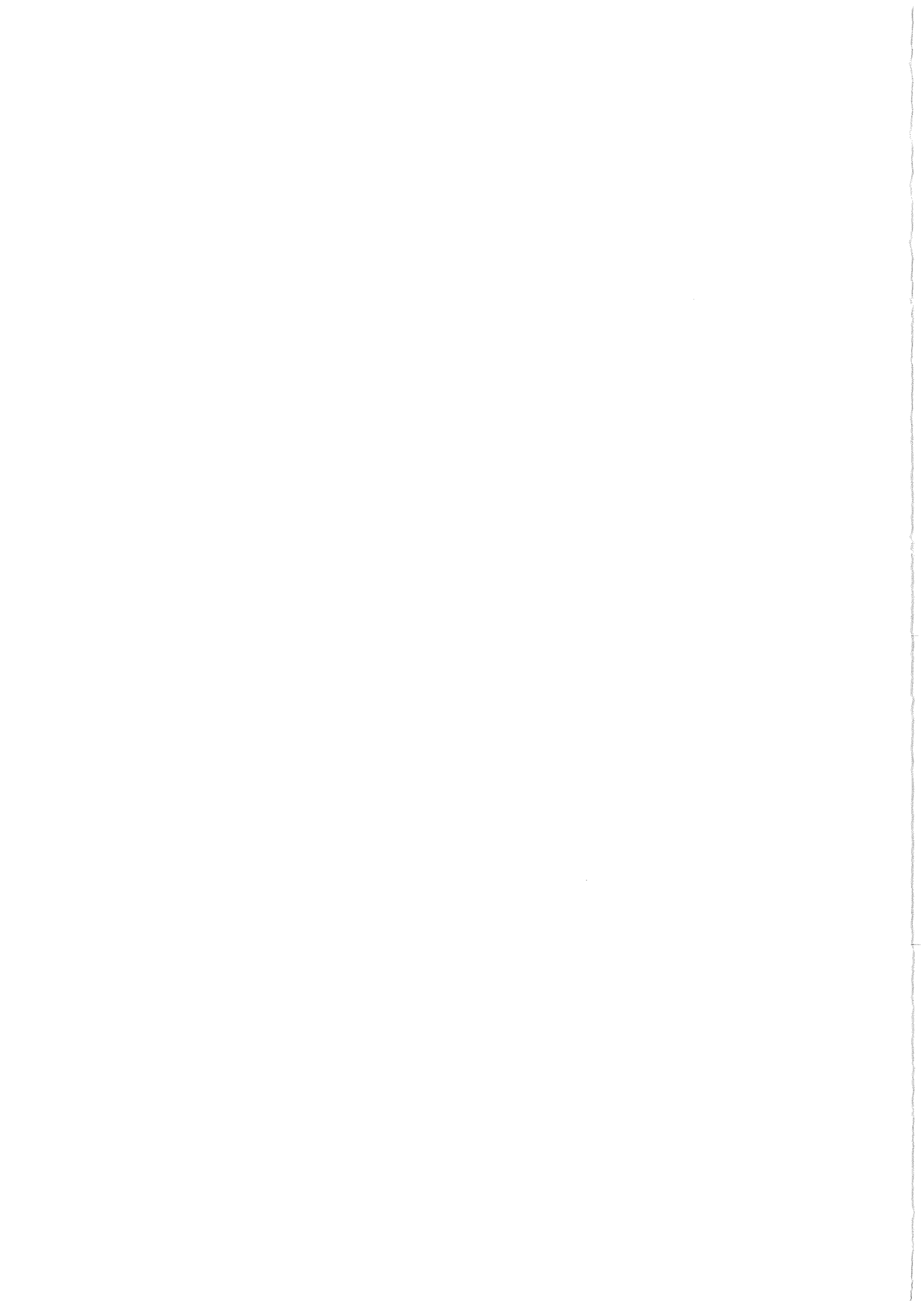
1. Transferts de concepts : auto-organisation et structures dissipatives dans la construction du nid chez les Termites *Belonogaster*.

2. Résolution collective de problèmes et intelligence en essaim.

a. L'assignation collective des tâches sans contrôleur central dans les colonies de guêpes primitives *Polistes dominulus*.

b. Une représentation collective sans représentation individuelle : la sélection des sources de nourritures chez *Lasius*.

c. Des constructions collectives sans plan d'organisation ni architecte : l'activité bâtisseuse des *Polistes*.



VERS UNE SPECIFICATION DE LA NOTION DE SYSTEME COOPÉRATIF¹

Anne BRYGOO* - Sylvie DESPRES**

*DRI - Equipe I S - Tour 55-65
Université Pierre et Marie Curie
4, place Jussieu
75252 Paris Cedex - France
tel : 44 27 58 77
email : abrygoo@ccr.jussieu.fr

**EHEI
Université René Descartes
45, rue des Saints Pères
75006 Paris - France
tel : 47 03 31 27
email : sd@ehei.ehei.fr

Résumé

L'intelligence collective dans le cas d'une interaction homme/homme, si tant est qu'elle existe, ne peut être que le fruit d'une coopération. Le cadre de notre étude est celui des systèmes artificiels. Après avoir rappelé un certain nombre des éléments qui caractérisent la coopération homme/homme, nous analysons ceux susceptibles de servir à la définition d'une coopération homme/machine. Dans un tel contexte, nous définissons la coopération dans la phase d'utilisation du système. Néanmoins, il serait réducteur de ne voir que deux partenaires dans cette coopération. En effet, l'envisager ainsi reviendrait à ignorer le rôle joué par l'ensemble des acteurs autour du système artificiel lors de sa conception. En outre, l'utilisation du système artificiel est une condition nécessaire à la coopération et par conséquent à la notion d'intelligence collective. Cette approche nous conduit à spécifier la notion de système artificiel coopératif, c'est-à-dire un système artificiel doté de capacité de coopération pour l'accomplissement de tâches communes.

1- Introduction

L'intelligence collective², telle qu'elle est envisagée dans cet appel à communication désigne le supplément d'intelligence qui naît de l'interaction entre individus, permettant à des groupes d'accomplir des tâches difficiles, de résoudre des problèmes et de s'adapter à des environnements variés et changeants sans coordination centrale.

¹Cette recherche a bénéficié des discussions menées au sein du groupe de recherche COOP du PRC-IA .

² Nous émettons de fortes réserves sur l'emploi d'une telle expression qui est en outre définie comme un "supplément d'intelligence". Mais nous adoptons pour la suite de cette contribution, la terminologie proposée dans l'appel à communication, à charge de discuter lors de ces journées sur le choix de ces expressions.

Cette production d'intelligence collective ne peut selon nous qu'émerger d'une coopération où l'accroissement des connaissances concernent chacun des partenaires. Cette coopération est caractérisée par une finalité commune aux différents partenaires qui participent ou agissent à sa réalisation. Un résultat tangible constitue l'aboutissement de la coopération qui pour une part est un supplément d'intelligence. En outre, l'étymologie du terme coopération met l'accent sur l'idée d'un partage des bénéfices issus de la coopération au prorata de la participation. Ainsi, des intérêts individuels et collectifs sont au centre de la coopération.

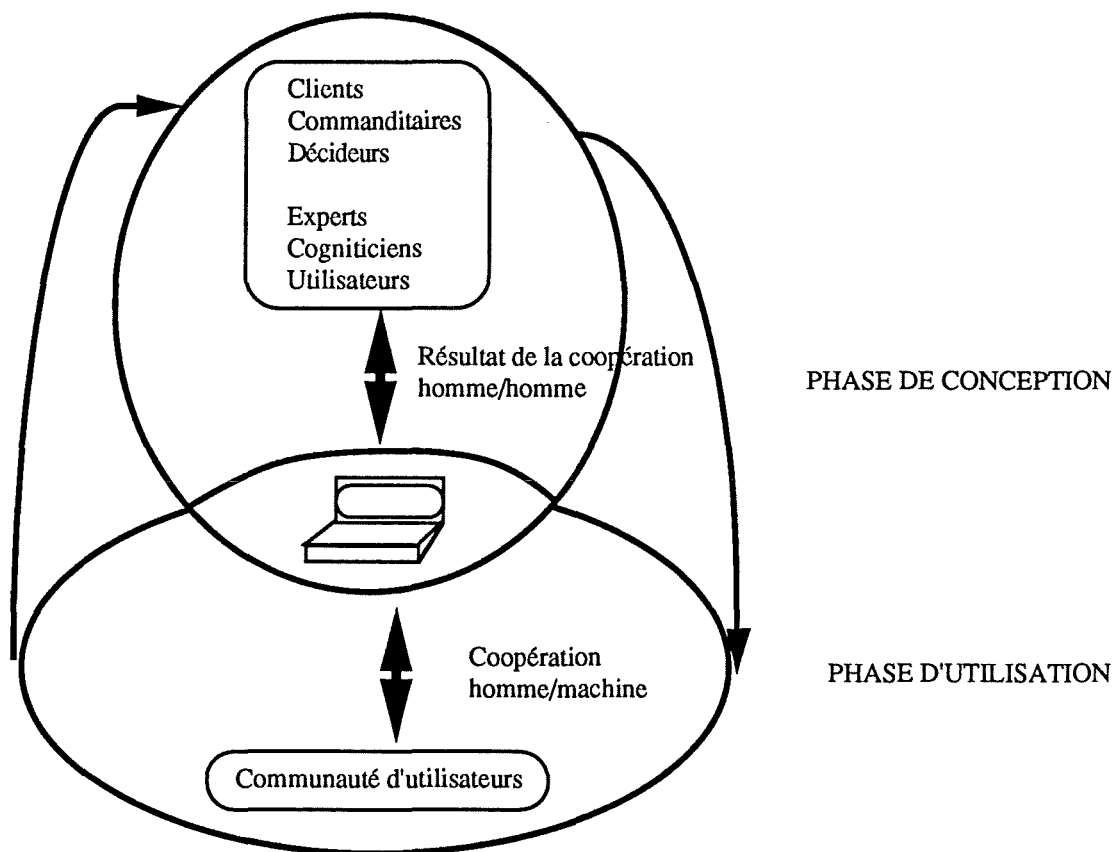
Notre étude se situe dans le cadre des systèmes artificiels. Par conséquent, les conditions de production d'intelligence collective doivent être réexaminées dans le cadre d'une interaction homme/machine. Il s'agit d'une situation particulière dans la mesure où l'un des participants au moins est une machine. Dans un tel contexte, une condition essentielle pour que naissent de l'intelligence collective de l'interaction homme/machine est l'utilisation du système artificiel dans son site d'implantation. Or, ils ne sont pas toujours utilisés ou mal utilisés. Pour améliorer cette situation, une approche nouvelle consiste à considérer comme indispensable l'instauration d'une réelle coopération entre l'utilisateur³ et le système artificiel [Brygoo, Després, Petolla, 1992].

Dans le cas d'une interaction homme/homme, la coopération n'est possible qu'une fois que les règles qui la régissent sont acceptées par tous. De façon analogue, dans le cas d'une coopération homme/machine des règles devront être établies et rendues explicites pour le système artificiel. La production de l'intelligence collective issue d'une telle coopération aura un impact au niveau organisationnel, social et individuel et se traduira différemment selon qu'il s'agit de l'opérateur, du système artificiel ou de l'ensemble des acteurs (concepteurs, décideurs, commanditaires, clients, ...) appartenant à la communauté constituée autour du système. Nous sommes ainsi conduits à spécifier la notion de système artificiel coopératif, soit un système artificiel doté de capacité de coopération pour l'accomplissement de tâches communes.

³ Si nous reprenons la distinction faite en ergonomie cognitive, l'utilisateur *utilise* le système tandis que l'opérateur *agit en utilisant* le système. La problématique de la coopération telle que nous l'envisageons, nous conduit naturellement à utiliser le terme opérateur.

2- Spécificité de la coopération opérateur/système artificiel

D'une façon générale un système artificiel possède deux statuts : celui d'objet de réalisation des concepteurs dans la phase de conception et celui d'interlocuteur des utilisateurs dans la phase d'utilisation. Nous schématisons ces deux statuts recouverts par le système artificiel et les deux phases dans la figure suivante.



Examinons maintenant les rôles respectifs de l'opérateur et du système artificiel dans l'interaction homme/machine dans cette dernière phase. L'opérateur utilise un système artificiel pour réaliser une tâche qu'il ne peut pas effectuer seul. Il y a donc obligation pour l'opérateur d'utiliser cet outil afin de réaliser sa tâche. Le système quant à lui ne peut exister qu'au travers de son utilisation. Les interfaces homme/machine sont conçues pour qu'il existe une interaction avec l'opérateur. La notion d'interaction est contenue dans l'idée de coopération, ainsi il ne peut pas exister de coopération sans interaction. C'est en ce sens que l'interaction constitue les prémices de l'installation d'une coopération. La coopération réclame la conscience d'un but commun, et par conséquent des échanges

"Vers une spécification de la notion de système coopératif"

Anne Brygoo - Sylvie Després

d'informations plus riches. Les participants possèdent un plus grand niveau d'autonomie pour la réalisation de la tâche. Le concept central pour la conception de la relation entre les systèmes artificiels et les opérateurs est celui d'aide. Le cadre dans lequel nous nous plaçons induit une reconnaissance de l'expertise de l'opérateur, il s'agit de fournir aux opérateurs un outil pour les aider dans leur activité. C'est dans la phase d'utilisation que la coopération se concrétise entre l'opérateur et le système artificiel, mais elle devra être traitée dès la phase de conception.

A partir de certaines études consacrées à la coopération homme/homme nous dégagons des éléments pour spécifier la coopération homme/machine.

Dans une interaction homme/homme, la coopération passe par une volonté de coopérer. Tandis que dans le cas d'interfaces homme/machine, la coopération passe par le besoin⁴ que l'un a de l'autre. Selon l'acteur de la communauté autour du système la nature des besoins diffère. Le besoin du système artificiel est dépendant des acteurs de cette communauté et ne pas tous les prendre en compte conduit à l'échec de la coopération.

Dans un article consacré à la créativité des groupes J.C. Abric [Moscovici, 1984] dégage de recherches expérimentales des résultats susceptibles d'éclaircir l'activité de groupe. Le fait d'être en groupe, l'interaction sociale et cognitive peut produire une baisse de résistance à la notion de changement et favoriser par là l'émergence de conduites nouvelles. L'hétérogénéité du groupe, aussi bien en attitudes qu'en aptitudes est une richesse potentielle puisqu'elle permet la confrontation avec des idées différentes et de l'utilisation de compétences autre que les siennes. C'est dans la mesure où plusieurs partenaires d'une même tâche vont être amenés à formuler des points de vue différents sur la perception de cette tâche et de sa résolution qu'ils pourront progresser. Ces points de vue différents vont les amener à coordonner leurs perspectives au départ divergentes pour accéder à une représentation la plus objective et la plus intégrée possibles des objets sur lesquels ils travaillent. Mais l'hétérogénéité peut aussi être source de blocages et de difficultés : inhibition face à des membres du groupe perçus comme plus compétents, blocage de la communication lié à des conflits d'attitudes. Les travaux de Hall et Watson [1970] permettent de mettre en évidence que l'hétérogénéité du groupe favorise la créativité pour autant que ces conflits sociaux soient pris en compte et régulés dans la situation. Dans le cas d'une coopération opérateurs/système, l'hétérogénéité du groupe d'opérateurs devra constituer une richesse potentielle dans la confrontation des solutions

⁴ La notion de besoin ne précède pas forcément la notion d'utilisation.

différentes proposées par le système ou les opérateurs. Les sources de blocage demeureront et se traduiront par des conflits qu'il conviendra de gérer.

Selon Guerin & al. [1991], l'activité d'un opérateur à un instant donné est le résultat d'un compromis complexe tenant compte de différents facteurs : des facteurs externes à l'opérateur, les objectifs fixés par l'entreprise, les moyens mis à disposition, les facteurs internes, les propriétés générales de l'organisme humain et les caractéristiques particulières, stables ou instantanées de l'opérateur considéré, les propriétés générales du raisonnement humain, les savoirs acquis par l'opérateur au cours de son histoire et l'orientation particulière de son raisonnement à un moment donné mais aussi la personnalité de l'opérateur et ses projets individuels.

Le compromis complexe évoqué se traduit dans le cas d'une interaction homme/machine par l'adaptation des utilisateurs aux systèmes artificiels dans la phase d'utilisation, mais il est nécessaire de le prendre en compte dès la phase de conception. Or les concepteurs tiennent rarement du rôle joué par chaque acteur et de son comportement face au processus informatique pourtant déterminant. Dans l'illusion informatique, F. Pavé [1989] parle de tricherie généralisée pour désigner la "capacité à coopérer" de chaque acteur, c'est-à-dire sa capacité à se réapproprier l'utilisation de l'informatique dans sa propre logique de fonctionnement [Assié, 1992]. Il convient en outre, d'étudier la répartition des tâches au cours de la coopération dans les phases de conception et d'utilisation. Dans cette dernière phase, le système est dans son site d'implantation et les utilisateurs dans leur environnement de travail.

De la même façon que l'on doit, d'un point de vue éthique, se poser le problème de l'appartenance du savoir-faire de l'expert et du contrôle de l'expert sur ses connaissances, la possibilité, pour l'utilisateur, de connaître l'existence d'un modèle le concernant, d'accéder au contenu de ce modèle, et de le modifier doit être discutée. Cette connaissance peut certes entraîner chez l'utilisateur une attitude tentant de cacher ses buts d'utilisation [Wahlster W. & al., 1986] et ce d'autant plus que de nombreux modèles contiennent la notion même d'évaluation, évaluation de la compétence par rapport au savoir de l'expert qui sert donc de référence et non par rapport à la tâche à effectuer par l'utilisateur. Dans le cadre d'une coopération entre le système et une communauté d'utilisateurs, de nouveaux problèmes concernant le contrôle sur les connaissances se trouvent posés. En effet, l'utilisateur considéré comme expert de sa tâche, se trouvera dépossédé par le système au même titre que l'expert, de savoirs privés et de leur évolution. En contrepartie, la coopération ne sera effective qu'en échange de la possibilité

"Vers une spécification de la notion de système coopératif"

Anne Brygoo - Sylvie Després

pour l'utilisateur de contrôler le modèle le concernant. Il s'agit en fait de recréer une transparence au niveau du système.

Dans des situations de travail, il est très fréquent que l'activité d'un opérateur s'articule avec celle de ses collègues. Les aspects collectifs de l'activité recouvrent différentes formes, par exemple la collaboration explicite pour la réalisation conjointe d'une même tâche, les aspects collectifs qui se manifestent seulement dans les résultats du travail, l'activité simultanée de travailleurs poursuivant des objectifs différents et les activités de régulation structurale. Dans le premier cas les différentes personnes engagées dans une action prélèvent de l'information sur le déroulement de l'action des autres, de manière à pouvoir ajuster leurs modes opératoires en temps réel [Guerin & al., 1991]. Dans le cas d'une coopération opérateurs/système artificiel le prélèvement d'informations sur le déroulement de l'action du système et de l'opérateur devra également être effectué. Une aide à la modélisation de cette coopération pourrait être trouvée dans le domaine des Computer Supported Cooperative Work (CSCW) [Seel, 1991], [Wilson, 1988], [Woods & al., 1988].

Quand on sait le rôle fondamental de la langue dans la coopération homme/homme, il convient de considérer celui de la place de la langue dans le cadre d'une coopération homme/machine.

3- Vers la spécification de la notion de système artificiel coopératif

Un système artificiel coopératif, c'est-à-dire un système doté de capacité de coopération pour l'accomplissement de tâches communes, devra donc posséder des fonctionnalités pour résoudre des problèmes⁵ de manière coopérative. Dans cette première tentative de spécification, nous limitons notre réflexion à la façon dont la mise en commun des connaissances appartenant aux différents partenaires peut se faire et au partage des tâches entre le système et les opérateurs. Le système devra posséder des fonctionnalités lui permettant de produire des explications, d'argumenter et de négocier. L'architecture du système dépendra directement des fonctionnalités précédemment définies. Dans ce contexte, les techniques de l'intelligence artificielle distribuée (IAD), nous semble pouvoir constituer une réponse appropriée à l'élaboration d'un tel système.

⁵ Il reste alors à caractériser les classes de problèmes susceptibles d'être résolus par ce moyen. Nous nous référons dans une première étude au texte de Gasser [1990]

3.1 Comment se fait la répartition des tâches ?

La coopération doit permettre une aide mutuelle évolutive et adaptative pour la réalisation d'un but commun dans des conditions claires et définies pour chacun des participants. Dans ce contexte d'assistance mutuelle, l'utilisateur doit avoir une part de responsabilité dans le succès de la réalisation de la tâche où il se trouve impliqué et le système doit être un élément déterminant pour la réalisation de la tâche allouée à l'opérateur. Dans la mesure où la finalité du système est son utilisation, le rôle qu'il joue dans l'assistance devient évident.

Le développement d'un référentiel opératif commun aux opérateurs et au système artificiel est une condition nécessaire à la mise en œuvre d'une coopération. Il s'agit en effet de mettre en commun des connaissances pour la résolution d'un problème ou l'accomplissement d'une tâche, et de répartir les tâches dans la phase d'utilisation entre le système et les opérateurs.

La reconnaissance des partenaires (système artificiel et les opérateurs) en tant qu'experts conduit à poser la question de la mise en commun des connaissances en terme de multiexpertise. Le système devra donc être doté de fonctionnalité de négociation pour gérer les conflits qui peuvent survenir entre les partenaires.

Le partage des tâches doit se faire au mieux des possibilités de chacun des partenaires impliqués dans la coopération. Woods [1986] constate ainsi que le raisonnement humain se caractérise par certaines faiblesses, telles que : la tendance à se fixer sur certains états du processus contrôlé, la difficulté à utiliser le raisonnement contre-factuel, la difficulté à prendre en compte des faits qui invalident une hypothèse, la difficulté à sélectionner et à intégrer les informations pertinentes, les difficultés à évaluer les conséquences d'interactions complexes. Les êtres humains sont différents des machines parce que leur charge de travail peut affecter leur performance. En particulier, la charge mentale [Parks, 1977; Boy, 1988] peut être prohibitive en résolution de problèmes ou en contrôle distribué. En outre, le modèle théorique de la performance d'un système homme-machine établi en fonction de différents niveaux d'autonomie et de différents niveaux de compréhension ou de connaissances sur le système homme-machine met en évidence que la performance du système homme-machine croît avec l'autonomie de la machine, jusqu'à un optimum, après lequel elle décroît.

De nombreux travaux portent sur l'étude de l'interaction homme/machine et devrait nous permettre de mieux définir les fonctionnalités nécessaires au développement d'un système artificiel coopératif. D'une façon générale, dans les domaines où il existe des systèmes informatiques les conditions d'associations de l'homme et de la machine sont

variées. Pélegrin [Boy, 1988] distingue trois types principaux : l'homme surveille, supervise, ou commande. Lorsque l'homme surveille, le système fonctionne en mode automatique tant qu'une intervention humaine ne désactive pas ce mode; si l'homme supervise, le système fonctionne en mode automatique, mais l'homme peut prendre le contrôle à tout instant; dans le cas où l'homme commande, le système est construit pour fonctionner avec une forte interaction de l'homme qui en assure la commande et le contrôle en permanence. Ces trois types d'associations induisent des modes d'interaction où le degré d'interactivité dépend fortement de la nature de l'association et entraînent la conception de différentes sortes d'interfaces. Ils suggèrent en outre l'idée que la notion de système implique celle de frontière qui induit celle de partage entre l'homme et la machine. La délimitation d'une frontière nécessite la définition ou à défaut l'estimation des interactions entre l'intérieur et l'extérieur du système. Ainsi, la décomposition d'un système en sous-systèmes induit des interactions entre ces derniers qui sont décidées et programmées par le concepteur et la plupart du temps méconnues de l'utilisateur. L'utilisateur se construit alors une image mentale de ces interactions fondée sur les tâches qu'il doit accomplir et non sur le système lui-même. L'analyse de la notion de coopération dans le cadre des interfaces se situe pour nous dans cette zone frontière. Dans le cas de systèmes dits intelligents, entre l'intérieur et l'extérieur du système, se situe un espace de communication avec l'utilisateur qui permet de gérer des dialogues, de produire des explications, d'argumenter des choix etc. Le degré de connaissances du problème de chacun des protagonistes donne lieu à trois types d'association qu'il est possible de différencier selon la position du système et de l'utilisateur relativement à leur degré de connaissances du problème : le système sait et domine, l'interlocuteur sait et domine, le système et l'interlocuteur savent sans domination. Dans le premier cas, l'explication produite par le système peut permettre un transfert des savoir-faire en fournissant des explications analytiques; dans ce cas l'aide fournit apparaît comme un commentaire permettant à l'utilisateur de résoudre son incompréhension. Dans le second cas, un dialogue est mis en place pour permettre à l'expert du domaine d'examiner, de rejeter des résultats du système dans le but d'améliorer le modèle construit par apprentissage. Le dernier cas survient quand le système dispose de connaissances acquises par des méthodes d'apprentissage par l'exemples. L'indétermination sur le résultat de l'interaction du système et de l'interlocuteur crée une situation d'investigation et de confrontation des connaissances.

"Vers une spécification de la notion de système coopératif"

Anne Brygoo - Sylvie Després

Transparence pour les partenaires de la coopération

La notion de transparence a deux aspects : voir au travers d'un corps et voir à travers ce corps. Il s'agit pour nous ici de la deuxième acception à savoir la qualité de ce qui laisse paraître la réalité toute entière et non la qualité d'un corps qui permet de distinguer avec une certaine netteté ce qui est derrière lui. La transparence est selon nous intimement lié à la multiexpertise mais également à celle de partage des tâches et par conséquent à la notion même de coopération.

Dans le cadre des SIAD, la notion de multiexpertise est mise en œuvre par l'utilisation de systèmes d'aide à la décision de groupe GDSS [Ram et al.,1990] qui constituent des outils d'aide à l'acquisition des connaissances à partir de plusieurs experts. Les GDSS présentent plusieurs avantages : la documentation automatique de l'expertise afin d'éviter les erreurs associées à la transcription (informations oubliées, biais ou erreurs dans l'interprétation) des interviews d'experts. En outre, un GDSS permet aux experts de travailler simultanément et d'accéder aux commentaires de chacun d'entre eux tout en préservant leur anonymat. Cet anonymat constitue un avantage majeur lorsque une résolution de conflit ou un vote sont nécessaires, il prémunit aussi contre une personnalité dominante qui peut imposer ses idées. Enfin, toute personne à accès à la connaissance fournie par les autres ce qui facilite en principe les résolutions de conflits dans la mesure où la discussion est possible au début de la phase d'acquisition des connaissances. La notion de coopération entre individus est réalisée, il y a un problème commun des expertises distinctes et une finalité commune qui consiste à produire une aide. La coopération aboutit au prix d'une transparence sur la connaissance de l'autre ce qui autorise des négociations pour la résolution de conflits.

Dans le cadre d'une coopération entre le système et une communauté d'opérateurs, de nouveaux problèmes concernant le contrôle sur les connaissances se trouvent posés. En effet, l'opérateur considéré comme expert de sa tâche, se trouvera dépossédé par le système au même titre que l'expert, de savoirs privés et de leur évolution. Le système coopératif devra donc être conçu de façon à rendre accessible aux opérateurs les connaissances accumulées sur eux au cours de la coopération. Il s'agit en fait de recréer une transparence au niveau du système. C'est à cette condition que la coopération pourra être réussie et par conséquent permettre la résolution du problème.

Il conviendra en outre de considérer la remarque de Pavé [1989] concernant la transparence : L'informatisation étant étroitement liée à la connaissance rationnelle, génère, de ce fait, de la transparence. Or, la complexité des systèmes sociaux peut parfois être qu'elle produise des situations dans lesquelles même une gestion transparente n'en

"Vers une spécification de la notion de système coopératif"

Anne Brygoo - Sylvie Després

permet pas une maîtrise satisfaisante. Néanmoins la transparence pour tous est productrice d'effets inattendus pour tous.

4- En conclusion

Nous sommes loin d'une spécification complète d'un tel système artificiel coopératif. Néanmoins, il nous semble que la réalisation et la simulation de systèmes artificiels capables de manifester de l'intelligence collective dépend de la capacité du système à coopérer avec une communauté d'opérateurs. Cette vue de la coopération dans la phase d'utilisation implique de penser les choses autrement dans la phase de conception.

5- Références bibliographiques

- Abrie J.C. - "La créativité des groupes" in Psychologie sociale PUF fondamentale, 1984.
- Assié S. - Des systèmes opérants aux systèmes de pilotage : changements de risques de l'informatisation. Colloque du CREIS Chances et risques de l'informatisation dans l'Europe de demain. Paris, 10-12 juin 1992.
- Boy
- Brygoo A., Després S., Petolla S. - Quels utilisateurs pour quel système à base de connaissances, in actes du congrès Ergo-IA-92, Biarritz, octobre 1992.
- Brygoo A., Després S., Petolla S. - Qu'est-ce-qu'un système coopératif d'assistance ?, rapport de recherche EHEI, à paraître.
- Boy G. - Assistance à l'opérateur. Une approche de l'Intelligence Artificielle. Toulouse, Teknea, 1988, 360p.
- Gasser L., Hunhs M.N. (eds) - Distributed artificial intelligence, Pitman Publishing-Morgan Kaufman, vol.2, 1989.
- Guérin F., Laville A., Danielou F., Duraffourg J., Kerguelen A. - Comprendre le travail pour le transformer. La pratique de l'ergonomie. Collections outils et méthodes. Editions ANACT 1991.
- Hall J., Watson W.H. - Effects of size and task type on group performance and member reactions, Sociometry, 33, 37, 1970.
- Parks, 1977
- Pavé F. - L'illusion informaticienne, Collection "Logique sociales", édition L'Harmatan, 1989.
- Ram et al., 1990

AFCET - Journées de Rochebrune
Janvier 1993

Proposition de communication

**L'intelligence collective entre le naturel, l'artificiel
et le social.**

**Laure BAZZOLI, Emmanuelle CONESA,
Jean-Pierre MICAELLI**

**Economie des Changements Technologiques
URA CNRS 945
Université Lumière - Lyon 2**

Le thème de l'intelligence collective intéresse le modélisateur de systèmes vivants car il voit là une propriété intrinsèque du vivant. Ce thème intéresse aussi le concepteur de systèmes artificiels dans la mesure où les systèmes intentionnellement dotés d'intelligence collective ont des capacités calculatoires et adaptatives intéressantes pour simuler ou gérer des phénomènes complexes. Mais qu'en est-il pour l'économiste ou, plus généralement, pour le spécialiste en sciences sociales ?

L'économie est une discipline qui explique, aux différents niveaux où s'observe l'activité économique, les causes et les effets économiques des opérations rationnelles de production et de consommation. Par rationnel, il faut entendre ce que Max Weber appelait la "rationalité en moyens" ou "*zweckrational*", c'est à dire la conduite par laquelle des moyens sont associés à fins stables et données. Or l'analyse sociale ou anthropologique montre que le contenu de la rationalité de l'acteur économique dépend étroitement du contexte géographique et historique dans lequel il se trouve. Hormis le cas d'école d'un individu agissant parfaitement seul sur une île déserte, cas connu sous le nom de "robinsonnade", l'économie suppose qu'il existe des interrelations profondes et complexes entre l'individu et l'institution ou la société auxquelles il appartient. La méthode qui nous permet d'appréhender ces interrelations n'est pas expérimentale mais compréhensive. En sciences sociales, l'appréhension de la réalité ne se limite pas à une observation empirique et à une mesure des phénomènes observés. Du fait de l'immersion de ses modèles et du sujet connaissant dans la réalité sociale, la compréhension du réel

consiste en une interprétation de la réalité : quels ont été les motifs de tel comportement, quelles en ont été les finalités, comment l'acteur a-t-il interprété la réalité ?

On ne sera dès lors pas surpris de l'origine proprement métaphysique des modèles de l'économie.

Deux métaphysiques sociales s'opposent en économie. Pour l'école atomiste, seul l'individu existe : la société n'est qu'un réseau de relations interindividuelles. Issue de la tradition aristotélicienne, l'école holiste ou "institutionnaliste" affirme au contraire que l'individu ne peut exister sans les institutions qui définissent son appartenance à une cité. Selon les postulats retenus, la définition de ce que recouvre la notion d'intelligence collective change radicalement.

Dans l'approche atomiste, l'intelligence collective est appréhendée comme la propriété qu'a un ordre social à s'établir et à perdurer indépendamment des actions individuelles. Selon Friedrich von Hayek, l'apparition de cet "ordre spontané" ou "cosmos" s'explique par une double procédure, caractéristique de l'intelligence collective sociale. A un moment donné de l'histoire d'une société, les individus sont liés entre eux par des relations derrière lesquelles leurs spécificités s'effacent. Le système qui assure le plus efficacement ce rôle est le marché dans la mesure où toutes les informations nécessaires à l'action rationnelle et tous les effets de la dite action s'expriment par des variations de prix. Le marché assure ainsi l'apparition de stabilités globales que des rencontres individualisées ou centralisées ne permettent pas d'obtenir. Une seconde procédure complète la première. En effet, au cours du temps, les générations sélectionnent par essais et erreurs les institutions qui garantissent un ordre social satisfaisant -l'économie de marché par exemple-, si bien qu'il est impossible de dire à un moment donné qui a conçu *ex nihilo* les institutions qui satisfont les besoins de la génération concernée. En résumé, l'approche atomiste distingue deux vecteurs principaux de l'intelligence collective : le marché et la tradition ou la coutume (au sens anglo-saxon de "law").

Dans une approche holiste ou institutionnaliste, l'intelligence collective est comprise comme la capacité qu'a une institution à adapter consciemment ses règles afin que le projet collectif qui la caractérise s'adapte mieux et, dans un contexte de concurrence entre institutions, plus originalement, aux évolutions de ses contraintes internes ou externes. Les deux vecteurs de l'intelligence collective relèvent alors de catégories téléologiques. Il s'agit de la gouvernabilité, ou capacité à gouverner, et de la réflexivité, où capacité à réfléchir sur son mode de contrôle sur les choses, les individus et les institutions. L'expression historique et géographique de ces propriétés se manifeste à travers la présence de règles politiques qui assurent l'obtention de compromis satisfaisants entre

individus, coalitions ou institutions opposées, mais aussi à travers le langage qui aide une communauté à avoir un sens partagé de son passé, de son présent et de son devenir.

La distinction entre les approches atomiste et holiste est d'ordre épistémologique et méthodologique. Elles représentent les bornes de l'intervalle à l'intérieur duquel oscille toute théorie de l'intelligence collective sociale. Leurs limites respectives sont aujourd'hui bien connues. L'histoire très récente montre qu'une approche atomistique n'est qu'une rétroprojection de siècles de reconnaissance de l'individu autonome comme fin de la société, et qu'une société de marché ne représente pas l'état premier d'une société. L'approche holiste ne cesse d'affirmer qu'il faut des institutions pour réguler toute société, mais ne peut déterminer précisément lesquelles, lorsque s'opèrent de vastes transitions politiques et économiques.

Il nous faut donc aujourd'hui tenter d'imaginer une approche de l'intelligence collective sociale qui tienne compte du caractère irréductible de l'individu, et des institutions ou de la société auxquelles il appartient. Nous qualifierons cette approche d'interactionniste.

Dans le cadre interactionniste, l'intelligence collective est appréhendée comme la capacité qu'a une institution à se définir comme un soi. Cette définition abstraite correspond à des faits sociaux actuels très précis. Mentionnons tout d'abord, au niveau des entreprises, l'existence d'un processus de rationalisation cognitive. Se trouvant confrontée à des problèmes explicites et stratégiques de gestion de la connaissance, l'entreprise ne doit plus seulement assurer la répartition de savoirs techniques en son sein, mais doit surtout s'intéresser aux démarches d'apprentissage, afin de faciliter l'acquisition individuelle de structures cognitives et communicationnelles qui permettent d'apprendre collectivement ce que l'on sait collectivement faire. Mentionnons ensuite, au niveau d'institutions diverses (entreprises, administrations publiques, hopitaux, organisations professionnelles,...), l'existence d'un processus de rationalisation éthique concomitant au processus de rationalisation cognitive. Un nombre croissant d'institutions, dont certaines sont spécialisées, ont ainsi vocation à définir de nouvelles règles de fonctionnement et d'évaluation de leurs projets et de leurs critères de performances du fait de la diffusion au sein de la société de nouvelles pratiques et de nouvelles valeurs.

Insister sur les différences conceptuelles entre l'intelligence collective naturelle, artificielle et sociale est important mais insuffisant. Nous voudrions en conclusion définir les rapports épistémologiques, méthodologiques ou pratiques plus précis qui peuvent exister entre elles. L'intelligence collective des systèmes naturels ou artificiels peut servir de métaphore, voire d'heuristique, aux sciences sociales dont le projet est de décrire le fonctionnement d'institutions ou de sociétés observables. Mais elle peut tout aussi bien

servir de technique, c'est à dire de modèle, de principes et d'outil concret, à une ingénierie cognitive qui ne serait plus seulement individuelle, mais viserait, en se combinant avec la théorie économique de l'organisation, à concevoir et gérer des organisations finalisées, dotées a priori d'intelligence collective. En tout état de cause, les modèles d'intelligence collective naturelle ou artificielle ne sauraient être vus comme de simples modèles de simulation d'une réalité expérimentale définissable sans d'autres problèmes que des problèmes de mesure. La difficulté du pontage entre les formes d'intelligence collective est alors considérable, puisqu'il s'agit d'arriver à combiner des approches compréhensives, nécessaires pour comprendre des faits historiques et géographiques précis, à une approche formelle, nécessaire à la modélisation, donc à un autre mode d'intelligibilité du monde.

Lyon, le Lundi 12 Octobre 1992

Détection collective

jean louis Dessalles
Télécom Paris - Département Informatique
46 rue Barrault - 75634 PARIS Cedex 13
Tel.: (1) 45 81 75 29 - Fax: (1) 45 81 31 19
E-✉: dessalles@enst.fr

- janvier 1993

résumé

Pour qu'une entité puisse être qualifiée d'intelligente, il faut qu'on soit en mesure de lui prêter des capacités de perception et/ou des capacités d'action suffisamment intégrées pour qu'elles soient adaptables. En particulier, elle doit être capable de *détecter* les situations qui appellent un changement de représentation ou une action appropriés.

Nous considérons ici plus particulièrement le cas d'une entité capable de percevoir de manière intégrée, et dont les différents niveaux d'intégration, y compris le niveau le plus élevé, restent distribués. Une telle entité peut être considérée de ce point de vue comme dépourvue de contrôle central. Le fonctionnement des différents niveaux, et notamment du niveau supérieur, tel que nous pouvons l'imaginer à partir de notre compréhension du système nerveux, peut nous fournir un modèle de certains processus collectifs intelligents, notamment les processus qui permettent à un système collectif de réaliser des opérations de détection sans centralisation d'information.

mots clés: détecteur, décision collective, émergence, connexionnisme, cohérence

1. Hirondelles et neurones

Un système qui ne disposerait pas du moyen de détecter les situations pertinentes dans lesquelles il doit agir ou modifier son état pourrait difficilement être qualifié d'intelligent. Il en est de même pour un système intelligent collectif. La question se pose alors de savoir comment une situation peut être collectivement détectée. De plus, lorsque la détection conduit à une décision (modification des représentations et/ou déclenchement d'action), nous devons comprendre comment la détection collective peut mener à la décision.

Prenons le cas d'une entité capable de réaliser une action qualitative particulière. Cette entité est supposée disposer de capteurs qui lui fournissent l'état de son environnement par l'intermédiaire

de paramètres qui prennent à chaque instant une valeur dans un ensemble éventuellement continu. Si cette entité est adaptée à son environnement, alors l'action n'est pas déclenchée au hasard, mais pour certains domaines de valeurs des paramètres. Le lien qui relie les paramètres à l'action est donc non linéaire, il comporte typiquement des seuils. Nous appelons détecteur un système non linéaire comportant deux états (actif et non actif), capable de s'activer dans des régions de l'espace des paramètres qu'il reçoit en entrée. Il peut transmettre son activation à d'autres détecteurs ou à des effecteurs. L'entité évoquée ci-dessus comporte nécessairement un détecteur. S'il s'agit d'une entité fortement intégrée, elle comportera de nombreux étages de détection.

Un neurone formel est un bon exemple de détecteur. Mais nous nous intéressons ici à des détecteurs collectifs, c'est-à-dire à des détecteurs formés de plusieurs éléments identiques et fonctionnellement équivalents. Donnons deux exemples de détecteurs collectifs.

Du point de vue du traitement de l'information, un humain peut être considéré grossièrement comme une collection de neurones. Personne n'imagine sérieusement que le niveau d'intégration le plus élevé, qui est éventuellement celui qui relie la perception d'une situation donnée à une action, puisse être assuré à un moment quelconque par un neurone unique. Au niveau neuronal, la reconnaissance d'une situation entraînant le déclenchement d'une action particulière est une reconnaissance collective.

Notre deuxième exemple provient d'une modélisation, sans doute sommaire et inexacte, du comportement migratoire des hirondelles. Mais peu importe l'adéquation entre le modèle et la réalité ornithologique, puisqu'il s'agit ici de montrer la faisabilité de certains types de détection collective.

Dans notre petit modèle concernant les hirondelles, la décision de migrer n'est pas non plus le fait d'une hirondelle unique. Les hirondelles du quartier se rassemblent un jour en grand nombre sur un toit ou sur des lignes électriques (alors qu'on ne les voit jamais se poser en temps normal). Elles ne se rassemblent qu'une seule fois, et ce jour-là est celui du départ. La détection du moment opportun apparaîtra dans notre modèle comme une détection collective, car nous décrirons ce phénomène comme une conséquence du comportement de chaque individu, sans qu'aucune centralisation de l'information n'ait lieu.

2. Détection collective et niveaux d'organisation

Le phénomène de détection collective est loin d'être marginal. Si l'on prend l'exemple d'une entité fortement intégrée comme le système nerveux, l'opération de détection n'intervient pas seulement au niveau d'intégration le plus élevé, mais aussi aux niveaux d'organisation intermédiaires. Le fait même de parler de niveaux d'organisation suppose qu'un élément du niveau n puisse être activé pour certaines configurations d'activation des éléments de niveaux $n-1$, et que les éléments de chaque niveau soient donc des détecteurs [dessalles 1992]. Nous pouvons émettre et explorer l'idée que ces détections intermédiaires, comme les détections du niveau le plus élevé, sont elles-mêmes collectives.

Le concept de détection collective devient alors essentiel pour la description des niveaux d'organisation d'un système intégré. Dans le cas du système nerveux, il permet de donner une plausibilité biologique à la notion de détecteur, qui n'est au départ qu'une notion abstraite postulée pour rendre compte de phénomènes perceptifs.

En effet l'organisation de la perception en niveaux de détection rend bien compte de phénomènes comme la perception d'objets émergents [dessalles 1992], elle est compatible avec le traitement hiérarchisé observé par exemple dans le système visuel, mais elle semble contradictoire

avec le fait généralement admis qu'aucune fonction perceptive n'est concentrée sur un seul neurone (inexistence des neurones "grand-mère"). Autrement dit le rôle de détecteur perceptif ne saurait être assuré par un neurone unique. Il s'agit alors de savoir comment les détecteurs sont matériellement représentés.

Dans la version localiste d'une hiérarchie de détecteurs câblés (figure 1), chaque détecteur de la couche n reçoit des connexions physiquement identifiables de détecteurs de la couche $n-1$. Une telle organisation permet d'assurer que les objets détectés par tous ces détecteurs dans une situation donnée sont structurés en une hiérarchie d'appartenance [dessalles 1992]. On peut ainsi expliquer des phénomènes comme la perception d'un objet émergent de niveau n , lorsqu'un détecteur de niveau n est soudainement activé par des détecteurs actifs de niveau $n-1$.

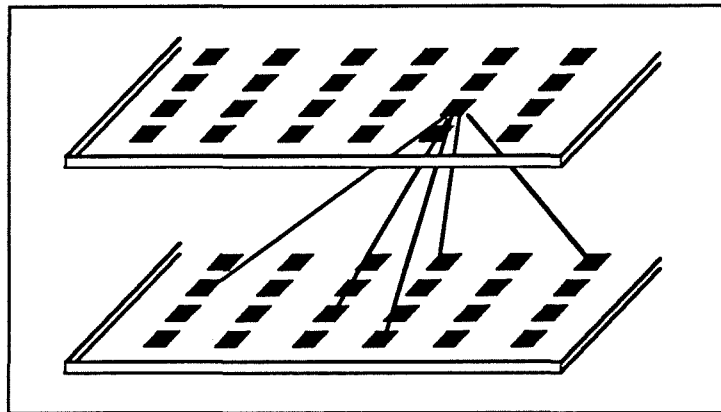


figure 1: hiérarchie de détecteurs câblés: chaque détecteur reçoit comme entrées les sorties de détecteurs du niveau immédiatement inférieur. On peut imaginer qu'un détecteur est une porte "AND", un neurone formel, ou comme suggéré ici, un groupe de neurones.

Pour prendre l'exemple du système visuel, notre perception des formes semble fortement hiérarchisée, comme les travaux classiques de D. Hubel & T. Wiesel ont commencé à le montrer il y a plus de vingt ans. L'intégration commence dès la rétine, et se poursuit jusqu'au néocortex. Schématisons-la par la chaîne suivante: pixels rétiniens → centres on-off → segments élémentaires orientés → contours → objets → scène. Les objets détectés dans une scène donnée par cette chaîne de détection sont bien dans une hiérarchie d'appartenance. Par exemple, les segments élémentaires d'un objet perçu appartiennent à cet objet. Cela est dû au fait que les détecteurs de contours sont alimentés par les détecteurs de segments élémentaires, comme on peut le vérifier en voyant "émerger" la spirale à droite de la figure 2. Les détecteurs de niveau contour, qui ne détectaient que la spirale carrée continue dans la figure de gauche, sont maintenant activés par la perception de segments éloignés en dimension 1 mais qui, avec l'angle de 89° , sont devenus voisins en dimension 2. Leur agencement provoque l'activation d'un détecteur de contour qui, lorsqu'elle est "soudaine", se traduit dans le sentiment perceptif que la spirale arrondie de droite émerge.

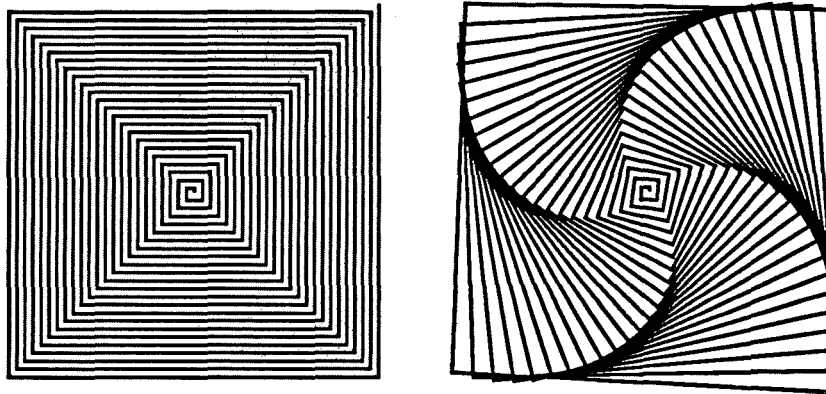


figure 2: La spirale quasi carrée (angle de 89°) créée par la tortue Logo à droite fait "émerger" une spirale arrondie à quatre branches (d'après [Lawler 1980]).

Le phénomène d'émergence est donc caractéristique d'une organisation en niveaux, et nous avons ainsi de bonnes raisons de penser que de nombreux systèmes naturels de perception intégrée sont réalisés par des détecteurs structurés en niveaux. Mais rien n'indique que le lien qui transmet l'activation d'un niveau vers le niveau supérieur adjacent soit de type localiste.

Si l'on considère le cas des systèmes neuronaux, des considérations biologiques et connexionnistes (distribution, fiabilité, taux d'intégration,...) laissent penser qu'au contraire, les détecteurs d'un niveau donné ne sont pas constitués par des neurones, mais par des groupes de neurones. Autrement dit, le schéma de la figure 1 explique ce qui se passe au niveau fonctionnel, mais une implémentation avec des détecteurs collectifs correspond à la structure décrite en figure 3.

Le lien qui permet à un détecteur de niveau $n-1$ d'activer un détecteur de niveau n n'est donc pas matérialisé par un axone. Les détecteurs des systèmes neuronaux naturels semblent distribués quel que soit leur niveau, et les neurones qui les composent réalisent donc des détections collectives.

Il nous faut donc tenter de donner un schéma plausible pour de tels détecteurs collectifs, en expliquant comment se manifeste l'activation et comment elle se transmet.

3. Détection collective et cohérence temporelle

Tout modèle de fonctionnement d'un détecteur collectif doit par définition respecter la contrainte de non-centralisation de l'information détectée. En particulier, l'activité du détecteur doit être portée par un phénomène physique défini au niveau collectif et dépourvu de signification au niveau d'un composant du détecteur. Il en est de même pour le système physique qui transmet l'activation: il doit lui aussi rester collectif.

Par quoi, donc, se traduit l'activation d'un détecteur collectif ? Pour quelques raisons que nous allons expliciter, nous suggérons que cette activation peut être caractérisée dans le cas neuronal par une activité cohérente des neurones qui composent le détecteur. Nous envisagerons ensuite l'extension éventuelle à d'autres cas de détection collective.

Dans le cas des neurones naturels, la cohérence se traduit par une émission synchronisée de potentiels d'action. Les neurones émettent spontanément des potentiels d'action, mais généralement sans coordination avec d'autres neurones. La cohérence temporelle est un bon candidat pour caractériser l'activation d'un détecteur collectif neuronal pour au moins trois raisons.

Tout d'abord, il s'agit d'une relation simple entre les neurones, et nous sommes "en droit" d'attendre une relation simple. En effet, le problème étant symétrique par nature, nous pouvons

exiger du motif d'activation qu'il soit lui-même symétrique, ou stochastiquement symétrique (i.e. rôle statistiquement équivalent pour chaque neurone). La synchronisation est l'une des relations symétriques les plus simples que l'on puisse imaginer. L'avantage d'une relation simple est que son déclenchement pourra être simple lui aussi. On comprend aisément comment la synchronisation des neurones d'un détecteur de niveau n peut être obtenue par synchronisation avec les neurones afférents de niveau $n-1$, autrement dit par l'action des détecteurs afférents actifs au niveau $n-1$. Nous détaillerons cet aspect plus loin.

Le deuxième argument vient de ce que la synchronisation neuronale est biologiquement vraisemblable. Elle est observée par exemple dans le cortex visuel, où l'on vérifie que les neurones répondant à une même zone du champ visuel, mais situés dans des aires différentes (comme l'aire V4 qui traite la couleur et l'aire V5 (aire MT) où les mouvements sont détectés), deviennent synchrones lorsqu'ils répondent au même objet du champ visuel [Zeki 1992]. Le même phénomène a été observé dans des zones sous-corticales comme le bulbe olfactif [Freeman 1991].

La troisième raison qui nous pousse à considérer la synchronisation comme un phénomène possible pour marquer le résultat d'une détection collective vient de notre connaissance de mécanismes inter-neuronaux qui la rendent possible. Par exemple, une stabilisation des synapses qui suit la loi de Hebb peut rendre un réseau de neurones artificiels *cohérent* : deux neurones d'un réseau cohérent seront soit toujours simultanément actifs, soit jamais actifs en même temps.

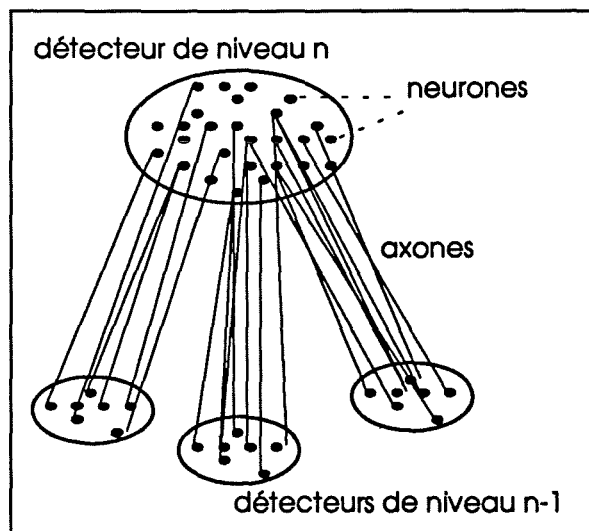


figure 3: implémentation neuronale de deux niveaux de détection collective. Chaque détecteur correspond à un carré élémentaire dans le schéma de la fig. 1, et un groupe d'axones implémente un trait de la fig. 1. Les neurones d'un détecteur reçoivent des synapses excitatrices de leurs voisins, ainsi que des synapses excitatrices en provenance de neurones appartenant à des détecteurs de niveau inférieur (ou à des capteurs). Si peu de détecteurs de niveau $n-1$ sont actifs de manière synchrone, peu de neurones du détecteur de niveau n peuvent sommer une activité suffisante, et cette activité "se perd" dans les connexions latérales. En revanche, l'activité d'un nombre suffisant de détecteurs de niveau $n-1$ permet au détecteur de niveau n de "s'embraser": ses neurones deviennent synchrones.

Il reste à expliquer comment l'activation est transmise d'un niveau à l'autre, et dans le cas des neurones, grâce à quel câblage. Il s'agit ensuite de se demander comment un tel câblage peut s'établir.

4. Déclenchement d'une détection collective neuronale

La notion de niveaux de détection est au départ une notion abstraite. Elle est postulée pour rendre compte notamment de certains effets perceptuels (émergence, illusions, etc.). Elle reste remarquablement compatible avec ce que l'on sait de l'organisation neuronale naturelle, mais sa plausibilité suppose que nous puissions imaginer un mécanisme vraisemblable pour le fonctionnement des détecteurs.

Nous sommes donc confrontés à la question de savoir comment un détecteur collectif peut être déclenché par l'activité de détecteurs collectifs du niveau inférieur, autrement dit nous devons préciser le type de câblage, au niveau neuronal, qui permet d'"alimenter" les détecteurs collectifs. Plus précisément, dans notre optique, il s'agit de comprendre comment l'activité synchrone de neurones de niveau $n-1$ peut déclencher l'activité d'un groupe de neurones donné au niveau n . Il s'agit en quelque sorte de définir ce qui caractérise les niveaux d'un point de vue structurel.

Des détecteurs collectifs cohérents

L'hypothèse la plus simple que nous pouvons faire semble être qu'un détecteur collectif est formé d'un ensemble cohérent de neurones: les connexions internes au détecteur sont excitatrices, et les connexions avec des neurones externes sont inhibitrices¹. Il peut s'ensuivre un phénomène de seuil, caractéristique de tout détecteur: si très peu de neurones du détecteur sont actifs de manière synchrone, alors les excitations qu'ils propagent ont peu de chance d'être sommées par d'autres neurones du détecteur. En revanche, sous l'effet d'une excitation synchrone extérieure, un nombre suffisant de neurones du détecteur peuvent devenir actifs et ainsi "embraser" le détecteur (cf. figure 3).

Un tel schéma suppose l'existence d'un câblage excitateur interne au détecteur, et d'un câblage excitateur en provenance des détecteurs afférents. Et la question se pose de savoir comment un tel câblage peut s'établir. Cette question est essentielle dans les systèmes naturels, pour lesquels il faut résoudre le contraste entre l'apparente complexité du câblage (10^{15} synapses) et la relative simplicité de son codage génétique (moins de 10^5 gènes).

Des détecteurs collectifs pseudo-locaux

Sans prétendre fournir une réponse précise, nous pouvons rappeler ici quelques éléments de réponse souvent mentionnés dans des contextes similaires. Le câblage excitateur interne peut être associé à la topologie des détecteurs. Si l'on pense à des détecteurs constitués de neurones proches (détecteurs collectifs pseudo-locaux), alors un système de connexions latérales rappelant le "chapeau mexicain" de T. Kohonen [1984] peut définir un ensemble de détecteurs collectifs avec un coût de codage architectural très faible.

De plus, le câblage excitateur afférent peut être lui aussi réalisé avec un faible coût grâce aux propriétés topologiques de tels détecteurs collectifs pseudo-locaux. Il suffit que des connexions en provenance d'un détecteur afférent se projettent à l'intérieur du détecteur cible. La mise au point fine peut être assurée par une sélection synaptique. Seules sont conservées les synapses qui sont fonctionnelles au sens de Hebb: lorsque tous les détecteurs afférents, ou une grande partie d'entre eux, sont actifs, le détecteur cible devient actif aussi, même si le câblage est grossier. Les synapses qui favorisent cette synchronisation sont alors stabilisées. En revanche les synapses qui se perdent dans des détecteurs voisins, non actifs, dégèrent. Une telle stabilisation peut s'opérer sans apport d'information extérieure, par exemple grâce à des générateurs aléatoires qui respectent la topologie [Shatz 1992]. Le coût de codage d'une telle architecture reste faible.

¹ Les connexions externes peuvent être aussi quelconques ou inexistantes, mais des connexions inhibitrices permettent d'augmenter le contraste, comme le note T. Kohonen [1984].

Ce schéma n'est pas le seul auquel on puisse penser, et surtout il n'implique pas nécessairement une architecture *neuronale* hiérarchique avec propagation ascendante, contrairement à ce qui était suggéré par la figure 3. Rappelons que dans notre modèle, la hiérarchie existe entre les détecteurs, mais rien n'est dit du niveau neuronal. En particulier, on peut imaginer que les mêmes neurones peuvent être impliqués dans des détecteurs de niveaux différents ! Les détecteurs de haut niveau étant par exemple caractérisés par des synapses excitatrices internes de poids plus faible.

5. Détection collective et décision

Dans le cas neuronal, ce qui a été dit pour la liaison entre deux détecteurs de niveaux adjacents n'a pas à être modifié pour rendre compte de la "décision" collective², c'est-à-dire de la connexion d'un détecteur du niveau le plus élevé avec l'un des effecteurs collectifs disponibles. Un tel système est un modèle simple de décision collective, et il peut rendre compte de certaines décisions du type perception → action.

L'analogie entre la description que nous venons de faire d'un mode de fonctionnement neuronal possible et d'autres situations de décision collective ne va pas toujours de soi, mais nous pensons qu'elle peut néanmoins se révéler féconde dans certains cas.

Elle ne va pas de soi parce que contrairement à ce qui semble être le cas dans le cerveau, les éléments qui constituent les détecteurs sont souvent les mêmes entités physiques que ceux qui constituent les effecteurs (penser aux abeilles, aux hirondelles, aux agents économiques, etc.). De plus, dans les cas de décision collective couramment évoqués, les mêmes éléments sont systématiquement impliqués dans des niveaux de détection différents.

Nous ne tenterons pas ici de pousser l'analogie trop loin, en nous contentant d'insister sur les effets de seuil qui définissent la détection, et sur la présence d'une synchronisation qui caractérise l'activité d'un détecteur collectif dans de nombreuses situations.

Pour voir comment fonctionne l'analogie, considérons à nouveau notre modélisation très simple de la décision de migrer des hirondelles. Lorsque l'automne s'avance, les oiseaux se rassemblent, sur un toit ou sur des lignes électriques par exemple, puis s'envolent ensemble pour un voyage de plusieurs milliers de kilomètres. Il s'agit de comprendre comment le jour du départ est déterminé, en supposant que les hirondelles ne disposent pas de moyen d'opter toutes exactement pour le même jour, d'après les indices environnementaux.

Nous supposons que les hirondelles ressentent une pulsion migratrice (qui s'accompagne de changements hormonaux dans beaucoup d'espèces) dont l'intensité dépend de paramètres comme le raccourcissement de la durée diurne (photopériode). Nous supposons aussi que les hirondelles ont une attirance pour tout rassemblement de taille K , qui croît avec K et avec la pulsion migratrice, et qui, si elle est suffisante, les pousse à aller se poser avec les hirondelles déjà posées. On peut dresser le parallèle suivant entre une détection neuronale et la situation présente:

² Nous employons le terme "décision" dans le sens: déclenchement d'action. Une décision collective est donc un mécanisme collectif pour déclencher une action collective. Par exemple, le mouvement d'un muscle du bras est une action collective, car il met en jeu des milliers de contractions de fibres musculaires. Comme indiqué plus haut, il s'agit, au niveau neuronal, d'une "décision" collective, puisque l'information détectée dans l'environnement et qui conduit à la contraction n'est jamais centralisée dans un neurone unique.

hirondelle	↔	neurone
ensemble des hirondelles du quartier	↔	détecteur collectif
hirondelle posée sur le toit	↔	neurone actif
rassemblement général	↔	activation du détecteur collectif
attirance pour un groupe posé	↔	synapse excitatrice interne
pulsion migratrice	↔	synapse excitatrice afférente (capteur)

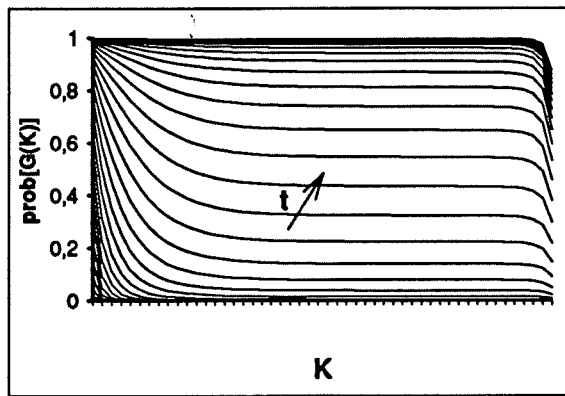
Des tentatives de rassemblement ont sans doute lieu à maintes reprises à l'approche de la période propice, mais comme nous allons le vérifier, la non-linéarité de la détection collective fait qu'on ne peut observer qu'un seul rassemblement de taille significative. Ce rassemblement sera général et marquera le jour du départ.

L'activation du détecteur collectif, ici le rassemblement, est bien un phénomène physique qui n'a pas de signification au niveau des hirondelles individuelles. Mais il faut aussi comprendre que la non-linéarité de ce détecteur se situe elle aussi au niveau collectif. Pour le voir, calculons la probabilité que le rassemblement atteigne la taille K ce jour-là. Dans la formule qui suit, N désigne le nombre des hirondelles du quartier, et $G(K)$ le fait que la taille du rassemblement atteigne K au jour t (nous omettons t pour alléger l'écriture). L'événement $G(K+1)$ contient donc $G(K)$. La probabilité que la taille du groupe passe de K à $K+1$ correspond à la probabilité qu'une au moins des $(N-K)$ hirondelles restantes se joigne au groupe. La probabilité $p_1(K)$ qu'une hirondelle se pose avec un groupe de taille K est une fonction supposée croissante de K et de l'avancement de la saison, t :

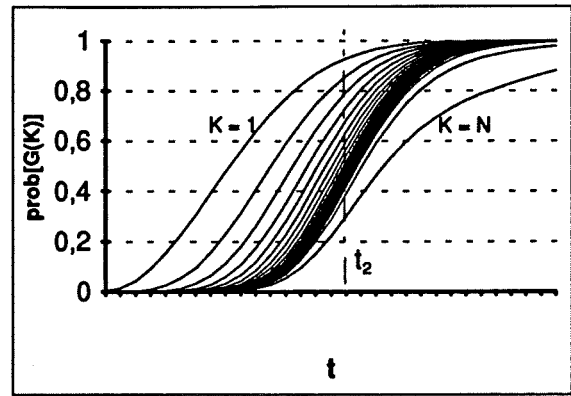
$$\text{Prob}[G(K+1) | G(K)] = \frac{\text{Prob}[G(K+1) \& G(K)]}{\text{Prob}[G(K)]} = \frac{\text{Prob}[G(K+1)]}{\text{Prob}[G(K)]} = 1 - [1 - p_1(K)]^{N-K}$$

La figure 4 montre $\text{Prob}[G(K)]$ en fonction de K et de t , ainsi que la probabilité d'accroissement du groupe $\text{Prob}[G(K+1)|G(K)]$.

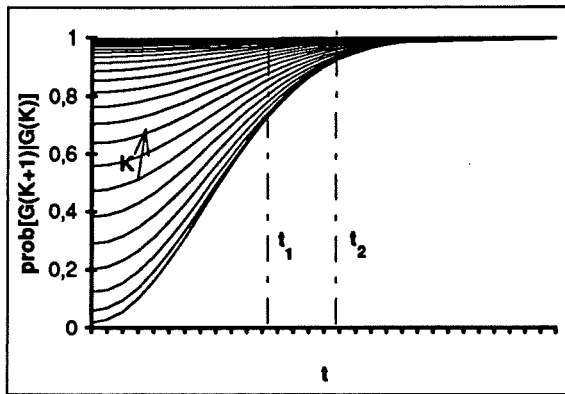
Le principal résultat montré par ces courbes est que le rassemblement est un phénomène "tout-ou-rien". Si un rassemblement a lieu, il est général. Il n'y a pas de gradualité pour le rassemblement (contrairement à ce qui se passe au niveau individuel, avec la fonction $p_1(K)$). Seule la probabilité du déclenchement est graduelle, quoique non-linéaire (figure 4b) : cette probabilité devient de plus en plus grande lorsqu'on approche de t_2 .



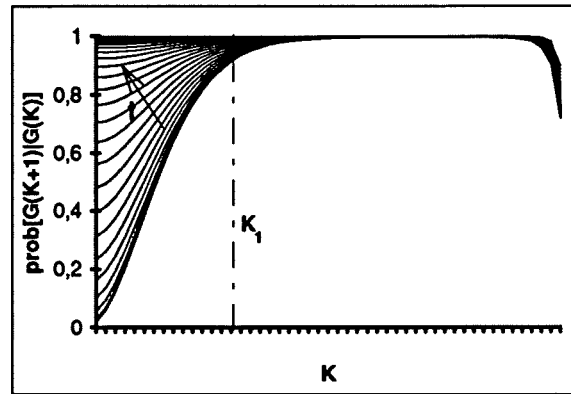
(a)



(b)



(c)



(d)

figure 4: ces courbes donnent $Prob[G(K)]$ en fonction de K et de t , ainsi que la probabilité d'accroissement du groupe $Prob[G(K+1)]/G(K)$. Le tracé dépend de la probabilité $p_j(K)$ qu'une hirondelle se joigne à un groupe de taille K au jour t . Pour cette figure, cette probabilité est supposée parabolique : $p_j(K) = A.K^2 + B.t^2 + C$. Les tracés (a) et (b) montrent que la probabilité d'un rassemblement dépend très peu de la taille de ce rassemblement. Le tracé (c) montre que tout rassemblement s'accroît presque sûrement dès que t dépasse la valeur t_2 . Après t_1 , un début de rassemblement fortuit a de fortes chances de conduire à un rassemblement général. Le tracé (d) met en évidence le fait que les rassemblements de taille intermédiaire (égale ou supérieure à K_1) sont fortement improbables: s'ils ont pu se produire, leur probabilité d'accroissement est très élevée.

Ce mécanisme permet au rassemblement de jouer le rôle d'un détecteur. Lorsque la période propice t_2 est proche, un début de rassemblement, causé par exemple par des individus précoces³, provoque le rassemblement général. Ce n'est pas le cas avant t_1 : les figures 4c et 4d montrent qu'un petit rassemblement a alors très peu de chances de croître. Dans la période $[t_1, t_2]$, le nombre des individus prêts à migrer devient significatif, ce que le mécanisme du rassemblement va détecter.

Ce détecteur fait jouer le même rôle à tous les individus. En effet, le rôle éventuel joué par les individus précoces est tout à fait marginal et n'apporte pas d'information. Il s'agit d'un bruit qui peut provoquer légèrement plus tôt le basculement vers le rassemblement général, en amenant le groupe rassemblé à un accroissement quasi sûr (le long d'une courbe de la fig. 4d). Nous sommes donc bien en présence d'un cas de détection collective.

³ Le modèle ne représente pas la variabilité de $p_j(K,t)$ selon les individus. En raisonnant de manière qualitative, nous pouvons néanmoins voir que cette variabilité ne remet pas en cause l'existence des seuils, ni même leur valeur.

Il est important de constater que ce phénomène collectif de détection a sa réalité autonome, c.à.d. qu'il n'est pas directement lié à l'existence d'une détection au niveau individuel. En effet, la forme de $p_I(K)$ importe peu, pourvu qu'elle soit fonction croissante de K et de t . Cette fonction peut même être linéaire (voir figure 5) ! Autrement dit, il y aura détection au niveau collectif même si $p_I(\cdot)$ ne présente strictement aucun seuil, c'est-à-dire même si les individus sont dépourvus de détecteur.

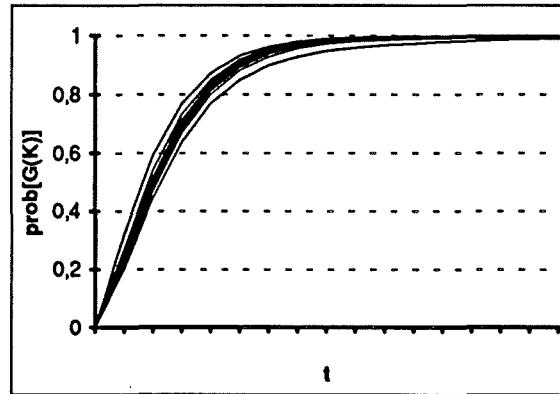


figure 5: cet ensemble de courbes est similaire à celui de la figure 4b, mais cette fois $p_I(K)$ a une croissance linéaire en K et en t : $p_I(K) = A.K + B.t + C$. Même dans ce cas limite peu réaliste, on observe encore un phénomène de détection au niveau collectif, en l'absence de tout seuil au niveau individuel : en effet, comme en fig. 4b, ces courbes sont quasiment confondues, ce qui fait du rassemblement un phénomène tout-ou-rien.

On voit ainsi comment une décision collective, celle de migrer, est obtenue à l'aide d'un détecteur collectif, le rassemblement. Comme dans notre modèle neuronal, ce détecteur fonctionne à travers l'activité synchrone des éléments qui le composent. Cette synchronisation repose sur un comportement des individus qui a dû être élaboré par la sélection naturelle : les individus actuels descendent de ceux qui n'ont pas fait l'erreur de migrer en petit groupe. Le comportement des individus (attirance pour le groupe qui augmente avec le temps) conduit à la constitution d'un détecteur collectif qui assure la non-linéarité de la décision de migrer : la probabilité de se rassembler, et donc de migrer, est une fonction non-linéaire (seuil) du nombre d'individus prêts à migrer.

6. Conclusion

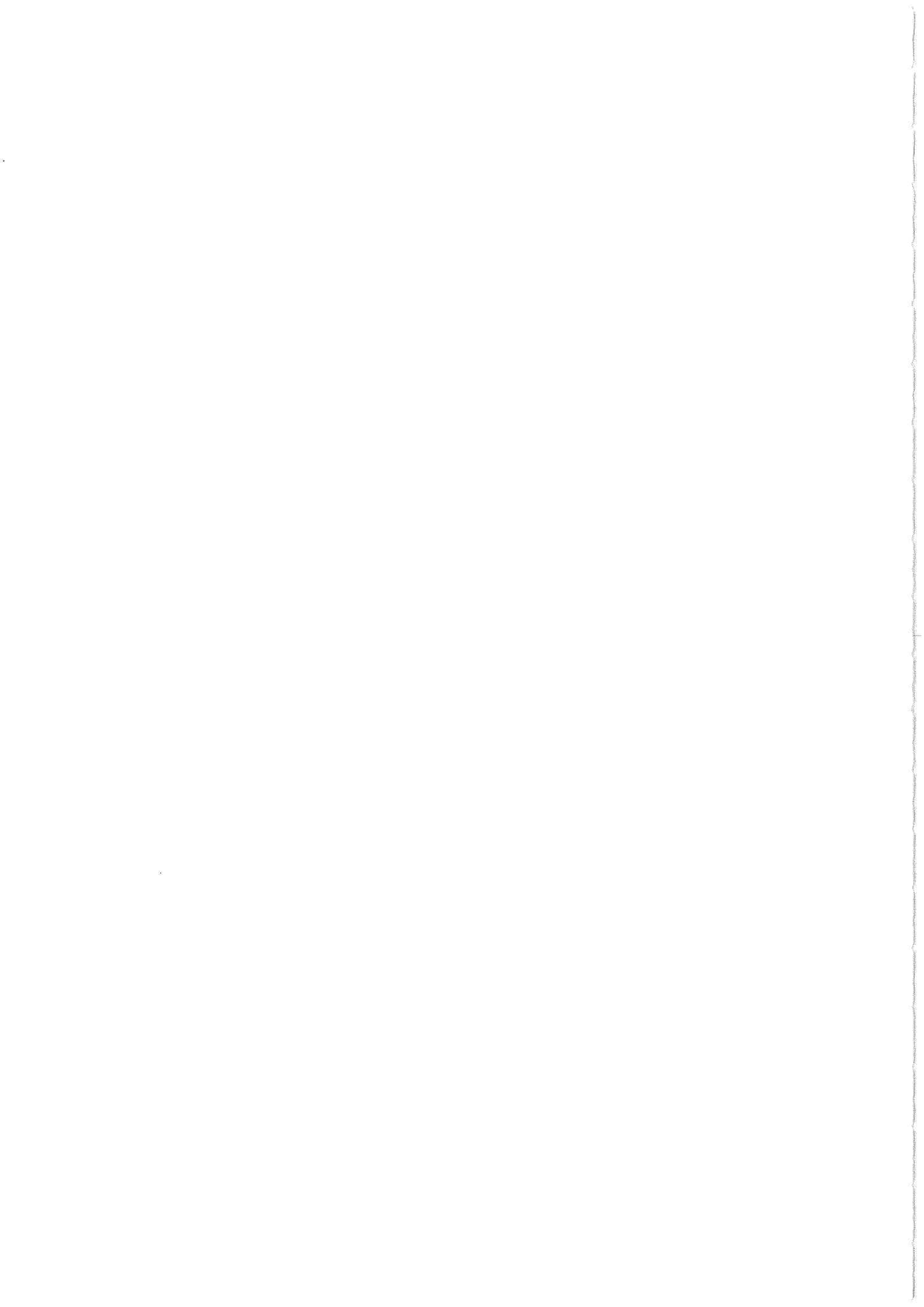
La notion de détecteur apparaît comme fondamentale dès que nous voulons rendre compte de l'organisation de systèmes qui sont capables de réaliser des fonctions en s'adaptant à leur environnement.

Dans le cas du système nerveux, les détecteurs apparaissent comme distribués, et les neurones qui les composent réalisent des détections collectives. Bien que les détecteurs soient au départ postulés comme des entités abstraites, nous avons pu suggérer une organisation neuronale biologiquement fondée, basée sur la cohérence temporelle et sur les propriétés topologiques des détecteurs, qui puisse assurer les fonctions de détections organisées hiérarchiquement.

Une décision collective observée dans des collections d'agents quels qu'ils soient suppose un mécanisme non-linéaire, typiquement un effet de seuil. On peut parler de détection collective lorsque l'existence de ce seuil est la conséquence du seul comportement des agents, sans qu'aucune centralisation de l'information détectée n'ait lieu. Il est alors parfois possible de s'inspirer d'une analogie avec les détecteurs collectifs neuronaux pour parvenir à une meilleure description du comportement collectif observé.

Références

- Dessalles Jean Louis (1992). Les aspects cognitifs de l'émergence. In Amy Bernard, Ducret J-J, Grumbach A., *Emergence dans les modèles de la cognition - Actes des journées*, TELECOM-Paris 92-S-003, Paris 1992, pp. 47-59
- Freeman Walter (1991). La physiologie de la perception. *Pour la Science* n° 162, 1991, pp. 70-78
- Kohonen Teuvo (1984). *Self-Organization and Associative Memory*. Springer Verlag, Berlin 1988
- Lawler Robert W. (1980). *Extending a Powerful Idea*. A.I. Memo 590, Massachusetts Institute of Technology, 1980
- Shatz Carla (1992). La maturation du cerveau. In *Pour La Science* n° 181, 1992, pp. 50-59
- Zeki Semir (1992). Les images visuelles. In *Pour La Science* n° 181, 1992, pp. 60-68



Rochebrune, Janvier 1993

**DYNAMIQUE COGNITIVE ET INTERACTIONS SOCIALES :
QUELQUES PHENOMENES DE "RESONANCE"**

Evelyne ANDREEWSKY
INSERM-TLNP
PARIS

La première épreuve de psychologie expérimentale connue (fort connue même, et maintes fois adaptée dans la littérature - i.e. "Le cercle de craie Augsburgois", Bertol Brecht) est associée au jugement de Salomon¹. Il s'agit bien de psychologie expérimentale, dans la mesure où dans cette épreuve, des *comportements observables* (rapports de force dans des *interactions* physiques) traduisent le phénomène psychologique (l'amour porté à un enfant) que l'on veut explorer.

Mais ce type d'épreuve ancrée dans des interactions sociales, est rarissime en psychologie. La *psychologie expérimentale* - dans son souci quasiment "paranoïaque" (qualificatif d'Henri Atlan, 1986), de calquer ses expériences sur le modèle des sciences "dures" - s'est en effet focalisée sur l'étude de phénomènes comportementaux "de laboratoire", *isolés* et *purifiés* autant que faire se peut de tout contexte "perturbateur", et notamment de toute interaction sociale. Dans ces conditions, ces phénomènes se présentent comme *objectifs* (dans la mesure où ils sont quantifiables et statistiquement reproductibles), et deviennent donc des objets d'étude *convenables*, susceptibles d'être abordés sur le mode scientifique "orthodoxe". Ces travaux ont produit un ensemble très grand de données comportementales que la psychologie cognitive vise à organiser et à expliquer avec la métaphore de l'ordinateur, en conceptualisant les mécanismes psychologiques - la dynamique cognitive - en termes de "programmes" produisant les phénomènes comportementaux.

Dans la mesure où l'on considère la "dynamique cognitive" comme un *programme*, le problème du *programmeur* se pose : opère-t-il du fond du gène, ou de quelque autre entité biologique ... ou peut-être encore s'agit-il d'un "programmeur distribué" ? Le renouvellement des métaphores du cognitif (*connexionisme, auto-organisation* ...) vise entre autre à évacuer ce problème. Mais programme ou autre métaphore, la dynamique cognitive n'en reste pas moins fondamentalement considérée comme *intrapersonnelle*, et donc en marge des interactions sociales.

Comment explorer les dimensions *interpersonnelles* de cette dynamique ? Certains comportements "pathologiques" peuvent aider à baliser cette exploration, dans la mesure où ils témoignent clairement de l'impact des interactions sociales - un peu à la manière dont le pont sur lequel passe une troupe au pas cadencé peut témoigner (parfois trop clairement) des "interactions fortes" - encore que longtemps insoupçonnées - entre troupes et ponts ! Nous présenterons quelques exemples de comportements pathologiques (au niveau individuel ou collectif) que l'on peut considérer comme résultant de *résonances* cognitivo-sociales (telles, par exemple, les "oscillations forcées" du *double-bind* du schizophrène, cf. G. Bateson, ou encore la "marée" de la foule en panique, cf. J. P. Dupuy). Ces exemples, *pseudo-expérimentations hors laboratoire*, invitent notamment à implanter le collectif dans les métaphores et modèles cognitifs.

1. Rappelons que pour départager deux femmes prétendant toutes deux être mère du même enfant, l'expérience dont il s'agit proposait une épreuve de force dont l'enjeu était cet enfant.



**INTELLIGENCE COLLECTIVE.
ÉMERGENCE DES CONCEPTS SOCIAUX :**
les théories naïves ou implicites concernant les individus.

Herminia Peraita.

Professeur de Psychologie
UNED Madrid

Journées de Rochebrune 1993

Ce travail a pour but d'analyser ce qu'il y a de commun, de collectif, d'une part, et d'individuel et personnel, de l'autre, dans la genèse, la structure et la transmission de certains types de théories implicites, considérées en tant que concepts sociaux et/ou représentations collectives. Dans ce cadre, on analysera les données d'un travail empirique en cours, sur la catégorisation des individus.

Signalons que les termes collectif et individuel ne seront pas considérés de façon traditionnelle, c'est-à-dire, collectif: synonyme de social, et individuel: synonyme de personnel, intransférable. Le collectif sera conceptualisé comme patrimoine linguistico-sémantico-cognitif "universel", tandis que l'individuel sera considéré comme un système ouvert, flexible, et soumis à des contraintes sociales.

Dans cette perspective, on cherche à établir une analogie avec le lexique, plus spécifiquement avec un modèle qui conçoit l'existence, pour chaque item lexical, d'une structure interne complexe, composée d'un ensemble de composantes essentielles, centrales, invariables et probablement universelles. Ce modèle est aussi componentiel, et rend compte de contraintes innées.

D'autre part, ce modèle comporte des composantes secondaires, variables, articulées aux précédentes. Ces composantes périphériques et accessoires; elles sont restreintes par les systèmes sociaux et les systèmes de croyances de l'individu.

D'où l'intérêt de l'analogie précédente : les items lexicaux présentent en effet une grande complexité structurelle, où l'on peut distinguer plusieurs niveaux de description, allant du collectif (phonologique, morphologique, syntaxique) à l'individuel (sémantico-cognitif, encyclopédique - lié aux systèmes de

croyances). On examinera, dans le cadre de cette analogie, la structure des théories implicites et la catégorisation des individus.

On analysera, entre autres, les aspects suivants :

1°. Le rôle de la culture/société en tant que cadre conceptuel de référence dans le processus de sélection et de transmission de l'information et des connaissances. Ce processus permet de fournir ou d'assigner des valeurs aux paramètres de la structure des items lexicaux, d'une part, ou des théories implicites de l'autre.

2°. Description de quelques structures schématiques des différentes catégories "ontologiques" qui constituent le monde .

Catégories d'objets :

naturels et artefacts.

Catégories aspectuelles-temporelles :

états, processus, événements, actions.

Catégories sociales :

institutions, personnes, théories scientifiques.

3 . Reconceptualisation de certaines dichotomies classiques : analytique vs. synthétique, nécessaire et suffisant vs. probable, en vue d'autres types d'organisation tels que invariants vs. variables, central vs. périphérique, et structure vs. contenu.

Berlin, Octobre 1992

**Eric Bonabeau
Rochebrune 1993**

CNET Lannion B-OCM/TEP

22 Lanion Cédex

Proposition de

**Wolf-Dieter Eberwein,
WZB-FIB**

**Reichpietschufer 50
1000 Berlin 30**

**Tel. 19-49-30-25 49 15 64 (bureau)
Fax 19-49-30-25 49 15 60**

Tel. 19-49-30-87 39 09 (privé)

(1) Stabilité politique, une Fonction des Décisions individuelles et collectives?

L'éclatement de l'empire soviétique représente sans doute un des phénomènes les plus fondamentaux auquel sont confrontés les sciences sociales. Une collectivité non prévue dans les systèmes communistes, des femmes et des hommes, les "sans-culottes" modernes, ont fait de l'histoire, quoi que juste pour un petit moment. Mais les conséquences ont été fondamentales. Non seulement ont ils contribué à l'écroulement du système bipolaire international dont tout le monde croyait qu'il serait de durée pratiquement illimitée. En plus, ils ont forcé les mandarins communistes en place hors du pouvoir, un pouvoir qu'ils croyaient posséder pour toujours.

Le politiste allemand, von Beyme (1990) a remarqué que les révolutions de l'automne 1989 doivent être considérées comme le *vendredi noir* (allusion au grand krach de la bourse américaine en 1929) des sciences sociales. Sans aller dans les détails son verdict n'est pas tout-à-fait justifié en ce qui concerne le développement théorique. Mais il a certainement raison en ce qui concerne les problèmes qu'ont ignorés les chercheurs parcequ'il ne les considéraient pas comme importants, notamment la population.

Les mouvements populaires qui se sont pratiquement manifestés d'un jour à l'autre à l'est n'étaient simplement pas considérés importants ni sur le plan théorique, ni sur le plan pratique. Aujourd'hui la situation est différente. Nous savons que la stabilité politique ni la survie d'un régime autoritaire peuvent être garanties à la longue sans l'existence d'un minimum de légitimité. Nous savons que la masse non bien définie peut pratiquement d'un jour à l'autre se transformer en acteur collectif, muni d'une intelligence collective.

Les développements en Europe de l'Est¹⁾ posent entre autres deux problèmes qui sont d'intérêt fondamental pour les sciences sociales:

1. disposons nous des fondations théoriques pour expliquer non seulement la stabilité politique comme elle évolue de jour en jour, mais aussi les conditions menant à des changements abruptes, soit révolutionnaires, à court terme?
2. Disposons nous de la méthodologie ou des méthodologies adéquates pour reproduire ces processus d'apparence chaotique dans des modèles plus rigoureux que ceux de nature verbales?

Je veux traiter de ces deux sujets car, à mon avis, ils sont étroitement liés l'un à l'autre. Dans ce texte je me limite à quelques remarques qui resteraient à expliquer en détail dans une communication approfondie. Il va de soi que je me limite dans le cas présent au problème que représente le changement fondamental de la base du pouvoir politique, en particulier du processus de la prédisposition à la contestation politique et de sa transformation en contestation politique manifeste.

(2) Aspects théoriques

Avant de m'aventurer dans une esquisse de la dimension théorique je veux mentionner un problème épistémologique. Si l'intérêt central est l'explication de changements abrupts d'un développement continu, soit incrémentaliste, cela présuppose une hypothèse fondamentale, celle que l'évolution politique obéit à certaines lois. Cela n'implique pas qu'elles soient de nature déterministe.

Je ne considère pas cette discussion de trop grand intérêt, car elle ne peut être résolue a priori. Si de Tocqueville, par exemple, explique la révolution de 1848 avec un accident, ce que d'ailleurs beaucoup de politistes semblent croire, il a peut être raison. Mais personne jusqu'à présent n'a pu démontrer de façon convaincante que d'autres explications soient clairement fausses. Jay Forrester mentionne un autre aspect du même problème avec ce qu'il appelle les effets inattendus si non contraires aux expectations des acteurs ou chercheurs. Dans ces deux cas ce sont juste des arguments qui ne démontrent que notre manque de connaissance approfondie du comportement de systèmes sociaux complexes. Du moment où nous connaissons le comportement d'un système dynamique, les effets inattendus n'existent plus.

Dans la discussion théorique trois développements, si l'on veut, me semblent d'importance primordiale. Le premier est relié à la fondation de la légitimité des systèmes politiques, développés sur le niveau théorique par Easton (1965;1974). Il fait une distinction entre la légitimité du gouvernement (*authority*), le régime constitutionnel (*regime*), et l'unité nationale (*polity*). Sur ces fondations j'ai essayé de relier les deux premiers niveaux de légitimité de façon interactive dans le cadre de deux types des systèmes politiques, le système démocratique et les système autoritaire, soit communiste (Eberwein, 1987). Ceci à été fait avec le but explicite de reproduire ces relations dans un modèle de simulation global, GLOBUS (Bremer, 1987). Important à mentionner est le fait qu'une telle conceptualisation n'est que valable si elle tient compte du contexte politique et économique et qu'elle soit fondée de façon empirique.

Le deuxième développement est relié à la conceptualisation de Hirschmann (1977) des réactions fondamentales dont dispose la population avec ce qu'il appelle *voice* (contestation), *exit* (à traduire peut-être avec exil volontaire), et *loyalty* (loyauté). En plus, comme Guy Bajoit (1988) l'a démontré, une quatrième forme de comportement doit être incluse, l'*apathie*. Ce que Hirschmann a certainement oublié est la distinction entre la prédisposition à l'action et

1) je comprends par l'Europe de l'Est - pour simplifier les choses - toute la région qui était incluse dans l'ancien Pacte de Varsovie.

l'action elle-même. Ceci est certainement le cas pour *voice* et *exit*. Malheureusement, la recherche n'a pas poursuivi l'explication et l'analyse de ces deux phénomènes simultanément, c'est-à-dire l'interaction entre la prédisposition à l'action et l'action politique elle-même.

Le troisième développement se relie aux théories des mouvements sociaux. En particulier, la conception de la structure d'opportunité politique et des cycles de contestation qu'a développé Tarrow (1991) me semblent être fondamentaux. Quoi que ces concepts aient été développés uniquement pour analyser les mouvements de contestations dans les pays industrialisés et démocratiques de l'OCDE, ils sont certainement de grande portée pour expliquer en partie les changements abrupts entre autres dans l'ancienne RDA (voir Eberwein et al., 1991).

Ceci dit, et sans aller dans les détails, il me semble que nous avons à notre disposition le répertoire théorique nécessaire pour développer un modèle systématique pour décrire et analyser les processus de l'effondrement de régimes politiques en fonction de mouvements de contestations collectifs. Cet argument, j'en suis conscient, doit certainement être spécifié. Néanmoins je crois que la direction du chemin théorique à parcourir me semble valable.²⁾

(3) Modèles de Processus Politiques

Ce qui semble être un problème fondamentalement théorique est un même temps un problème méthodologique: peut-on développer sans grands problèmes un modèle de simulation qui est capable de reproduire avec des équations différentielles non linéaires ces changements abrupts dont nous avons observé un certain nombre de variations en Europe de l'Est?

Avant de répondre à ce sujet je dois encore une fois faire référence à un problème fondamental en sciences sociales. En ce qui concerne la recherche empirique systématique, dite quantitative en science politique, nous poursuivons de coutume comme but la prédiction. C'est-à-dire, nous assumons qu'un modèle économétrique par exemple servirait non seulement à expliquer le passé mais aussi à prévoir le futur par voie d'extrapolation. Cette conception me semble problématique.

Même si nous avons recours au phénomène de *path dependency* démontré dans la théorie mathématique des systèmes, si je le comprends bien, cela implique que mêmes dans certaines marges il existe toujours toute une variété de possibilités de développements d'un système politique et de sa chance de survie. Si, en plus, nous avons bonne raison de croire que les processus politiques (dans notre cas) sont non linéaires et complexes cela implique logiquement que la réalisation actuelle de l'état de stabilité d'un système politique ne représente qu'une réalisation parmi d'autres, tant que nous ne connaissons pas les relations existantes d'un système politique, qu'il soit concret ou abstrait

En conséquence la prévision ne devrait pas vraiment être le sujet le plus central des sciences sociales, à moins que nous nous contentions de prévisions à court terme. Beaucoup plus important est une approche expérimentale qui nous permettrait d'évaluer plusieurs variations d'évolution d'un système politique. C'est le cas aussi longtemps que nous ne disposons ni d'une fondation théorique satisfaisante ni de la fondation empirique de ces systèmes théoriques. Il me semble, en conséquence, que la simulation représente un instrument méthodologique de grande valeur, pour la recherche comme pour l'évaluation de politiques (policies) envisagées.

Ceci dit, je veux caractériser plusieurs alternatives de modélisation en science politique qui me semblent être d'importance primaire. Il est néanmoins important de mentionner un problème qui est loin d'être trivial. Pour représenter le processus de stabilité politique et la possibilité d'un changement abrupt de changement de régime politique (comme par exemple la transition

2) voir en détail l'analyse de Eberwein/Groenen, 1992.

d'un régime autoritaire à un régime démocratique), le modèle doit inclure tous les éléments nécessaires qui déterminent l'évolution incrémentale comme le changement abrupt. Je reviendrai à la fin de ce petit exposé à cette question.

En science politique nous pouvons faire, entre autres, une distinction entre deux types de modèles de simulation pour décrire des processus politiques: ceux décrivant des développements à long terme, et ceux s'adressant aux développements à court terme. Dans la première classe nous trouvons typiquement les modèles globaux comme GLOBUS (Eberwein, 1987) ou le type de modèle pour les pays de l'OCDE (Widmaier, 1990). Ce type de modèle pose un problème: il ne permet pas de changements fondamentaux politiques. En fait, ces modèles sont tous construits de façon à ce qu'ils évoluent autour d'un certain niveau de stabilité politique. Ceci suit logiquement de l'intérêt des chercheurs se concentrant sur l'évolution incrémentale.

La deuxième classe de modèles des processus politiques tient compte de la dynamique à court terme, c'est-à-dire de l'interaction dynamique entre contestataires et le pouvoir politique, que cela soit l'interaction entre la police et des contestataires ayant recours à la violence (Salert and Sprague, 1980), l'interaction entre mouvement révolutionnaire et le pouvoir gouvernemental (Lichbach, 1990), ou le changement abrupt de la relation entre l'élite politique et une contre-élite d'un côté, la population et le changement de ses préférences privées et publiques envers l'élite au pouvoir et la contre-élite (Kuran, 1989). Ces deux derniers modèles (Lichbach, Kuran) ont le "petit" désavantage d'être purement théoriques. On pourrait ajouter à cette liste le modèle de Axelord (1986) qui a analysé l'effet de l'évolution de normes déterminant le comportement d'individus, donc d'un collectif d'individus.

De cet aperçu suit une première conclusion, veut dire, qu'il semblerait naturel de s'engager dans la voie de combiner les éléments différents dans un modèle intégral pour décrire et analyser le processus politique. Ce qui semble sur le plan analytique est comblé de problèmes loins d'être bien connus. Pour illustrer cet argument prenons le cas de l'ancienne RDA. Selon nos données et d'une première analyse il semble que l'ancien régime des vieux à Berlin-Est se vit confronté pratiquement en quelques semaines avec une vague croissante de personnes choisissant *exit* ou *voice*. Il y a probablement eu un processus de renforcement entre ces deux dynamiques entre le mois de septembre et de novembre 1989 comparable à la diffusion d'une infection contagieuse. C'est une des raisons pour lesquelles le régime n'a pas eu vraiment le temps de réagir avec les moyens répressifs comme en 1953.

Le mécontentement qui a éclaté dans la RDA ne peut-être expliqué sans tenir compte de l'évolution à long terme de l'insatisfaction de la population. La base de la frustration de la population peut-être reliée au développement socio-économique. La situation s'est détériorée dans les années 70 de façon plus ou moins continue. Pourquoi la chute du système politique en RDA a eu lieu en novembre 1989 reste une énigme.

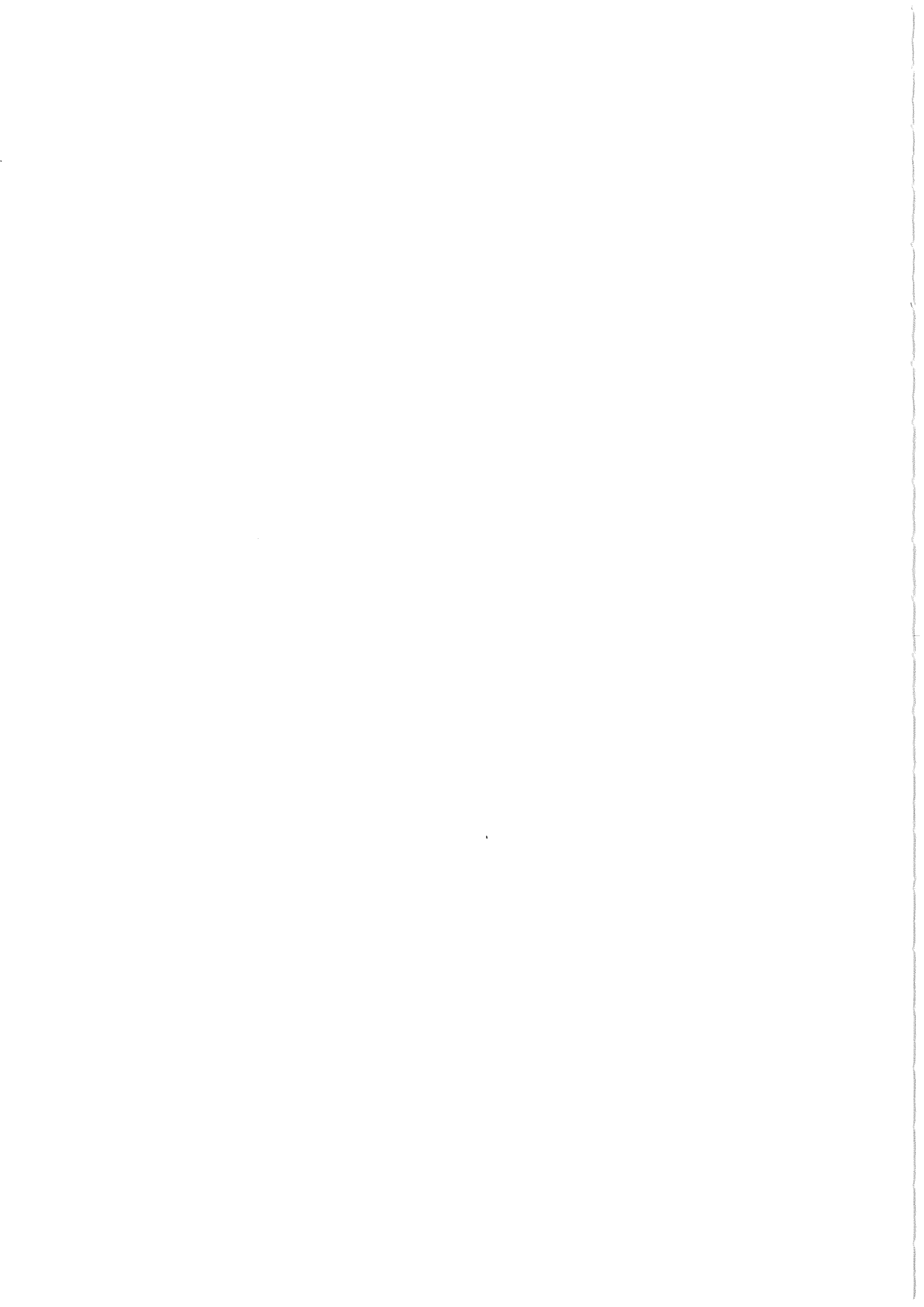
Cette illustration nous permet de tirer une première conclusion. Pour développer un modèle tenant compte de ces deux aspects nous devons combiner des éléments de différents niveaux d'aggrégation temporelle et de collectifs. Dans le premier cas ce sont des semaines, dans le deuxième cas ce sont des années, dans le premier cas nous devons utiliser des groupes d'individus comme acteurs, dans le deuxième cas le collectif population est suffisant. Mais comment combiner les deux?

Une deuxième conclusion en suit de la première. Même si nous avons résolu ce problème nous n'avons que traité de l'écroulement d'un système politique existant. Nous n'avons pas encore adressé le sujet du processus de transition d'un régime politique à l'autre ni le processus de transformation après l'établissement formel d'un nouveau système. En plus, comment relier les changements dans le domaine économique avec ceux dans la société et le système politique? Quoi qu'il en soit, il semble clair qu'il existe un potentiel explosif menant à de nouvelles situations "chaotiques" dans les temps à venir.

Je mentionne ces problèmes pour faire le point que le problème central aujourd'hui ne consiste pas en premier lieu d'illustrer avec des exemples réels une approche mathématique. Ce qui est nécessaire par contre est une exposition du problème théorique, si possible fondé empiriquement, et de la, de développer un modèle (ou plusieurs) qui est capable de reproduire les relations substantielles. Pour ne pas s'éloigner complètement de la réalité la base des données, en plus, doit être élargie.

Références

- Axelrod, Robert (1986) "An Evolutionary Approach to Norms" *American Political Science Review* 80/4:1095-1111.
- Bajoit, Guy (1988) Exit, voice, loyalty ... and apathy. *Revue française de sociologie*, avril-juin, XXIX-2, 325-345.
- Bremer, Stuart A. (ed) *The GLOBUS Model - Computer Simulation of Worldwide Political and Economic Developments*, Boulder/Frankfurt: Westview und Campus Verlag.
- Beyme, Klaus von (1990) "Die vergleichende Politikwissenschaft und der Paradigmenwechsel in der politischen Theorie." *Politische Vierteljahresschrift* 31/3:457-74.
- Easton, David (1976) "Theoretical Approaches to Political Support" *Canadian Journal of Political Science* 9:431-448.
- Eberwein, Wolf-Dieter (1987) The Domestic Political Process. In: S. A. Bremer (ed) *The GLOBUS Model - Computer Simulation of Worldwide Political and Economic Developments*, Kap. 3. Boulder/Frankfurt: Westview und Campus Verlag, 159-282.
- Eberwein, Wolf-Dieter, and W. Groenen (1992) Transition to Democracy and Transformation Processes in Eastern Europe - The Political Dimension. In: W.-D. Eberwein (ed) *Transformation Processes in Eastern Europe - Perspectives from the Modelling Laboratory*. Frankfurt a.M.: Peter Lange, 9-62
- Eberwein, Wolf-Dieter, Carsten Johnson, Jörg Stangel und Katrin Gärtner (1991) *Vom Aufstand der Massen zum Ende der DDR*. Veröffentlichungen der Forschungsgruppe Internationale Beziehungen, Wissenschaftszentrum für Sozialforschung Berlin, P 91-308.
- Hirschmann, Albert O. (1977) *Exit, voice, and loyalty. Responses to decline in firms, organizations, and states*. Cambridge, MA.: Harvard University Press.
- Kuran, Timur (1989) "Sparks and Prairie Fires: A Theory of Unanticipated Political Revolution" *Public Choice* 61:41-74.
- Lichbach, Mark I. (1987) "Deterrence or Escalation? The Puzzle of Aggregate Studies of Repression and Dissent" *Journal of Conflict Resolution* 31/2:266-297.
- Salert, Barbara and John Sprague (1980) *The Dynamics of Riots*. Ann Arbor: Inter-University Consortium for Political and Social Research.
- Tarrow, Sidney (1991) "Aiming at a Moving Target': Social Science and the Recent Rebellions in Eastern Europe" *PS: Political Science and Politics* 24/1:12-18.
- Widmaier, Ulrich (1990) "Political Stability in OECD nations". In W. D. Eberwein and K. W. Deutsch *Global Modeling*. International Political Science Review 11/2: 219-242.



Marie-Noëlle SARGET
Ecole des Hautes Etudes en Sciences Sociales (C.E.M.S.)

LES CONDITIONS D'EMERGENCE DE L'INTELLIGENCE COLLECTIVE
DANS LES SYSTEMES POLITIQUES

Le concept d'intelligence collective désigne une émergence des interactions inter-individuelles . Mais dans les sociétés humaines , une telle émergence ne peut être considérée comme nécessaire et automatique . Ce concept n'est susceptible d'avoir un intérêt en sciences humaines et en sociologie politique qu'à condition , d'une part , que l'on tienne pleinement compte des caractéristiques propres aux sociétés humaines , d'autre part , que l'on analyse les processus précis par lesquels se développe l'intelligence collective .

L'exposé comportera trois parties :

- I - Remarques à propos de l'utilisation du concept d'intelligence collective en sociologie politique
- II - Expérience politique et apprentissage de l'intelligence collective
- III - Essai de modélisation des processus d'accroissement de l'intelligence collective dans les systèmes politiques .

I - Remarques à propos de l'utilisation du concept d'intelligence collective en sociologie politique

- En sciences humaines , on court plus qu'ailleurs le risque d'émettre des jugements de valeur dans la désignation de ce qui est intelligence ou ce qui ne l'est pas . Ainsi , est-il toujours justifié d'affirmer que les comportements de coopération requièrent et génèrent plus d'intelligence collective que l'hostilité ou les antagonismes ? Rien n'est moins sûr ... On peut , par exemple , considérer , comme l'ont fait certains sociologues , que les guerres sont des régulateurs démographiques ! Et ce qui nous paraît généralement le fruit d'une intelligence collective supérieure , comme le développement économique et technologique des sociétés occidentales , peut n'être que pure folie aux yeux de certains philosophes , d'écologistes, ou de sociétés traditionnelles .

- L'intelligence collective n'est pas toujours diffusée de façon homogène . En effet , les systèmes sociaux , bien qu'interdépendants , sont relativement autonomes les uns par rapport aux autres . On peut ainsi constater que les progrès de la démocratie peuvent coexister avec une crise durable de l'économie , ou , à l'inverse , que les progrès économiques peuvent s'accompagner de régimes dictatoriaux ou de la pire corruption .

- On constate des phénomènes de désapprentissage de la démocratie , de la paix sociale et de la vie commune : des pays comme le Chili , l'Uruguay , l'Argentine , le Pérou ,

la Yougoslavie , etc. , ont basculé vers des guerres civiles larvées ou déclarées après avoir connu des périodes de paix ou de démocratie prolongées . Si ces dernières peuvent être considérées comme des périodes d'intelligence collective , on est obligé de constater que celle-ci ne connaît pas nécessairement un progrès linéaire et cumulatif . On pourrait même , au vu des évolutions de longue durée , postuler une sorte de loi d'alternance des périodes d'essor et des périodes de régression de l'intelligence collective dans certaines sociétés , alors que l'évolution serait plus linéaire pour d'autres sociétés .

- Pour qu'il y ait intelligence collective , il faut que la collectivité en question soit constituée , et que la majorité des individus qui la composent y participent . La nation est une forme possible de collectivité , qui peut être déchirée par les luttes des ethnies , des clans , des religions , des classes , ou des partis politiques ; ces luttes peuvent dégénérer en spirales de division , et se traduire par la "libanisation" de la société . Mais on ne peut exclure que de tels processus entropiques ne précèdent l'émergence de néguentropies nouvelles , à partir de groupes issus de la segmentation , ou à partir de l'ensemble de la société en question .

II - Expérience politique et apprentissage de l'intelligence collective

L'intelligence collective ne survient pas par les seules vertus des idées , mais aussi par l'expérience , qui caractérise principalement les sociétés humaines , par opposition aux sociétés animales , par exemple , où prédomine l'instinct . Cette expérience est infiniment riche et variée .

- L'expérience de la non-coopération et des antagonismes peut permettre aussi l'accroissement de l'intelligence collective

. par l'expérience de la souffrance collective

exemple : En Espagne et au Portugal , après les dictatures , les acteurs politiques ont manifesté une nette volonté de coopération et d'attachement à la démocratie , alors qu'ils étaient précédemment réticents à toute action commune.

. par l'expérience de l'échec :

exemple : Au Chili , le refus de coopération entre la Démocratie Chrétienne et l'Unité Populaire a interdit la réalisation de leurs programmes , qui étaient très voisins , et permis le coup d'état du Général Pinochet . 15 ans après , la réflexion menée sur ce point a conduit la Démocratie Chrétienne et les ex-partis de l'Unité Populaire à faire alliance contre le régime Pinochet et à le battre aux élections .

. par épuisement des combattants

exemple : L'Europe saignée à blanc par les guerres franco-allemandes de 1870 , 1914 et 1940 est enfin prête

dans les années 50 à s'engager dans la CECA , puis dans le Marché Commun .

2 - Il peut y avoir aussi apprentissage par la pratique des institutions : ainsi ,on a pu constater dans de multiples pays une amélioration avec le temps des pratiques démocratiques à partir de l'adoption d'une constitution démocratique ; il semblerait même que les diverses pratiques frauduleuses de vote elles-mêmes permettent à long terme un apprentissage progressif des procédures électorales et le développement d'une aspiration à l'élimination des fraudes . De même , l'existence d'institutions destinées au règlement pacifique des conflits , comme les cours de justice nationales ou internationales , permet d'envisager un recours plus restreint à la force .

3- Pour que l'intelligence collective s'accroisse avec l'expérience , il faut qu'il existe un débat collectif d'où sortiront une prise de conscience et une information utilisables, de forme rationnelle , une conception nouvelle susceptible de générer des comportements innovateurs dans les domaines économique , social ou politique . On peut considérer que c'est pour cela que certains réclament actuellement qu'un débat soit ouvert en France à propos du comportement des français pendant la dernière guerre mondiale . La réflexion collective sur l'expérience vécue est nécessaire au progrès de l'intelligence collective .

4 - Le développement de l'intelligence collective suppose une large diffusion chez tous les acteurs du système politique . Cette diffusion peut avoir lieu par des moyens variés :

. l'information : on constate que bien souvent , le parti au pouvoir applique des mesures d'abord préconisées par d'autres partis , ce qui signifie qu'il a été apparemment convaincu par leur argumentation .

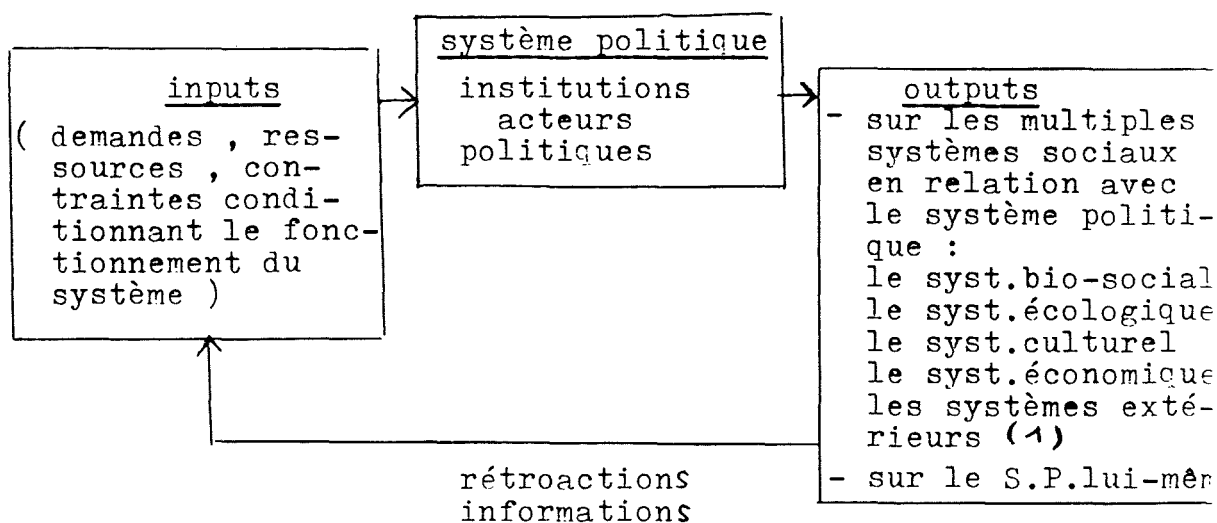
. là encore , rien ne vaut , semble-t-il , l'expérience pour modifier les idées . Ainsi , après quelques années de pratique du pouvoir , le Parti Socialiste français admet que les nationalisations ne constituent pas une réponse valable aux problèmes de l'économie contemporaine et reconnaît le rôle du marché . S'il était resté dans l'opposition , donc dépourvu de l'expérience que constitue l'exercice du pouvoir , une telle évolution n'aurait sans doute pas eu lieu aussi rapidement . De même , le gouvernement Chirac lors de la cohabitation a-t-il dû édulcorer les aspects ultralibéraux de ses projets . L'expérience paraît ainsi être le meilleur remède aux délires de l'idéologie .

5 - Un autre remède à ces délires est l'existence d'une opposition supposée contrôler le parti au pouvoir , qui joue le rôle d'un garde-fou . Il est vrai que l'opposition et l'ensemble des contre-pouvoirs peuvent jouer le rôle inverse et s'opposer à tout changement , à toute moderni-

sation au profit des intérêts à court terme de corporations ou de groupes de pression . Mais s'opposer à un rythme excessif de modernisation peut constituer un signe d'intelligence collective et garantir le succès de celle-ci en évitant des réactions de rejet de type iranien . Ce dernier exemple montre que l'évaluation de l'intelligence collective est particulièrement difficile en sciences humaines , parce qu'elle suppose la prise en compte de tout un contexte , et celui du passé comme de l'avenir des processus en cause .

III - Essai de modélisation des processus d'accroissement de l'intelligence collective dans le système politique

La modélisation répondra au schéma classique du système politique dans le cadre de la théorie des systèmes :



(1) Les 4 premiers systèmes mentionnés incluent ; avec le système politique , l'ensemble des systèmes de la société globale . (voir J. W. LAPIERRE, l'Analyse des Systèmes politiques, PUF, 1973)

les acteurs peuvent faire un mauvais diagnostic , ne pas comprendre le sens de l'information qui leur est transmise , choisir des remèdes qui aggravent le mal . On se trouve alors devant des spirales d'échecs .

L'orientation des processus nous paraît souvent commandée par le rôle joué par la culture nationale dans la constitution des schèmes cognitifs . Elle constitue un filtre qui intervient à tous les niveaux de la conscience des acteurs politiques et sociaux . La compréhension et l'assimilation de l'information , les leçons tirées de l'expérience , varient en fonction de la culture politique antérieurement acquise , qui consiste en structures mentales cristallisées dans la longue durée . Il n'est donc pas possible d'ignorer l'histoire : la culture politique d'un pays donné n'est pas la même en 1900 qu'en 1930 ou 70 , car l'expérience - qui est de l'histoire vécue - la modifie en permanence .

CONCLUSION

Qu'est-ce qui commande la positivité du processus ? L'interaction des individus et des groupes humains n'entraîne pas automatiquement un accroissement de l'intelligence collective . Cela dépend du contenu de cette interaction , donc du contenu de l'information ou de l'expérience transmises

qui sont elles-même fonction d'une part , des capacités cognitives de l'émetteur et du récepteur du message , d'autre part , de leur système de valeur et du sens qu'ils entendent donner à leur action .

En effet , l'homme n'est pas seulement une machine rationnelle , homo sapiens , mais il peut être aussi , comme le rappelle Edgar Morin , homo demens . Il serait naïf d'oublier qu'intelligence et rationalité ont été mises au service de la destruction de l'humanité , et que sciences et techniques ont permis les méthodes d'extermination les plus efficaces .

=====

LE LEXIQUE: CRISTALLISOIR DES CONNAISSANCES COLLECTIVES

Pierre-Yves Raccah
idl-CNRS
4, rue du Roi de Sicile
75004 Paris
tél.: (1) 42 71 69 76
fax: (1) 42 77 90 90
e-mail: pyr@ccr.jussieu.fr

La position qui passe pour idéologiquement neutre en sémantique linguistique est, encore à notre époque, celle qui assimile les mots à des étiquettes désignant le contenu de pots de confiture, qui seraient des concepts. Selon cette métaphore pâtissière, la signification d'un mot ne serait autre que cette confiture que les concepts contiennent (elle existerait donc indépendamment des langues, qui ne seraient rien d'autre que des systèmes de rangement...); parler reviendrait à donner la recette d'un mélange de confitures, que le destinataire s'empresserait de faire sur la suggestion du locuteur. En gardant à l'esprit cette métaphore, et en supposant que les énoncés témoignent de l'éducation de leurs locuteurs, on dispose d'un éclairage brillant pour saisir le fameux dicton concernant la culture... Mais rien de plus ne me paraît pouvoir être éclairé par cette métaphore. En effet, d'une part, elle conduit à des modèles naïfs, qui ne peuvent ni rendre compte de phénomènes élémentaires qui caractérisent les langues (humaines, naturelles, ...), ni décrire, avec une quelconque plausibilité, le fonctionnement cognitif (humain, naturel, ...); et, d'autre part, elle masque les liens entre les signifiés et la langue, et annule ainsi toute possibilité théorique de décrire les rapports entre le système linguistique du lexique et les idéologies dans lesquelles et à partir desquelles il s'est progressivement construit.

Je reprendrai brièvement la critique de la naïveté des modèles pâtissiers, critique que j'ai déjà tartinée à plusieurs reprises, et que, fort heureusement, je ne suis pas seul à étaler; puis, en analysant un peu les rapports entre le lexique et les idéologies, je proposerai une autre métaphore, plus laborantine: celle du cristallisoir. Je montrerai que mes thèses préférées (caractère argumentationnel du sens, caractère inférentiel et graduel de la signification, et autres réactifs) relèvent de cette seconde métaphore et évitent les travers des modèles naïfs, tout en permettant de formuler des hypothèses riches sur les rapports entre le système lexical, les connaissances et la société.

1. CRITIQUE DES MODÈLES PÂTISSIERS

La logique fascine les linguistes et les philosophes du langage au moins autant que le langage fascine les logiciens... Lorsque le linguiste en attend trop, son usage peut pourtant conduire à des impasses, dues à un usage impropre, puis à des déceptions (lorsque l'on s'aperçoit de l'impasse sans s'être aperçu de l'impropriété de l'utilisation...). Cette fascination a conduit à deux types de situations:

- ♣ les uns s'acharnent à sortir de l'impasse au moyens de systèmes de logique de plus en plus sophistiqués, assumant le risque de déplaire aux logiciens purs, qui ne voient dans ces systèmes que des aberrations, et d'impatienter ceux des linguistes qui ne sont pas fascinés par la logique, lesquels trouvent ad hoc les constructions proposées, et découvrent sans cesse des contre-exemples ou des phénomènes dont le linguiste "logicisé" ne peut pas rendre compte.

- ♣ les autres, après quelques tentatives infructueuses pour décrire tel ou tel phénomène au moyen de tel ou tel système de logique formelle (que ces tentatives soient le fruit de leur travail ou émanent d'autres...), finissent par abandonner tout espoir de recours à la logique pour décrire les phénomènes linguistiques, et se retranchent dans une position selon laquelle il serait impensable et même aberrant, pour ne pas dire condamnable, d'envisager un tel recours.

Avant d'aller plus loin, je dirai quelques mots sur l'ambiguïté du mot "logique", ambiguïté qui complique encore la situation. Le mot "logique" est, en effet, tantôt employé pour désigner, de façon un peu confuse, les "lois objectives de la pensée", tantôt utilisé comme synonyme de "système formel de représentation". Dans le premier sens, *la* logique est vue comme un jeu de règles à caractère universel et normatif, auquel doit se conformer tout enchaînement de pensées digne de ce nom. En ce sens, il n'y a qu'une logique. Mais cette Logique (accordons-lui la majuscule...) ne peut certainement pas être utilisée pour décrire la signification, car il s'agit plutôt de préceptes, de règles de comportement. Dans le deuxième sens du mot, *une* logique est un des systèmes formels de représentation qui respecte tout ou partie des lois de *la* Logique. Le conflit entre les deux positions -engouement ou rejet- me paraît reposer, entre autre, sur une confusion entre ces deux acceptions. Il semble en effet difficile de revendiquer qu'une théorie de la signification ne repose pas sur les lois de la logique; d'autre part, -je tenterai de le montrer dans mon

exposé- l'idée selon laquelle une langue naturelle serait réductible à un système formel de représentation, est inacceptable.

Une célèbre phrase du logicien Richard Montague, remontant au début des années 70, et surtout les interprétations qui l'ont malmenée, illustre bien ces deux positions:

«There is, in my opinion, no interesting theoretical difference between natural languages and the artificial languages of logicians;...»¹

Si je ne partage pas l'opinion exprimée crûment par Montague, j'admets néanmoins une idée qui peut sembler proche de cette dernière: il s'agit de l'idée selon laquelle une théorie scientifique de la signification, pour être adéquate doit *avoir une structure* semblable à celle des théories formelles et, en particulier, aux langages de logique formelle. Il ne s'agit plus ici d'une position empirique concernant la structure des langues, mais d'une position épistémologique, plus précisément méthodologique, relative à la structure des théories concernant les langues.

Je voudrais tout d'abord comparer le rôle que trois notions classiques jouent dans les langages artificiels avec celui qu'elles jouent dans les langues naturelles. Les différences que je tenterai de mettre en lumière pourront apparaître comme des différences fonctionnelles, liées à l'usage: je tenterai de montrer que, au contraire, il s'agit de différences *essentiels*, qui obligent à considérer langues et langages comme deux genres différents, et non pas comme deux espèces du même genre.

1.1.1 Le statut des énoncés

Dans un langage artificiel, chaque énoncé est présenté comme indépendant de la situation d'énonciation et, en particulier, du locuteur. On pourrait exprimer la même idée en disant que les situations d'énonciations sont implicitement quantifiées universellement. Il s'ensuit que, dans les langages formels, la différence de statut entre *phrase* et *énoncé* n'est pas pertinente. En fait, quand les logiciens parlent de *référence*, *dénotation*, *vérité*, ils restent au niveau de la proposition-phrase, sans aucun besoin de relativiser ce qu'ils en disent à un sujet parlant. Ainsi, le sens de tout énoncé de la phrase

(1) La racine carrée de cinq n'est pas un nombre rationnel est indépendant de la situation d'énonciation: il dépend exclusivement de la définition des termes utilisés, définition considérée comme universelle.

¹ Montague, "Universal Grammar", THEORIA 36 p. 373, 1970.

En revanche, les énoncés des langues naturelles sont reliés à des locuteurs et à des situations; et, de plus, cette relation détermine (en partie) le sens de l'énoncé. Ainsi, la phrase (2)

(2) Le chien est encore sur le tapis

donne lieu à différents énoncés, de sens distincts, en fonction des rapports entre le locuteur et le chien, entre le locuteur et le tapis, etc.

Des observations de ce type mettent en lumière une différence essentielle entre la sémantique des langages formels et celle des langues naturelles: la première se limite à construire le sens (unique) que le destinataire (universel) a à découvrir dans chaque énoncé-phrase. La seconde doit expliciter les instructions, marquées dans les phrases, que le destinataire applique pour construire les sens des énoncés de ces phrases, en fonction des différentes situations d'énonciation.

1.1.2 Statut de la notion de vérité

Beaucoup d'encre a coulé à propos de l'utilité de la notion de *vérité* pour la description de la signification des phrases ou pour celle du sens des énoncés des langues naturelles. Cette question ne concerne pas seulement ceux qui veulent utiliser les systèmes de logique formelle pour décrire la signification des phrases, mais aussi ceux qui veulent rendre compte des aspects référentiels des énoncés (sans pour autant s'intéresser aux formalismes). Lorsque j'énonce

(4) Le chien est sur le tapis

je parle de quelque chose dans le monde, et il est légitime de se demander si ce que j'en dis est vrai. Néanmoins, pour qui désire comprendre les structures des langues (naturelles), l'importance de cette question est très relative, si on la compare à sa position centrale dans les langages artificiels. En effet, quand je profère (4) (et ceci est encore plus vrai pour (3)), je fais plus qu'attribuer une propriété à un objet; pour s'en rendre compte, il suffit de comparer (4) avec (5)

(5) Le tapis est sous le chien

On remarquera que, dans toutes les situations où un énoncé de (4) est vrai, un énoncé de (5) est aussi vrai, et *vice versa*, et pourtant, (4) et (5) n'ont pas la même signification. Ces phrases font donc plus que nous renseigner sur les conditions de vérité de leurs énoncés, et il est donc légitime de considérer la notion de vérité comme secondaire, *du point de vue des structures des langues*. Mais ce qui m'intéresse ici est que, même en ce qui concerne les aspects du sens pour la description desquels la notion de vérité pourrait être utile, cette

notion de vérité n'est pas la même que celle qui est utilisée dans les langages formels: cette dernière s'oppose à la *fausseté*, tandis que la notion qui pourrait être utile pour la description des langues s'oppose au *mensonge*. Pour comprendre l'importance de la différence que cette situation crée entre les langues et les langages, il suffit de constater que *fausseté* et *mensonge* ne coïncident pas nécessairement: on peut trouver des cas de *fausseté* dans lesquels il est difficile de parler de *mensonge*, et trouver des cas de *mensonge* dans lesquels on ne peut pas parler de *fausseté*. Ainsi, le congressiste domicilié à Issy-lès-Moulineaux, qui, à Tokyo, déclare à un de ses collègues australiens "J'habite Paris", ne *ment* pas bien qu'il ne dise pas la *vérité*; de même, le passant, fumeur de cigarillos, à qui un clochard demande une cigarette, répondra sans doute, pour avoir la paix, qu'il n'en a pas, ce qui sera *vrai*, mais restera un *mensonge* (s'il lui reste des cigarillos).

1.1.3 Variabilité de la signification à l'intérieur de même la phrase

Un dernier phénomène général que je veux mentionner ici est le fait que, dans les langues, deux occurrences du même mot dans un énoncé doivent généralement être interprétées sur la base de significations différentes, même quand elles jouent le même rôle informationnel. Pour illustrer cette propriété, j'emprunterai un exemple utilisé par Oswald Ducrot dans un séminaire de 1991. Il s'agit d'un vers d'*Andromaque*, de Racine, dans lequel le roi Pyrrhus annonce à Andromaque qu'il ne l'attendra plus et qu'il a décidé de tuer Astyanax, le fils d'Andromaque, comme les grecs le lui demandent:

(6) Je meurs si je vous perds mais je meurs si j'attends

L'articulation avec *mais* ne permet pas que les deux occurrences de *je meurs* soient interprétées dans le même sens: s'il en était ainsi, elles ne pourraient évidemment pas avoir des orientations argumentatives opposées, comme le requiert la présence de *mais*. D'autre part, la présence de *mais* requiert aussi que l'orientation de la deuxième occurrence de *je meurs* soit présentée comme plus importante pour son locuteur. Il en résulte que Pyrrhus, en prononçant (6), présente la "seconde mort" comme plus importante, plus terrible que la première: c'est pour l'éviter (lui préférant la première) qu'il décide de tuer Astyanax. Il est clair que les langages formels ne permettent pas ce genre de choses. Que dire, en effet de

(7) La racine carrée positive de X est paire si X est multiple de 4, mais la racine carrée positive de X est paire si X est pair

Malgré ma présentation en trois paragraphes distincts, les trois propriétés générales que je viens de rappeler sont liées. Par exemple, on ne pourrait pas parler de *mensonge*, s'il n'y avait pas de locuteur, responsable de son

énoncé. Ces différences ne sont alors que des symptômes d'une différence plus profonde entre langue et langage, différence qui a laissé indifférents la plupart des linguistes et la quasi-totalité des logiciens et des informaticiens. La prise en compte de ces phénomènes permet une conception des langues dans laquelle le sens est construit² par le destinataire, ce dernier utilisant pour ce faire, ses connaissances et croyances sur le monde, conjointement aux règles linguistiques évoquées par la phrase et les mots, règles qui constituent leur *signification*.

2. Une alternative aux théories "pâtisseries"

Je propose une alternative aux théories naïves, dans laquelle l'accent est mis sur le caractère inter-subjectif de la langue elle-même (et pas seulement de l'interaction langagière). Cette alternative s'appuie sur une réflexion épistémologique qui prend en compte les contraintes décrites au paragraphe précédent; elle vise à une description des phénomènes de langue (donc, au niveau de la *phrase*), qui rende compte des observations empiriques (qui ne peuvent donc se situer, au mieux, qu'au niveau de l'*énoncé*³). Après avoir présenté les grandes lignes de cette réflexion épistémologique, je propose le modèle théorique appelé Théorie de l'Argumentation dans la Langue (TAL), qui constitue une manière **parmi d'autres** de prendre en compte les résultats de cette réflexion: choisir ce cadre théorique implique l'adoption d'hypothèses spécifiques, caractéristiques de ce cadre théorique, et qui s'ajoutent à celles qui découlent de la réflexion épistémologique⁴. Je parlerai enfin de la place que la TAL donne au lexique et de la manière dont ce dernier, dans la TAL, cristallise les connaissances et les croyances en vigueur dans une communauté linguistique.

2.1 Effets de sens et signification

L'étude empirique des énoncés permet, au linguiste, non pas d'en observer directement le sens, mais d'en proposer une construction, à partir des effets observables de ces sens. Ces effets peuvent être de tout ordre:

- effets référentiels;
- effets illocutoires;
- effets argumentatifs;
- effets littéraires et poétiques;
- effets émotionnels;

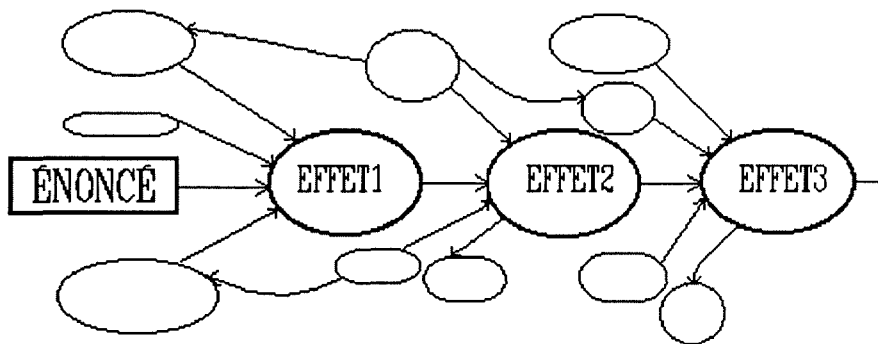
² Ou plutôt "sculpté", comme me l'a suggéré Evelyne Andréewsky, pour éviter de laisser entendre que le sens est assemblé à partir de "petits bouts de sens".

³ En toute rigueur, l'*énoncé*, lui-même, est déjà une construction théorique, fondée sur des "observations" plus empiriques, telles que l'occurrence d'énoncé, l'effet de sens, etc.: cette question est abordée plus bas.

⁴ D'une manière générale, l'épistémologie ne peut pas se substituer à la science...

- effets hypnotiques (pourquoi pas ?);
- effets curatifs, religieux, ou autres (pour ceux qui y croient...);
-

Selon les points de vue, tel type d'effet de sens -si tant est qu'on l'admette- peut apparaître comme "*plus proche*" du sens que tel autre, en ce sens, par exemple, qu'une chaîne causale reliant l'énoncé à l'effet en question serait plus courte et contiendrait moins de points auxquels pourraient intervenir d'autres causes:



Après avoir précisé la place et le rôle que, selon cette conception, la sémantique linguistique doit jouer, je montrerai la nécessité d'une description argumentationnelle de la signification, et l'autonomie d'une telle description vis à vis d'une description informationnelle. Je proposerai un modèle théorique et descriptif qui rende justice à ces deux aspects de la signification.

2.2 Le rôle de la sémantique linguistique

Quelque soient les divergences concernant le "classement" des effets de sens, l'ensemble des disciplines s'intéressant aux énoncés sont d'accord pour admettre que ce sont les effets de sens les plus "*proches*" de l'énoncé (dans la chaîne causale mentionnée plus haut) qui relèvent, en quelque sorte par définition, de la linguistique. Cette dernière serait donc censée "coller à la langue" et proposer des outils utilisables par les autres disciplines. Cette position, qui fournit une sorte de *cahier des charges* à la linguistique, n'a pas un caractère d'évidence suffisant pour justifier qu'elle reste implicite dans la quasi-totalité des approches; je l'adopterai néanmoins, sans en approfondir ici l'analyse.

De ce point de vue, la sémantique linguistique a pour tâche:

- de mettre en évidence ce qui, dans les phrases de la langue, instruit sur la construction (ou la sculpture...) du sens des énoncés de ces phrases⁵, et
- de décrire les mécanismes de cette construction (ou sculpture...), étant entendu qu'il ne s'agit pas ici de mécanismes psychologiques dont on postulerait qu'ils seraient mis en oeuvre dans l'activité humaine de compréhension: du point de vue du linguiste, il s'agit bien d'une *re-construction* formelle (et surtout pas d'une *sculpture*, si l'on ne veut pas être ésotérique...) du résultat de cette activité de construction (ou de sculpture...). Cette *re-construction* entretient néanmoins des liens avec la description psycho-linguistique des mécanismes humains mis en oeuvre dans l'activité de compréhension⁶.

2.3 Quelques mots sur la TAL

2.3.1 *Information et argumentation*

Le courant très nettement dominant en sémantique linguistique considère les effets référentiels comme tellement "*proches*" du sens, qu'ils seraient pratiquement identifiables à ce dernier. C'est le cas de toutes les sémantiques utilisant, de près ou de loin, la logique mathématique comme système de représentation du sens: de Davidson à Montague et de Chomsky à Sowa, en passant par Shank ou Kamp, la signification des phrases (qui n'est généralement, chez ces auteurs, pas distinguée du sens des énoncés) est assimilée à une formule d'un système de logique⁷, destinée à "*capturer* l'information objective *contenue* dans les phrases".

Je montrerai pourquoi une description sémantique des phrases ne peut limiter son ambition à rendre compte de ceux des effets de sens des énoncés qui relèvent de l'information objective: des contraintes sur l'orientation argumentative des énoncés sont inscrites dans la langue elle-même, et ne sont pas exprimables en termes d'information. Pour cette *démonstration* (la nécessité d'une composante argumentationnelle autonome dans la sémantique des phrases), j'étudierai les contraintes sur l'argumentation introduites par certains connecteurs et opérateurs, et montrerai qu'elles sont indépendantes de la situation d'énonciation. D'où la nécessité d'inclure ces contraintes dans la description sémantique des phrases. Je montrerai ensuite que ces contraintes

⁵. Le sens étant établi, comme suggéré ci-dessus, par l'observation et la hiérarchisation des effets de sens.

⁶. Voir [Racah 1984] pour un développement sur ce sujet.

⁷. Calcul des prédicats, pour les plus faibles; théorie des types, logiques modales, intensionnelles, non monotones, pour les plus élaborées; réseaux de toutes sortes, pour les ordinateurs, leurs ingénieurs et beaucoup de leurs théoriciens....

n'ont aucune influence sur l'information objective qu'expriment les énoncés, et ne sont nullement influencées par elle. D'où l'autonomie de la composante argumentationnelle.

Considérons les phrases

(8) Jean est intelligent mais brouillon

et

(9) Jean est brouillon mais intelligent

et examinons les relations entre certains des effets de sens de leurs énoncés possibles. Les énoncés possibles de (8) et de (9) se présentent comme décrivant des propriétés ou des dispositions d'un individu nommé "Jean". Les propriétés attribuées à Jean en (8) semblent bien les mêmes que celles qui lui sont attribuées en (9). Néanmoins, certains des effets de sens des énoncés de (8) et de (9) diffèrent: par exemple, un énoncé de (8) peut suggérer que Jean (malgré ses qualités) risque de ne pas être à la hauteur d'une certaine tâche, tandis que, dans la même situation, un énoncé de (9) suggérerait, au contraire, que (malgré ses défauts) Jean serait sans doute à la hauteur de cette tâche. Convenons d'appeler *orientation argumentative* ce type de suggestions auxquelles les énoncés invitent dans telle ou telle situation. Nous observons que les phrases (8) et (9), qui, du point de vue de leur forme, ne diffèrent que par l'ordre des adjectifs reliés par la conjonction *mais*, sont marquées pour des orientations argumentatives opposées, et ce, quelle que soit la nature de ces orientations (laquelle dépend de la situation d'énonciation). Si le rôle de la sémantique est bien celui qui a été précisé plus haut, les descriptions sémantiques de (8) et de (9) doivent alors rendre compte de cette contrainte sur les orientations argumentatives. En l'occurrence, c'est dans la description de *mais* que doit figurer la description de cette contrainte. L'analyse des phrases contenant des opérateurs comme *peu*, *un peu*, *presque*, *à peine*, etc. ou d'autres connecteurs comme *néanmoins*, *pourtant*, *justement*, confirme cette nécessité d'inclure les contraintes sur l'orientation argumentative dans leur description sémantique. Une composante *argumentationnelle* (si l'on accepte de l'appeler ainsi) est donc nécessaire dans la sémantique des phrases.

Pour montrer que la description des contraintes sur l'orientation argumentative est autonome vis à vis de la description informationnelle, c'est à dire, que les contraintes argumentationnelles ne peuvent pas être décrites exclusivement au moyen de concepts relevant de l'information objective, je distinguerai deux types de lecteurs: ceux qui sont déjà convaincus que la

signification des mots et des phrases ne se résume pas à leurs conditions de référence⁸, et les autres.

Pour les premiers, le raisonnement est simple: l'orientation argumentative repose sur certains -au moins- de ces éléments de signification qui ne relèvent pas des conditions de référence (trivial pour ce groupe de lecteurs), donc les contraintes sur ces orientations ne peuvent pas être décrites exclusivement au moyen de concepts relevant des conditions de référence.

Pour les autres, nous allons supposer un instant que la description de la signification de certains mots est épuisée par celle de leurs conditions de référence⁹. Supposons donc que la signification de *intelligent* et celle de *brouillon* soient totalement décrites par leurs conditions de référence. Si l'on tient compte de la démonstration précédente, à savoir, que les descriptions sémantiques des phrases (8) et (9) doivent rendre compte des contraintes sur les orientations argumentatives de leurs énoncés, deux possibilités semblent se présenter, selon que l'on considère que (8) et (9) ont ou n'ont pas les mêmes conditions de référence.

Dans la première hypothèse, on ne peut évidemment pas attribuer la différence argumentationnelle à une différence informationnelle objective, puisque nous sommes précisément dans l'hypothèse dans laquelle il n'y a pas de différence dans les conditions de référence. C'est ce qu'il fallait démontrer.

La deuxième hypothèse soulève deux problèmes sérieux:

- a) en premier lieu, il est difficile d'admettre que (8) et (9) puissent différer par leurs conditions de référence; et
- b) même si on se forçait à l'admettre, l'idée selon laquelle cette différence expliquerait la différence dans les orientations argumentatives est indéfendable, puisque, comme on l'a vu, cette dernière est une différence de points de vue concernant le même fait.

2.3.2 Le dispositif de description I: les *topoi*

Ces observations et les remarques qui en découlent imposent des contraintes importantes sur ce que devrait être une théorie adéquate de la signification. Je ne développerai pas ici cette question (voir, à ce sujet, [Anscombe et Ducrot 1983], [Racah 1986 et 1987]). Je me bornerai à résumer les hypothèses des premières versions de la Théorie de l'Argumentation

⁸. Ce concept est destiné à préciser la notion un peu vague d'*information objective*.

⁹. Il s'agit, bien sûr, d'une concession bien provisoire; cf. plus bas.

dans la Langue (TAL); d'autres hypothèses, plus récentes, seront exprimées au paragraphe 3.

La TAL se caractérise par l'hypothèse suivante: lorsqu'un énoncé E se présente comme destiné à suggérer la conclusion C, il le fait en vertu d'une règle d'inférence *graduelle*, présentée comme *partagée* par l'ensemble des interlocuteurs, et présentée comme *générale*. Ces règles, qui sont de la forme

// plus (ou moins) P, plus (ou moins) Q //

sont appelées *topoi* (au singulier: *topos*). Je ne m'attarderai pas sur cette hypothèse (cf. [Racah 1990] pour un exposé plus détaillé), sauf pour préciser qu'il s'agit d'une hypothèse très forte (réduire tous les garants des argumentations à la forme ci-dessus), qui, comme toute hypothèse constitutive d'une théorie, ne peut être validée que par le pouvoir descriptif (et l'adéquation) de la théorie (ou des théories) qu'elle engendre.

Le sens des énoncés étant décrit au moyen de ces *topoi*, il est alors possible de formuler une description sémantique des connecteurs et des opérateurs ayant des effets sur l'orientation argumentative: cette description est formulée en termes de contraintes sur les *topoi* utilisables comme garants. Ainsi, la description sémantique de *mais*, dans une phrase de type *P mais Q*, indiquera d'une part que le *topos* utilisé pour l'énoncé de Q doit avoir un conséquent opposé à celui du *topos* utilisé pour l'énoncé de P; et, d'autre part, que le *topos* utilisé pour l'énoncé global est celui qui aurait été utilisé dans un énoncé de Q tout seul (cf. [Racah 1987] pour une description plus technique).

2.3.3 *Raisonnement et argumentation*

Ces observations -et beaucoup d'autres- montrent que la signification des phrases ne se réduit pas à des éléments d'information mais intègre des indications sur le potentiel argumentatif de leurs énoncés, indications fondées sur des règles d'inférence graduelles. Ce fait pourrait ne concerner que le linguiste, car même si les moyens linguistiques permettant de transmettre des connaissances sont régis par ces modèles argumentatifs, il ne s'ensuit pas immédiatement que les connaissances elles-mêmes le soient. La démarche que je propose requiert une hypothèse supplémentaire, découlant de préoccupations cognitives. Il s'agit de l'hypothèse selon laquelle la langue est, en quelque sorte, un miroir des représentations cognitives des locuteurs. Si l'on admet cette dernière hypothèse (que je ne justifierai pas ici), on est amené à penser que certains aspects, au moins, des structures de nos connaissances sont homomorphes aux structures linguistiques de l'argumentation. L'étude des connecteurs et des opérateurs argumentatifs a pu, ainsi, enrichir notre compréhension du raisonnement et, par conséquent, fournir des bases nouvelles pour sa modélisation en intelligence artificielle. Mais cette hypothèse ne concerne pas

seulement les enchaînements argumentatifs marqués par les connecteurs et autres articulateurs: des travaux de description du lexique sont actuellement en cours, qui révèlent le rôle essentiel de l'inférence argumentative dans la construction (la *sculpture*) du sens.

2.3.4 *Inférence argumentative et implication*

Avec Sylvie Bruxelles, j'ai proposé une description argumentative du connecteur français *si*, qui rend compte d'un certain nombre d'observations incompatibles avec la description classique en termes d'implication, tout en admettant, comme cas limite, l'usage implicatif des mathématiciens (cf. Bruxelles-Racah 1987). En guise de première application de la théorie de l'argumentation, je souhaite illustrer rapidement l'intérêt qu'il y a, dans certains cas, à remplacer une règle de production par un topos.

Pour résumer la description que Sylvie Bruxelles et moi faisons de *si*, on peut considérer qu'un énoncé contenant *si* s'appuie sur une chaîne de topoi,

// plus P1, plus P2 // , // plus P2, plus P3 // , ... , // plus Pn, plus Q //

pour construire et utiliser le topos

// plus P1, plus Q //

où P1 est le champ topique de l'antécédent de l'énoncé et Q le champ topique de son conséquent.

Ainsi, selon notre description, lorsqu'un locuteur énonce

si A, B

il se présente comme utilisant A comme argument pour B, en se fondant sur une chaîne de mouvements argumentatifs partant de A et se terminant en B.

Cette description rend immédiatement compte du caractère graduel et du caractère relativement réversible des raisonnements "naturels" utilisant *si*, caractères illustrés tous deux par l'analyse de l'énoncé suivant:

(10) Si tu abîmes le vase, tu seras puni

supposé destiné à un enfant. Il est clair, en effet, que l'enfant comprend (à juste titre) que s'il n'abîme pas le vase, il ne sera pas puni (toutes choses restant égales par ailleurs); que s'il l'abîme peu, il aura une petite punition, et que s'il l'abîme beaucoup, il aura une forte punition. D'après notre description, (10) "annonce" que le locuteur utilise un topos du genre:

// Plus un enfant se comporte mal,

plus désagréables pour lui sont les réactions des adultes //

ce qui rend compte de la gradualité des raisonnements "naturels" utilisant *si*: si l'enfant décèle, dans (10), une règle comme la règle ci-dessus, il comprendra du même coup que, plus le vase sera abîmé, plus la punition sera sévère. D'autre part, le topos ci-dessus correspond à la même croyance que le topos

// Moins un enfant se comporte mal,

Moins désagréables pour lui sont les réactions des adultes //

ce qui rend compte de la réversibilité de (10) et, plus généralement des énoncés "naturels" utilisant la conjonction *si*. Il s'agit d'un type de réversibilité: celle qui inverse les sens de parcours des champs topiques sans en intervertir l'ordre, et donc sans inverser la direction de l'inférence; je parlerai de ce deuxième type de réversibilité au paragraphe 2.3.

2.4 Règles de production et règles topiques

Une telle description du connecteur français *si* conduit à penser que la représentation d'énoncés contenant *si*, sous forme de règles de production constitue une simplification abusive des éléments d'expertise que l'on essaye de modéliser. En effet, s'il est vrai que, dans certains cas, de tels énoncés peuvent avoir une valeur implicative, il s'agit de cas particuliers, dans lesquels les champs graduels auxquels renvoient l'antécédent et le conséquent de ces énoncés ne possèdent que deux valeurs, correspondant au vrai et au faux (voir Prade 1988, pour une représentation des topoi utilisant la logique floue). Ainsi, lorsqu'un élément de connaissance est formulé par une phrase française de type

si A, B

sa représentation par l'implication logique de A à B masque la richesse sémantique d'une telle formulation. Cette richesse apparaît, au contraire, si on représente

- A comme une valeur d'un champ graduel P;
- B comme une valeur d'un champ graduel Q;
- le lien entre A et B comme une instance d'un lien topique entre P et Q.

Ce principe de représentation permet de rendre compte du fait que, si l'énoncé de l'expert permet de conclure de A à B, et si A' est plus P que A, alors, ce même énoncé permet de conclure de A' à B', B' étant au moins autant Q que B.

L'exemple suivant permettra de fixer les idées. Considérons l'énoncé:

(11) Si la houle est forte, la digue doit être épaisse

Une représentation de (11) sous forme d'implication, telle que

FORTE(houle) IMPLIQUE EPAISSE(digue)

ne permet pas d'inférer directement ce que l'on doit attendre dans le cas d'une houle moyennement forte ou très forte: il faudrait disposer de représentation du genre:

FORCE(houle,moyenne) IMPLIQUE EPAISSEUR(digue,moyenne), et
FORCE(houle,élevée) IMPLIQUE EPAISSEUR(digue,élevée), et

....

et ainsi de suite, pour chaque valeur possible du paramètre FORCE. Si, en revanche, on représente (11) par quelque chose du genre:

// Plus la digue est exposée, plus elle doit être résistante //

en considérant le fait la houle est forte comme un degré dans le champ EXPOSITION DE LA DIGUE, et le fait la digue est épaisse comme un degré dans le champ RESISTANCE DE LA DIGUE, les inférences désirables peuvent être obtenues.

Les topoi permettent donc de "capturer", de manière directement utilisable, les connaissances concernant les variations de paramètres (objectifs ou subjectifs), et leurs conséquences sur d'autres paramètres, sans toutefois permettre de calculer les valeurs correspondantes. En ce sens, l'approche topique est voisine de celle qui est proposée par la physique qualitative et, plus généralement, par l'étude du raisonnement qualitatif.

3 Lexique et champs topiques

3.1 Enchaînement et lexique

Jusqu'à une date assez récente, la majeure partie de la recherche sur l'argumentation concernait exclusivement les articulations linguistiques, marques des enchaînements, traces des raisonnements implicites des sujets parlants. Même les travaux fondés sur le cadre théorique de l'*Argumentation dans la Langue* proposé initialement par Anscombe et Ducrot et brièvement décrit ci-dessus, se sont centrés sur l'étude des *articulateurs argumentatifs* (opérateurs et connecteurs), en dépit de l'hypothèse caractéristique de ce cadre, selon laquelle, précisément, "il y a de l'argumentation dans la langue".

En se préoccupant fondamentalement de l'enchaînement, la sémantique de l'argumentation restait une "macro-sémantique"; et puisqu'on s'intéressait surtout à ces traces de l'activité d'inférence, la proximité de cette conception de la sémantique avec la pragmatique était suffisante pour que l'amalgame soit inévitable.

En effet, la description des articulateurs apparaissant dans une phrase, au moyen de contraintes sur les topoi susceptibles d'être mis en oeuvre lors de l'énonciation de la phrase, suppose que l'on ait une idée des topoi pouvant être mis en oeuvre lors de l'énonciation des sous-phrases articulées. Or, puisque aucune recherche n'était prévue au départ sur l'émergence de ces topoi au fil de la construction de la phrase, il était nécessaire de se placer d'emblée sur le terrain de l'énoncé: il fallait supposer que l'on s'était donné une situation d'énonciation, caractérisée par les topoi utilisables dans cette situation. D'où l'aspect pragmatique de ce type d'approche.

Néanmoins, les hypothèses de la théorie de l'Argumentation dans la Langue n'imposent pas une approche exclusivement pragmatique. Elles permettent, au contraire un traitement sémantique poussé de l'argumentation, à condition d'accorder une certaine attention aux phénomènes lexicaux. C'est du moins ce que j'espère montrer. Bien entendu, comme nous l'avons vu plus haut, l'argumentativité d'un énoncé, simple ou complexe, ne peut pas être déterminée par la sémantique seule, puisque l'orientation argumentative dépend des topoi *effectivement* mis en oeuvre lors de l'énonciation; ces topoi reflètent à la fois les croyances et les intentions des locuteurs et ne peuvent donc pas être déterminés par la seule analyse de la phrase. Mais les mots choisis contraignent les topoi possibles (ou encore, du point de vue opposé, les topoi que l'on veut utiliser contraignent le choix des mots): n'importe quel topos ne peut pas être associé à n'importe quelle phrase. On pourrait parler ici de *sous-détermination*, dans un sens positif (c'est à dire, dans un sens où sous-déterminer implique un certain degré de détermination).

De ce point de vue, l'étude linguistique de l'argumentation ne peut pas se limiter à l'analyse des morphèmes utilisés pour articuler les arguments et les conclusions. Cette analyse ne constitue qu'un des aspects, une des directions, de la recherche linguistique sur l'argumentation, certes importante, mais qui ne doit pas masquer un deuxième aspect: la (sous-)détermination des topoi au fil de la construction des phrases.

3.2 Le dispositif de description II: (champs topiques lexicaux)

Je tenterai de montrer qu'une description lexicale faisant apparaître des champs topiques associés aux mots, permet non seulement d'étayer la construction des contraintes sur l'argumentation que les énoncés d'une phrase mettent en jeu, mais aussi de rendre compte de leurs connotations habituelles, et de décrire, de ce fait, la manière dont le lexique cristallise les croyances et connaissances collectives, nécessaires à la communication langagière à l'intérieur d'un groupe linguistique. L'idée générale de cette démarche peut être schématisée de la façon suivante.

Un topos peut être conçu comme un couple de champs topiques, couple dont le premier terme est l'antécédent du topos et le deuxième terme, le conséquent. Un champ topique est, en gros, une "façon de voir" une entité, une propriété ou une relation. Cette façon de voir est, elle-même, déterminée par la façon dont on voit une autre entité, une autre propriété ou une autre relation: c'est à dire par un autre champ topique. On peut ainsi représenter un champ topique par une chaîne de champs topiques emboîtés les uns dans les autres, de telle sorte que chaque champ topique est caractérisé d'une part, par un champ conceptuel (l'entité, la propriété ou la relation), et d'autre part, par le champ topique qu'il contient, lequel est lui-même caractérisé par un champ conceptuel et par le champ topique qu'il contient, et ainsi de suite jusqu'à un *champ topique élémentaire*. Ce dernier étant un principe de valuation, introduit une gradation dans le champ topique qui le contient. Nous sommes ainsi amenés à la définition récursive suivante:

Le couple (X,Y) est un champ topique si et seulement si l'une des deux conditions suivantes est respectée:

- (i) X est un champ conceptuel et
Y est une valeur (bien ou mal)
- ou
- (ii) X est un champ conceptuel et
Y est un champ topique.

Cette définition permet de construire des chaînes croissantes de champs topiques à partir de champs topiques élémentaires, contenant une valuation: le principe de gradation introduit par la valuation contenue dans le champ topique élémentaire se répercute sur la totalité du champ topique qui le contient. C'est ce qui permet de considérer un champ topique non seulement comme une façon de voir un champ conceptuel, mais aussi comme un principe de gradation pour d'autres champs topiques.

- A chaque champ topique CT (= <CC,CT' >) correspond un topos, qui lui est canoniquement associé: il s'agit du couple (CT,CT').
- A chaque mot est associé un ou plusieurs champs topiques¹⁰. Un champ topique associé à un mot est dit *intrinsèque* à ce mot. Un *topos intrinsèque* à

¹⁰ Il y a deux manières (combinables) d'associer plusieurs champs topiques à un mot. Dans le premier cas, les différents champs sont *conjointement* associés au mot: la signification du mot intègre ces différents champs topiques; les enchaînements peuvent ne concerner qu'un de ces champs, mais peuvent aussi en intégrer plusieurs. Dans le deuxième cas, les champs sont *exclusifs* les uns des autres: le mot est argumentativement ambigu, phénomène qui se produit assez souvent, même dans le cas de mots informativement non-

un mot est le topos canoniquement associé à un champ topique intrinsèque à ce mot.

- Un énoncé d'une phrase contenant un mot M peut utiliser soit un topos intrinsèque à M: il s'agit alors d'un *énoncé doxal*, soit un autre topos, qui peut être contraire au topos intrinsèque à M (on parle alors d'*énoncé paradoxal*), ou simplement différent. Dans tous les cas, l'antécédent du topos est un champ topique intrinsèque de M.

3.3 Exemples

Les trois exemples suivants, bien que brièvement analysés, permettront d'illustrer ces derniers points:

(12) Jean a beaucoup travaillé, il doit être fatigué

(13) Jean a beaucoup travaillé, il doit être en pleine forme

(14) Jean a beaucoup travaillé, Marie a dû s'ennuyer

Sylvie Bruxelles, Oswald Ducrot et moi-même considérons que la description sémantique de *travailler* doit faire apparaître que, pour qu'une activité puisse être considérée comme du *travail*, elle doit (entre autre) impliquer une certaine fatigue (pour s'en convaincre, il suffit de remarquer que, pour s'opposer à l'idée selon laquelle quelqu'un aurait travaillé, il suffit de faire remarquer qu'il ne s'est pas fatigué...). Si l'on admet cette hypothèse externe, on comprendra que nous associons au lexème travailler le champ topique (ACTIVITE,*fatigue*), où ACTIVITE est le champ conceptuel principal de travailler et *fatigue* est le champ topique *du point de vue duquel* l'activité est considérée. Le topos canoniquement associé à ce champ topique est

// (ACTIVITE,*fatigue*),*fatigue* //

qui peut se lire aussi:

// plus on agit, plus on se fatigue //

Les énoncés de la phrase (12) sont *doxaux* en ce qu'ils utilisent un topos intrinsèque à l'un de ses lexèmes: un locuteur de (12) n'ajoute rien à la *doxa* codée dans la langue. Les énoncés de (13) peuvent provoquer une première réaction d'étonnement, qui se dissipe dès que l'on comprend que le locuteur se présente comme s'opposant à la doxa suggérée par le mot travailler: le topos utilisé

ambigus (cf. Racciah 1987); c'est le cas de *tiède*, qui peut être associé soit au champ de la chaleur (comme dans «Cette bière est tiède») soit au champ du froid (comme dans «Ce café est tiède»). Enfin, on peut imaginer (je n'en ai pas trouvé, mais je n'en ai pas cherché...) des cas d'associations combinées.

// plus on agit, plus on est en forme //

est opposé à l'un des topoi intrinsèques de *travailler*: les énoncés de (13) sont *paradoxaux*. Enfin, les énoncés de la phrase (14) font appel à un topos du genre

// plus quelqu'un travaille, plus son entourage s'ennuie //

qui pourrait être obtenu par la concaténation de deux topoi:

// plus on travaille, moins on s'occupe de son entourage //

et

// moins on s'occupe de quelqu'un, plus il (elle) s'ennuie //

Ces énoncés proposent un point de vue qui n'est pas habituellement "codé", du moins me semble-t-il, dans le mot *travailler*. Néanmoins, et c'est ce qui me fait parler de "cristalliseur", dans une communauté linguistique qui utiliserait régulièrement ce topos, au point que le point de vue auquel il correspond deviendrait doxal pour cette communauté (on pourrait dire ici "non marqué", pour reprendre la terminologie classique des linguistes), l'inférence de *travailler* à *ennui de l'entourage* devenant automatique, elle devrait être décrite comme appartenant à la signification même du mot, pour cette communauté.

En adoptant ce point de vue, on est amené à concevoir les mots comme des "réservoirs d'inférences" (mais il ne s'agit plus de pots de confiture...), des indicateurs de points de vue (les étiquettes des pots...): ce que ces pots contiennent n'est plus une sorte de signification référentielle (la confiture...), mais des manières de voir le monde, manières qui contraignent la "sculpture du sens". Ces pots se remplissent de ces *recettes*, au fur et à mesure que de nouveaux points de vue sont intégrés dans la collectivité qui les utilise. Il me semble, mais je n'ai pas encore les moyens de le prouver, que ces pots ne se vident pas: lorsque, dans une collectivité, une manière de voir a été associée à un mot, elle y reste associée, même si, par la suite, un autre point de vue lui est associé. Si l'on tient absolument à conserver une métaphore pâtissière, disons que la signification est un mille-feuille...

BIBLIOGRAPHIE SOMMAIRE

Anscombre Jean-Claude et Oswald Ducrot (1983):

L'argumentation dans la langue, Bruxelles, Mardaga.

Bruxelles Sylvie, Denis Carcagno et Corinne Fournier (1987):

"Vers une construction automatique des topoi à partir du lexique", à paraître dans les actes de CC20, Gand.

Bruxelles Sylvie et Pierre-Yves Raccach (1987):

"Information et argumentation: l'expression de la conséquence", Paris, COGNITIVA 87.

Davis Harley (1989):

"Using models of dynamic behaviour in expert systems", Neuvièmes Journées Internationales d'Avignon".

Dieng Rosé (1987):

"Génération de topoi à partir de règles d'un système expert", à paraître dans les actes de CC20, Gand.

Dieng Rose et Brigitte Trousse 1989:

"Utilisation de connaissances graduelles en intelligence artificielle", Neuvièmes Journées Internationales d'Avignon".

Ducrot Oswald:

(1973): *La preuve et le dire*, Paris, Mame.

(1980): *Les Echelles argumentatives*, Ed. de Minuit.

(1988): "Topoi et formes topiques" *Bulletin d'études de linguistique française*, vol 22; Tokyo.

Fournier Corinne et Pierre-Yves Raccach (1988):

"Argumentation and artificial intelligence: from linguistic models to knowledge management"; Communication à ALLC, Juin 1988. In Yaacov Choueka (ed.) *Computers in literary and linguistic research*, Genève, Slatkine 1990.

Prade Henri (1988):

"Raisonnement avec des règles d'inférence graduelle", *Revue d'Intelligence Artificielle*.

Raccach Pierre-Yves:

(1984a): "Argumentation in Representation Semantics", in *Proceedings of the tenth Conference in Computational Linguistics (COLING)*, Stanford, Cal. Juin 1984.

- (1984b): "Où voulez-vous en venir?" Colloque de neuro-psycho-linguistique et sciences cognitives, St Riquier, Mai 1984; actes publiée en 1986 dans STS.
- (1987): "Modelling argumentation and modelling with argumentation", Rapport CNRS-idl, Paris; à paraître dans Argumentation, Bruxelles.
- (1992): "Expertise et connaissances implicites: de la gradualité des structures cognitives", *RIS*, 6:1-2, pp. 11-26.

COMPORTEMENTS COLLECTIFS EN I.A. DISTRIBUEE

Marie-Pierre GLEIZES, Pierre GLIZE
IRIT - Université Paul Sabatier
118 Route de narbonne, 31062 TOULOUSE Cedex
Email: glize@irit.fr Tel: 61 55 61 76

RESUME

Pour parler d'"intelligence collective", il faut au préalable supposer l'existence d'un savoir réparti entre plusieurs entités et dont l'agrégation conduit à un comportement collectif cohérent et non extérieurement guidé. SYNERGIC, dont les principes sont développés dans cet article, est un outil d'intelligence artificielle distribuée qui est employé dans trois applications d'aide à la décision: en télémédecine, en recherche documentaire et en prévision des crues.

L'approche distribuée permet une méthode de développement de projet totalement incrémentale:

- les agents d'une application sont ajoutés à la collectivité au gré des besoins des concepteurs sans une vue globale préalable,
- le savoir d'un agent, qui est naturellement conçu dans sa globalité par un expert, est totalement réparti dans un réseau de cellules pour son traitement

Nous présentons dans cet article comment, d'une distribution éclatée de compétences, peuvent résulter "émerger" des comportements collectifs cohérents dans une société d'agents et dans une société de cellules. Ces principes et techniques d'I.A.D sont illustrés par des exemples de fonctionnement d'applications opérationnelles.

MOTS-CLES

Intelligence artificielle distribuée, coopération, réseau, parallélisme, auto-organisation, émergence

1. INTRODUCTION

L'Intelligence Artificielle Distribuée (I.A.D) s'intéresse aux domaines pour lesquels un seul résolveur de problème, une seule machine ou un seul lieu de traitement semble inapproprié (Davis, 1981). Dans un système d'IAD, chaque agent ne dispose que d'une connaissance parcellaire qui ne requiert qu'un mécanisme de raisonnement simplifié, compensé par la complexité des interactions entre agents pour déterminer les structures de

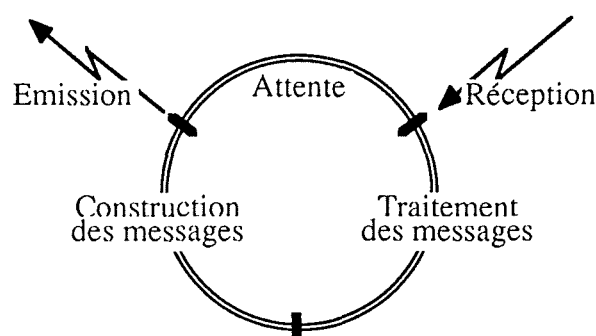
communication ou définir la connaissance commune dont ils disposent. L'importance du raisonnement de chacun des agents est en synergie avec l'interaction entre ces agents: la connaissance est donc simultanément cognitive et sociale.

La description statique d'une collectivité correspond à son organisation et est décrite par le réseau de ses relations, tandis que sa dynamique permet d'étudier son comportement en situation.

1.1. Le modèle d'exécution

Dans nos travaux, nous employons des principes de distribution à tous les granules de connaissances, car le savoir de chaque agent est lui-même complètement distribué dans des cellules. Depuis 1986, nous déclinons ces principes sur un générateur de systèmes multi-agents appelé SYNERGIC.

Le modèle d'exécution est commun à l'ensemble des entités autonomes de SYNERGIC (cellules et agents). Il est constitué d'un cycle de base décomposé en trois parties comme indiqué sur le schéma joint.



Modèle d'exécution des agents et cellules

Une entité qui reçoit des messages les prend en compte en fonction de ses capacités cognitives: dans une cellule cela correspond plutôt à un réflexe, tandis qu'un agent réalise en général des inférences beaucoup plus complexes. Le traitement achevé, l'entité diffuse vers son environnement des informations relatives à son état.

Les interactions sont réalisées exclusivement par des envois de messages.

1.2. Le principe de la communication

Le comportement rationnel et finalisé des entités, présenté ci-dessus garantit l'efficacité et la pertinence des interactions. Cette cohésion sociale est principalement due à l'intégration, dans les principes de communication, de maximes de la conversation développé initialement par Grice (Grice,75). Les transactions particulières que constituent l'échange d'informations entre agents ou cellules obéissent à ces maximes. En résumé, l'entité émettrice tente d'être aussi vraie, aussi pertinente, aussi claire qu'elle peut, tandis que l'entité réceptrice postule que c'est bien le cas des informations reçues.

Maximes de quantité. Si une entité a traité un message qui ne modifie aucunement son état interne, alors il n'y aura pas de propagation vers les entités voisines. Cela évite les redites et minimise le nombre de messages.

Maximes de qualité: que la contribution soit vraie. L'émission de messages n'est effectuée que lorsque l'entité ne peut plus raisonner (état stable); ce qu'elle sait est donc temporairement véridique.

Maxime de relation. Chaque entité s'engage à communiquer ses informations, ce qui garantit la cohérence des informations distribuées.

Maximes de manière pour éviter l'obscurité d'expression et l'ambiguïté. Il n'y a pas d'ambiguïté car une cellule donne son identificateur dans le message tandis qu'un agent rajoute la chaîne symbolique "de l'agent X" à l'information transmise. Le message transmis entre agents est clair car l'émetteur traduit si nécessaire le message dans la terminologie du destinataire.

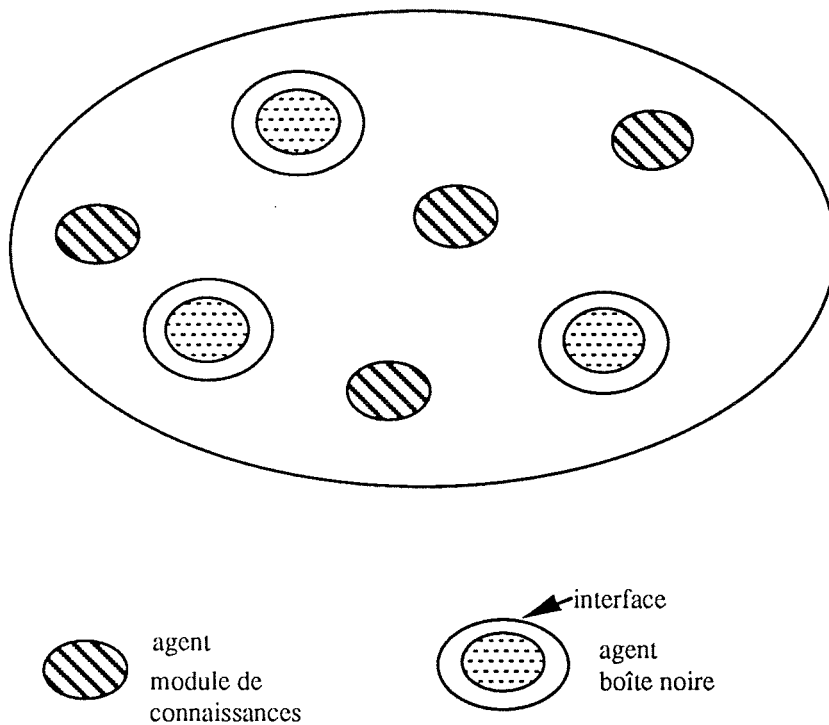
2. COMPORTEMENTS D'UNE SOCIÉTÉ D'AGENTS

2.1. La représentation d'une société d'agents

SYNERGIC (Gleizes, 92) est un environnement de développement de systèmes multi-agents, dans lequel les agents sont:

- soit des boîtes noires: capteurs, programmes d'affichage d'images, programmes mathématiques...
- soit des modules de connaissances qui utilisent les techniques de raisonnement propres à SYNERGIC.

Les agents peuvent être identiques ou très différents de par leur granularité, leur rôle, leur comportement.



La société d'agents dans SYNERGIC

Dans ce papier, nous nous intéressons uniquement aux agents modules de connaissances. Ce sont des agents cognitifs qui en général, correspondent à un domaine d'expertise. Dans le système Télémac qui est une application de SYNERGIC au domaine médical, les agents modules de connaissances sont le module du généraliste, des fonctions vitales, des traumatismes thoraciques, des douleurs abdominales, des douleurs thoraciques, des pertes de connaissances brèves et des blessures (tête, mains, bras, jambes, dos).

Ces agents sont constitués des connaissances relatives à leur domaine d'expertise et des connaissances d'acointances. L'acquisition des connaissances du domaine est réalisée avec la collaboration d'un expert du domaine. Chaque module est constitué de manière indépendante c'est-à-dire que le système peut fonctionner sans avoir tous ses agents opérationnels au départ. Ainsi, la société peut évoluer en intégrant de nouveaux agents et si un agent venait à tomber en panne, le système continuerait en mode dégradé. Les accointances représentent la connaissance d'autres agents du domaine. Ce terme a été emprunté à la terminologie des langages d'acteurs (Agha, 85). Ce type de connaissance permet de mettre en oeuvre les interactions entre les agents. En effet, pour tous les agents,

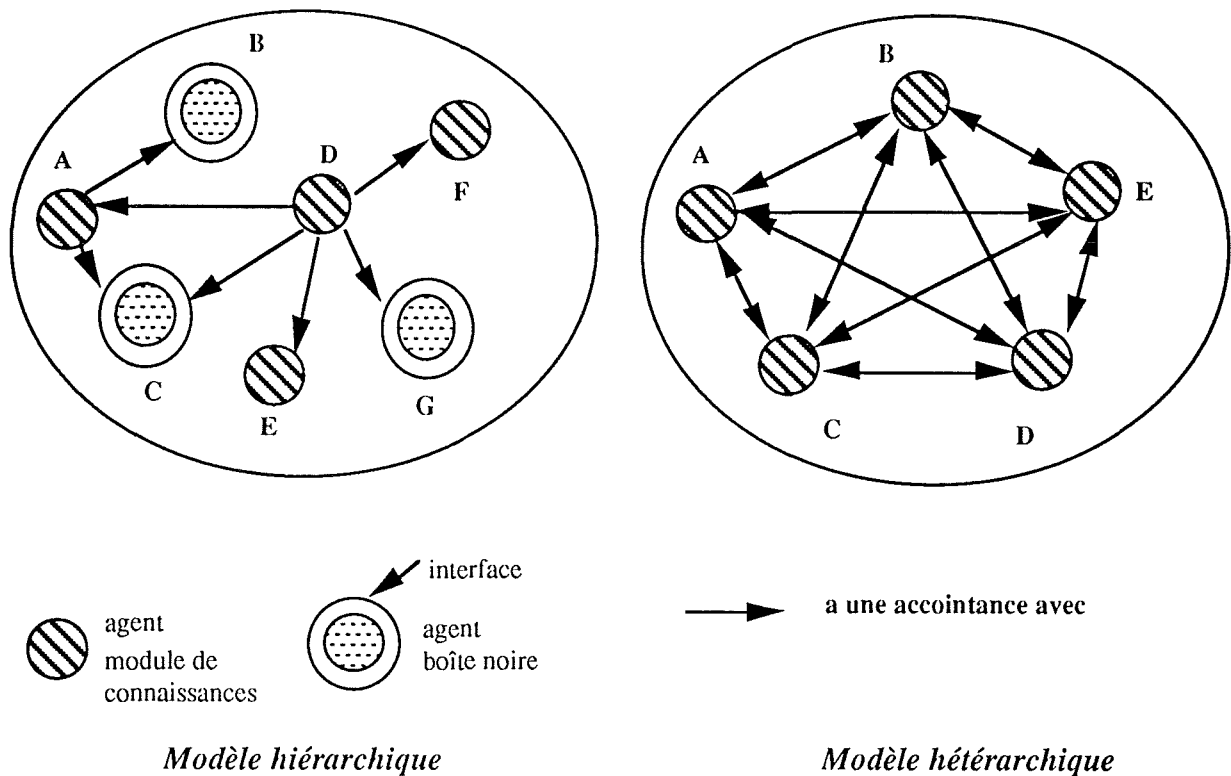
elles décrivent sous la forme de connaissance, les relations statiques qui existent entre un agent et les autres agents de la société. Ainsi les agents sont les noeuds de ce réseau de connaissances.

Comme Weisbuch le souligne (Weisbuch,83), une organisation sociale est comprise en tant qu'organisation fonctionnelle: un ensemble d'éléments multiples est considéré comme organisé s'il est capable d'un comportement collectif distinct de celui d'un magma constitué des mêmes éléments.

2.2. Aspects statiques

En paraphrasant Toynbee, on peut dire qu'une société d'agents est constituée par le réseau complet de leurs relations. Cette topologie est définie avec SYNERGIC par les accointances des agents. Cette approche cognitive de l'organisation sociale autorise la modélisation de sociétés hiérarchiques aussi bien qu'hétérarchiques:

- dans le domaine médical (Télémac), le médecin généraliste joue le rôle de chef d'orchestre relativement aux spécialistes. C'est lui qui permet de coordonner les différentes interventions des spécialistes au cours du diagnostic d'une maladie chez un patient;
- dans le jeu des loups et de l'agneau, les loups ont pour objectif d'encercler l'agneau. Les loups ont des connaissances et des rôles similaires.



2.3. Aspects dynamiques

Le cycle de vie d'un agent correspond à l'enchaînement de trois actions:

- traiter les informations qu'il vient de recevoir. Le fait qu'il reçoive des informations le crée ou le réveille et il effectue le traitement pour lequel il a été conçu;
- communiquer à d'autres agents des informations. C'est grâce à ses connaissances d'accointances qu'il réalise cette étape;
- se mettre en attente; il est inactif, son activation sera due à une réception d'informations.

L'ensemble des agents d'une application se comporte comme s'ils interagissaient par partage de tâches et de résultats, bien que rien ne soit spécifiquement programmé dans ce

sens. Ce résultat est dû aux accointances qui permettent la communication de faits ou de problèmes:

- un agent qui a déduit un résultat le communique à tous les agents qu'il sait potentiellement intéressés par cette information; même si elle ne lui a pas été explicitement demandée;
- un agent qui a engendré une tâche la communique aux agents susceptibles de la résoudre.

Par exemple dans Télémac, le généraliste possède des accointances qui lui permettent d'invoquer des agents pour résoudre une tâche.

si état du patient de l'agent Traumatismes-Thoraciques = blessé
et siège du problème de l'agent Traumatismes-Thoraciques = 2
et fonction-neurologique du patient de l'agent Traumatismes-Thoraciques
alors diagnostic du thorax de l'agent Traumatismes-Thoraciques

Cette règle d'accointance lui permet de savoir que le module Traumatismes-Thoracique est susceptible de trouver le diagnostic du thorax. Pour cela, le généraliste lui communique les informations suivantes: l'état du patient, le siège du problème et sa fonction neurologique et la tâche à résoudre: le diagnostic du thorax. Cette règle permet de créer le message au module des Traumatismes-Thoraciques.

Le généraliste possède aussi des accointances pour communiquer une information à d'autres agents.

age du patient de l'agent Traumatismes-Thoraciques
age du patient de l'agent Douleurs-Abdominales

...

L'âge du patient, s'il est connu par le généraliste, peut ainsi être communiqué au module des Traumatismes-Thoraciques et des Douleurs-Abdominales.

Le comportement de cette société d'agents repose entièrement sur les communications entre ses composantes. Le contenu de la communication, c'est-à-dire les informations à communiquer et à qui les communiquer, est défini par les accointances. La communication est déclenchée lorsqu'un agent a fini de raisonner localement, c'est-à-dire quand ses connaissances ont atteint un état stable. La propagation systématique d'informations communes garantit ainsi la cohérence de la solution globale fournie par le système.

Au cours du fonctionnement du système, seuls certains chemins dans le graphe des relations statiques entre les agents seront empruntés. Ce chemin correspond à l'ordre d'intervention des agents qui, en fait, est fonction de la connaissance d'accointances inférée. Ces inférences dépendent elles-mêmes de la résolution c'est-à-dire des connaissances du domaine inférées.

Un observateur d'une application peut juger que l'ordre non préétabli des interventions reproduit un réel processus interdisciplinaire alors que seules les accointances distribuées guident les interactions.

3. COMPORTEMENTS D'UN AGENT

L'existence d'un agent ainsi que son autonomie cognitive sont garanties par deux éléments:

- sa mémoire privée qui contient le savoir expert et les données et buts d'un problème;
- le processus informatique permettant la manipulation de ces informations qui n'est pas tributaire de l'existence d'autres agents.

3.1. La connaissance distribuée d'un agent

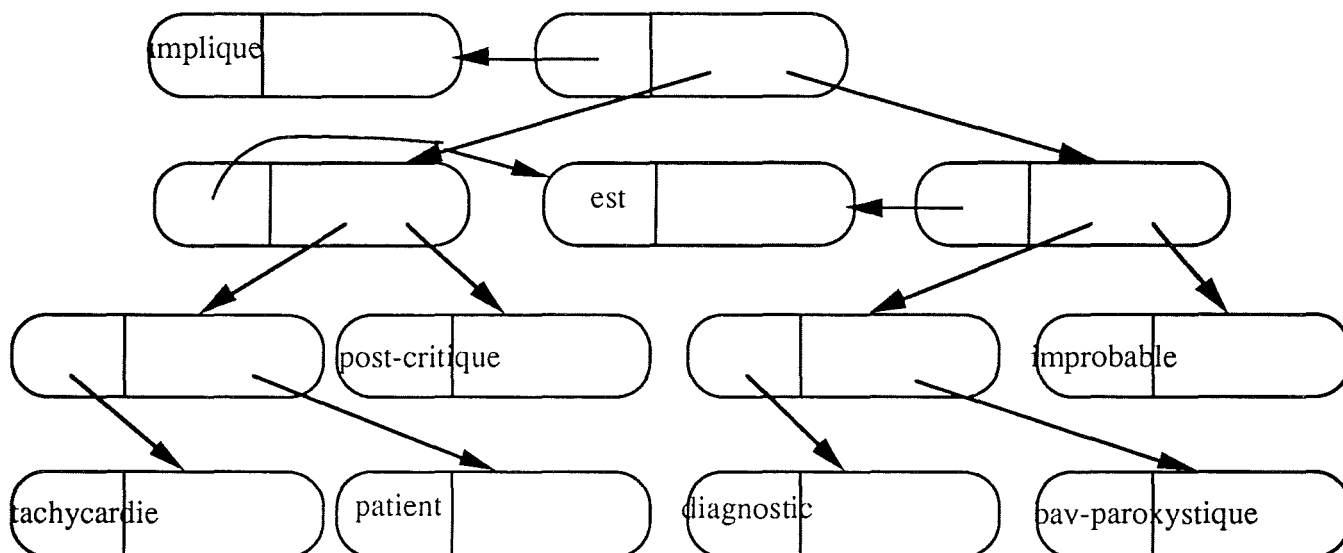
La représentation est à la fois locale et distribuée. Elle est locale parce que toute information, même complexe comme une règle de production, est représentée par l'état

d'une seule cellule. Mais elle est aussi distribuée parce que la sémantique d'une cellule fait souvent appel à plusieurs niveaux de décomposition en cellules plus élémentaires.

La règle de production suivante donne un exemple du principe de représentation distribuée employée dans SYNERGIC.

/* REGLE 157 */

Si la tachycardie du patient est post-critique
alors le diagnostic de bav-paroxystique est improbable



La partie gauche d'une cellule définit la relation entre ses fils, pointés par la partie droite. Par exemple, une cellule règle a pour relation *implique* et pour fils le *prémisse* et le *conséquent*). Chaque donnée est ainsi encapsulée par un traitement simple qui est directement associé au type de relation défini dans la cellule.

3.2. Les métaconnaissances

Dans toute application réaliste à base de connaissances, des métaconnaissances permettent l'élaboration de plans ou de stratégies de résolution. Des heuristiques sont aussi fréquemment employées pour obtenir une solution à la fois efficace et pertinente. Pour garantir la généralité de l'outil, elles ne doivent pas être programmées dans le moteur mais décrites dans des structures spécifiques.

Dans SYNERGIC, l'expert peut les intégrer dans des niveaux hiérarchiquement ordonnés (Rasmussen,85), ce qui assure une bonne lisibilité du savoir formalisé. Elles sont prises en compte dans ce système distribué par interactions entre sous-modules (appelés niveaux). Ces sous-modules sont aussi des cellules avec un traitement spécifique. Ils communiquent entre eux des problèmes ou résultats provenant de connaissances (cellules de plus faible granularité) qui leur sont associées. Le nombre de niveaux dépend uniquement de l'application:

- les plus bas intègrent des informations détaillées, utiles à l'exécution de l'action,
- les plus hauts engendrent et ordonnent des sous-problèmes pour les niveaux inférieurs (Pitrat,90). C'est au plus haut niveau (généralement 3 ou 4) que débute la résolution pour analyser le contexte du problème.

3.3. Aspects statiques

Un module de connaissances peut contenir plusieurs milliers de cellules. Chaque élément d'information (allant jusqu'aux chaînes de caractères) est encapsulé par une procédure. Chaque procédure est un réflexe prédéfini par le type de cellule. Il existe une trentaine de relations prédéfinies (dans lesquelles nous retrouvons les prédicats de la logique) et une relation par défaut.

La règle d'activation de ce modèle est une procédure locale que chaque cellule suit en mettant à jour son état interne en fonction du contexte d'activation des cellules voisines.

Selon les modifications survenues aux cellules voisines, une cellule est "excitée" ce qui peut entraîner la modification de son état interne (réalisé par le réflexe associé); propagé à son tour vers ses cellules voisines. Le dialogue à ce stade ne consiste qu'en la propagation de l'état et est déclenché à la fin de la réalisation de la tâche qui lui est assignée.

Parallélisme et mémoire active sont donc les deux notions importantes pour le traitement d'un module de connaissances. C'est de l'assemblage de ces micro-comportements que peut émerger un comportement global cohérent

3.4. Aspects dynamiques

Le réseau de cellules est totalement asynchrone. Ce qui implique que le délai entre chaque traitement est variable. Il est même fréquent que certaines cellules ne soient pas activées durant une session. Le rôle du "moteur d'inférences" est de gérer l'ensemble des messages entre cellules. Une liste, constituée au départ d'une session, est alimentée par l'émission de messages et diminuée lorsque une cellule prend un message qui lui est destiné. Comme ces cellules ne sont pas techniquement des processus informatiques, la prise en compte des messages est séquentielle.

L'intérêt de cette approche est que la parallélisation du traitement est implicite: chaque cellule pourrait être un processus, c'est un des axes de développement actuels. En effet, pour optimiser le traitement sur une machine multiprocesseur, nous étudions maintenant une technique de compilation des bases de connaissances qui puisse associer à chaque cellule un processus système. Ce travail en cours de développement sous Unix, doit associer le traitement d'une cellule à un "thread". Cela supprimera du même coup le "moteur d'inférences" dont le seul rôle est de gérer la messagerie inter cellules.

Ce modèle de réseau de cellules possède la propriété d'auto-organisation (Bourret,91) car aucun programme externe ne dicte explicitement la conduite du réseau sous-jacent. L'excitation provoquée par une cellule voisine peut entraîner la modification de son état interne; cette modification se propage à son tour aux cellules voisines. Lorsque le réseau est stabilisé (tous les messages sont traités), les informations sont cohérentes.

3.5. Le cycle de production émergent

Il n'y a donc véritablement pas de moteur d'inférences au sens classique du terme, mais seulement des interactions entre cellules. Pourtant le comportement d'un module de connaissances est similaire à celui que l'on aurait avec un moteur non monotone inférant en largeur simultanément en chaînage avant et arrière. Rappelons que le cycle de production classique se décompose en cinq phases: restriction, filtrage, résolution des conflits, inférence, vérification de la condition d'arrêt.

La restriction dans SYNERGIC est réalisée par l'organisation en niveaux. La cellule associée au niveau supérieur communique un sous-but au niveau inférieur. Il est ensuite propagé à l'ensemble des connaissances du niveau qui peuvent le déduire.

Parmi l'ensemble des règles contenant des faits ou buts, le filtrage correspond à l'activation des cellules de type implique. Cette activation est implicite par la propagation depuis les cellules faits ou des buts. Bien entendu les cellules règles sont filtrées de façon totalement asynchrone.

La résolution de conflits est réduite à sa plus simple expression, car toute cellule règle activable est inférée.

L'inférence est la demande, par une cellule implique, de l'activation de son premier fils (raisonnement par le but) ou de son second fils (raisonnement par la donnée).

L'arrêt est obtenu lorsqu'il n'y a plus de messages à traiter dans un niveau.

4. CONCLUSIONS ET PERSPECTIVES

L'approche présentée se situe entre les techniques traditionnelles d'intelligence artificielle et les réseaux neuronaux (Barthès,91), (Low,91), (Touretzky,88). C'est un outil de développement d'applications d'intelligence artificielle distribuée car la présentation des connaissances est symbolique et les comportements observés modélisent des sociétés

d'agents coopérants. C'est aussi un outil proche des techniques des réseaux neuronaux par la représentation interne totalement distribuée et les traitements locaux associés. Les notions d'agents réactifs et cognitifs habituellement évoquées ne sont pas significatives dans notre approche dans la mesure où il y a un passage continu de l'un à l'autre: un agent cognitif est l'agrégation de cellules réactives. Cette approche permet un parallélisme massif et explicite du calcul à travers le réseau, associé au caractère local de l'information traitée par chaque noeud.

Les comportements dynamiques cohérents d'un agent ou d'une société d'agents sont supposés vus par un observateur extérieur au système. Ces émergences observées dans des artefacts peuvent être considérées comme une apparition inattendue d'un phénomène qui n'avait pas semblé inhérent aux différentes parties d'un système. Ces phénomènes collectifs montrent qu'un tout peut être supérieur à la somme de ses parties. Néanmoins, comme le souligne Minsky (Minsky,88), une recherche plus poussée fait cependant généralement apparaître que les phénomènes émergents peuvent tout à fait être expliqués dès lors que l'on prend également en compte les interactions entre ces parties, ainsi que les particularités et lacunes des perceptions et des attentes de l'observateur. C'est ce qui justifie dans SYNERGIC la possibilité de distribuer un savoir global dans des cellules et agents interdépendants.

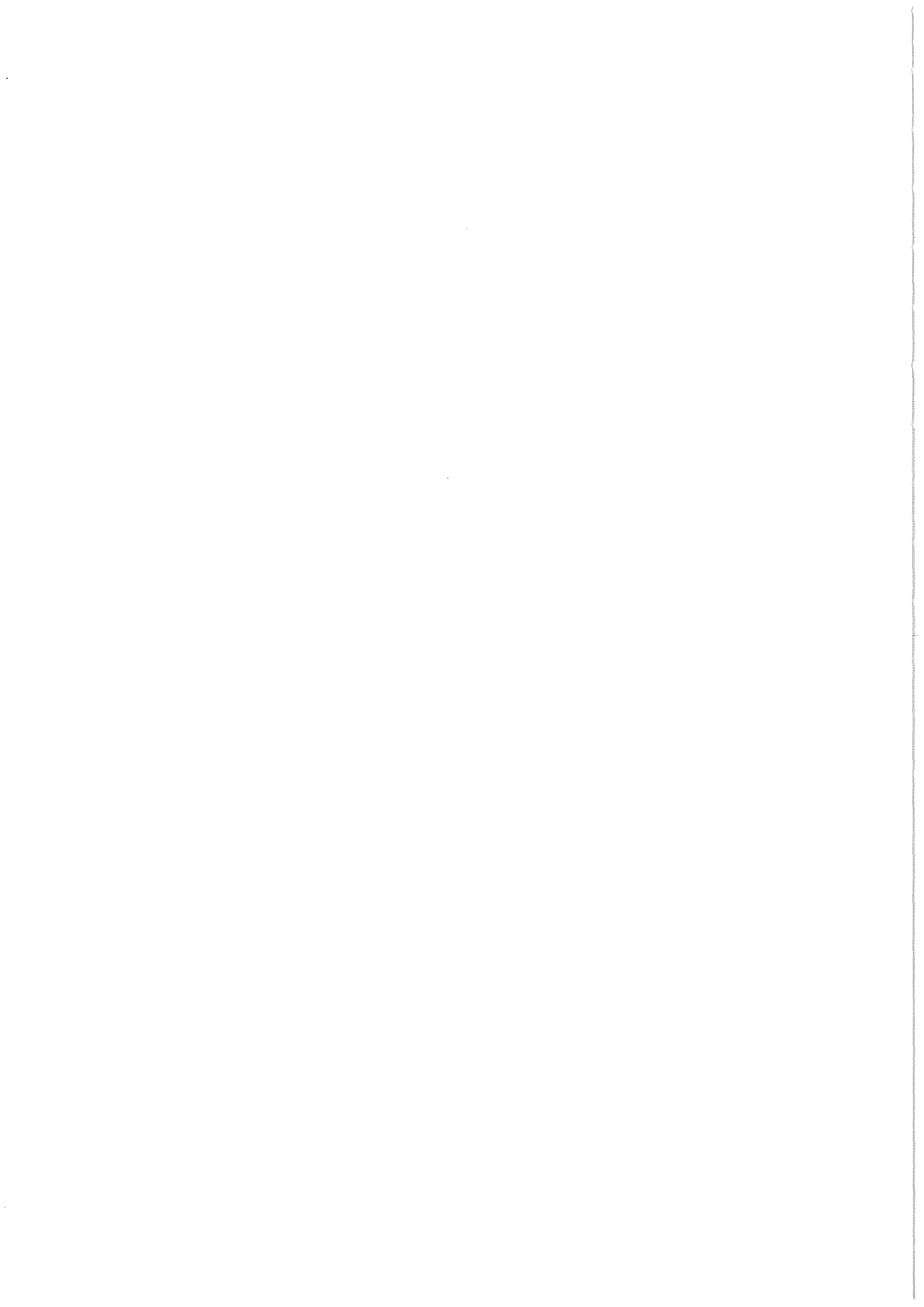
Certaines lacunes devraient néanmoins être comblées, notamment la possibilité de perception sociale d'un agent: actuellement, seule existe la connaissance d'autres agents comme individus à travers la description d'acointances. Il n'est donc pas possible de raisonner sur l'organisation pour aboutir par exemple à une restructuration en vue d'accroître l'efficacité de la collectivité. Cette limitation n'est pas conceptuelle car toute interaction y est cognitive, le concept de société et d'appartenance d'un agent à cette société peut donc être intégré.

Il faudrait aussi ajouter un mécanisme de relaxation avec lequel un agent pourrait diffuser une demande à travers les agents qu'il connaît actuellement. Cette notion de relaxation est également analysée pour modifier les interactions entre les cellules d'un agent conduisant à un apprentissage. La difficulté est que cet apprentissage est similaire à celui d'un réseau neuronal et n'est pas symbolique dans la mesure où il doit être distribué.

5. BIBLIOGRAPHIE

- Agha G., Hewitt C.
Concurrent Programming Using Actors: Exploiting Large-Scale Parallelism
in Readings in Distributed Artificial Intelligence - Bond A.H. and Gasser L Eds
Morgan Kaufmann Publishers , 1985
- Atlan H.
Entre le cristal et la fumée
Editions du Seuil, 1979
- Barthès C., Chabaud C., Gleizes M.P., Glize P.
Ergonomics inference: study and example
11^{ème} Congress of International Ergonomy Association - Paris, 1991
- Barthès C., Cohen P., Glize P., Machonin A.
Neural computation in knowledge based systems
International conference on Digital Signal Processing - Florence, 1991
- Bourret P., Reggia J., Samuelides M.
Réseaux neuronaux - Une approche connexionniste de l'I.A.
Editions Teknea, 1991
- Cannizzo-Puyet P., Carpuat B., Gleizes M.P., Glize P.
TELEMAC: système multi-agent en télé-médecine
Convention Intelligence Artificielle - Paris, 1991

- Smith R.G., Davis R.
 Frameworks for cooperation in Distributed Problem Solving
 IEEE Transactions on systems, man and cybernetics - Vol11 N°1, 1981
- Dumouchel P., Dupuy J.P.
 Colloque de Cerisy - L'auto-organisation de la physique au politique
 Edition du Seuil - Avril 1983
- Dupuy J.P.
 Ordres et désordres - Enquête sur un nouveau paradigme
 Edition du Seuil, 1982
- Fodor J.A.
 La modularité de l'esprit
 Les éditions de minuit, 1986
- Gleizes M.P., Glize P., trouilhet S.
 Distributed problem-solving by autonomous cooperative cognitive agents
 Expersys 92 - Editions IITT, 1992
- Grice H.P.
 Logic and conversation
 Syntax and semantics, Vol3
 P.Cole et J.L.Morgan Eds - N.Y. Academic Press, 1975
- Low B.T., Lui H.C., Tan A.H., TEH H.H.
 Connectionist expert system with learning capability
 IEEE Transactions on knowledge and data engineering - Vol3 N°2, 1991
- Minsky M.
 La société de l'esprit
 Interéditions - Paris, 1988
- Pitrat J.
 Métaconnaissance. Le futur de l'intelligence artificielle
 Editions Hermès, Paris, 1990
- Politzer G.
 L'informativité des énoncés: contraintes sur le jugement et le raisonnement
 Intellectica, Vol1 N°11 - 1991
- Rasmussen J.
 The role of hierarchical knowledge in decision-making and system management
 IEEE Trans. on SMC - Vol15 N°2 - Avril 1985
- Touretzky D.S., Hinton G.E.
 A distributed connectionist production system
 Cognitive science - Vol12, 1988
- Varela F.J.
 Autonomie et connaissance - Essai sur le vivant
 Editions Seuil - 1988
- Von Foerster H., Zopf G. W.
 Principles of self-organization
 Pergamon Press, 1962
- Weisbuch G.
 Quelques approches physiques de l'organisation
 In Colloque de Cerisy - L'auto-organisation de la physique au politique
 Dumouchel P., Dupuy J.P., Eds - Edition du Seuil - Avril 1983



Approcher la notion de collectif

Equipe Miriad*

*Laforia,
4, Place Jussieu, Boîte 169
France, PARIS 75252 Cedex 05
miriad@laforia.ibp.fr*

1 Introduction

Cet article traite du phénomène d'*intelligence collective*, c'est-à-dire de ce supplément d'intelligence constaté dans un *collectif* d'agents. L'étude de ce phénomène, abordé dans d'autres travaux [Agre & Gasser 92] sous les noms d'intelligence en essaim [Deneubourg 89], de fonctionnalité émergente [Steels 91] [Conte & Castelfranchi 92] ou de coopération multi-agents [Durfee & Lesser 87], pose deux questions fondamentales :

1. Quel est le niveau de description adéquat qui permet d'exprimer et de caractériser ce supplément d'intelligence ?
2. Comment apparaît-il ?

A partir de là, deux démarches peuvent être employées : la démarche ascendante ou constructive est celle qui consiste à construire et agencer les éléments qui vont donner lieu à ce supplément d'intelligence ; la démarche descendante ou descriptive est celle qui consiste à fournir des modèles qui permettent de décrire et d'expliquer le supplément d'intelligence observé.

Nous proposons dans cet article¹ un ensemble de primitives pour définir un langage de spécification opératoire de systèmes multi-agents. Ces primitives jettent les bases d'une terminologie nouvelle en IA distribuée, permettant une étude aussi bien constructive que descriptive des systèmes multi-agents. En effet, ces primitives unifient les concepts liés aux contraintes de l'observateur (description intensive du groupe d'agents observé, démarche descendante) et ceux du concepteur (description extensive des agents formant le groupe qui sera observé, démarche ascendante). De plus, nous considérons que l'étude de l'intelligence collective doit se faire indépendamment de toute architecture d'agent ou de système multi-agents.

Il y a donc deux lectures possibles de cet article : en allant du début de l'article à sa fin, on adopte une démarche ascendante ; en remontant de la fin vers le début de l'article, on adopte une démarche descendante.

Dans le cas d'une lecture ascendante l'article se présente comme suit : nous allons de l'*agent* vers le *groupe*, puis, à partir de la notion essentielle d'*engagement*, nous définissons le

* L'équipe MIRIAD (Méthodologie des Interactions et Recherches en Intelligence Artificielle Distribuée) dirigée par Jacques Ferber, Anne Collinot et Patrice Carle, comprend près d'une quinzaine de chercheurs dont Nadia Abchiche, Thierry Bouron, Stéphane Bura, Christophe Cambier, Claude Delaye, Stéphane Desvignes, Alexis Drogoul, Christophe Dubreuil, Christophe Iffenecker, Eric Jacopin, Didier Mousseau, Denis Tabourdeau, Karim Zeghal. Cet article a été rédigé suite à plusieurs séances de travail sur le thème de la coopération et du groupe.

¹ Les propositions avancées dans cet article ne doivent pas être considérées comme ayant un caractère définitif. Elles reflètent l'état courant des réflexions de l'équipe. Nous considérons cependant que ce travail est suffisamment avancé pour être confronté aux réalisations et aux points de vue des autres chercheurs de la communauté.

collectif. Le comportement de ce collectif et son évolution sont ce que nous désignons respectivement par *coopération* et par *intelligence collective*.

2 Primitive d'Agents

Définition 1 (Agent) [Ferber 89] On appelle *agent* une entité réelle ou abstraite capable d'agir sur elle-même et sur son environnement.

L'environnement peut inclure d'autres agents. On verra plus loin (section 3) des situations d'interactions entre l'agent et son environnement, notamment en cas de formation de groupe.

Le comportement d'un agent définit ses façons d'agir sur lui-même ou son environnement. Il dépend d'un certain nombre de facteurs, dont sa représentation de cet environnement et la perception limitée qu'il en a, ses buts, et ses capacités. Qu'il s'agisse de concevoir ou d'observer des agents, il nous apparaît nécessaire de pouvoir regrouper sous un terme générique l'ensemble de ces facteurs. Pour cela nous introduisons la notion de *déterminant*.

Définition 2 (Déterminant d'un agent) On appelle *déterminant* d'un agent l'ensemble² minimal de ses caractéristiques *structurelles*, *environnementales* et *comportementales* ; c'est l'ensemble nécessaire et suffisant de ses caractéristiques qui permet de prescrire ou d'expliquer complètement ses façons d'agir. On notera D_A le déterminant d'un agent A.

Les caractéristiques environnementales d'un agent caractérisent les états des représentations internes et externes de l'agent. Elles sont directement liées à la représentation que se fait l'agent de son environnement et de lui-même. Les caractéristiques structurelles d'un agent déterminent l'ensemble de ses comportements possibles, alors que les caractéristiques comportementales contraignent l'ensemble de ces comportements, en accord avec les caractéristiques environnementales. Il faut cependant noter que la séparation entre partie structurelle et partie comportementale, souvent très nette pour les systèmes artificiels, ne l'est pas toujours pour le vivant. C'est le cas pour la fourmi : le système nerveux fait partie de sa structure et contrôle l'ensemble de ses comportements, alors que, pour un avion, l'équivalent du système nerveux est le couple (pilote, plan de vol).

La table 1 propose des caractéristiques structurelles, comportementales et environnementales pour différents types d'agents.

Agent	Caractéristiques structurelles	Caractéristiques comportementales	Caractéristiques environnementales
BB1	Tableau noir	Plan de contrôle pour un ensemble de KS donnés	Événements
Automate	Alphabet	Fonction de transition pour l'alphabet de l'automate	Etats
Avion	Capacité, vitesse, etc	Pilote et plan de vol	Position
Fourmi	Code génétique	Dépendent du contenu génétique et de l'environnement	Phéromones
Plante	Code génétique	Tropismes	-

Table 1.

² Dans la suite, le terme d'ensemble ne fait pas référence à une définition mathématique stricte et ne préconise aucune structure particulière.

L'ensemble des caractéristiques qui composent le déterminant de l'agent peut évoluer au cours du temps. Cette évolution correspond à la modification des caractéristiques de l'agent, ou à l'ajout ou la suppression de caractéristiques. Un cas typique d'évolution est celui de la spécialisation progressive des fourmis au sein d'une colonie comme résultat de l'interaction des fourmis avec leur environnement [Drogoul et Ferber 92]. L'évolution d'un agent se traduit par la variation de son déterminant :

Définition 3 (Différentiel d'un déterminant d'agent) On appelle *différentiel* du déterminant d'un agent l'évolution de ce déterminant au cours du temps. On note Δ_{DA} le différentiel du déterminant de l'agent A.

Par exemple, si l'on reprend le cas des colonies de bourdons étudié par Hogeweg et Hesper [Hogeweg et Hesper 86], les bourdons ouvriers vont à un moment donné de l'année, suite à l'exclusion de la reine, se mettre à pondre des œufs, alors que leur comportement par défaut exclut la ponte. Cette évolution du comportement résulte d'une évolution des caractéristiques environnementales (dans ce cas, la disparition de la reine lève une inhibition chez les ouvriers). Un autre exemple est celui de la modification d'un agent par lui-même. Si on considère un agent de type blackboard qui ramasse des cubes, la KS (l'opérateur) qui implémente la prise d'un cube, peut être spécialisée en modifiant sa précondition de façon à ce que l'agent ne ramasse que des cubes de la même couleur du premier cube qu'il a ramassé.

La possibilité de l'évolution du déterminant de l'agent amène naturellement à se poser la question de la datation de cette évolution. Ainsi, on parle du déterminant initial de l'agent selon le point de vue adopté. Si l'on est observateur de l'évolution du déterminant alors l'instant initial sera celui où l'observation commence. Si l'on est le concepteur du système, alors le déterminant initial est caractérisé par l'instant de démarrage de l'exécution du système. Mais l'évolution du déterminant peut être telle que le comportement devient inexplicable par suite de changements trop radicaux du déterminant. Il faut donc tenir compte de l'instant initial, quel que soit le point de vue adopté : le déterminant de l'agent à l'instant initial constitue ce que nous appelons l'*inné* de l'agent. La différence entre deux déterminants associés à deux instants distincts représente l'*acquis* de l'agent entre ces deux instants ; cet acquis provient essentiellement de l'interaction entre l'agent et son environnement. Par exemple, la spécialisation du ramassage de cube en ramassage de cube d'une couleur donnée constitue un acquis pour notre agent ramasseur de cubes : son déterminant après la prise du premier cube est différent de son déterminant de conception. Il faut cependant noter que l'acquis tel que nous le définissons n'est pas forcément définitif. Il est toujours possible de régresser !

3 Primitives de Groupes

En considérant les déterminants d'un ensemble d'agents pris au hasard, il est peu probable qu'ils aient des points communs. Par exemple, les cinq types d'agents de la Table 1 n'en ont sûrement pas. Mais si l'on ne considère qu'un seul « type » d'agent, par exemple tous les avions, tous les automates ou encore toutes les fourmis, alors ces agents possèdent des caractéristiques communes. Les avions ont tous en commun ce qui leur permet de voler, les automates n'existent pas sans leur fonction de transition et les fourmis sans leur code génétique. Mais si la structure d'un ensemble d'agents permet de les reconnaître comme (structurellement) semblables, leur comportement permet également de les regrouper : les avions atterrissant à Orly sont des avions particuliers, car leur plan de vol se termine à Orly. La division du travail dans une colonie de fourmis donne un classement comportemental des fourmis (soigneuses, gardiennes, fourrageuses, etc...). Ainsi, ce sont non seulement ses caractéristiques structurelles, mais également l'étude de son plan de vol, de sa position, qui nous permettent de reconnaître un avion et de mieux appréhender son comportement en vol. Les caractéristiques structurelles, comportementales ou environnementales faisant partie du déterminant de l'agent, c'est donc en fonction d'un ensemble d'éléments de son déterminant que l'on pourra classer un agent. Un tel ensemble est appelé un ensemble d'indices ou de caractéristiques indicielles car un

élément de cet ensemble nous donne un indice sur le *groupe* auquel appartient l'agent. En ne privilégiant pas a priori l'étude (conceptuelle ou observatoire) d'un groupe particulier, n'importe quel élément d'un déterminant d'agent peut induire l'appartenance de cet agent à un groupe :

Définition 4 (Indice) Tout élément du déterminant d'un agent, que cet élément soit structurel, comportemental ou environnemental est susceptible d'être un indice.

Le fait qu'un élément soit un indice n'a de sens que pour un type d'interaction donné. La formation d'un groupe se fait sur la base des indices associée à un type d'interaction. Par exemple, dans le cas du trafic aérien, l'anticollision est un des types d'interaction auquel on s'intéresse en priorité. Les capacités d'évitement d'un avion constituent les indices pour identifier les avions qui peuvent respecter les consignes d'anticollision. Dans le cas d'un ensemble de robots devant participer à la collecte de minerai, le fait que leur structure ou leur comportement se complètent n'a de sens que par rapport à des caractéristiques environnementales communes, qui sont les indices du groupe, à savoir les puits de forage et la présence de minerai. Les robots foreurs produisent des puits de forage qui peuvent être perçus par d'autres robots foreurs. Ces mêmes robots foreurs remontent le minerai à la surface, et le rendent ainsi accessible aux robots transporteurs qui vont alors le rapporter vers l'endroit convenu.

Définition 5 (Groupe) Un groupe est un ensemble d'agent dont les déterminants ont une intersection non nulle. Cette intersection constitue l'ensemble des indices qui caractérisent le groupe.

Par construction, un agent appartient à un groupe si et seulement si son déterminant possède toutes les caractéristiques indiciaires du groupe. Un agent peut donc quitter un groupe ou y adhérer, parce que certaines de ses caractéristiques sont devenues celles du groupe ou parce qu'il a perdu au moins une des caractéristiques indiciaires du groupe. Par exemple, un robot qui perd ses moyens de locomotion ne fera plus partie du groupe des robots transporteurs. De façon similaire, un robot qui perd toute possibilité de percevoir le minerai n'appartient plus au groupe de robots chargés de collecter le minerai.

Définition 6 (Déterminant de groupe) L'ensemble d'indices communs aux agents du groupe est appelé le déterminant du groupe. Le déterminant d'un groupe G , D_G , doit être vu comme le dénominateur commun aux déterminants D_A des agents A appartenant à G .

4 Primitives de Collectifs

Dans la section précédente, nous avons donné une condition nécessaire à l'interaction d'un ensemble d'agents. Cette condition - former un groupe - est statique dans le sens où elle ne porte que sur les potentialités d'interactions rendues possibles par la présence de caractéristiques communes. Dans cette section, nous cherchons à obtenir une condition suffisante à la mise en œuvre effective d'une interaction, c'est-à-dire à garantir le fait qu'il y ait ou non interaction entre les agents d'un groupe. Pour cela, nous devons tenir compte de la dynamique du groupe, c'est à dire du comportement *actif* de chacun des agents. En fait, il s'agit de considérer les activations / désactivations des caractéristiques comportementales, souvent liées à l'activation / désactivation (présence / absence) de caractéristiques environnementales et structurelles.

Il faut considérer les indices d'un agent comme des caractéristiques potentielles de son comportement. En effet, il est possible que le comportement actif d'un agent (i.e. le comportement effectivement mis en œuvre à l'instant considéré) ne soit pas le produit de l'ensemble de ses indices. Autrement dit, à un instant donné (et dans un contexte donné), il se peut que certains indices ne soient pas *activés*. Par exemple, bien que l'indice comportemental d'évitement soit prioritaire pour un avion, seul l'indice environnemental dénotant la distance minimale qui le

sépare des avions qu'il perçoit, active le comportement (i.e. fait effet de déclencheur du comportement) indicé par l'indice d'évitement. Un avion survolant Orly ne donne pas lieu à l'activation du comportement d'évitement d'un avion survolant Londres, soit parce que ce dernier ne le perçoit pas, soit parce que la distance les séparant est supérieure à la distance minimale.

On associe donc à un indice i d'un agent, une valeur $v_t(i)$ qui indique le statut d'activité de l'indice i dans le comportement de l'agent tel qu'il apparaît à un instant t donné :

$$\begin{aligned} v_t : I &\rightarrow \{0,1\} \\ i &\rightarrow v_t(i) \end{aligned}$$

où $v_t(i) = 1$ si i joue un rôle actif dans le comportement de l'agent, et $v_t(i) = 0$ si i n'intervient pas dans le comportement de l'agent. L'avion survolant Londres peut se diriger vers Orly ; lors de son arrivée, il influencera le comportement d'évitement des avions atterrissant à Orly. La valeur d'activité $v_t(i)$ évolue donc dynamiquement et traduit la dynamique du groupe.

Définition 5 (Indice actif) Un indice i est un indice actif à un instant t si $v_t(i) = 1$. Pour les indices structurels et environnementaux, c'est la présence ou l'absence de ces indices qui déterminent leur valeur d'activité.

La notion d'indice actif est pertinente aussi bien du point de vue du concepteur que de celui de l'observateur. Le concepteur, *connaissant* le détail du fonctionnement de l'agent, peut affirmer que le comportement courant *est exactement* celui qui est attendu étant donnés les indices. Un observateur cherchant à *modéliser* le fonctionnement de l'ensemble des agents, peut constater que le comportement courant *pourrait se justifier par* l'activation des indices.

Définition 6 (Indice actif de groupe) Un indice d'un groupe est dit actif si et seulement si cet indice est actif pour tous les agents de ce groupe.

L'ensemble des indices actifs constitue la partie active du déterminant que l'on appelle déterminant actif par abus de langage. On note D^*_A le déterminant actif d'un agent A et D^*_G le déterminant actif d'un groupe G . Le déterminant actif d'un groupe est l'intersection des déterminants actifs des agents de ce groupe.

On cherche maintenant à situer un agent A , à un instant donné, par rapport aux possibilités d'interactions indicées par son appartenance à un groupe G , c'est-à-dire à caractériser son degré d'*engagement* au sein du groupe. Il semble donc naturel d'examiner le déterminant actif D^*_A de l'agent A considéré. Si D^*_A contient les indices de D_G qui rendent possible un type d'interaction, alors on dit que l'agent A est *engagé* vis-à-vis de ce type d'interaction. Si D_G est inclus dans D^*_A , alors on dit que A est *engagé* vis-à-vis de tous les types d'interactions indicés par le groupe G .

Définition 7 (Engagement) On définit la fonction d'engagement \mathcal{E} pour tout agent A appartenant à un groupe G par :

$$\begin{aligned} \mathcal{E} : D^* \times D &\rightarrow D \\ (D^*_A, D_G) &\rightarrow \mathcal{E}(D^*_A, D_G) \end{aligned}$$

telle que :

$$\mathcal{E}(D^*_A, D_G) = D^*_A \cap D_G$$

En effet, on n'exclut pas a priori la possibilité pour un agent de posséder des indices actifs qui ne font pas partie du déterminant du groupe.

\mathcal{E} détermine l'engagement de A par rapport au groupe G . $\text{Card} [\mathcal{E}(D^*_A, D_G)]$ est le degré d'engagement de l'agent A au sein du groupe G . On dit que A est engagé dès que $\text{Card} [\mathcal{E}(D^*_A, D_G)] \geq 1$.

Etant donnée une fonction d'engagement \mathcal{E} , on cherche à identifier les agents d'un groupe impliqués dans le (ou les) engagement(s) dénoté(s) par \mathcal{E} . Pour cela nous introduisons la notion de relation d'engagement :

Définition 8 (Relation d'engagement) Etant donnée une fonction d'engagement \mathcal{E} , une relation d'engagement $R_{\mathcal{E}}$ est définie entre les agents d'un groupe G par :

Soient A et B deux agents appartenant au groupe G ,

$$A R_{\mathcal{E}} B \Leftrightarrow \mathcal{E}(D^*_A, D_G) \cap \mathcal{E}(D^*_B, D_G) \neq \emptyset$$

$\mathcal{E}(D^*_A, D_G) \cap \mathcal{E}(D^*_B, D_G)$ est appelé le déterminant de la relation $R_{\mathcal{E}}$.

Le déterminant de $R_{\mathcal{E}}$ s'exprime plus simplement par $D^*_A \cap D^*_B \cap D_G$.

Définition 9 (Collectif) Etant donnée une fonction d'engagement \mathcal{E} définie sur un groupe d'agents G , toute relation d'engagement $R_{\mathcal{E}}$ définit un collectif d'agents dans le groupe G . Le déterminant de $R_{\mathcal{E}}$ est le déterminant du collectif.

Un agent A appartient à un collectif d'agents du groupe G si et seulement si $\mathcal{E}(D^*_A, D_G)$ contient le déterminant du collectif. On peut noter qu'un agent peut appartenir simultanément à plusieurs collectifs.

Considérons maintenant l'engagement d'un agent au sein d'un collectif. Lorsqu'un agent ne possède pas à un instant donné les indices actifs nécessaires à son engagement dans un collectif, il peut les acquérir et donc faire évoluer son déterminant actif. Cette évolution du déterminant peut-être imposée par l'environnement ou délibérément souhaitée par l'agent lui-même. Cette évolution, qu'elle soit imposée ou délibérée, définit une forme d'intentionnalité, dans le sens où la modification du déterminant se fait *pour* un engagement donné.

Définition 10 (Intentionnalité) Un collectif est intentionnel si et seulement si la formation de ce collectif résulte d'un différentiel non nul des déterminants actifs de tous les agents engagés dans ce collectif.

Prenons par exemple le cas du tri collectif dans lequel un ensemble de robots doivent trier des cubes de trois couleurs différentes. L'intentionnalité dans cet exemple peut être caractérisée par une focalisation des robots sur les cubes de la couleur du premier cube qu'ils ont transporté, sur la décision du concepteur de leur attribuer une couleur particulière ou sur une entente préalable entre les robots. Le comportement de focalisation sur une couleur est un indice comportemental commun à ces robots, qui leur permet à chacun d'adhérer à un collectif particulier (un collectif par couleur) au sein du groupe des robots qui trient. La formation de ces trois collectifs résulte d'une évolution des comportements des robots, évolution permise par l'appartenance des robots au collectif indicé par l'indice comportemental de focalisation.

Définition 11 (Non Intentionnalité) Un collectif est non intentionnel si et seulement si tous les déterminants actifs des agents engagés dans ce collectif ont un différentiel nul.

Par définition, l'engagement sous-jacent à la formation d'un collectif intentionnel est dit intentionnel. Inversement, l'engagement sous-jacent à la formation d'un collectif non intentionnel est dit non intentionnel. La table 2 considère un collectif de deux agents A et B et définit les 3 types d'engagement possibles étant donnée l'évolution de leurs déterminants actifs respectifs.

Collectif {A, B} $D^*_A \cap D^*_B \neq \emptyset$	$\Delta_{D^*_A} = 0, \Delta_{D^*_B} = 0$	$\Delta_{D^*_A} \neq 0, \Delta_{D^*_B} \neq 0$	$\Delta_{D^*_A} = 0, \Delta_{D^*_B} \neq 0$ ou $\Delta_{D^*_A} \neq 0, \Delta_{D^*_B} = 0$
Type d'engagement	<i>Engagement non intentionnel</i>	<i>Engagement intentionnel</i>	<i>Engagement "hybride"</i>

Table 2.

5 Primitives de Coopération et Intelligence Collective

Dans cette section, nous tentons d'approcher les notions de coopération et d'intelligence collective à la lumière des primitives d'agents, de groupes et de collectifs que nous avons introduites dans les sections précédentes.

Définition 12 (Coopération) La *coopération* d'un collectif d'agents est définie par le déterminant de ce collectif et par les déterminants des sous-collectifs inclus dans ce collectif.

Par définition on parle de coopération intentionnelle dans le cas d'un collectif intentionnel, et de coopération non intentionnelle dans le cas d'un collectif non intentionnel.

Par exemple, dans le cas du tri collectif, la coopération intentionnelle des agents qui trient les cubes est définie par l'indice de focalisation sur une couleur et par les indices qui déterminent chacun un sous-collectif spécialisé dans une couleur. Le déterminant de la coopération (comportement de focalisation, focalisation effective dans l'une des trois couleurs, par exemple bleu, vert et rouge) ne permet pas de reconstruire un collectif capable de trier, si ce n'est pour une situation identique (mêmes couleurs). Si on met les robots en face de cubes roses et violets, ils reconstruiront d'eux-mêmes d'autres collectifs et feront émerger une fonctionnalité de tri.

Définition 13 (Intelligence Collective) On appelle *intelligence collective* l'évolution de la coopération d'un collectif ; l'intelligence collective est donc le différentiel du déterminant du collectif considéré.

Nous faisons ainsi apparaître le supplément d'intelligence d'un collectif. En effet, il y a plus dans le collectif que dans chacun des agents du collectif pris séparément. Toujours dans l'exemple du tri collectif, les robots ne possèdent aucune fonctionnalité de tri, et pourtant si on se place du point de vue de l'observateur, on croit voir un ensemble de robots en train de trier. Ce supplément d'intelligence apparaît ici sous la forme de l'émergence d'une fonctionnalité non programmée, le tri. Cette définition est pertinente du point de vue de l'observateur. Celui-ci pourra donc décrire ce qu'il observe en ajoutant un indice supplémentaire (le tri) au déterminant du collectif. Cependant, du point de vue du concepteur, il n'y a pas de création de code rendant tangible ce supplément d'intelligence (un algorithme de tri) et permettant ainsi de le reproduire.

6 Conclusion

Nous avons tenté dans cet article d'approcher la notion d'intelligence collective à la lumière d'une nouvelle terminologie. L'originalité de notre démarche tient dans notre volonté de supprimer le clivage traditionnel entre les points de vue de l'observateur et du concepteur. Nous avons cherché à nous donner un outil pour décrire et construire nos systèmes. Cette approche nous semble pertinente dès lors que l'on s'intéresse à l'intelligence collective artificielle, puisqu'il s'agit alors de construire des systèmes qui pour un observateur (voire le concepteur en position d'observation) exhibent des comportements non programmés explicitement, qu'il leur faut décrire.

7 Références

[Agre & Gasser 92]

P. Agre and L. Gasser, *Knowledge and Action at Social and Organizational Levels*, AAAI'91 Fall Symposium Series Reports, AI Magazine, vol. 13 n° 1, Spring 1992.

[Conte & Castelfranchi 92]

R. Conte and C. Castelfranchi, *Mind is not enough: Precognitive bases of social interaction*, Proceedings of the Simulating Societies Symposium, University of Surrey, UK, 1992.

[Deneubourg 89]

J.L. Deneubourg and S. Goss, *Collective patterns and decision-making*, Ethology, Ecology and Evolution, vol. 1, pp. 295-311, 1989.

[Drogoul & Ferber 92]

A. Drogoul and J. Ferber, *Multi-Agent Simulation as a Tool for Modeling Societies: Application to Social Differentiation in Ant Colonies*, Proceedings of MAAMAW'92 (forthcoming "Decentralized AI IV"), 1992.

[Durfee & al. 87]

E.H. Durfee, V.R. Lesser and D.D. Corkill, *Coherent Cooperation among Communicating Problem Solvers*, IEEE Transactions in Computers, pp. 1275-1291, 1987.

[Ferber 89] J. Ferber, *Objets et agents : une étude des structures de représentation et de communications en Intelligence Artificielle*, Thèse de Doctorat d'Etat, Paris VI, 1989.

[Hogeweg et Hesper 86]

P. Hogeweg et B. Hesper, *Knowledge seeking in variable structure models*, in: Elras, Oven et Zergler (eds.) *Simulation in the artificial intelligence era*, North Holland, 1986.

[Steels 91]

L. Steels, *Towards a Theory of Emergent Functionality*, From Animals to Animats, MIT Press, p. 451, 1991.

Auto-organisation et intelligence collective.

Anne Nicolle-Adam
LAIAC Université de Caen 14032 Caen Cedex
tel : 31 45 56 69
mail : Anne.Nicolle@univ-caen.fr

1. Introduction

Malgré leurs faibles capacités de mémoire immédiate (5 à 7 chunks) et de calcul, les êtres humains ont de bonnes performances pour résoudre des problèmes très complexes comme la vie en société, la conception de machines, la recherche scientifique. La résolution des problèmes complexes est améliorée par le travail collectif, presque en parallèle d'un grand nombre de sujets et par l'échange sur ce travail. Je me propose de mettre des machines dans ce travail collectif et pas de leur demander d'être intelligentes toutes seules, comme on l'envisage habituellement en intelligence artificielle. Il est important que ces machines soient autonomes, qu'elles aient des capacités d'auto-organisation mais aussi qu'elles aient la capacité de mener des échanges langagiers avec d'autres sujets. Le transfert est un modèle de l'échange langagier compatible avec l'auto-organisation. Les ordinateurs ont des capacités de mémoire et de calcul moins faibles que les humains, les faire entrer dans un travail collectif pourrait permettre de construire interactivement de meilleures solutions à des problèmes très complexes pour lesquels l'optimum n'est pas à la portée des sujets humains et pour lesquelles on cherche donc des solutions adéquates [Simon 87]. Je cherche donc à concevoir un modèle de machine très simple qui puisse améliorer ses performances par sa propre expérience d'une part, avec l'observation des résultats de ses actions pour elle-même et de leurs effets sur le contexte, et par des échanges langagiers sur le modèle du transfert d'autre part, pour profiter des connaissances des autres sujets.

2. Le mode du transfert

La parole permet d'évoquer des choses absentes, des actions passées, d'imaginer des plans d'action. Pour désigner les choses, elle leur donne un nom qui est conventionnel ou utilise un groupe nominal, dont l'expression peut être extrêmement variée. Les expressions langagières provoquent chez celui qui écoute un transfert de perceptions et sensations, qui transforme les perceptions auditives de la parole entendue en perceptions et sensations à propos de ce qui est dit. Dans le mode du canal, proposé comme modèle de l'échange langagier de manière dominante en linguistique, le sujet qui parle impose sa parole de manière mécanique, comme un tuyau apporte de l'eau. Ce mode de communication me semble bien représenté dans le domaine des relations homme-machine par les travaux de Pitrat, qui impose des connaissances Maciste. Il ne peut pas les refuser même si elles le détruisent [Pitrat 90 p 200]. Il y a bien un filtrage syntaxique des données qui évite des erreurs, mais la parole de l'homme n'est jamais mise en doute sur le fond. Dans le mode du transfert, proposé par la psychanalyse comme modèle de l'échange langagier hors contraintes, le sujet entendant est actif et prend ce qui l'intéresse dans la parole de l'autre [Coursil 92]. L'auto-organisation du sujet entendant est prioritaire : il n'y a pas de sens intrinsèque à la parole, mais un sens pour chacun, à chaque moment. Donnons quelques exemples du décalage entre la parole comme bruit (ou la lecture comme perception visuelle) et la parole écoutée et comprise :

"Il n'y a pire sourd que celui qui ne veut pas entendre." (proverbe populaire)

Le jeu du téléphone : un des joueurs dit une phrase à l'oreille de son voisin qui la répète au suivant. Même si

on donne comme consigne de parler très distinctement, une phrase un peu longue est profondément transformée après avoir fait le tour de la table.

Chacun peut faire l'expérience de relire plusieurs fois le même texte, il n'en retient pas les mêmes choses à chaque lecture.

Concevoir des échanges langagiers avec les machines sur le mode du transfert est compatible avec l'auto-organisation de la machine, puisque rien ne lui est imposé du dehors, contrairement au modèle du canal utilisé pour communiquer l'expertise aux systèmes experts. La machine doit pouvoir choisir ce qui l'intéresse dans ce qui lui est dit pour sa propre construction. Elle doit pouvoir proposer des solutions, que son interlocuteur peut critiquer sans pouvoir imposer les siennes : il doit la convaincre. Ce modèle peut renouveler la problématique du dialogue homme-machine dans le sens où il ne s'agit plus de transmettre des connaissances mais d'aider à une évolution vers l'expertise, dans un sens comme dans l'autre.

3. L'auto-organisation

Les travaux récents sur la Vie Artificielle [Bourgin 92] ont mis en évidence l'importance de l'autonomie et de l'auto-organisation. J'ai été frappée par les modèles très simples de la croissance des plantes permettant de construire des images de végétaux très réalistes. Il est intéressant de proposer un modèle analogue pour la croissance de nos connaissances. Au lieu de donner à priori une grande quantité de connaissances, comme dans le projet CYC [Lenat & Feigenbaum 91], ou de métaconnaissances comme dans le projet Maciste [Pitrat 90], je cherche à construire un modèle minimal qui va se déployer. Le modèle de machine comme sujet, fondé sur l'auto-organisation, doit prendre en compte la réflexion du sujet sur le monde et sur lui-même. La possibilité d'appréhender sa propre pensée transforme l'être en sujet, il lui faut pour cela transformer ses perceptions et ses actions en objet d'étude. La manière caractéristique d'appréhender le monde pour un sujet pensant est la réification (construction d'une représentation déclarative d'un objet, d'une propriété, d'une action ou de toute autre notion) qui permet le raisonnement abstrait et l'échange langagier.

4. Un exemple

Pour montrer comment un modèle peut se déployer pour construire une représentation réflexive, considérons un système de fiches décrivant des livres ou des avis sur ces livres, ou des avis sur les avis et ainsi de suite. Le modèle de fiche de livre, noté <fiche-livre> peut être représenté par :

```
[ auteur : <prénom> <nom>
  titre : <string>
  année : <entier ≤ année en cours>
  fiches d'avis : ( <fiche-avis>* ) ]
```

Le modèle de fiche d'avis, noté <fiche-avis> peut être représenté par :

```
[ auteur : <prénom> <nom>
  avis : <string>
  sur la fiche : <fiche-livre> ou <fiche-avis>
  année : <entier ≤ année en cours>
  fiches d'avis : ( <fiche-avis>* ) ]
```

Un modèle de fiche est son propre modèle, mais dans un modèle de fiche, les constantes d'une fiche concrète sont remplacées par des contraintes sur les constantes admissibles alors que dans un modèle logique elles seraient remplacées par des variables. Pour définir en intension les constantes qui peuvent prendre cette place, nous proposons une notation de domaines dans un autre article [Nicolle-Adam 92]. Ce système de description de fiches est récursif puisqu'un modèle peut apparaître dans les contraintes de sa

propre définition. Il est possible de développer autant de niveaux d'avis qu'il est nécessaire pour représenter une discussion sur des livres sans notations particulières.

Fiche de livre :

```
f0 : [ auteur : Pattie Maes
      titre : Computational reflection
      année : 1987
      fiches d'avis : ( f1 f2 ) ]
```

Fiches d'avis :

```
f1 : [ auteur : Anne Nicolle
      année : 1992
      avis : fondamental
      sur la fiche : f0
      fiches d'avis : ( f3 ) ]
f3 : [ auteur : Anne Nicolle
      année : 1992
      avis : à développer
      sur la fiche : f1 ]
f2 : [ auteur : X
      année : 1992
      avis : trop difficile à mettre en oeuvre
      sur la fiche : f1
      fiches d'avis : ( f4 ) ]
f4 : [ auteur : Anne Nicolle
      année : 1992
      avis : à justifier
      sur la fiche : f2 ]
```

Une fiche d'avis peut porter aussi bien sur un livre que sur un avis. Le système des fiches d'avis comporte des feuilles, les fiches n'ayant pas de fiches d'avis, qui arrêtent naturellement le système, comme dans une conversation, quand les points d'accord ou de désaccord ont été mis en évidence. Ces feuilles sont de même type que les noeuds et de même type que les modèles : le système est réflexif. Il ne nécessite pas pour sa gestion de complexité autre que celle qui vient de l'univers de référence.

Il est facile de construire une fonction de lecture de fiches qui lit le flot d'entrée et construit des fiches si on lui donne des renseignements sur des livres ou des avis et qui ignore le reste. Les informations que la machine va sélectionner dans son environnement externe pour faire évoluer l'environnement interne de fiches sont celles qui entrent dans ses attendus, c'est à dire celles qui ont un sens pour le modèle de fiches. Le reste est ignoré par le système de fiches.

La machine à construire ne doit pas ignorer ce qu'elle ne comprends pas encore. Elle doit utiliser un modèle analogue au niveau méta (voir §5.2.), pour construire des modèles de ce qu'elle ne connaît pas encore à partir d'un modèle des modèles et des régularités rencontrées. Pour cela, il faut avoir un modèle de l'activité de catégorisation [Dubois 91].

5. Un modèle réflexif d'objets

ObjVlisp [Cointe 84] a proposé un modèle réflexif pour les classes et métaclasses d'un langage à objets. Airelle [Adam-Nicolle 90] a généralisé ce modèle à la définition des attributs et des sélecteurs. La

méthode mise à l'oeuvre dans cette généralisation et dans les autres projets réalisés en Airelle [Brutus 92] est le déploiement du point fixe réflexif, à la manière d'un fractal pour représenter progressivement tout ce qu'on veut représenter. Dans l'exemple ci-dessus, le point fixe réflexif est le modèle des fiches d'avis qui se déploie pour construire un système de fiches. Il peut prendre en compte de nouveaux livres et de nouveaux avis avec ses modèles. Le déploiement du point fixe d'Airelle est de niveau méta par rapport à cet exemple, c'est à dire qu'il porte sur les représentations des représentations. Il peut donc construire de nouveaux modèles pour prendre en compte de nouvelles représentations. Il s'appuie sur la réification, le contrôle explicite à tous les niveaux et la réflexivité des représentations et du contrôle.

5.1. La réification

La réification peut s'appliquer au monde extérieur comme au monde de la pensée, la réification des représentations et des raisonnements permettant de raisonner au niveau méta. Elle présente de manière simplifiée une notion dont la complexité propre est importante par oubli du contingent pour se focaliser sur l'essentiel. La réification est le procédé qui permet de "découper" un objet dans l'univers, de choisir les points de vues "pertinents" sur cet objet dans le contexte du problème ou de la famille de problèmes qu'on veut modéliser et d'en faire un "objet" au sens informatique, c'est à dire une structure de donnée avec des opérations d'accès et de manipulation. Les objets de l'univers, les sujets, les actions, les pensées peuvent ainsi devenir des "objets" d'étude, puis des "objets" d'un système informatique. Le parallèle entre la notion informatique de conception par objets et le modèle proposé pour décrire l'activité scientifique des physiciens par M. Mugur-Schächter doit être souligné. Elle propose de distinguer les concepts suivants :

- . un découpeur, qui isole un objet encore non spécifié.
- . un regard, qui se fixe sur un attribut et détermine sa pertinence pour cet objet
- . une observation, qui permet d'attribuer une valeur à un attribut d'un objet
- . un observateur, qui réalise les actions impliquées par les concepts précédents.

Avec ces concepts, elle construit un modèle de l'activité scientifique et des erreurs dans les modèles qui est transférable terme à terme à la conception par objets [Mugur-Schächter 89].

5.2. Le contrôle

Dans les programmes classiques, le contrôle est programmé, c'est un algorithme qui se déroule. Dans les programmes interactifs, le contrôle est assuré par l'utilisateur qui fait avancer les choses à sa guise et les algorithmes programmés sont vus comme des modules élémentaires quelle que soit leur complexité. Pour automatiser les problèmes trop complexes pour être traités par un algorithme, il faut faire jouer à la machine le rôle d'un utilisateur de système interactif. Elle voit les algorithmes comme des objets dont elle a une représentation déclarative simplifiée. Pour réaliser cela, il faut concevoir une architecture logicielle à plusieurs niveaux.

On appellera architecture d'un système computationnel la description abstraite du contrôle (boucle de calcul ou d'interaction) et des relations entre les structures de données, procédures, règles ou objets et le contrôle. Citons quelques exemples d'architectures proposées en intelligence artificielle : les agendas de AM [Davis & Lenat 82], les réflexes, les agents [Ferber 89], le tableau noir dans ses multiples versions [BS 88].

Le contrôle d'un système auto-organisé peut être réalisé par un algorithme ou être lui-même un système complexe. Dans le contexte de l'auto-organisation animale un algorithme convient, mais pour représenter la pensée qui organise aussi bien le contrôle de la relation au monde que le contrôle de la relation à soi-même, le contrôle est un système complexe, le contrôle du contrôle est un système

complexe et ainsi de suite. L'organisation du contrôle doit alors être réflexive [Maes & Nardi 87]. Un tel système décrit les computations du méta-système pour organiser les computations du système et il s'organise lui-même de la même manière qu'il organise le système qu'il contrôle [Bricon-Souf 92].

5.3. La réflexivité et le problème de l'arrêt

Le modèle réflexif que nous proposons est basé sur les hypothèses suivantes :

- 1- Du point de vue des représentations, il n'y a pas besoin de plus d'un niveau méta : tout méta-système peut être son propre méta-système à condition que ce méta-système soit assez riche pour se représenter, c'est à dire contienne au moins la théorie des nombres [Hofstadter 79].
- 2- La représentation du monde étant incomplète, le système doit faire les hypothèses minimales pour clore ses représentations, et pouvoir les remettre en cause s'il rencontre une contradiction, comme dans l'exemple ci-dessous.
- 3- Du point de vue des raisonnements, il peut y avoir empilement de plusieurs niveaux méta. Pour se prendre comme objet sans tomber dans l'indécidabilité, le méta-système doit se dédoubler. Un raisonnement réflexif part d'un point de cohérence et se copie autant de fois que nécessaire pour résoudre le problème. S'il fait des modifications sur l'une des copies, il doit les prendre en compte en se fusionnant au retour s'il n'y a pas de contradiction. S'il y a contradiction, le méta-système doit provoquer le déploiement du modèle primitif. S'il y a répétition infinie des copies, la copie de niveau méta qui contrôle l'empilement des copies peut faire l'hypothèse de la répétition infinie, le démontrer, et en conclure qu'il s'agit d'un paradoxe.

Un exemple : Soient deux sujets A et B ayant chacun une image de l'autre. Notons $I(x y)$ l'image que x a de y. Chaque sujet a une image de lui même et une image de son interlocuteur.

<i>A</i>	<i>B</i>
<i>image de A = I(A A)</i>	<i>image de B = I(B B)</i>
<i>image de B = I(A B)</i>	<i>image de A = I(B A)</i>

Dans l'image que A a de B, il y a aussi :

l'image de B pour B

l'image de A pour B

Notons © l'inclusion d'une image dans un image.

$$I(A I(B B)) \text{ © } I(A B)$$

$$I(A I(B A)) \text{ © } I(A B)$$

Si personne ne ment, et si tout le monde sait tout, il est raisonnable de faire l'hypothèse que l'image de B dans l'image que A a de lui-même est la même que l'image de A pour A

axiomes :

$$1)- I(A A) = I(A I(A A))$$

$$2)- I(A B) = I(A I(A B))$$

Si ce modèle provoque une contradiction dans un raisonnement parce que B ment, on peut déployer le modèle sur un niveau en remplaçant l'axiome 2 par :

$$2')- I(A B) = I(A I(A I(A B)))$$

Prendre en compte dans le modèle l'inconscient de A peut se faire en remplaçant l'axiome 1 par :

$$1')- I(A A) = I(A I(A I(A A)))$$

Avec ce modèle, le nombre d'images à considérer est fini et petit contrairement aux modèles logiques

des mondes possibles. Le remplacement d'un axiome par un autre est une opération du méta-système, il ne nécessite pas l'intervention d'un sujet extérieur et il est donc compatible avec l'auto-organisation.

5.4. Le modèle

Avec ces hypothèses, un modèle de système computationnel pour explorer la complexité peut être construit à partir d'un interprète réflexif très simple et très abstrait : une boucle infinie

perception - décision - action.

Il y a deux différences avec une boucle **oplevel** d'un interprète comme lisp :

- la perception choisit ses données dans le contexte
- la décision est non déterministe.

Dans les modèles "animaux" de la Vie Artificielle, on trouve une boucle analogue pour gérer le rapport de l'animal à son environnement. Pour en faire un modèle de sujet pensant, il faut la rendre réflexive, c'est à dire faire qu'elle s'applique au contexte à certains moments et qu'elle prenne comme objet son propre fonctionnement à d'autres moments. On passe du modèle de l'animal au modèle du sujet en déployant une fois la boucle de l'animal par réflexion du monde intérieur dans la boucle de rapport au monde. Le modèle de contrôle réflexif par tableau noir proposé par Nathalie Souf [Bricon-Souf 92] peut gérer ces deux boucles sans intervention extérieure.

Le sujet construit et interprète des objets qui représentent son interface avec le monde extérieur et la réflexion du monde intérieur. L'activité de perception est une activité de catégorisation des sensations. Le langage permet de se mettre d'accord sur la catégorisation, donc de la préciser et de l'argumenter. Par le transfert, la catégorisation faite par chaque individu s'améliore car elle peut tenir compte de davantage d'arguments, fondés sur les expériences des autres. Un des buts de Comperobot [Vivier & Bricon-Souf 90] est justement d'utiliser le dialogue enfant-machine pour amener les enfants à paraphraser et donc pour étudier ce processus de négociation sur la désignation des objets, de leur position et de leur localisation par rapport à d'autres objets.

Les représentations sont catégorisées grâce à deux opérations, généralisation-spécialisation d'une part et abstraction-instanciation d'autre part.

- généraliser, c'est remplacer une constante par un domaine pour faire une classe plus générale que la classe de départ. C'est une opération de factorisation entre des descriptions génériques.
- abstraire, c'est trouver une analogie et faire un modèle abstrait de ce qui est semblable. C'est une opération qui permet de passer du spécifique au générique. Un modèle abstrait peut servir dans des univers de référence très différents pour décrire des objets ayant des propriétés abstraites communes. Les modèles abstraits sont des classes sur lesquelles des opérations de généralisation et de spécialisation sont possibles. En faisant des modèles abstraits de modèles abstraits, on arrive aux objets mathématiques (structures algébriques, équations...).

Ces opérations et la manière de représenter les connaissances dans les machines en en tenant compte ont été mises en oeuvre dans des réseaux sémantiques, des frames, des graphes conceptuels. Le savoir faire informatique dans ce domaine est important [Sowa 91].

La mise en commun des résultats de ses opérations constitue la connaissance commune. Elle est constituée de résultats sur les classes et sur les modèles, mais aussi d'une argumentation permettant à chacun de se les approprier par transfert, et pas avec des arguments d'autorité. L'oubli de cette argumentation dans les systèmes experts, même basés sur des connaissances, en fait des machines à répéter.

Le modèle primitif des représentations est fondé sur la réification, c'est à dire sur la recherche d'un objet d'étude, d'un attribut pour cet objet, d'une valeur pour cet attribut. Avec la réflexion, les

attributs et les types des valeurs sont aussi des objets d'étude. En construisant les représentations comme déploiement de ce modèle primitif qui se raffine, on conserve la trace des raisons qui amènent à complexifier le modèle et donc l'argumentation : à chaque déploiement une nouvelle différence est réifiée et prise en compte. L'uniformité des représentations permet de ne pas introduire dans le système de complexité liée à sa propre gestion : c'est la complexité du monde qui se reflète dans le système, qui est construit à partir de briques syntaxiquement et sémantiquement simples.

Avec ce modèle, les connaissances sont incomplètes par hypothèse car c'est cette hypothèse qui permet le déploiement permettant d'acquérir de nouvelles connaissances. Pour raisonner, on complète temporairement les connaissances en faisant l'hypothèse de la simplicité maximale. L'expérience révèle les contradictions dues à des simplifications abusives, les échanges langagiers permettent d'avoir accès à l'expérience des autres sujets et à la connaissance partagée, ce sont les deux ressorts du déploiement du modèle.

6. Machines et processus langagiers

Ce modèle est mis en oeuvre et expérimenté dans plusieurs domaines. Le projet le plus intéressant pour ce qui est de l'intelligence collective est Compérobot, dont le sujet est l'étude théorique et expérimentale des échanges langagiers entre des enfants et des compères, les compères étant d'abord des adultes, puis des adultes se faisant passer pour des machines, puis des machines. C'est un projet interdisciplinaire entre psychologues du langage, linguistes et informaticiens. Les exigences sur le comportement des machines pour les expériences prévues par les psychologues nous ont conduit à concevoir le modèle proposé ci-dessus. Décrire le langage est trop complexe pour être fait de manière exhaustive, un modèle minimal et adaptatif s'impose, encore faut-il le faire tourner, ce qui n'est pas encore possible en octobre 92. Notons que dans ce projet, le langage est à la fois outil de communication entre l'enfant et la machine et objet d'étude. Ce n'est pas le cas des autres systèmes de dialogue homme machine, où le langage est seulement un outil et qui peuvent utiliser des modèles moins compliqués.

7. Références

- **Adam-Nicolle 90**
Anne Adam-Nicolle, Le métalangage Airelle, Revue Française d'Intelligence Artificielle, volume 4 n° 1 - 1990.
- **Bourgine 92**
"Towards a practice of autonomous systems." Paul Bourgine Editeur, Pergamon Press 1992
- **Bricon-Souf 92**
Nathalie Bricon-Souf, "Contrôle et réflexivité." Communication au colloque d'intelligence artificielle du LAFORIA, septembre 1992. Cahiers du LAIAC n°92-3 Université de Caen.
- **Brutus 92**
Philippe Brutus, "Génie Logiciel à base d'Objets et de Reflexivité pour le développement d'applications à Interface Avancée automatique." Thèse de doctorat de l'Université de Caen. 1992
- **Brooks 91**
R.A. Brooks, "Intelligence without representation." Artificial Intelligence vol 47 n°1-3 janvier 1991.
- **BS 88**
"Blackboard Systems." Addison Wesley. 1988
- **Cointe 84**
P. Cointe, " Implémentation et Interpretation des Langages Orientés Objets. Application aux Langages SMALLTALK, OBJVLISP et FORMES", Thèse d'Etat, LITP, Université Paris 6 (Décembre 1984)

- **Coursil 92**
Jacques Coursil, "Grammaire analytique du français contemporain. Essai d'Intelligence Artificielle et de Linguistique générale." Thèse de doctorat de l'Université de Caen. 1992
- **Davis & Lenat 82**
Randall Davis et Douglas B. Lenat
"Knowledge-Based Systems in Artificial Intelligence"
Mc Graw-Hill (1982)
- **Dubois 91**
Danièle Dubois, "Symbolisation, systèmes symboliques." Ecole d'été de l'ARC - langues et langage : carrefour des sciences de la cognition. 1991
- **Ferber 89**
Jacques Ferber, "Objets et agents : une étude des structures de représentations et de communications en Intelligence Artificielle", Thèse de Doctorat d'Etat - Paris 6 - 1989.
- **Hofstadter 79**
D. Hofstadter, "Gödel, Escher, Bach Les brins d'une guirlande éternelle." Basic Books, Inc. 1979.
Edition Française Interéditions 1985.
- **Lenat & Feigenbaum 91**
D.B. Lenat & E.A. Feigenbaum, "On the thresholds of knowledge." Artificial Intelligence vol 47 n°1-3 janvier 1991.
- **Maes & Nardi 87**
Maes et Nardi Eds : "Meta-level architectures and reflection." North-Holland, Amsterdam 1987
- **Mugur-Schächter 89**
M. Mugur-Schächter, "Esquisse d'une méthode générale de conceptualisation relativisée. " dans "Arguments pour une méthode, autour de Edgar Morin" Editions du Seuil Paris 1989.
- **Nicolle-Adam 92**
Anne Nicolle-Adam, "Etude d'une notation de domaines" Cahiers du LAIAC n°1-92 Université de Caen
- **Pitrat 90**
Jacques Pitrat, "Métaconnaissance" Hermès 1990.
- **Simon 87**
Herbert A. Simon, "Sciences des systèmes, sciences de l'artificiel."
Collection AFCET DUNOD 1987
- **Sowa 91**
J.F. Sowa, "Principles of semantic Networks" Morgan-Kaufman 1991
- **Vivier & Bricon-Souf 90**
Jean Vivier, Nathalie Bricon-Souf, "Etude d'un dialogue enfant-machine." Cahiers de linguistique sociale n°16. Université de Rouen - SUDLA 1990

Coordination *via* l'environnement vs coordination *via* une organisation, une problématique récurrente dans l'analyse des phénomènes d'intelligence collective

Manuel Zacklad (1) (2) & Jean Erceau (1)

(1) ONERA
Groupe d'Intelligence Artificielle (DMI)
29, av. Division Leclerc
BP 72 - 92322 CHATILLON CEDEX - FRANCE

(2) UNIVERSITÉ DE TECHNOLOGIE DE COMPIEGNE
Département du Génie Informatique
URA CNRS 817
BP 649 - 60206 COMPIEGNE CEDEX - FRANCE

1 - Introduction

Dans les recherches actuelles en Intelligence Artificielle Distribuée (IAD) il est de coutume d'opposer deux conceptions dans la modélisation des agents et de leurs interactions. On oppose ainsi agents "cognitifs" vs agents "réactifs" [Erceau et Ferber 91] ou approche "sociale" vs approche "biologique" [Ferber 89]. Dans la première tendance, on assimile les agents à des acteurs sociaux. La coopération repose sur des phénomènes de négociation, de planification, d'échange, d'explication et de partage de buts et de plans. L'intelligence collective, que l'on peut sommairement assimiler à la capacité que possède un groupe d'agents à résoudre des problèmes, est alors le résultat de la coordination de ces agents au moyen d'un plan plus ou moins délibéré et partagé. Dans l'autre tendance, on considère que les agents individuels ne possèdent pas de représentation compréhensive de la situation voire même qu'il ne possèdent pas de représentation du tout. Le comportement intelligent du groupe est alors le résultat des interactions entre des agents qui sont seulement dotés de quelques automatismes simples. Le comportement collectif cohérent et efficace va *émerger*, d'une dynamique d'interactions au sein du groupe par des phénomènes de propagation et de diffusion. Or les agents, dans beaucoup d'applications, ne sont jamais totalement et exclusivement cognitifs ou réactifs. Au delà de cette dialectique nous voulons proposer une approche complémentaire qui, nous l'espérons, aura une bonne valeur heuristique pour comprendre et expliquer le fonctionnement de l'intelligence collective. Son but est d'avancer dans la réponse à cette question de base : quels sont les mécanismes de coordination qui permettent aux agents réactifs et cognitifs d'avoir un comportement collectif intelligent ?

Les deux pôles de ce système d'oppositions visent à traduire deux modalités fondamentales de la coordination dans un groupe d'agents: la coordination utilisant le médium de l'environnement et la coordination utilisant le médium de l'organisation. Cette opposition ne recouvre pas l'opposition réactif/cognitif. Nous verrons, à travers une série d'analogies que nous allons emprunter successivement à plusieurs disciplines, que la coordination via l'environnement se retrouve aussi bien dans la métaphore biologique, qui mettait en scène des agents réactifs, que dans la métaphore économique, qui met en scène des agents sociaux dotés de représentations. Cette opposition est d'un ordre différent qui met d'avantage l'accent sur le mécanisme par lequel une action collective cohérente se construit, que par les propriétés intrinsèques dont sont dotés, ou ne sont pas dotés, les agents qui la réalisent.

2 - Premier aperçu des deux formes de coordination

Dans ce papier nous chercherons à éclairer cette opposition entre les deux modalités de la coopération à travers une série d'analogies empruntées à la psychologie cognitive, à l'éthologie et à l'économie. Mais commençons par donner une première définition de la coordination et ensuite de ces deux modes de coordination.

2.1 - Une définition de la coordination des actions rapportée au degré de standardisation.

Selon [Mintzberg 82], l'action collective doit répondre à deux exigences fondamentales et contradictoires : la division du travail des agents entre les différentes tâches à accomplir et la coordination de ces tâches pour l'accomplissement du travail :

"La structure d'une organisation peut être définie simplement comme la somme totale des moyens employés pour diviser le travail entre tâches distinctes et pour ensuite assurer la coordination nécessaire entre ces tâches." (p.18). Pour l'auteur, la coordination recouvre aussi bien les dimensions relatives à la communication qu'au contrôle (identification et sélection des actions).

A la suite des travaux de [Simon 57] et [March et Simon 58], Mintzberg décrit cinq mécanismes de coordination fondamentaux correspondant aussi à des phases d'évolution temporelle de l'organisation avec le temps (p. 18 à 23):

- 1) L'ajustement mutuel qui réalise la coordination du travail par simple communication informelle.
- 2) la supervision directe qui est le mécanisme de coordination par lequel une personne se trouve investie de la responsabilité du travail des autres.

Les trois autres mécanismes de coordination sont des variantes de la standardisation. La standardisation est le moyen par lequel "la coordination des diverses parties est incorporée dans le programme (de travail) dès la conception" le besoin de communication continue s'en trouvant ainsi réduit ([March et Simon 58], cité par Mintzberg). L'auteur distingue trois formes de standardisation :

- 3) La standardisation des procédés qui a lieu quand "les procédés de travail sont standardisés et lorsque le *contenu* du travail est spécifié ou programmé" (p. 21).
- 4) La standardisation des résultats qui a lieu quand il est "possible de standardiser les *résultats* du travail (par exemple en spécifiant à l'avance les dimensions du produit, ou la performance à atteindre).
- 5) La standardisation des qualifications, qui a lieu quand est spécifiée la *formation* de celui qui exécute le travail.

Remarquons que, par cette définition Mintzberg caractérise les organisations à travers différents modes de coordination alors que nous posons comme hypothèse de travail que l'organisation est un mode particulier de coordination entre agents. Ce désaccord terminologique vient de la différence entre les préoccupations d'origine. Mintzberg travaille dans le domaine de la sciences des organisations. Pour lui, l'organisation est le phénomène premier qu'il s'agit d'expliquer en faisant intervenir les modalités de la coordination. Pour nous, le phénomène premier est le phénomène d'action collective. Si ce phénomène peut sembler organisé, il ne résulte pas nécessairement de l'existence d'une *organisation délibérée* entre les agents. Par contre, l'action collective, si elle est collective, doit être coordonnée. Nous considérons donc que l'institution d'une organisation entre les agents est un mode particulier de coordination, même si l'observateur peut avoir l'impression qu'une certaine organisation préexiste alors qu'elle ne fait qu'émerger dans le cas d'une coordination par l'environnement. Pour être plus précis nous aurions pu opposer une organisation "spontanée" via l'environnement à une coordination via une organisation "instituée". Mais l'utilisation du terme d'organisation spontanée présente aussi des difficultés et celle d'organisation instituée est assez lourde. Nous maintiendrons pour l'instant, faute d'une meilleure dénomination, l'opposition environnement vs organisation

2.2 - Une nouvelle caractérisation de la coordination des actions selon le médium : environnement vs organisation.

L'analyse de Mintzberg analyse les modes de coordination à partir de la standardisation progressive des plans d'action. Il y a un continuum de sophistication de la coordination allant de l'ajustement mutuel à la standardisation des qualifications. Mais la standardisation des actions n'est profitable que pour des environnements suffisamment stables. Face à des environnements très changeants, dans les problèmes de conception créative, par exemple, l'ajustement mutuel reste irremplaçable. Mais il y a une autre dimension importante qui influence la coordination et qui est pour une part indépendante de la dimension du degré de standardisation des actions. Nous appellerons cette dimension le médium de coordination en opposant deux médias: l'environnement et la structure de l'organisation.

Ces deux modes peuvent d'abord être appréhendés du point de vue des réseaux de communication entre les agents. Dans la coordination par l'environnement, les communications relatives à une tâche ne sont pas adressées à un autre agent particulier mais sont médiatisées par une transformation physique de l'environnement commun. Par exemple, dans leur ethno-modélisation d'une société de fourmis, [Drogoul et Ferber 92] considèrent que la communication par l'environnement entre les fourmis, réalisées par des signaux chimiques, n'est pas intentionnelle parce que le signal peut activer des réactions différentes selon son "lecteur". Mais nous irons plus loin. Dans un cas général de coordination par l'environnement, il n'est même pas nécessaire d'envisager les transformations de l'environnement sous l'angle de la communication. Même si la finalité communicationnelle est totalement absente de la transformation de l'environnement, elle pourra servir de support à la coordination des actions des agents.

A l'opposé, dans la coordination par l'organisation, les réseaux de communication sont précisément définis par la structure de l'organisation. Des transformations sur l'environnement interviennent aussi dans ce mode de coordination, mais elles ne conditionnent pas directement les actions des agents. Les actions des agents sont déterminées par le rôle que l'organisation leur assigne. De plus, tous les agents n'ont pas accès ou ne traitent pas directement les signaux de l'environnement. Pour l'essentiel, les signaux qui conditionnent l'action, proviennent de messages spécialement émis par d'autres agents. Seuls certains d'entre eux perçoivent directement l'environnement et cette perception peut elle-même être restreinte à certaines facettes, sur lesquelles l'organisation les amène à se focaliser.

Voici brièvement exposées les principales caractéristique de ces deux média de coordination. A travers des analogies empruntées à trois domaines nous allons essayer d'en percevoir la généralité sous la diversité apparente des problématiques.

3 - Analogies avec la psychologie cognitive et l'évolution du contrôle de l'activité dans l'apprentissage.

Dans le cadre d'une analogie avec la psychologie cognitive, l'intelligence collective peut être définie à partir d'une société d'agents "mentaux" au sens de la société de l'esprit de Minsky 86]. Nous proposons de montrer que les phénomènes d'apprentissage peuvent être vus comme le passage d'une coordination par une organisation fonctionnelle des agents mentaux à une coordination par l'environnement du sujet. Nous commencerons par un rappel sur la notion de "niveau de contrôle" qui permet de rendre compte des phénomènes d'apprentissage chez l'homme.

En psychologie, on considère que le niveau de contrôle des actions évolue avec l'apprentissage. Au début, les actions de l'apprenant reposent sur une représentation explicite des procédures de traitement et de leur déroulement. Puis avec l'accroissement des performances, les automatismes vont prendre le relais du contrôle par des représentations. Par exemple, Rasmussen 84 dans des études consacrées au contrôleurs de processus industriels, distinguent trois niveaux de contrôle de l'activité: un contrôle par automatismes, par règles et par des connaissances de haut niveau que nous appellerons contrôle représentationnel.

L'apprentissage ou l'accroissement d'expertise, consiste à passer de situations qui sont presque des situations de résolution de problème, où l'activité est contrôlée des représentations complexes des connaissances, à un contrôle de l'activité par des automatismes déclenchés et contrôlés directement par des perceptions provenant de l'environnement. Si l'on adopte les analogies de Minsky, qui considère les savoir-faire comme des agents délibérant et se coordonnant "dans la tête" du sujet, on peut considérer que le passage d'un contrôle représentationnel à un contrôle par des automatismes se produit de la manière suivante. Tout d'abord, le contrôle représentationnel repose sur des agents cognitifs qui coopèrent en utilisant une organisation fonctionnelle représentée dans la mémoire de travail du sujet. Par la suite, avec les progrès de l'automatisation des actions, ces agents mentaux "internes", qui jusqu'alors coopéraient entre eux dans une organisation, vont progressivement devenir plus autonomes. Leur coordination reposera d'avantage sur les caractéristiques physiques de l'environnement de l'action qui se modifient au fur et à mesure, au cours

de son execution. On peut le comprendre en utilisant la théorie d'[Anderson 83]. Le contrôle des actions sera réalisé au moyen de programmes plus longs mais plus opaques, opaques parce que les raisonnements intermédiaires ont été compilés et ne sont plus représentés explicitement dans la mémoire du sujet. Par contre, les déclencheurs internes de ces macro-actions seront plus spécialisés selon des indicateurs contextuels extérieurs.

Du point de vue de notre problématique, la société des agents mentaux va passer d'un mode de coordination à l'autre, les agents utilisant de plus en plus l'environnement pour contrôler l'action au détriment d'une organisation fonctionnelle des connaissances maintenue dans une mémoire de travail interne. Progressivement, l'environnement va se substituer en partie à la mémoire de travail pour devenir, selon les termes de Newell et Simon 72, une *mémoire de travail externe*. La coordination entre les agents mentaux, ne sera alors plus médiatisée par leur rôle fonctionnel défini dans l'organisation et correspondant au plan d'action représenté par le sujet, mais par des caractéristiques de l'environnement dont les transformations résulteront des actions individuelles et apparemment indépendantes des agents.

Ce fonctionnement des agents mentaux au moyen d'une structure de données commune à laquelle chacun a accès et qu'il peut modifier indépendamment des autres est d'ailleurs une vision ancienne en psychologie, qui nous conforte dans notre idée selon laquelle la coordination par l'environnement est une modalité essentielle dans l'analyse de l'action collective. Dans les travaux sur la reconnaissance de la parole, Selfridge 59] avait proposé un modèle dans lequel cette structure de données, un "tableau-noir", permettait à différents démons spécialisés dans la reconnaissance de caractéristiques particulières des signaux d'interagir entre eux. A priori cependant, le tableau noir était une structure de représentation interne. Les bases de ce modèle ont été reprises en IA dans la définition des architectures "black-board", initialement utilisées, là aussi, pour la reconnaissance de la parole. On sait le très large succès que ces architectures ont eu ensuite dans de nombreux domaines.

Par rapport à notre problématique, on remarquera cependant deux points. Premièrement, du point de vue des agents "mentaux", l'analogie que nous utilisons consiste à considérer qu'il y a un passage progressif d'une organisation fonctionnelle des agents représentée mentalement par le sujet à l'exploitation d'une mémoire environnementale externe. Ce passage s'accompagne d'une transformation structurelle des agents, c'est à dire des programmes, qui vont, par exemple, payer la simplification de leurs mécanismes d'interaction mutuelle par une augmentation en volume et en

nombre. Cette analogie est différente de l'analogie originelle des black-boards, selon laquelle, à notre avis, la structure de données constitue, pour les agents mentaux, un *environnement mnésique* leur offrant, sur un plan cognitif, la possibilité d'une coordination par l'environnement. Deuxième point, toujours par rapport aux black-boards, on sait que l'utilisation de cette structure de données s'accompagne généralement d'une spécialisation fonctionnelle des différentes zones du tableau ou, pour nous, de l'environnement. De même les sources de connaissances du black-boards possèdent une spécialisation dans les zones du tableaux auxquelles elles accèdent en "écriture" ou en "lecture". Dans ce cas, on a alors un compromis entre les deux modes fondamentaux de coordination que nous présentons, puisque l'environnement intègre lui-même dans sa structure les bases de l'organisation fonctionnelle des rôles des agents.

4 - Analogies avec l'éthologie et la biologie.

Avec l'essor de l'éthologie, les comportements sociaux des animaux ont été reconnus dans toute leur richesse et leur complexité. Il existe aujourd'hui des classifications sophistiquées des différentes sociétés animales. Dans l'analyse de l'intelligence collective animale, on peut se situer sur deux échelles temporelles distinctes et parfois concurrentes dans les explications biologiques. Dans un cas, l'intelligence animale collective peut avoir comme but d'assurer la survie d'un groupe particulier situé sur un territoire donné, pendant une période de temps limitée à la vie de ce groupe. Cependant, l'observateur est souvent confronté à des comportements de sacrifice individuel d'un individu pour le groupe dont il n'est pas aisé de comprendre la justification dans une logique de sélection darwinienne appliquée aux individus. On fait alors appel à une autre explication selon laquelle la raison de la cohésion du groupe est à chercher dans la proximité génétique des individus qui le compose. Selon les termes de la sociobiologie, les individus seraient, en quelque sorte, les instruments du collectif de gènes cherchant à se reproduire. Le sacrifice d'un individu pour le groupe s'explique par la présence importante, dans ce groupe, de parents ayant des gènes communs avec lui. Servant l'intérêt des gènes, l'intelligence est alors plutôt l'intelligence de l'espèce. Selon certains chercheurs, l'espèce pourrait avoir une stratégie visant à créer des sous-structures génétiques servant d'unité de base sur lesquelles opérera la sélection naturelle (cf. Marler 73]p. 211). Dans l'approche de l'intelligence collective des sociétés animales, il peut être pertinent de garder en perspective ces deux plans: celle de l'intelligence collective du groupe d'animaux confronté au problème de leur survie "ici et maintenant" et celle de l'intelligence collective de l'espèce qui cherche à augmenter ses chances de

survie à long terme en développant une socialité groupale, agrégeant les individus et les soumettant à un ordre commun.

4.1 - Coordination par l'environnement: la diversité des média disponibles.

L'étude des communications animales montre qu'il existe plusieurs environnements de coordination potentiels selon les canaux de communication. En effet, les signaux peuvent être chimiques, visuels, auditifs ou dans certains cas électriques et "l'animal doit donc effectuer un choix dans les limites qui lui sont imposées par son équipement sensori-moteur" (Marler p. 212). Marler évoque les avantages et les inconvénients de l'utilisation de ces différents canaux qui définissent autant de média de coordination différents entre les individus. Ainsi, dans le cadre d'une coordination visant à "maintenir la structure sociale de l'espèce, en sauvegardant l'espacement voulu à l'intérieur et à l'extérieur du groupe" (p. 212), un média de coordination auditif sera particulièrement intéressant pour des espèces en déplacement constant, comme les oiseaux ou les singes des forêts, "qui ont besoin de communiquer pour signaler les déplacements de chaque groupe et maintenir ainsi les distances de chacun" (p.212). Le canal auditif est efficace de nuit comme de jour et est moins sensible aux obstacles que le canal visuel. Par contre, pour une communication à faible distance, le canal visuel offre les avantages suivants: localisation plus facile de l'émetteur dans le cas des situations de foule, codage spatial plus aisée que par la communication auditive, possibilité de variation indépendante des expressions faciales et corporelles... En revanche, l'efficacité de ce canal disparaît dès que la nuit tombe. Ce n'est pas le cas des signaux chimiques, qui ne sont pas sensibles à l'éclairage et qui possèdent une durabilité que n'ont pas les signaux auditifs. Ce caractère de durabilité est partagé avec certains signaux visuels, qui opèrent par une transformation durable apportée à l'environnement physique, comme les traces de pas, les pistes, les nids, ou, chez l'homme, des dispositifs comme la signalisation routière.

L'exemple des communications animales a l'intérêt de montrer que les média de coordination environnementaux peuvent être très divers, offrant plusieurs possibilités de choix pour l'établissement de la coordination.

4.2 - La coordination par l'organisation chez les abeilles.

Les mêmes colonies d'insectes qui offrent les plus beaux exemples de coordination par l'environnement, fournissent aussi des exemples impressionnants de coordination par l'organisation. Chez les abeilles, les rôles sont très spécialisés et le partage du travail

entre ouvrières, soldats et nourricières est parfaitement réglé pour un développement harmonieux de la ruche. Un des aspects les plus intéressants de ces sociétés, est la capacité dont est doté un agent centralisateur de faire évoluer dynamiquement la structure de l'organisation par la gestion des contingents d'individus qui prendront tel ou tel rôle.

En effet, chez les abeilles, c'est la reine qui programme le rôle futur d'une larve en fonction du stade de développement de la ruche et des conditions environnementales. Centralisatrice des informations vitales sur les évolutions de l'environnement et de la ruche, qu'elle programme en retour, la reine est le "centre stratégique" du groupe dont elle régule les évolutions. L'exemple de la ruche est riche d'informations à plus d'un titre. D'une part, on peut constater que l'évolution de l'organisation s'appuie sur une "polyvalence" originelle des agents qui peuvent être programmés pour tenir tel ou tel rôle. D'autre part, on voit que la coordination par l'organisation permet une forme d'anticipation des besoins futurs par une planification à plus ou moins long terme de son fonctionnement. En dehors du modèle Darwinien qui prédit le développement de caractéristiques génétiques avantageuses des agents par un mécanisme de sélection naturelle au moment de la reproduction des "survivants", on a pas de modèles d'anticipation des besoins du groupe dans le cas d'une intelligence collective se coordonnant par l'environnement. C'est aussi une des différences importantes entre les deux modes de coordination. Si l'on s'intéresse aux capacités d'apprentissage et d'évolution dans le cadre de ces deux modes de coordination, on observe, à partir de ces exemples, que dans le cas d'une coordination par l'environnement "pure", celui-ci a aussi une influence directe sur l'évolution des caractéristiques fonctionnelles des agents individuels. Dans l'autre cas, cette adaptation aux variations de l'environnement est médiatisée par la perception d'un agent centralisateur (ou selon d'autres mécanismes en fonction de la structure de l'organisation). L'adaptation ne se produit pas nécessairement par une modification des fonctionnalités individuelles des agents, mais peut se faire par une régulation des rôles existant à l'intérieur de l'organisation.

5 - Les analogies économiques

Le meilleur exemple de la coordination *via* l'environnement dans le domaine social est celui de la coordination des agents économiques à travers un marché. Où est l'intelligence collective dans le fonctionnement du marché ? Elle est dans la capacité de régulation qui existe entre les agents producteurs et les agents consommateurs. Les premiers doivent calculer la quantité de biens à produire, l'importance de leurs investissements, le niveau de leur marge bénéficiaire. Les seconds doivent gérer leur

budget en fonction de leurs préférences et de leurs besoins. Consommateurs et producteurs peuvent posséder des rôles bien distincts mais peuvent aussi être interchangeables. Chaque agent peut être producteur d'une certaine catégorie de bien consommée par d'autres agents. Globalement, le groupe d'agents, à la fois producteurs et consommateurs, aura un comportement intelligent si l'exploitation des ressources environnementales est adaptée au besoins de consommation du groupe. Comment le comportement des agents "producteurs" s'adapte-t-il aux besoins et aux désirs des agents "consommateurs" ? Il n'existe à notre connaissance qu'une seule alternative dans les sociétés complexes: la coordination *via* le marché ou *via* le "plan" (organisation) dans une économie planifiée.

Pour l'économiste, le marché est bien sûr d'abord une institution. Mais cette institution va avoir toutes les caractéristiques, à un niveau abstrait, d'un environnement partagé par les agents producteurs et consommateurs qui leur permet de se coordonner. En économie, le marché est avant tout une structure de gestion des informations relatives à l'offre et à la demande. Pour ajuster production et consommation, les agents utilisent des signaux privilégiés qui sont les prix. Pour l'économiste, le prix sur un marché, est un signal adressé indirectement par un agent producteur au groupe des agents consommateurs (et éventuellement aussi au groupe des agents producteurs). Il le modulera en fonction du taux de réponse qu'il obtiendra en retour, c'est à dire des actes d'achat.

Il est intéressant de constater que le mode de coordination par le marché est une institution sociale recherchée pour ses avantages spécifiques et protégé explicitement par des lois que font respecter des organisations. Par exemple, si les producteurs d'un bien particulier cherchent à s'organiser faussant le jeu de la "libre concurrence", ils seront sanctionnés dans une économie recourant au marché pour coordonner les agents économiques.

Comme c'est le cas pour les sociétés animales, si le marché est bien adapté pour réagir rapidement face à des variations soudaines des conditions environnementales, il ne nous semble pas permettre, en lui-même, une anticipation permettant de faire face à des risques de transformation en profondeur de l'environnement, à cause de l'absence de perception globale des évolutions. Pour des raisons multiples, qu'il serait trop long de développer ici, une coordination par l'organisation nous semble permettre sans doute plus facilement le développement de ces mécanismes d'anticipation.

6 - Un exemple d'hypothèse sur les performances relatives des deux modes de coordination: la réaction à des fluctuations des conditions environnementales

Dans cette partie, nous présentons une hypothèse sur les performances d'un groupe du point de vue de la perception de l'environnement selon le mode de coordination choisi. Cet aspect des performances du groupe n'en est qu'un parmi d'autres et nous envisageons d'élargir le spectre des comparaisons possibles dans un prochain travail.

Nous avons proposé ailleurs [Zacklad 87], de considérer qu'un groupe engagé dans la résolution de problème pouvait être appréhendé comme un système de traitement de l'information de groupe que nous notons STIg (Système de Traitement de l'Information de groupe). Le STIg peut être crédité d'un comportement rationnel décrit par un système de buts, de perceptions et d'actions. Les perceptions du STIg sont médiatisées par l'équipement perceptif de certains agents individuels qui peuvent se spécialiser dans l'analyse de signaux particuliers et communiquer à d'autres agents du groupe les résultats de cette analyse. Toutefois, nous n'avons alors envisagé cette description que pour des groupes se coordonnant par l'organisation, caractéristique que nous appréhendons par la notion de "plan de groupe". Dans l'analyse d'une coordination par l'environnement la notion de STIg semble plus délicate à justifier.

La description du groupe en terme de STIg, dans le cas d'une coordination par l'organisation, laisse à penser que ce mode de coordination permet de réaliser une économie globale des efforts perceptifs dirigés vers l'environnement, puisque seuls quelques agents ont leur attention tournée vers lui. Par contre, pour chaque agent, des efforts spécifiques de perception et d'action à destination des autres agents sont imposés par le fonctionnement de l'organisation alors qu'ils ne le sont pas dans le cas d'une coordination par l'environnement. Ainsi, si le STIg permet une économie globale des activités perceptives dirigées vers l'environnement, ses temps de réaction pourront être plus importants à cause du travail de coordination interne entre agents.

Au contraire, dans la coordination par l'environnement, le fait que chaque agent doive posséder une perception "égalitaire" de l'environnement, évoque une constance des activités perceptives dirigées vers lui. Ces efforts d'attention soutenus de chaque individu pourraient se faire au détriment, sur un plan global, d'une analyse plus en profondeur de l'information perçue par le groupe. En contrepartie, les réponses à l'égard des variations de l'environnement semblent pouvoir être plus rapides, puisque chaque agent est directement en relation avec lui.

Ces propriétés contrastées nous amènent à formuler une hypothèse qui reste à valider par des simulations ou des analyses de cas :

L'hypothèse de l'influence des deux modes de coordination sur l'efficacité des réactions du groupe selon les variations de son environnement.

Face à des environnements caractérisés par des variations rapides mais facilement interprétable par un agent individuel, la coordination par l'environnement peut être préférable pour construire rapidement une action collective. Par contre, si l'environnement est difficilement compréhensible par un agent (volume, complexité ou ambiguïté des signaux), la coordination par l'organisation peut être préférable parcequ'elle permet de spécialiser des agents particuliers sur certaines fréquences du "spectre" des signaux environnementaux (d'un point de vue spatial, temporel ou sémantique), et de construire une "image" plus globale des variations.

7 - Conclusion : mise à l'épreuve du système d'opposition proposé et approfondissement.

Dans ce papier, nous avons ébauché un système de comparaison des groupes multi-agents différent du paradigme "cognitif/réactif". Des agents réactifs peuvent se coordonner par l'organisation et des agents cognitifs choisir une coordination par l'environnement. La composition avec ces deux modes de coordination nous semble être un des déterminants essentiel, mais non toujours pris en compte, des phénomènes d'intelligence collective. Notre projet de recherche actuel est d'utiliser ce système d'opposition dans une étude de l'état de l'art des réalisations en IAD. Nous espérons parvenir à enrichir les classifications qui sont proposées aujourd'hui Gasser et Hill 90]en mettant en évidence des regroupements nouveaux expliquant les performances collectives des agents. Cette confrontation devrait également nous permettre d'améliorer notre définition de ces modes de coordination qui a encore un statut d'hypothèse de travail.

Par ailleurs, parmi les différentes directions dans lesquelles nous souhaitons poursuivre la réflexion, l'une a trait aux développement des phénomènes d'apprentissage dans le contexte d'une intelligence collective. Une première opposition existe entre des phénomènes d'apprentissage réalisés par une reproduction sélective des agents les mieux adaptés, ou par la reproduction des comportements adaptatifs les mieux adaptés qu'ils mettent en oeuvre, et un phénomène d'apprentissage résultant de modifications structurelles survenant, non chez les agents, mais dans leur organisation. Une troisième

forme d'apprentissage peut se traduire par le passage d'une modalité de coordination à l'autre. L'examen de ces modalités d'apprentissage devrait permettre d'enrichir la connaissance des critères de comparaison et de choix entre les deux modes de coordination.

8 - Bibliographie.

[Anderson 83] Anderson J.R., The architecture of cognition. Cambridge, Mass., Harvard University Press, 1983.

[Drogoul et Ferber 92] A. Drogoul et J. Ferber, Ethomodeling: a multi-agent behavioral simulation model, 1992, Rapport interne du LAFORIA, n° 02.92, Institut Blaise Pascal, Univ. Paris VI et Paris VII.

[Erceau et Ferber 91] Erceau J. et Ferber J., L'intelligence Artificielle Distribuée, La Recherche 233, vol, 22, pp. 750-758.

[Ferber 89] Ferber, J., Objets et Agents, une étude des Structures de Représentation et de Communication en Intelligence Artificielle, Thèse d'Etat, LAFORIA, Université de Paris 6.

[Gasser et Hill 90] L. Gasser et R.W. Hill Jr., Coordinated Problem Solvers, Annual Review of Computer Science, 1990, 4:203-53.

[March et Simon 58], March, J.G. et Simon H.A., Les organisation, Dunod, 1969 (trad. frçse).

[Marler 73], Marler, P., Les communications animales, in La Recherche en Ethologie, pp. 200-248, Ed. du Seuil, 1979.

[Minsky 86] Minsky, M., The Society of Mind, Basic Books, 1986.

[Mintzberg 82] Mintzberg, H., Structure et Dynamique des Organisations, Les Editions d'Organisation, 1982 (trad. frçse).

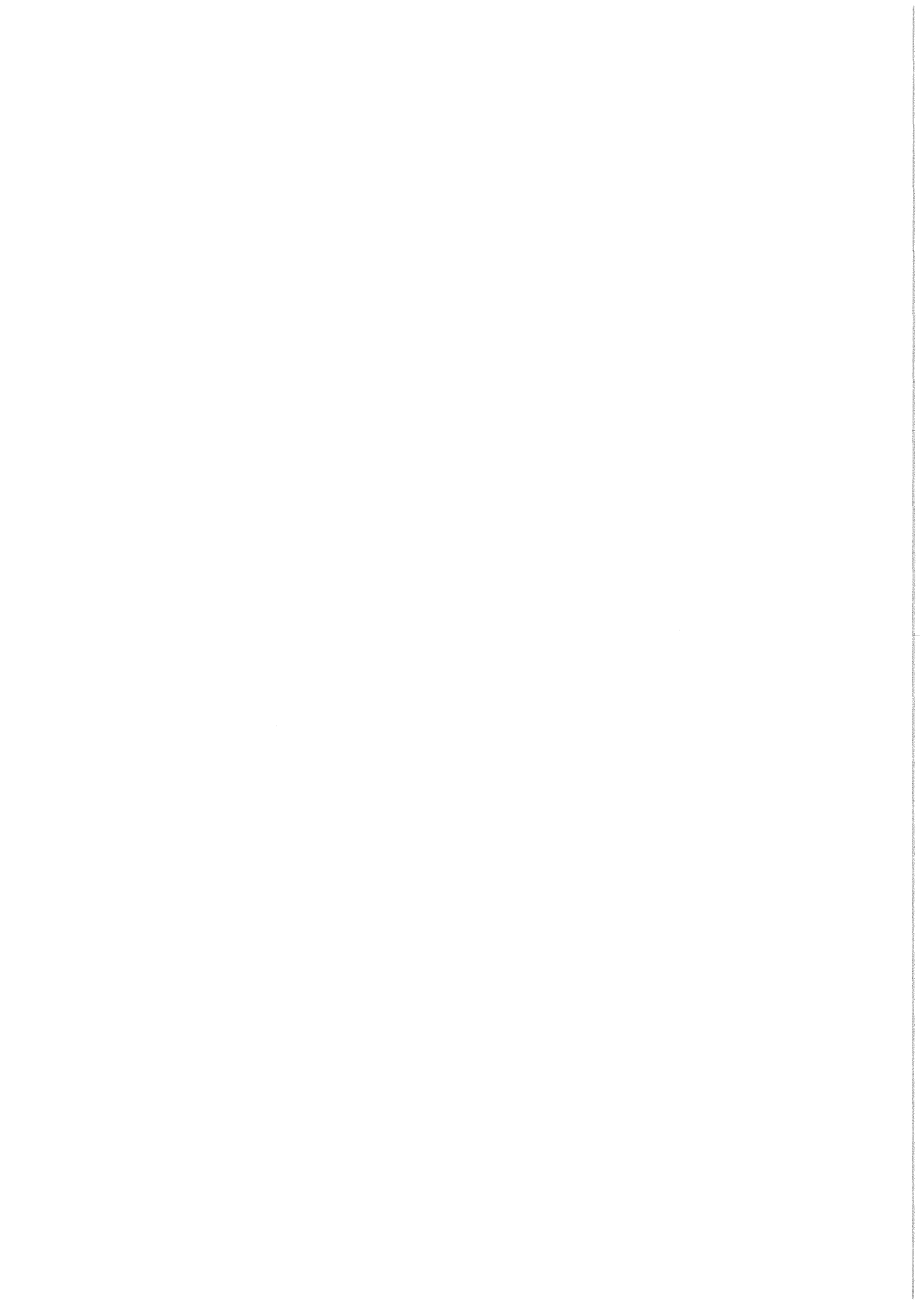
[Newell et Simon 72] Newell A. et Simon H.A., Human Problem Solving, Englewoods-Cliffs, N.J., Prentice Hall, 1972.

[Rasmussen J. et Lind, M. 82] A model of human decision making in complex systems and its use for design of system control strategies. Roskilde, Danemark, RISO, M-2349, 1982, cité par J.M. Hoc, Psychologie Cognitive de la Planification, PUG, Grenoble, 1987.

[Selfgrige 59] Selfridge O., Pandemonium: a paradigm for learning. In Symposium on the mechanization of thought processes. London: HM Stationery Office, 1959, cité par Lindsay Norman, Traitement de l'information et comportement humain, Edition Etudes Vivantes Ltée, Saint-Laurent, Québec, 1977.

[Simon 57] Simon H.A., Administrative Behavior, 2nd ed. Macmillan, 1957.

[Zacklad 87] Zacklad M., Contribution à une psychologie cognitive des activités de résolution de problème en groupe, Rapport de DEA Processus Cognitifs, Equipe de psychologie cognitive du traitement de l'information symbolique, Université de Paris VIII, 1987.



COOPERATION ENTRE AGENTS
ECONOMIQUES, PHYSIQUES OU COGNITIFS :
L'ALGORITHME "BUCKET-BRIGADE"

P. Collard J. Biondi
Université de Nice-Sophia Antipolis
CNRS-I3S
bat. 4, rue A. Einstein 06560 Valbonne, France
Tel : (33) 92-94-26-17 (33) 92-94-26-55
Fax : (33) 92-94-28-98
E-mail : pc@unice.fr jb@unice.fr

Résumé

Dans le cadre des systèmes adaptatifs basés sur les approches évolutionnistes, les *systèmes de classifieurs* [Holland 75] sont des systèmes qui apprennent incrémentalement par interaction avec leur environnement ; ils mettent en jeu des mécanismes d'évolution (opérateurs génétiques) et d'apprentissage (algorithme Bucket Brigade). Les interactions entre le système et son environnement prennent la forme d'une récompense et résulte d'une tentative de résolution du problème par le système. Les systèmes de classifieurs incorporent un *système de productions* dans lequel les productions sont activées en parallèle, un *algorithme d'attribution des récompenses* sur les différentes règles de production ("les classifieurs") et un *algorithme génétique* qui permet de découvrir de nouveaux classifieurs [Holland 75] [Goldberg 89]. Dans ce papier, nous nous intéressons à l'*algorithme Bucket Brigade* qui permet de résoudre le problème de l'attribution de récompenses et de blâmes [Huang 89]. Cet algorithme gère, pour chaque action utilisée dans une séquence, une valeur scalaire, la force, qui représente l'intérêt relatif de l'action pour le système ; c'est cette valeur qui est utilisée en phase de résolution pour déterminer, en cas de conflit, l'action qui sera déclenchée. L'intérêt de cet algorithme est de résoudre le problème d'attribution de récompense en utilisant uniquement des informations locales indépendantes du domaine [Belew 89]. Nous présentons trois interprétations de cet algorithme ; elles fournissent des points de vue complémentaires qui permettent de mieux comprendre le mécanisme mis en œuvre et qui mettent en évidence les comportements résultants. Ainsi, un système de classifieurs peut être vu comme un système adaptatif composé d'agents interagissant par l'intermédiaire de mécanismes, soit *économiques*, soit *physiques*, soit *psychologiques* et de mécanismes *évolutionnistes* ; ces interactions permettent d'accomplir des tâches complexes dans des environnements variés sans coordination centrale.

Mots clés : Apprentissage, Attribution de récompenses, Algorithme Bucket Brigade.

INTRODUCTION

Les systèmes mettant en œuvre les algorithmes génétiques sont des systèmes adaptatifs basés sur les principes évolutionnistes : dans une population, les individus les mieux adaptés à leur environnement ont une plus grande chance de survivre et de se reproduire. Ce processus de sélection naturelle s'applique à des chromosomes plutôt qu'aux êtres vivants et provoque la reproduction plus importante des chromosomes qui codent des structures utiles. Ce principe est mis en œuvre dans les algorithmes génétiques (A.G.) au moyen des opérateurs de reproduction, de croisement et de mutation qui vont être appliqués à des structures codées représentant les individus (solutions potentielles) d'une population étudiée. Le champ d'étude des A.G. couvre principalement les domaines de l'optimisation et de l'apprentissage. Dans les deux cas, il est nécessaire de disposer d'un codage des solutions potentielles (ceci afin de pouvoir utiliser les opérateurs génétiques) et d'une évaluation de chaque individu. La sélection, et donc la reproduction, est le lien entre ces individus (chromosomes ou classifieurs) et la performance de leurs structures codées (phénotype). Les systèmes de classifieurs mettent en œuvre, d'une part les opérateurs génétiques pour leur composante évolution et d'autre part un algorithme d'attribution des récompenses (la Bucket Brigade) pour leur composante apprentissage. L'adaptation est réalisée par une succession de phases d'apprentissage et d'évolution.

Lors de la phase d'apprentissage, le système donne une réponse à une sollicitation de son environnement et reçoit une évaluation de sa réponse. Cette évaluation est prise en compte par l'algorithme d'apprentissage qui va modifier les forces associées aux classifieurs. Dans l'algorithme Bucket Brigade, les étapes de résolution et d'apprentissage sont étroitement liées et correspondent à la phase d'exploitation des connaissances.

Lors de la phase d'évolution, les opérateurs génétiques sont appliqués sur les meilleurs classifieurs afin de créer de nouveaux individus. Cette phase permet l'exploration de nouvelles régions dans l'espace problème.

Nous nous intéressons dans ce papier aux mécanismes de l'algorithme de la Bucket Brigade qui permet l'apprentissage d'une séquence d'actions. Cet apprentissage est, par nature, une tâche complexe. La difficulté vient essentiellement du fait que l'on doit gérer des mesures locales, celles qui sont associées aux actions, à partir d'informations globales concernant les séquences. Cette difficulté est accentuée par le fait que, dans une séquence, les premières actions (les plus éloignées de l'issue et par conséquent les moins impliquées dans le comportement global de la séquence) sont celles qui doivent être apprises en premier. Nous commencerons par présenter formellement l'algorithme Bucket Brigade ; cette formalisation nous servira de référence. Nous donnerons, ensuite, trois interprétations dans des disciplines aussi variées que l'Economie, la Physique et la Psychologie. Les deux premières, outre leur intérêt pédagogique, permettent de suggérer des améliorations ou des variantes de l'algorithme. La dernière interprétation met en évidence les principes psychologiques sous-jacents et permet, ainsi, de caractériser les comportements résultants. Cette dernière démarche est précieuse pour valider le mécanisme mis en œuvre dans l'algorithme en tant que processus d'apprentissage "naturel".

1. PRESENTATION FORMELLE

L'algorithme "Bucket-Brigade" (B.B) est le plus souvent utilisé dans les Systèmes de Classifieurs (S.C.) introduit par J.H. Holland. Nous commençons donc par présenter succinctement ces systèmes ; puis, nous donnons une formulation formelle de la B.B sous la forme d'un algorithme en pseudo-Pascal.

1.1. LES SYSTEMES DE CLASSIFIEURS

Nous présentons la structure d'un système de classifieurs de façon à comprendre le mécanisme mis en œuvre dans l'algorithme B.B. Un S.C. est composé d'un ensemble de règles de productions (rule set), appelées classifieurs. Une règle comporte une prémisse et une conclusion exprimées sous la forme d'une chaîne de longueur constante composée de symboles (allèles) pris dans un ensemble fini (alphabet). Par exemple, on aura le classifieur $[100*1] \rightarrow [101*1]$ de longueur cinq défini sur l'alphabet $(0,1,*)$. A chaque étape, la liste des messages (message space) recueille les messages reçus de l'environnement ou des autres classifieurs ; un message a la même syntaxe que la prémisse ou le conséquent d'un classifieur. Le système est doté d'une interface composée de détecteurs et d'effecteurs qui permet de recevoir ou d'envoyer des messages à l'environnement. A chaque classifieur est associée une force ; initialement, toutes les forces ont la même valeur. En phase de résolution, les classifieurs qui "match" les messages de la liste des messages, entrent en compétition pour réaliser la ou les inférences ; la force des classifieurs est alors utilisée pour résoudre les conflits. Pour simplifier la présentation, nous supposons que la liste des messages contient un unique message et qu'un seul classifieur est déclenché à chaque étape. Cette simplification ne nuit pas à la généralité de l'exposé dans la mesure où nous nous intéressons, ici, uniquement au mécanisme d'attribution de récompenses. Une interaction est donc caractérisée par la séquence des classifieurs utilisés, ou celle des actions correspondantes.

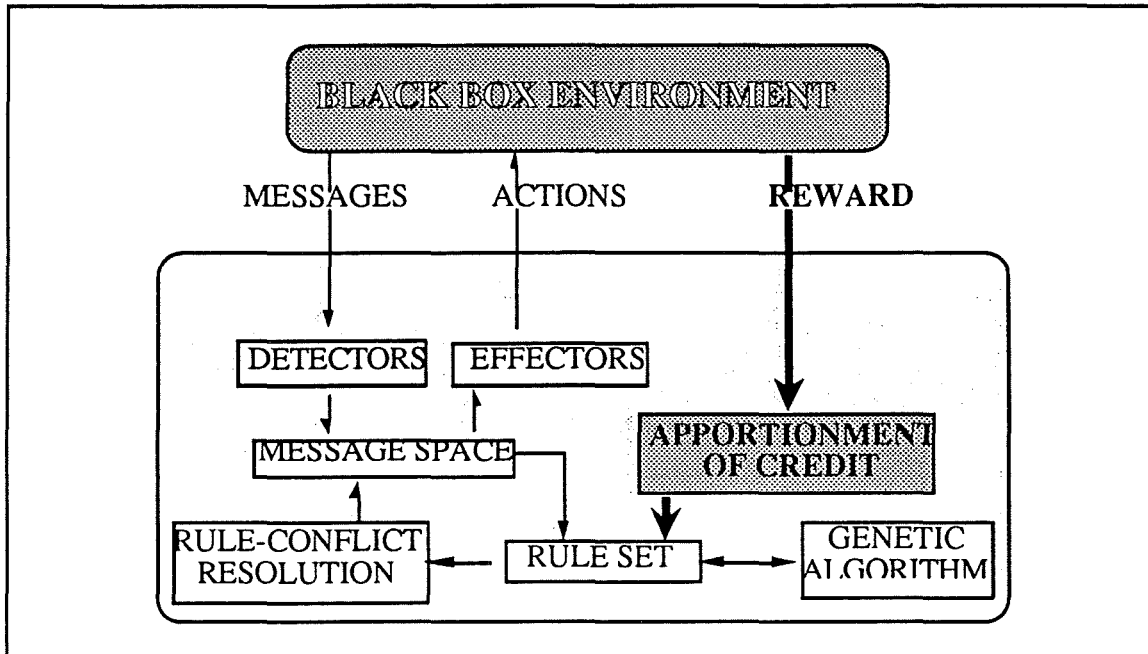


Figure 1 : Architecture d'un Système de Classifieurs.

1.2. L'ALGORITHME "BUCKET BRIGADE"

Le but de ce paragraphe est de donner une présentation formelle de l'algorithme B.B qui servira de référence par la suite. Nous commencerons par fixer les notations puis nous précisons comment l'activité du système est répartie entre les phases de résolution et d'apprentissage.

Nous noterons $\{A_t^1, \dots, A_t^i, \dots, A_t^n\}$ la séquence des actions exécutées lors de l'interaction (t) et $\{S_t^1, \dots, S_t^i, \dots, S_t^n\}$ la séquence des forces (Strength) correspondantes. L'algorithme B.B. fonctionne selon un cycle composé de deux phases, une première phase de résolution-exécution

(feedforward) et une seconde phase d'apprentissage (feedback). Lors de la première phase, les actions (ou étapes) sont exécutées séquentiellement de la première à la dernière ; lors de la phase d'apprentissage les forces associées aux actions sont mises à jour en fonction de la récompense (reward) externe fournie par l'environnement ; cette récompense est toujours propagée de la dernière action exécutée vers la première, on parle de rétro-propagation de la récompense. On peut distinguer deux types de système selon la répartition temporelle des phases de résolution et d'apprentissage. Dans le premier type, les deux phases sont indépendantes et consécutives. On commence par résoudre le problème en utilisant les valeurs courantes des forces puis on apprend en modifiant ces mêmes forces en fonction de la récompense externe. Dans le second type, les deux phases sont concomitantes ; à chaque étape de la résolution, c'est-à-dire après l'exécution d'une action, le système apprend en modifiant les forces : dans ce cas, la phase de rétro-propagation est implicite : elle se manifeste progressivement au cours des différentes interactions. L'algorithme B.B. est du second type.

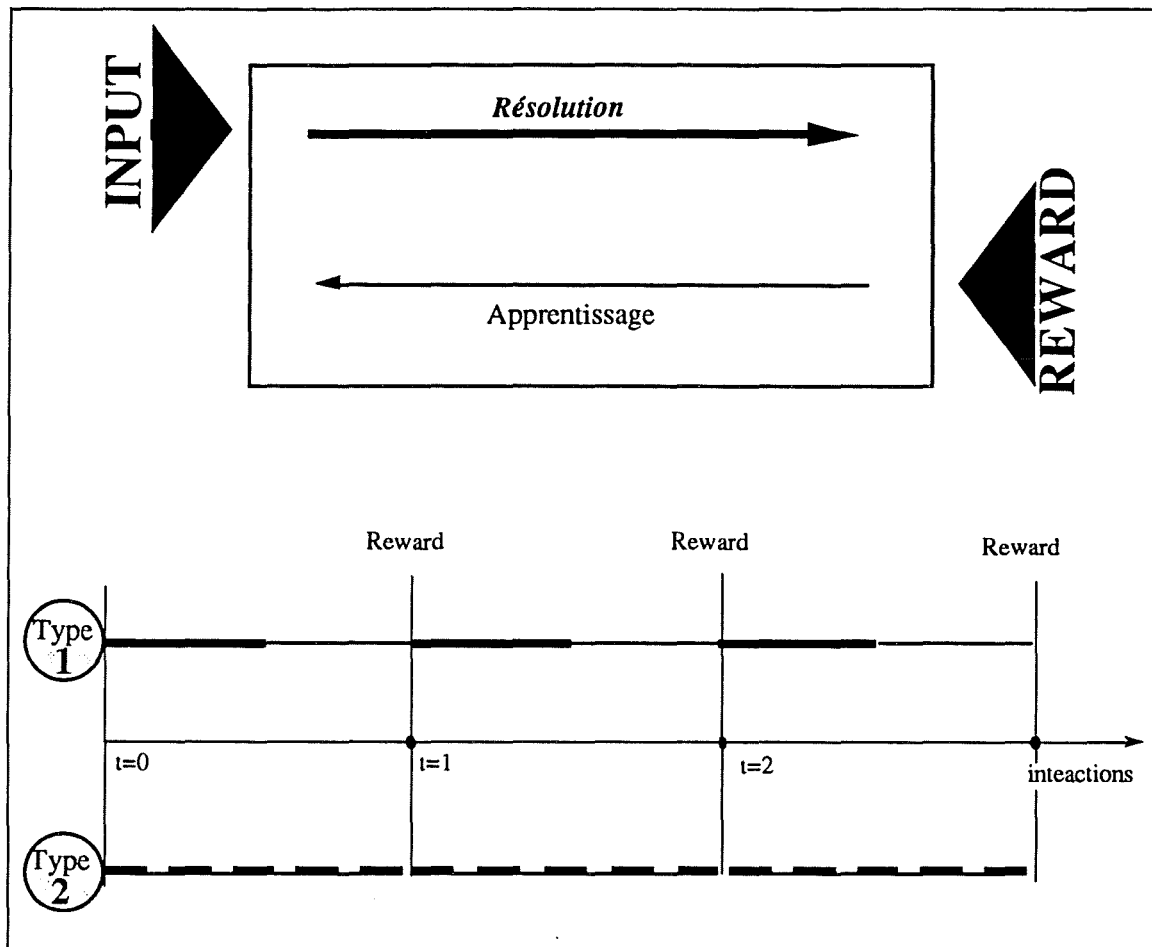


Figure 2 : Les deux types de mécanisme d'apprentissage par rétro-propagation.

Nous donnons maintenant une présentation formelle de l'algorithme B.B. sous la forme d'un algorithme en pseudo-Pascal.

```

procedure BucketBrigade

  procedure Actualiser(var x, y)
  const  $\alpha=0.2$ 
  begin
    x :=  $\alpha*x + (1-\alpha)*y$ 
  end

begin
  t := 0
  while not (FinApprentissage)
  begin
    act( $A_t^1$ )    {exécuter la première action de la séquence}
    i := 1
    while not (FinSequence)
    begin
      i := i+1
      act( $A_t^i$ )    {exécuter l'action courante de la séquence}
      actualiser( $S_t^{i-1}, S_t^i$ ) {mise à jour de la force de l'action précédente}
    end
    react(Reward)
    actualiser( $S_t^n, Reward$ ) {mise à jour de la force de la dernière action}
    t := t+1
  end
end

```

La procédure *act* exécute l'action passée en argument ; elle a pour effet de bord de modifier le contexte courant.

La procédure *react* renvoie la récompense (Reward).

La procédure *actualiser* réalise la mise à jour d'une force passée en premier argument en utilisant l'information passée en second argument. Cette dernière information est, dans le cas général, la force du successeur et dans le cas particulier de la dernière action la récompense (Reward). Cette mise à jour est réalisée par le calcul d'un barycentre qui fait intervenir un coefficient α , a priori constant, dont la valeur se situe en général entre 0,1 et 0,4. C'est la combinaison d'un processus cyclique : les interactions qui engendrent les séquences (représentées par la boucle englobante), et d'un processus qui est localement rétro-actif : la mise à jour de la force de l'action précédente (représenté par l'appel *actualiser*(S_t^{i-1}, S_t^i)), qui engendre au cours des interactions une rétroaction globale sur tous les éléments de la séquence.

2. INTERPRÉTATION ECONOMIQUE

Nous allons commencer par donner l'interprétation "classique" de l'algorithme B.B. Cette première interprétation utilise une métaphore économique [Golberg 89]. L'algorithme B.B. modélise le comportement d'une société économique : les actions sont considérées comme des agents économiques, la force d'une action représente le capital de l'agent. Initialement,

tous les agents possèdent le même capital. A chaque étape, les agents candidats, qui correspondent aux actions dont la partie condition "match" le contexte courant, participent à une vente aux enchères ; l'objet de la vente est le droit de participer à la séquence. Ces agents doivent émettre une offre (Bid) pour concourir. Le (ou les gagnants) sera celui qui aura proposé la meilleure offre ; il aura alors le droit d'agir pour modifier l'état courant du système. Le mécanisme de mise à jour des forces repose uniquement sur le fait que l'agent retenu pour agir doit payer son offre à l'agent qui a gagné l'étape précédente ; ce qui est justifié par le fait que ce dernier est directement responsable de la situation qui a permis à l'agent retenu de concourir. Un agent qui intervient dans une séquence verra donc son capital diminué de son offre et augmenté de l'offre de son successeur. En particulier, le dernier agent qui intervient dans une séquence est rétribué directement par l'environnement, le salaire correspond à la récompense. Le premier agent de la séquence, quant à lui, paye directement son offre à l'environnement qui a sollicité le système. Une séquence d'actions peut donc être interprétée comme une chaîne d'intermédiaires entre celui qui paie la récompense et celui qui reçoit l'offre du premier agent. On obtient ainsi la relation de mise à jour :

$$S_{t+1}^i = S_t^i - Bid_t^i + Bid_t^{i+1}$$

Si l'on suppose que l'offre est proportionnelle au capital :

$$Bid = \beta * S$$

où β est un coefficient constant compris en général entre 0.1 et 0.4

on obtient finalement :

$$S_{t+1}^i = (1 - \beta) * S_t^i + \beta * S_t^{i+1}$$

Enfin, si l'on pose $\alpha = (1 - \beta)$ on constate que l'on retrouve la formulation formelle présentée au §1.

Cette métaphore a le mérite de proposer une interprétation cohérente qui permet de mieux comprendre le mécanisme de la B.B. Cependant, elle impose des limites. Par exemple, les récompenses et les forces doivent être, par nature, positives. Un blâme sera donc représenté, au pire, par une récompense nulle ; ainsi, lors des transactions, le capital d'un agent ne pourra que décroître. Les agents qui ne sont jamais utilisés se verront donc avantagés par rapport aux agents actifs. Cette situation incitera les agents à rester inactifs ! Pour résoudre ce problème, il a été proposé d'appliquer aux agents un système de taxes. Lors de chaque vente aux enchères, tous les agents doivent payer une taxe proportionnelle à leur capital (c'est l'impôt sur le capital).

L'intérêt d'une interprétation métaphorique n'est pas seulement d'ordre pédagogique : ce peut être, également, une source d'inspiration. Le domaine de l'économie est suffisamment riche pour nous suggérer des variantes ou des améliorations de l'algorithme B.B. Par exemple, on pourrait réinterpréter l'impôt sur les transactions ou sur les revenus, introduire des tranches et des seuils d'imposition ou encore introduire le hasard en suggérant que des agents puissent jouer au loto. On pourrait également imaginer qu'un agent loue les services d'un ou plusieurs autres agents pour réaliser le travail qui lui a été demandé ; il y aurait alors sous-traitance. Cette dernière idée suggère une décomposition d'une séquence en sous-séquences. Cette solution a été développée par S.W. Wilson ; elle permet, en particulier, de résoudre les problèmes posés par les longues séquences [Wilson 87].

3. INTERPRETATION PHYSIQUE

Dans cette deuxième interprétation, chaque action est assimilée à un récipient, disons un seau, qui contient un liquide, disons de l'eau. L'algorithme B.B tient son appellation de cette interprétation! La force d'une action représente le volume d'eau contenu dans le seau. Initialement, tous les seaux contiennent la même quantité d'eau ; en particulier, ils peuvent être vides. A chaque étape, parmi tous les seaux disponibles, on choisit celui qui contient le plus grand volume d'eau. Le mécanisme de mise à jour des forces repose ici sur le fait que lorsqu'un seau est choisi, on doit verser une partie de son contenu dans le seau sélectionné à l'étape précédente. Le dernier seau choisi dans une séquence reçoit de l'environnement une quantité d'eau qui correspond à la récompense alors que le premier se déverse dans un "puit perdu". Un des avantages de cette interprétation est qu'elle permet de visualiser, bien qu'elle soit implicite, la phase de rétro-propagation de la récompense par une cascade de seaux qui se déversent les uns dans les autres. Dans cette interprétation, la récompense et les forces représentent un volume d'eau ; par conséquent, nous sommes, comme pour l'interprétation économique, amenés à poser que leur valeur est positive. Pour des raisons analogues à celles que nous avons évoquées précédemment, on doit supposer, qu'à chaque transvasement, tous les seaux, même ceux qui ne participent pas à la compétition, perdent un peu de leur contenu. Il suffit d'admettre que tous les seaux fuient !

L'intérêt de cette métaphore, outre qu'elle met en évidence le mécanisme de rétro-propagation, est qu'elle permet de visualiser l'origine de l'eau dans les seaux ainsi que l'ordre dans lequel les seaux ont été remplis à partir des différentes sources que représentent les récompenses. Autrement dit, elle permet de visualiser la structure de chaque force ; c'est à dire l'organisation de ses composants selon leur origine. Sur la figure 3, nous avons représenté l'évolution du contenu des seaux au cours des différentes interactions. Nous avons "colorier" l'eau d'une couleur différente pour chaque interaction et considérer un coefficient de proportionnalité α égale à $\frac{1}{2}$. Ceci nous permet de caractériser *quantitativement* et *qualitativement* le contenu de chaque seau ; en précisant son volume et les différentes origines de l'eau. Par exemple, on peut observer dans le seau du milieu en bas que l'eau issue de la première récompense est interclassée avec celle issue de la deuxième. On remarque également que le dernier seau présente une invariance d'échelle : il contient, après quatre interactions, la quantité $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8}$. Cette organisation suggère une structure fractale !...

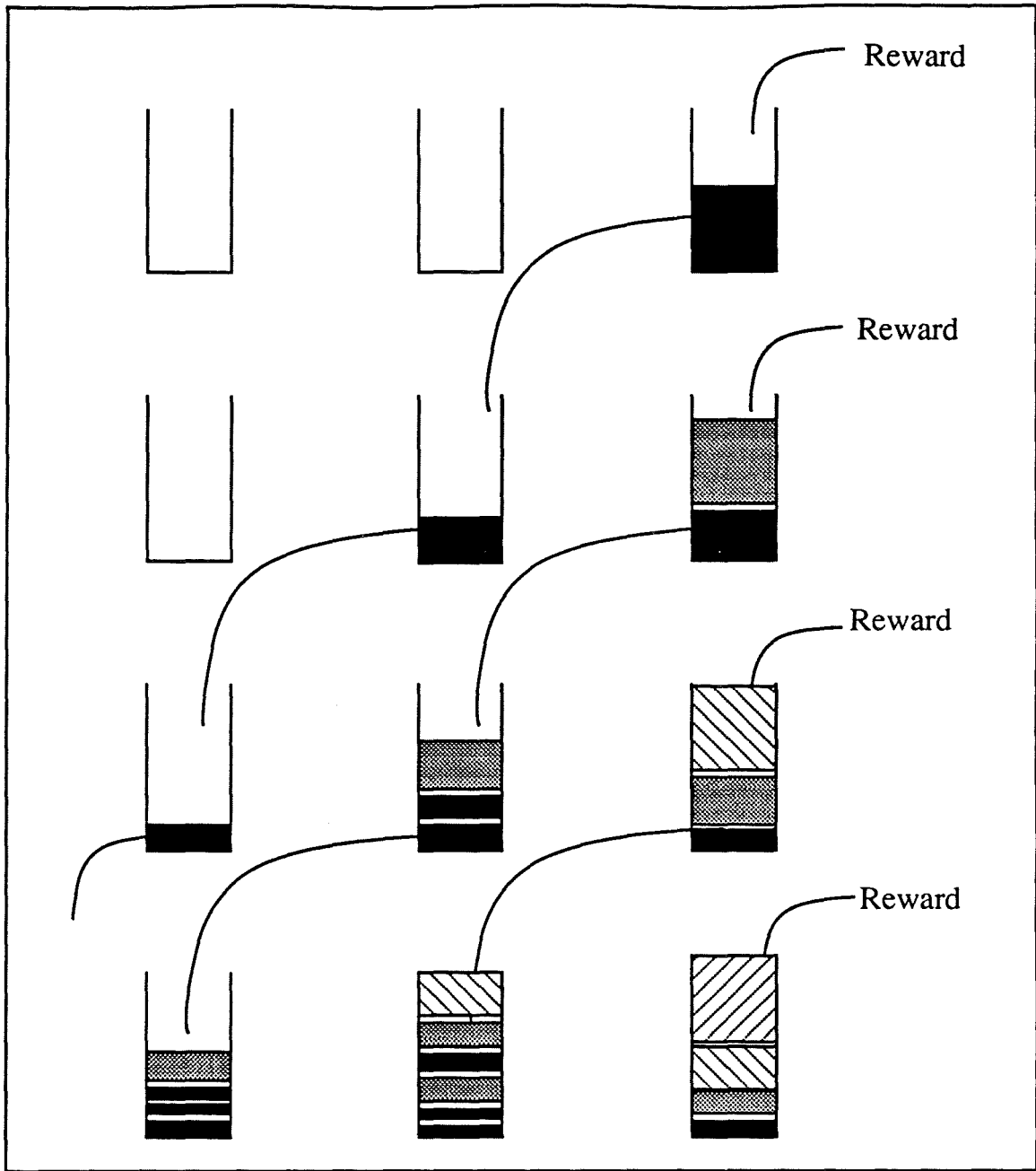


Figure 3 : interprétation Physique de l'algorithme Bucket Brigade

4. INTERPRETATION PSYCHOLOGIQUE

Nous allons dans cette partie énoncer un certain nombre de principes qui relèvent de la psychologie comportementale et montrer que ces principes sont mis en œuvre dans l'algorithme B.B. Ces principes indiquent quels sont les composants du système (les agents cognitifs) qui doivent supporter les conséquences d'un comportement. Ils permettent, également, de donner un sens précis à la notion de conséquence. Enfin, ils indiquent comment répercuter les conséquences d'un comportement global dans le cas où plusieurs agents sont responsables de l'issue [Biondi 90] [Collard 90].

4.1 LE PRINCIPE DE SUFFISANCE TEMPORELLE ET LA LOI DE L'EFFET

Commençons par citer D. Hume, pour nous rappeler une évidence : "Il suffit de constater qu'aucune relation n'impressionne davantage l'imagination et n'enchaîne mieux les idées que la relation de cause à effet entre deux objets".

Le *principe de suffisance temporelle* postule qu'à partir d'une relation de succession temporelle d'événements observés, on établit une relation de dépendance implicative entre les événements (principe élémentaire de causalité). D'un point de vue externe au système, ce principe intervient entre le stimulus et la réponse, ou entre la réponse et le renforcement. Par exemple, ce principe conduit à la création de relations implicatives dans le modèle proposé par A. Grumbach ; c'est ce qu'il nomme le principe d'induction séquentielle-implicative [Grumbach 87]. L'application de ce principe est nécessaire pour l'établissement d'un lien associatif entre deux éléments, lien qui pourra être renforcé selon la loi de l'effet.

La *loi de l'effet* a été énoncée pour la première fois par Thorndike sous la forme suivante : «une connexion est renforcée ou affaiblie par l'effet de ses conséquences» [Ehrlich 75]. Il faut noter qu'après 1930, Thorndike propose une loi de l'effet tronquée dans laquelle seule la satisfaction influe sur une connexion. Notons que la loi de l'effet nécessite la mise en œuvre du principe de suffisance temporelle.

Plus récemment, A. Dickinson, dans une argumentation sur le caractère général des mécanismes d'apprentissage chez l'animal, déclare que « ... les relations de causalité ont des propriétés universelles qui dépassent chaque cas particulier ; par exemple les effets ne se produisent jamais sans cause, et ils ne se produisent jamais avant cette cause. Ces sortes de contraintes peuvent nous paraître évidentes et même triviales, car elles font partie des propositions fondamentales par lesquelles nous donnons sens au monde, et par lesquelles nous organisons les relations entre événements dans l'espace et le temps. Leur grande familiarité ne doit pas cependant nous masquer le fait qu'elles constituent des propriétés fondamentales partagées par toutes les relations causales.» [Dickinson 80]

MISE EN ŒUVRE DANS L'ALGORITHME BUCKET BRIGADE

La prise en compte de ces principes permet de préciser la notion de force associée à une action. Dans l'interprétation psychologique nous considérons chaque action comme un agent cognitif. Nous étudions une action sous son aspect «causale». Supposons que l'agent A représente la relation causale (e_1, e_2) où e_1 représente le contexte d'activation de l'agent et e_2 le contexte résultant de son activité. Dans cette interprétation, la force S associée à A indique dans quelle mesure la présence de e_1 est suffisante pour annoncer e_2 ; on dira que S est une mesure de *suffisance*.

- une valeur de S positive mesure la confiance que le système appreni a dans la relation causale (e_1, e_2)
- une valeur négative de S a exactement la même signification mais pour la relation ($e_1, \bar{C}e_2$) ou $\bar{C}e_2$ représente le complémentaire de l'événement e_2 (occurrence du contexte e_2) dans l'ensemble des événements possibles.

On peut justifier cette interprétation des valeurs négatives de la force d'un point de vue logique en rappelant, d'une part, que la mesure S ne varie que si l'événement e_1 est réalisé (occurrence du contexte e_1) et en remarquant, d'autre part, que si e_1 est vrai alors $\text{non}(e_1 \Rightarrow e_2)$ est équivalent à $(e_1 \Rightarrow \text{non}(e_2))$.

Nous pouvons, maintenant, énoncer un principe d'apprentissage, ou plutôt ce que l'on peut considérer comme un prototype de principe :

"Si un agent A est impliqué dans une séquence alors faire varier sa force S d'une valeur ΔS "

Ce principe, mis en œuvre dans l'algorithme B.B., repose sur le postulat de l'existence d'une relation de cause à effet entre une réponse et les agents activés pour fournir cette réponse.

Pour obtenir un principe applicable, il faut préciser la fonction qui permet de calculer la variation ΔS . En toute généralité, on peut penser que cette fonction dépend de deux variables la récompense et la force elle-même. Essayons de justifier, a priori, cette double dépendance. La dépendance vis-à-vis de la récompense est évidente : la force doit augmenter si le système fournit une «bonne» réponse et diminuer dans le cas contraire. La dépendance de ΔS vis-à-vis de l'agent amène à des interprétations moins évidentes ; cette dépendance peut être relative à chacun de ses attributs : âge, nombre d'activations, date de la dernière activation, force, etc...

4.2 L'EFFET DE SURPRISE/ACCOUTUMANCE

Concernant l'apprentissage chez l'animal d'une association de cause à effet, le psychologue R.A. Rescorla conclut «... l'ensemble de ces résultats conforte l'idée que la quantité d'apprentissage d'un animal à l'égard de la relation entre un stimulus neutre et un choc dépend du degré de surprise créée par le choc» [Rescorla 71].

Suivant la même idée, nous établissons que la variation de la quantité d'apprentissage est fonction non seulement de la récompense mais également du *niveau cognitif* du système. Plus précisément, on peut énoncer les principes suivants :

- *Principe de surprise* : l'effet d'une «émotion» est d'autant plus important qu'elle est en opposition avec celle attendue.
- *Principe d'accoutumance* : l'effet d'une «émotion» est d'autant moins important que celle-ci était attendue.

Ces deux principes établissent donc la variation quantitative de la connexion comme étant fonction non seulement des conséquences immédiates de ses actions mais également des exercices passés et donc des effets attendus pour le comportement présent [Rescorla 72].

MISE EN ŒUVRE DANS L'ALGORITHME BUCKET BRIGADE

Ces deux principes sont mises en œuvre dans l'algorithme B.B. En effet, si nous considérons la dernière action de la séquence, sa force est mise à jour selon la formule :

$$S_{t+1}^n = \alpha * S_t^n + (1 - \alpha) * \text{Reward}$$

la variation de cette force

$$\Delta S = S_{t+1}^n - S_t^n = (1 - \alpha) * (\text{Reward} - S_t^n)$$

est proportionnelle à la différence entre la récompense et la force qui est à l'origine de cette récompense ; il existe une relation directe de cause à effet entre la valeur de la force et la valeur de la récompense. La variation de la force est donc d'autant plus importante que la récompense infirme cette force ; de même, elle est d'autant plus faible que la récompense la confirme.

Si l'on considère maintenant une action quelconque de la séquence, sa force est mise à jour selon la formule :

$$S_{t+1}^i = \alpha * S_t^i + (1 - \alpha) * S_t^{i+1}$$

la variation

$$\Delta S = S_{t+1}^i - S_t^i = (1 - \alpha) * (S_t^{i+1} - S_t^i)$$

est proportionnelle à la différence entre la force du successeur dans la séquence et la valeur courante de la force. La force du successeur est donc considérée ici comme la valeur attendue ; c'est-à-dire la valeur qui sert de référence pour déterminer si il y a surprise ou accoutumance. Pour justifier ce point il faut rappeler que lors de l'apprentissage toutes les forces de la séquence solution convergent vers une même limite égale à la récompense ; la force des actions proches de l'issue convergeant plus rapidement que celles des premières actions de la séquence. Ainsi, dans une séquence solution on peut considérer que la force d'une action est moins "bonne" que celle de son successeur dans le sens où elle est plus éloignée de la valeur limite ; il paraît donc raisonnable de prendre cette dernière comme référence. La variation de la force est donc d'autant plus importante que la force du successeur infirme la force ; de même, elle est d'autant plus faible que cette valeur confirme la force. Ces deux principes, surprise et accoutumance, confèrent aux S.C. une grande *versatilité* : un tel système est capable de se remettre en cause dans la mesure où il prend «conscience» de son inadéquation avec l'environnement.

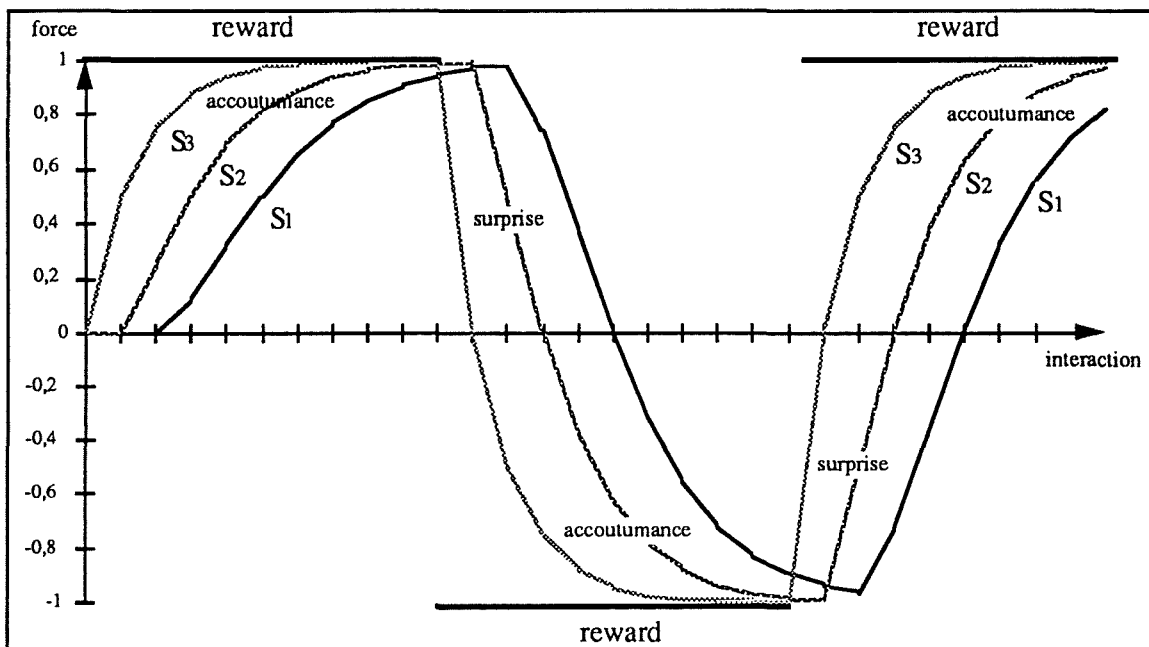


Figure 4 : Effet de surprise et d'accoutumance sur une séquence de trois actions.

On peut observer sur la figure 4 les effets de surprise et d'accoutumance. Les trois courbes représentent l'évolution de la force des trois actions d'une séquence à laquelle on applique des séries de récompenses identiques. On commence par appliquer une série de 10 récompenses (Reward = +1) suivie de 10 blâmes (Reward = -1) et enfin, de nouveau, 10 récompenses. On peut remarquer que chaque changement d'attitude de la part de l'environnement déclenche des effets de surprise, qui correspondent à une phase de remise en cause profonde pendant laquelle les forces évoluent rapidement. Cette phase est suivie d'une période de confirmation pendant laquelle le système s'accoutume au comportement constant de l'environnement, les forces évoluent alors lentement.

4.3 LA LOI DE PROPAGATION DE L'EFFET

Cette loi consiste à propager la valeur qualitative d'une émotion aux événements qui lui sont liés [Grumbach 86] ; elle s'appuie sur le principe de transitivité du renforcement ou renforcement secondaire chez Hull¹. Nous allons élargir ce principe à une connexion indirecte pour laquelle il existe plus d'une action à exécuter pour fournir une réponse à un stimulus.

LE PRINCIPE DE RELATIVITE DE L'IMPLICATION

Jusqu'à présent, nous avons traité toutes les agents qui participent à l'élaboration d'une réponse sur un même plan ; en fait, dans la pratique, on constate que chaque agent est plus ou moins impliquée dans l'issue. Nous devons, par conséquent, tenir compte du degré d'implication des agents. Cependant, dans le cas général, il est impossible de définir une «mesure d'implication» qui nous permettrait de déterminer la récompense perçue par chacun des éléments responsables de l'issue. Pour fixer les idées, nous allons supposer qu'une telle mesure existe, soit \mathcal{I} ; elle indique le degré d'implication d'un élément vis-à-vis de l'issue. On peut alors exprimer qualitativement le «principe de relativité de l'implication» en disant que les choix corrects (resp incorrects) sont d'autant plus vite appris (resp mieux évités) qu'ils sont «proches», au sens de la mesure \mathcal{I} , de l'issue. Ce que l'on peut exprimer formellement par :

$$\text{si } \mathcal{I}(x) < \mathcal{I}(x') \text{ alors } \Pi(\text{Reward}, x) < \Pi(\text{Reward}, x')$$

où $\Pi(\text{Reward}, x)$ représente l'influence de la récompense sur l'élément x . G. George introduit, à ce sujet, la notion de «gradient de but» [George 83] : «lorsque la conduite instrumentale comporte une succession de choix comme le parcours d'un labyrinthe, les choix corrects sont d'autant plus vite appris qu'ils sont plus proches de la sortie». Dans notre contexte nous pouvons définir la fonction \mathcal{I} comme une fonction croissante du rang de l'agent dans la séquence. Par exemple, on pourra poser $\mathcal{I}(A_r) = r$ où r représente le rang.

LE PRINCIPE DE LA CAUSE TRIVIALE

Dans de nombreux cas, étant donné un ensemble de causes relatives à une situation, il existe une cause qui est directement impliquée dans l'issue ; par exemple, dans le cas d'un jeu, la dernière action exécutée (qui correspond au dernier coup joué) est directement responsable du succès ou de l'échec de la partie. L'idée exprimée par le «principe de la cause triviale» est

¹ Chez Hull ce principe n'est pas associé à l'idée de propagation mais à la notion de conditionnement de second ordre, activation «médiata» : «une connexion Stimulus-Réponse peut être établie et renforcée non par la réduction directe d'un besoin mais par la présence d'un stimulus qui a été associé plusieurs fois à une réponse conduisant à la réduction d'un besoin» [EHR 75].

qu'une telle action (ou groupe d'actions), si elle existe, doit subir la sanction dans son intégralité. Soit A une telle action, on peut exprimer ce principe par l'égalité :

$$\Pi(\text{Reward}, A) = \text{Reward}$$

Dans le cas d'une séquence de n actions on aura donc $\Pi(\text{Reward}, A_n) = \text{Reward}$; en d'autres termes, la dernière action subit intégralement la récompense ou le blâme.

MISE EN ŒUVRE DANS L'ALGORITHME BUCKET BRIGADE

La rétro-propagation implicite de la récompense réalisée par l'algorithme B.B. met en œuvre la loi de propagation de l'effet ainsi que les principes de relativité de l'implication et de la cause triviale.

- la loi de propagation de l'effet est trivialement appliquée puisque la récompense (ie. la valeur quantitative de l'émotion) est propagée sur les agents de la séquence (i.e. les événements qui sont liés à l'émotion).

- Le principe de relativité de l'implication est appliqué : l'influence d'une récompense sur un agent est d'autant plus forte que cet agent est proche de l'issue ; en particulier, le dernier agent de la séquence subit intégralement la sanction.

On peut constater sur la figure 4 que l'effet de surprise/acoutumance porte sur la force S_3 du dernier agent de la séquence et que cet effet est, dans une moindre mesure, rétro-propagé sur les forces S_2 et S_1 .

4.4 LA LOI DE L'EXERCICE

Nous allons maintenant énoncer une dernière loi : la *loi de l'exercice* ; elle établit que «les connexions entre la situation et la réponse sont renforcées par l'exercice et affaiblies lorsque l'exercice (pratiquement, la répétition) est arrêté» [Ehrlich 75]. Nous nous intéressons, ici, uniquement au renforcement d'une connexion laissant de côté le dernier aspect de cette loi concernant le phénomène d'oubli. Alors que les lois de l'effet et de surprise/acoutumance précisent l'amplitude de la variation de la force d'une connexion, la loi de l'exercice établit l'effet cumulatif de l'exercice sur cette force. Il faut noter que la répétition ne favorise l'apprentissage que dans le cas où il y a également intervention de l'effet ; les deux lois sont indissociables pour la réalisation de l'apprentissage. Par nature, le mécanisme mis en œuvre dans l'algorithme B.B. ne pourra donner une valeur significative à une force, que si on l'applique de façon répétitive sur cette force : l'apprentissage dépend du nombre d'applications des principes.

L'interprétation psychologique de l'algorithme Bucket Brigade conforte l'hypothèse qu'il existe des Principes d'Apprentissage Généraux [Biondi 90] [Collard 90] indépendants du domaine de connaissance qui permettent l'apprentissage à un niveau sub-symbolique.

CONCLUSION

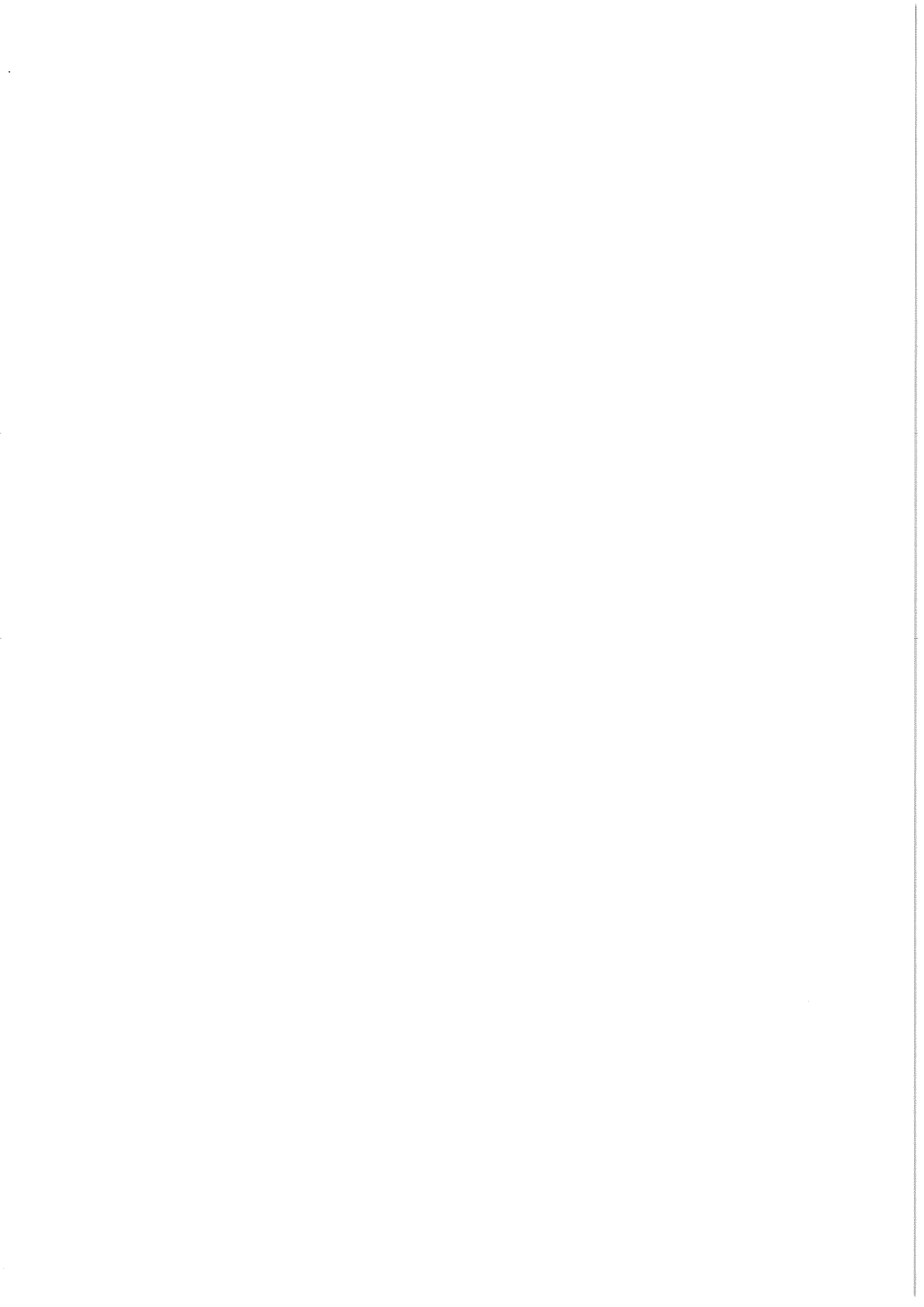
Nous venons de présenter trois interprétations de l'algorithme Bucket Brigade dans des domaines différents. Notre objectif était de mettre en évidence le mécanisme de rétro-propagation d'une récompense dans un contexte multi-agents. Nous avons vu comment un processus local, dans le temps et dans l'espace, a permis la propagation globale d'une récompense et donc l'adaptation du système à son environnement par auto-organisation de ses structures codées celles-ci pouvant être interprétées comme des agents économiques, physiques ou cognitifs. Ce processus est temporellement local au sens où il utilise des informations qui ont été mises à jour récemment ; de même, il est spatialement local au sens où il met uniquement en jeu les forces associées à des agents voisins [Dorigo 91] sans qu'il soit nécessaire de disposer d'une coordination centrale. Les métaphores économique et physique ont fourni un support

pédagogique à l'exposé. Nous avons vu également qu'elles pouvaient être génératrices d'idées ; nous avons fait quelques propositions dans ce sens. Enfin, l'interprétation psychologique a permis d'établir le lien entre un mécanisme formel et ce qui pourrait être considéré comme des mécanismes naturels de l'apprentissage.

BIBLIOGRAPHIE

- [Belew 89] R. K. Belew
Back-propagation for the Classifier System.
In proceeding of an International Conference on Genetic Algorithms and their Applications 1989.
- [Biondi 90] J. Biondi
MAGE : un modèle d'apprentissage général.
Thèse d'Etat, Université de Nice-Sophia Antipolis 1990.
- [Biondi 92] J. Biondi & P. Collard
Attribution de récompenses dans un système de classifieurs, Apprentissage versus robustesse.
Journées Intelligence Artificielle et Cognition : Apprentissage, évolution, adaptation
Rochebrune, France, Janvier 1992.
- [Collard 90] P. Collard
L'apprentissage discriminant dans MAGE : un modèle d'apprentissage général.
Thèse de doctorat, Mention Informatique, Université de Nice-Sophia Antipolis 1990.
- [Collard 92] P. Collard
Back-specialization in a Classifier System
The Fourth North Carolina Symposium on Artificial Intelligence and Advanced Computing
Technology, Raleigh, U.S.A., March 1992.
- [Dorigo 91] M. Dorigo
Message-Based Bucket Brigade: An Algorithm for the Apportionement of Credit Problem
EWSL-91, Porto, Portugal, 1991.
- [Dickinson 80] A. Dickinson
Contemporary Animal Learning Theory.
Cambridge University Press, Londres, 1980.
- [Ehrlich 75] S. Ehrlich
Apprentissage et mémoire chez l'homme.
Puf 1975.
- [George 83] C. George
Apprendre par l'action.
Ed. PUF, 1983.
- [Goldberg 89] D.E. Goldberg
Genetic Algorithms in Search, Optimization & Machine Learning. Ed. Addison-Wesley, 1989.
- [Grefenstette 88] J.J. Grefenstette
Credit assignment in rule discovery system based on genetic algorithms.
Machine Learning, 3 (2/3), (pp 225-245).

- [Grumbach 86] A. Grumbach & A. Nguyen Xuan
Une approche cognitive d'apprentissage humain par l'action. Actes 2ème JFA, Orsay 1986.
- [Holland 75] J.H. Holland
Adaptation in natural and artificial systems.
Ann Arbor : The University of Michigan Press, 1975.
- [Holland 85] J.H. Holland
Properties of the bucket-brigade algorithm.
In proceeding of an International Conference on Genetic Algorithms and their Applications
Pittsburgh. Carnegie-Mellon University, July 1985.
- [Holland 86] J.H. Holland
Escaping brittleness : the possibilities of general-purpose learning algorithms applied to parallel
rule-based systems.
In R.S. Michalski, J.G. Carbonell & T.M. Mitchell (Eds) Machine Learning, an artificial intelligence
approach. Volume II. Los Altos, California : Morgan Kaufmann.
- [Huang 89] D. Huang
Credit-Apportionment in Rule-based Systems : Problem Analysis and Algorithm Synthesis,
PhD thesis, University of Michigan, 1989.
- [Rescorla 71] R.A. Rescorla
Variations in the effectiveness of reinforcement and nonreinforcement following prior
inhibitory conditioning. Learning and Motivation,2,113-23.
- [Rescorla 72] R.A. Rescorla & A.R. Wagner
Variations in the effectiveness of reinforcement and nonreinforcement. In Classical
Conditioning II : Current Research and Theory
Ed. A.H. Black and W.F. Prokasy, pp64-99, New York : Appleton-Century-Crofts.
- [Riolo 87] R.L. Riolo
Bucket Brigade Performance.
In proceeding of 2nd International Conference on Genetic Algorithms and their Applications
Hillsdale, New Jersey : Lawrence Erlbaum Assoc, 1987.
- [Sutton 86] R.S. Sutton
Learning to predict by methods of temporal difference. Machine Learning, 1988.
- [Wilson 87] S.W. Wilson
Hierarchical credit allocation in a classifier system.
Proceedings of the Tenth International Joint Conference on Artificial Intelligence
Los Altos, CA : Morgan Kaufmann.
- [Wilson 88] S.W. Wilson
Bid competition and specificity reconsidered. Complex Systems 2 (6), (pp 705, 723)



REFLEXION METHODOLOGIQUE SUR LES SYSTEMES D'INFORMATION COOPERATIFS DANS LES ORGANISATIONS

A.M.ALQUIER-BLANC, Université TOULOUSE I, place Anatole
France, 31042 Toulouse Cedex

RESUME

L'intelligence collective ou processus cognitif d'une Organisation est une émergence de processus cognitifs individuels. Il s'agit d'un processus constructiviste. Les processus cognitifs individuels sont liés à la spécialisation des individus et des unités organisationnelles, et aux langages spécialisés qu'ils développent. Ces spécialisations forment l'Organisation avec des ruptures cognitives, linguistiques et conceptuelles.

Nous appellerons Système d'Information Coopératif¹ la négociation cognitive, sémantique et conceptuelle entre ces langages spécialisés. Pour concevoir des Systèmes d'Information Coopératifs (architecture client/serveur,...), nous modéliserons ces ruptures à travers la Théorie des Catastrophes, ce qui nous permettra d'en donner des formes canoniques catastrophiques.

Les concepts s'appliquent également à une méthode d'audit et schéma directeur d'informatisation.

REFLEXION METHODOLOGIQUE SUR LES SYSTEMES D'INFORMATION COOPERATIFS DANS LES ORGANISATIONS

Hypothèses préalables

- Une Organisation est un système auto-organisé (concepts d'auto-information, et d'infor-gétique [Le Moigne, 1])
- Constructivisme du Système d'Information par rapport à l'organisation ("l'information in-forme l'organisation qui la forme" [Le Moigne, 1]), hypothèse qui a deux sens:
 - représentation de l'un par l'autre: homéomorphisme entre Système d'Information et Organisation.
 - constructivisme de l'un par l'autre
- Homéomorphisme des ruptures dans les actes de langage et dans la structure d'une Organisation, ruptures qu'une analyse morphologique nous permettra d'étudier
- Le processus cognitif de l'Organisation (PCO) est une négociation sémantique entre les langages spécialisés des unités organisationnelles, négociation qui est la base de ce que nous avons appelé Système d'Information Coopératif.
- La négociation sémantique permet l'émergence de connaissances globales (connaissances collectives).

Analyse morphologique des systèmes complexes par la boîte noire

Une analyse morphologique d'un système complexe de grande taille nécessite une approche par la boîte noire. Pour cela, il nous faut une théorie du comportement d'un système:

- bâtie sur des variables externes en nombre faible et faciles à caractériser, pour une identification des composants
- utilisant des variables externes interprétables ou traductibles dans les méthodes causales, pour permettre le passage de la boîte noire (global et systémique c'est à dire avec capacité d'auto-poïèse) à la boîte blanche (local et causal)
- liée si possible à des types de structures internes, pour une reproductibilité du diagnostic
- bâtie sur une typologie structurale caractérisée de façon globale, pour une identification des composants et de leur comportement.

La théorie des catastrophes (TC) permet une certaine compréhension des processus morphologiques sans avoir recours aux propriétés spéciales au substrat des formes, ou à la nature des forces agissantes. La boîte noire est décrite comme si ses mécanismes étaient régis par une "volonté interne", décrite comme un potentiel.

Cette fonction potentiel permet de localiser un état du système et un point à l'intérieur du système. Dans notre cas, nous supposerons une fonction potentiel organisationnelle [Lewin, 59]. Elle sera interprétée ici par des grandeurs linguistiques, permettant le passage du niveau global (boîte noire) au niveau local (boîte blanche) et inversement.

La TC remplace donc l'hypothèse du déterminisme* par la propriété empiriquement vérifiable de "stabilité structurelle* ". La

TC décrit un ensemble inévitable de formes et comment naissent ces formes.

la boîte noire définie à partir de la Théorie des Catastrophes

La théorie des catastrophes développe une analyse topologique de la stabilité des formes:

- Tout objet(système) peut être représenté par un attracteur C d'un système dynamique dans un espace M de variables internes.
- Un tel objet n'est stable que si l'attracteur correspondant est structurellement stable.
- Un attracteur A est défini par son domaine d'existence (ensemble des points réguliers), et son bassin (ensemble des trajectoires du champ qui tendent vers A).
- Toute morphogenèse peut être décrite par la disparition des attracteurs représentant les formes initiales et leur remplacement par capture par les attracteurs représentant les formes finales. Ce processus, appelé catastrophe, peut être décrit sur un espace P de variables externes.
- Tout processus morphologique structurellement stable est décrit par un système de catastrophes structurellement stables sur P .
- Une chréode C est la topologie globale du (ou des) attracteur(s) associé(s) et celle des catastrophes qu'il(s) subi(ssent). En particulier, pour un attracteur donné, la signification est définie par la géométrie de son domaine d'existence sur P et la topologie des catastrophes de régulation qui limitent ce domaine.
- Tout processus naturel se décompose en îlots structurellement stables, les chréodes. L'ensemble des chréodes et la syntaxe multidimensionnelle qui régit leurs positions respectives constituent un modèle sémantique du processus. Une chréode C est considérée comme un mot de ce modèle sémantique.
- Les bassins des différents attracteurs sont séparés par des hypersurfaces de type "ligne de crête". L'évolution qualitative des surfaces séparatrices est réalisée dans le cadre de la théorie des catastrophes (si la dynamique locale est une dynamique de gradient, et moyennant soit la convention de Maxwell soit la convention du délai maximum).

La Théorie des Catastrophes Élémentaires (TCE) ne considère que des attracteurs de type simple, pour lesquels la propriété de stabilité structurelle est immédiate. Ces surfaces séparatrices ne présentent alors qu'un petit nombre de singularités stables: les catastrophes élémentaires.

- La structure est un invariant algébrique. Le nombre de discontinuités de la surface de régulation ne dépend que du nombre de variables de contrôle, seules influentes sur la stabilité structurelle. Si $r \leq 4$, on a les 7 catastrophes dites élémentaires, pour lesquelles il n'y a pas plus de 2 variables internes.

Les invariants peuvent bifurquer l'un dans l'autre par un passage catastrophique.

- Une singularité catastrophique est étudiée par son déploiement universel, qui est une manière de déployer toute l'information

intrinsèque renfermée en une singularité, moyen unique pour décrire toutes les déformations possibles du germe de la fonction. Le déploiement universel d'une singularité admet deux invariants numériques: la dimension interne de la singularité, nombre minimum de variables internes, et la codimension, différence entre la dimension de l'objet et la dimension de l'espace dans lequel il est plongé. Ainsi la morphologie est indépendante de la complexité du substrat. La codimension permet de raisonner indépendamment du substrat.

- Dans la morphogenèse, l'identité d'un être matériel est définie comme un domaine connexe d'espace-temps à 4 variables: 3 variables d'espace + 1 de temps. Ces 4 variables de contrôle indépendantes correspondent à des morphologies liées aux 7 catastrophes élémentaires.

Pour Thom, les applications de la TC à des domaines non physiques supposent la définition d'un être abstrait, définition qui est de nature sémantique, donc qui fait appel à la compréhension d'un concept. La TC apporte la spatialisation de concepts abstraits; mais tout n'est pas modélisable par la TC: le cogito ne l'est pas.

Stabilité et morphogenèse structurelles d'une Organisation: approche catastrophique

Nous utiliserons donc la Théorie des Catastrophes en tant que théorie des Organisations.

Nous nous placerons dans le cas des attracteurs de type simple où les surfaces séparatrices ne présentent qu'un petit nombre de singularités stables: les catastrophes élémentaires (théorie des catastrophes élémentaires ou TCE).

Les hypothèses suivantes permettent de poser que la TC peut être utilisée comme outil d'analyse de la stabilité et de la morphologie structurelles d'une Organisation:

- la structure observée d'une Organisation est considérée comme une forme structurellement stable au sens de la TC (à l'échelle de l'analyse en tout cas).

La division du travail y partage les tâches en fonctions confiées à des groupes de personnes. La structure d'une Organisation, résultat de la division du travail et de la coordination nécessaire entre les différentes tâches du fait de cette division, est l'agencement de différentes fonctions, assurées par des groupes de personnes. La nature de l'invariance structurelle tient à la division du travail qui fixe et stratifie des comportements pour les rendre reproductibles par rapport à des événements extérieurs (le phénomène bureaucratique est un extrême de cette stratification).

L'hypothèse d'une infinité de stimuli, pour lesquels n'existe par contre qu'un nombre fini de réflexes correcteurs p_j est vérifiée. En effet, l'Organisation soumise à un stimulus/événement (e_i) parcourt une chréode, revenant au point d'équilibre (état de repos non excité) par un parcours bien défini (procédure prévue ou procédure effective) caractéristique du réflexe correcteur $p(e)$. L'Organisation répond par

une procédure descriptible à l'avance (assimilable au concept de réflexe) à un événement/ stimulus auquel elle est soumise, réponse dont l'objectif est effectivement d'annuler la perturbation causée par le stimulus.

Les stimuli en provenance de l'environnement de l'Organisation (le stimulus est équivalent au concept d'événement du Système d'Information), forment bien une infinité a priori continue, un continuum sémantique pluri-dimensionnel (W), même si leur interprétation par l'Organisation ramène cette sémantique à celle définie ou pratiquée par l'Organisation et ses processus de reconnaissance de stimuli/événements.

- une description de la structure d'une Organisation est possible par une géométrisation de son "champ organisationnel" [Lewin, 59].

Ces processus réflexes d'une Organisation proviennent des processus mis en oeuvre dans les sous-ensembles qui la composent. Ces sous-ensembles ont une finalité définie par un découpage organisationnel qui délimite pour chaque sous-ensemble son domaine d'action, en même temps que son espace de liberté au sens de Crozier. La compétence de l'Organisation divisée en sous ensembles délègue à chaque sous ensemble un ensemble de tâches, processus de travail ayant pour objectif de produire, soit des éléments matériels (téléviseurs,...), soit des éléments abstraits (fiches d'état civil,...). Le découpage en sous-ensembles est contrebalancé par des mécanismes de coordination qui ont pour rôle d'assurer le maintien et l'existence de la structure globale, par la réalisation de finalités plus larges que celles des sous-ensembles. Ces sous-ensembles se caractérisent par leur stabilité structurelle à un moment donné du fonctionnement de l'Organisation.

- Un processus organisationnel, conséquence de la décomposition des tâches par sous-ensembles, peut être considéré, selon la théorie des catastrophes, comme la matérialisation d'un champ organisationnel régulateur¹

(p --> p(e))

Le "champ organisationnel" est supposé être un système stable dense, et la stabilité structurelle d'une Organisation peut être spécifiée en termes catastrophiques.

Ce champ a subi au cours de la mise en place de la structure (et notamment des processus organisationnels) toute une série de décompositions, de catastrophes, dûes à la formation d'organes traitant le champ de façon spécialisée.

L'espace des stimuli/ événements sera partagé en "bassins d'attraction" associés à chacun des réflexes p_j ; le terme de *figure de régulation globale* du système pour désigner cette configuration sera conservé.

- chaque sous ensemble est considéré comme équivalent de l'objet de base de la théorie des catastrophes. Il est représenté par un attracteur d'un système dynamique dans un espace M de variables internes. Un sous-ensemble peut donc être modélisé comme un objet dont l'état à un moment donné est un élément dans un espace d'états, point

critique de cet espace fonction de conditions de stabilité. Cette stabilité suppose l'existence d'homéomorphismes locaux qui définissent une forme structurellement stable par une G-équivalence entre des formes voisines.

- Le champ régulateur

$$(p \rightarrow p(e))$$

peut se reconstituer en tant que forme globale d'activités, à condition que les activités de liaison/ coordination soient de nature à le permettre.

Ainsi, dans la structure globale du Système-Organisation, de la même façon qu'en théorie des catastrophes, on différenciera:

- des cellules organisationnelles spécialisées qui ne peuvent garder les états (e); et ne savent mettre en oeuvre que des (p)

- des cellules non spécialisées qui ne gardent que les états (e) et perdent ou ignorent les états (p). Ces cellules sont dotées de mémoires. Sont alors des hypothèses valides: 1) leur identification à des cellules coordinatrices de l'appareil coordinateur (par exemple le SI coopératif); 2) le rôle de l'appareil coordinateur, qui est la conservation de la sémantique globale et non des sémantiques locales, du même type que le modèle de fonctionnement du système nerveux de Stafford Beer comme modèle d'une Organisation; 3) l'hypothèse de structuration en parallèle des cellules coordinatrices avec les cellules spécialisées.

- des cellules à mi-chemin qui perdent la compétence.

Toute coordination est de l'un des types ci-dessus.

Statut épistémologique de l'approche par la Théorie des Catastrophes

La TC est une modélisation des phénomènes à ruptures, mais le type de variable modélisé lui donne suivant le cas un statut plus épistémologique ou plus explicatif. Son application au domaine des Organisations est des deux sortes. Nous aurons ainsi deux types de modélisation des variables.

Dans notre cas, les types de variables sont non quantifiables (simplement ordonnables) et leur mesure n'a guère de sens.

Les variables internes sont: partage en tâches et coordination entre tâches. Le cas échéant, on pourra les considérer comme des variables agrégées dont les vecteurs composants sont par exemple les paramètres de conception ou variables dépendantes de [Mintzberg, 1], et détailler la modélisation des vecteurs composants.

La différence de champ d'application dépend des variables de contrôle:

- dans le premier champ d'application explicatif, il s'agit d'ordonner les données de l'expérience (diagnostic sur un Système d'Information existant et sa fonction organisationnelle). Nous n'avons dans ce cas pas besoin d'une modélisation fine des variables de contrôle, et nous choisirons comme variables de contrôle les 4 dimensions (3 spatiales et 1 temporelle) des stimuli/ événements liés au Système d'Information. Nous n'avons donc pas besoin de chercher à identifier les variables de contrôle de façon significative par rapport aux entrées du système, seules les morphologies archétypes des sous-ensembles nous intéressent. La sémantique

n'intervient qu'aux frontières. Les ruptures catastrophiques entre points critiques sont alors liées à des besoins, dysfonctionnements organisationnels-types qu'il importe de caractériser.

- dans le deuxième champ d'application, plus épistémologique, il s'agit de modéliser (certains) des facteurs de changement possibles d'une Organisation, à travers la fonction symbolique de son Système d'Information. Le gain de compétence par mise en oeuvre différenciatrice de structures organisationnelles spécialisées pour obtenir plus de cycles de réaction, un plus grand nombre de degrés de liberté, et un accroissement de la variété des attracteurs, est par exemple un phénomène catastrophique qui correspond à une morphogenèse de cellules organisationnelles. L'étude des presque attracteurs ou confineurs et des bifurcations possibles de l'espace-produit (l'Organisation) permet d'étudier les changements organisationnels comme la régulation catastrophique du système global. Il s'agit alors de modéliser les changements et le passage d'une structure à une autre: points de rupture, maillons les plus faibles, effondrement localisé ou généralisé.

Dans ce cas, les variables de contrôle sont primordiales: elle permettent de modéliser des attracteurs. Leur choix est celui du modélisateur; nous choisirons quant à nous les facteurs des théories de la contingence en Théories des Organisations, dont l'hypothèse de congruence est de même nature que les hypothèses de stabilité structurelle de la TC. Dans sa synthèse de la recherche dans ce domaine, [Mintzberg, 1] identifie:

- 4 types de flux internes à une Organisation: flux d'autorité, de matériel, d'information et de processus de décision

- 11 facteurs de contingence pour une Organisation, qu'il agrège en 4 variables intermédiaires (intelligibilité, prévisibilité du travail, diversité du travail, vitesse de réponse). Les variables de contrôle peuvent être ces 4 variables intermédiaires: ressources humaines (formation, compétence des opérateurs,...), technologies (degré de contrôle que les instruments exercent sur ceux qui les utilisent), contrôle externe (pouvoir externe², environnement), et le temps (assimilable à la notion de vitesse de réponse ci-dessus). La méthode d'audit que nous proposons a pour objectif d'intégrer les 4 aspects des variables de contrôle.

Les niveaux cognitif et linguistique du Système d'Information

L'hypothèse de Thom est que l'une des premières fonctions du langage a été la description des processus spatio-temporels ayant lieu dans le voisinage de l'individu ou du groupe social. Cette hypothèse s'applique aux Organisations, car cette fonction est liée à l'existence même des Organisations complexes.

Associées aux tâches et aux processus spatio-temporels des sous-ensembles, existeront donc des processus informationnels. Ces processus informationnels sont au coeur de processus cognitifs dont la modélisation se fait par l'insertion de deux phases antérieures au niveau conceptuel:

	Cognitif	Perceptions (taxinomies)	Actions (actinomies)
Force illocutoire	Linguistique	taxèmes	actèmes
Intentionnalité	Conceptuel	Données	Traitements
	Organisationnel	MLD	MOT
	Technique	MPD	MOpT

La structure et les concepts d'un sous-Système d'Information associé à un sous-ensemble de l'Organisation sont une catégorisation effectuée dans l'univers du sous-ensemble, sous-morphologie (E), catastrophes (au sens de THOM) entre signifiant et signifié résultant de la fonction "nommer" du Langage et de la fonction "organiser l'action" de l'Organisation.

Autrement dit, les niveaux cognitif et linguistique ci-dessus sont formés par des catégorisations catastrophiques, et le processus collectif émerge de la négociation linguistique et cognitive que mène effectivement l'Organisation à partir ces catégorisations.

La modélisation catastrophique du Système d'Information

On a vu que les catastrophes correspondent aux frontières entre actants (ici les sous-ensembles), ce qui correspond à notre définition du Système d'Information Coopératif de l'Organisation.

Les activités d'une Organisation correspondent à un ensemble de processus spatio-temporels. Ces processus peuvent être exprimés par des expressions langagières au sens large. Cette remarque nous permet de reprendre le processus d'analyse morphologique de Thom appliqué au langage naturel, mais en étendant les définitions au contexte du langage d'une Organisation à travers les concepts que nous en avons donnés.

La démarche proposée peut être résumée ici:

- Les activités de l'Organisation (processus spatio-temporels) sont exprimables par des textes. Un texte (sélectionné ou rédigé par un analyste) représente un processus spatio-temporel de cette Organisation.
- Tout texte se décompose en phrases atomiques ou nucléaires (un seul verbe d'action y figure). Une phrase nucléaire est le constat d'un conflit entre régimes locaux (les actants), qui se disputent un domaine de l'espace-temps à 4 dimensions.
- Le graphe d'interaction du processus spatio-temporel décrit par une phrase atomique rentre dans l'un des 16 types de graphes répertoriés dans [Thom, 4], qui correspondent aux 16 morphologies

d'interactions-types. Nous montrons ci-dessous les extensions au contexte des Organisations.

Le type topologique des interactions détermine généralement la structure syntaxique de la phrase qui la décrit (ceci nous aidera dans la constitution d'une bibliothèque de structures-types). Les verbes itératifs sont considérés comme une catégorie à part.

- La fonction potentiel (F) est la réalisation de l'objectif du système considéré, production de biens matériels ou abstraits. Les processus par lesquels les cellules organisationnelles (les actants) réalisent des tâches, en coordination éventuellement avec d'autres actants, sont des points réguliers du potentiel (F).

La théorie des catastrophes élémentaires définit des ruptures-types entre les unités, singularités du Système de Savoir divisé en unités fonctionnelles, singularités qui se retrouvent dans le SI. Ces accidents sémantiques élémentaires archétypes conduisent à des problèmes informationnels et organisationnels prototypiques qu'il semble possible de répertorier. Ils sont interprétables de façon temporelle (avec plusieurs échelles de temps) ou spatiale.

Les morphologies archétypes [Thom, 4] et [Thom, 5] sont exposées ci-après. Les premières morphologies sont liées à une seule variable interne: l'organisation du travail. Les 4 dernières morphologies sont liées à des actions à la fois d'organisation du travail et de coordination (2 variables internes):

Morphologie sémantique archétype	Interprétation de la catastrophe	Morphologie archétype du Système d'Information
1- Organisation du travail		
être, avoir	Etat du sous-ensemble; Points réguliers (identifiables par la fonction, la finalité de l'actant)	Données représentatives, Base de Données Traitements procéduraux (de type répétitif) sur les données
finir (pli)	Frontière spatiale du sous-ensemble; Définition de la limite de compétence de l'actant; Fin de processus temporel	Fin de processus à une certaine échelle de temps; Informations finales (sans suite), ex: Archives; <u>Problème</u> : traces ultérieures des événements ou des processus pour une autre échelle de temps
commencer (pli)	Frontière spatiale du sous-ensemble; Définition de la compétence de l'actant; Début de processus temporel (événement autonome)	Début de processus à une certaine échelle de temps; Informations initiales (sans prédecesseur), ex: déclenchement de la paie à date fixe; Evènements extérieurs; <u>Problème</u> : traces des événements ou des processus pour une autre échelle de temps

changer (fronce)	Frontière entre deux actants dont le fonctionnement est couplé; Découpage d'une action globale, découpage informationnel, spatial ou temporel	Informations ou processus liés; <u>Problèmes:</u> Changement de sémantique (à évaluer); Typologies d'émission et de réception éventuellement différenciées
capturer (fronce)	Réception par un sous-ensemble d'événements provenant de l'environnement; Structure syntaxique de type sujet/ objet	Déclenchement d'un traitement local par l'arrivée d'un événement extérieur; Fin d'un parallélisme de traitement; <u>Problème</u> éventuel de délai ou de récupération physique
émettre (fronce)	Déclenchement d'événements par un sous-ensemble en direction de l'environnement; Structure syntaxique de type sujet/ objet	Emission d'un événement; Début de parallélisme;
rejeter (queue d'aronde)	Rejet d'un événement	Evènement non traité par le sous-ensemble (non reconnu, refusé, ...)
traverser (queue d'aronde)	Passage à travers un sous-ensemble; Le sous-ensemble n'est pas du type messenger actif	Transmission d'informations ou de processus, transmission éventuellement inutile, liée à la géographie,... N'a pas besoin de traitement local
émettre de façon transitoire (queue d'aronde)	Phénomène transitoire dans une émission d'événements par un sous-ensemble	Parallélisme interrompu, avorté, incomplet
faillir (queue d'aronde)	Phénomène transitoire dans une suite d'action globale (processus intéressant plusieurs sous-ensembles)	Changement de sémantique qui conduit à une interruption d'action
secouer (queue d'aronde)	Phénomènes répétitifs, échelle de temps à préciser	Phénomènes qui peuvent être spécifiques à la transmission d'informations indépendamment de la répétition de tâches
donner (papillon)	verbes trivalents: source - message - récepteur	Typiquement, la transmission. <u>Problème:</u> code du message, typologies de réception et d'émission

2- Organisation du travail + coordination

envoyer (ombilic elliptique)	Capture ou don par messenger interposé, acteur de la transmission;	La transmission entre les sous-ensembles est faite par messenger interposé; <u>Problèmes:</u> - spatiaux et temporels sur le porteur du message (difficulté de connexion destinataire/ émetteur) - élimination éventuelle du messenger - sémantique du messenger - déclenchement par messenger Ex: le ticket de réservation de terrain de tennis est transmis par le joueur entre le service des sports et le gardien des terrain
prendre (ombilic hyperbolique)	Réception ou émission à l'aide d'un instrument émis par le sujet, c'est à dire déclenchée par lui	Capture ou émission à l'aide d'un instrument (à analyser) d'informations détenues par d'autres; <u>Problème</u> de l'instrument en matière d'information: l'instrument peut ne pas être à disposition, l'instrument peut être un filtre,...

lier ou comparer (ombilic hyperbolique)	Le sujet peut se saisir de l'objet avec l'instrument à condition que ce soit à propos d'un but déterminé (la jonction sujet/ instrument/ objet n'est stable qu'à l'atteinte du but)	Aide à la décision: la finalité donne sens l'instrument et à l'objet L'instrument peut être un filtre: agrégation, synthèse,...
couper ou lier (ombilic parabolique)	Préhension ou capture partielle de l'objet par instrument émis par le sujet	La capture d'information a été partielle (problème d'instrument, d'impossibilité de connaître la sémantique de l'information complète,...)

Conclusion opérationnelle

Le Système d'Information Coopératif sera subdivisé selon le type de sémantique en: transmission (communication sans sémantique), coordination (communication entre processus portant sur la sémantique des données), coopération (communication des processus avec sémantique de ces processus), variables intermédiaires qui sont directement utilisables au sein d'une méthode de conception client/serveur.

Ces variables intermédiaires sont également utilisables au sein d'une méthode d'audit et/ou schéma directeur des Systèmes d'Information.

VARIABLE INTERMEDIAIRE (SYSTEME D'INFORMATION COOPERATIF)	LINGUISTIQUE DES ORGANISATIONS (THEORIES DES ORGANISATIONS)
Communication-type	Sémantique

1- Transmission

Transmission de messages	Sémantique simple
Transmission d'informations de ou vers SI collectif	Sémantique négociée/ imposée
Transmission d'informations de ou vers autres départements	Ruptures sémantiques et/ou temporelles

2- Coordination- Coopération

Transmission de tâches	Sémantique simple
Transmission de tâches de ou vers SI collectif	Sémantique négociée/ imposée <i>Suivi de courrier, suivi de processus</i>
Transmission de tâches de ou vers autres départements	Ruptures sémantiques et temporelles <i>Suivi de dossiers</i>

Communication-type	Sémantique, pragmatique	Outils
--------------------	-------------------------	--------

1- Transmission

Communication de messages	Sans sémantique sur le contenu du message; Localisation des émetteurs et récepteurs, définition du support et du type de code	Messagerie
---------------------------	--	------------

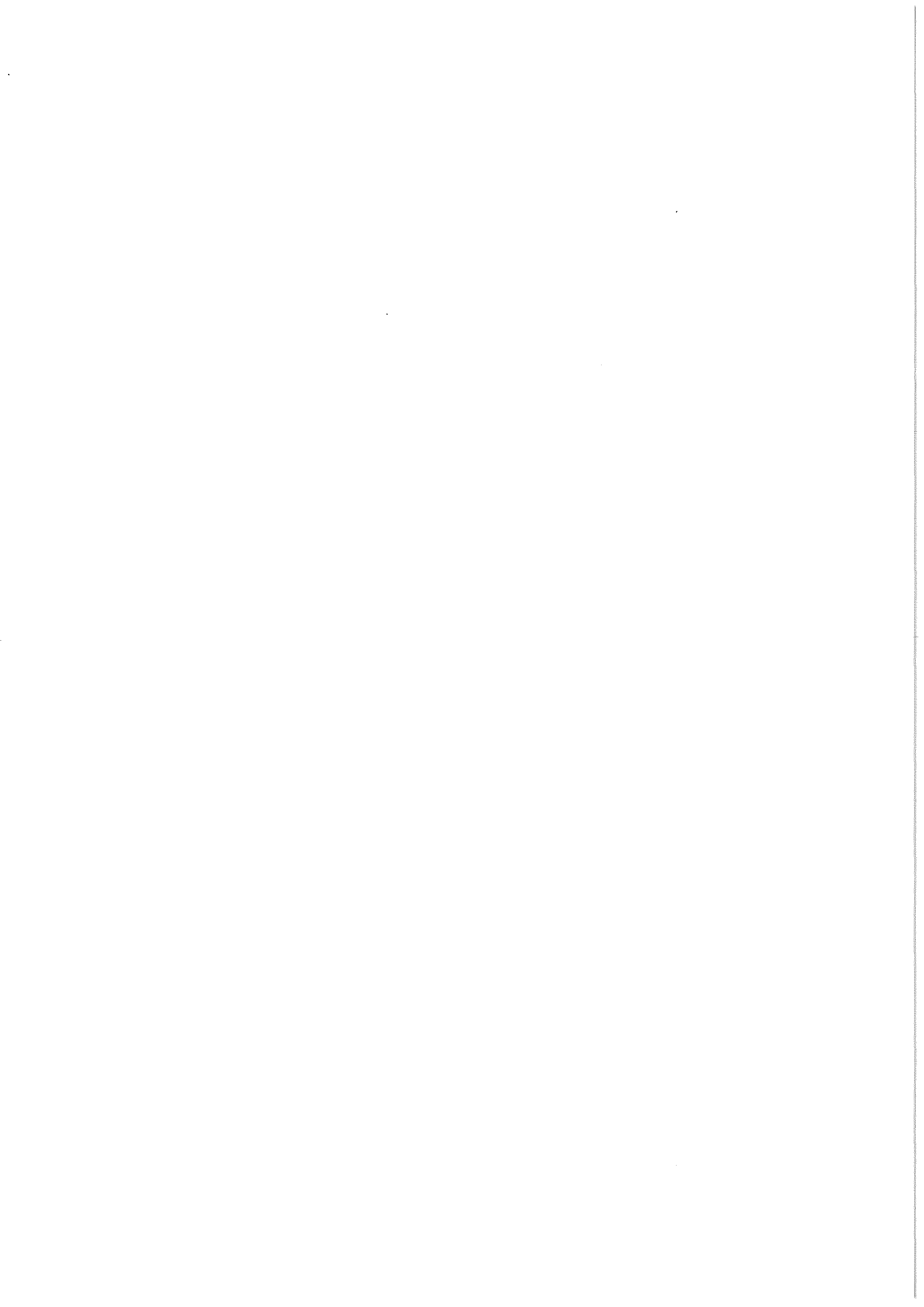
2- Coordination

Communication d'informations de ou vers SI collectif Données communes (SI collectif) sur objets de gestion	Sémantique négociée/ imposée par SI collectif Ruptures sémantiques, temporelles Localisation des émetteurs et récepteurs, définition du support et du type de code	Banques de données sur objets de gestion communs Suivi de courrier <i>Polysémie du SI collectif</i>
Communication d'informations de ou vers autres départements	Ruptures sémantiques, temporelles, de pragmatique Localisation des émetteurs et récepteurs, définition du support et du type de code	Suivi de transmissions dans suivi de courrier <i>Polysémie</i>
3- Coopération		
Communication de tâches de ou vers SI collectif	Sémantique négociée/ imposée par SI collectif Ruptures de pragmatique	Suivi de processus
Communication de tâches de ou vers autres départements	Ruptures sémantiques, temporelles, de pragmatique	Suivi de dossiers

BIBLIOGRAPHIE

- [Mintzberg, 1] Mintzberg H.
"The structuring of organizations: a synthesis of the research"; Prentice Hall, 82
"Structure et dynamique des Organisations"; Editions d'Organisation, 82
- [Thom, 1] Thom R. "Apologie du logos"
Hachette, 90
- [Thom, 2] Thom R. "Esquisse d'une sémiophysique"
Interéditions, 88
- [Thom, 3] Thom R. "Paraboles et catastrophes"
Flammarion, 83
- [Thom, 4] Thom R. "Modèles mathématiques de la morphogénèse"
Ed. C. Bougois, 80
- [Thom, 5] Thom R. "Stabilité structurelle et morphogénèse"
Interéditions, 72, 77
- [Thom, 6] Thom R., Lejeune C., Duport J.P. "Morphogénèse et imaginaire"
CIRCE, 78
- [Petitot] Petitot J., Concordia A. "Morphogénèse du sens"
PUF, 85
- [Le Moigne, 1] Le Moigne J.L. "Représentation des processus d'AUTO-IN-FORMATION de l'Organisation sociale"
Représentation du réel, Saint Etienne, 88
- [Alquier, 90] Alquier A.M. "Cognitive modelling and design for Knowledge dynamics in Automated Information Systems"
Cognitiva, Madrid, 90

- [Lewin, 59] K.Lewin - "Psychologie dynamique"
PUF, Paris, 59
- [Searle, 85] J.R.Searle - "Intentionality. An essay in the philosophy of mind"; Cambridge University Press, 83
"L'intentionnalité"; Editions de Minuit, 85
- [Searle, 83] J.R.Searle - "Sens et expression"
Cambridge University Press, 79
- [Alquier, Prince, 90] A.M.ALQUIER, V. PRINCE, "Cognitive modeling and design for knowledge dynamics in Automatized Information Systems
Cognitiva ; Madrid - novembre 90
- [Alquier, 90-1] A.M. ALQUIER - "Model and methodology to design knowledge dynamics in automatized information systems"
8° international congress of Cybernetics and Systems ;
New York - juin 90
- [Alquier, 90-2] A.M. ALQUIER - "Morphology of Organizations: dynamic cognitive model based on system theory and linguistics"
European conference of System Dynamics ; Milan - octobre 90
- [Lerbet, 84] G.LERBET - "Approche systémique et production de savoir"
Editions Universitaires, 84
- [Le Moigne, 87] J.L. Le Moigne, "Systémographie de l'entreprise"
Revue de Systémique, vol 1, n°4, 87
- [Le Moigne, 88] J.L. Le Moigne, "Auto- représentation des informations"; Colloque sur la représentation du réel, St Etienne - mai 88
- [Morin, 86] E. MORIN - "La connaissance de la connaissance" -
Livre 1°; Seuil, 1986



De l'internalisation à l'autonomie, et à l'action collective: un modèle mathématique.

P.Boldini. Ecole Nationale Supérieure des Télécommunications de Bretagne.

1. INTRODUCTION

Les systèmes biologiques, sociaux et cognitifs sont les produits d'évolutions qui mêlent intimement influences externes et structurations internes. L'intelligence de tels systèmes est fonction de leur autonomie, c'est-à-dire de leur capacité à internaliser les contraintes extérieures et à réagir de manière spécifique.

Cette contribution présente un modèle mathématique très homogène des mécanismes d'évolution d'un système autonome, il met en évidence les conditions suffisantes pour que les mécanismes d'internalisation se mettent en œuvre, et les formes "d'action collective" nécessaires à l'évolution du système. Par son inspiration catégorique il est proche, mais moins complet, du modèle de systèmes évolutifs avec mémoire de Ehresmann & Vanbremeersch ([2]). Notre parti-pris est de mettre ici en évidence les aspects algébriques et logiques du modèle; nous conseillons au lecteur curieux de se reporter à [2] pour voir comment des constructions catégoriques modélisent les processus de plus bas niveau.

A ce stade de développement, le modèle est interprété essentiellement en termes sociologiques, car il est le fruit d'une réflexion inspirée par les travaux de Pierre Bourdieu sur la genèse et la structure des champs dans l'espace social. Ces travaux sont exemplaires d'un "structuralisme génétique" qui seul nous permet de comprendre comment des systèmes fonctionnellement hiérarchisés échappent à toute explication réductionniste "par le bas", à tout déterminisme naïf; car si les structures s'engendrent bien les unes à partir des autres et retraduisent à un niveau supérieur les contraintes du bas, cette retraduction se fait en termes spécifiques qui ne sont intelligibles et efficaces qu'au niveau engendré.

L'homogénéité du modèle tient au fait que toutes les notions habituellement utilisées pour décrire l'évolution des systèmes, sont les objets d'une même catégorie au sens mathématique du terme. Pour résumer sommairement le

modèle nous dirons que les états du système sont des faisceaux, et que l'évolution du système se fait dans "l'espace-temps" constitué par la catégorie de ces faisceaux; cette dernière formant un topos elle possède une logique interne, l'évolution peut donc s'interpréter dans les termes de cette logique intrinsèque.

Pour terminer, il est épistémologiquement instructif, de voir des concepts non-triviaux de géométrie algébrique participer à l'intelligibilité de phénomènes complexes.

2. LE PREFAISCEAU, L'OBSERVATEUR, ET LA STRUCTURE INTERNE

Les systèmes complexes se caractérisent par une structuration locale et par des comportements collectifs émergents. Les concepts de pré-faisceau, de faisceau, de fibre et de germe, issus de la Géométrie Algébrique nous permettent de caractériser en termes locaux des organisations globales, et de définir localement les changements qui affecteront le système dans sa globalité.

Nous sommes malheureusement obligés d'infliger au lecteur quelques constructions mathématiques¹ assez délicates, car c'est de leur interprétation que le modèle tire sa force explicative.

2.1. Préfaisceau

DEFINITION 2.1. Un *préfaisceau d'ensembles* F sur un espace topologique $(X, \mathcal{O}(X))$ est un foncteur contravariant entre la catégorie $\mathcal{O}(X)$ des ouverts de X (le seul morphisme étant l'inclusion) et la catégorie des ensembles.

On peut considérer aussi des préfaisceaux de groupes, d'anneaux, etc...

Comme F est contravariant, l'application $F_{UV}: F(U) \rightarrow F(V)$ existe lorsque $V \subseteq U$ et est appelée *restriction*. Un élément s de $F(U)$ s'appelle une *section* sur U , et on note $F_{UV}(s) = s|_V$.

Pour nous, un état du système est un préfaisceau, plus précisément c'est un préfaisceau pour l'observateur qui distingue un espace, structuré localement par une topologie; les places dans l'espace étant déterminées par un ensemble de relations différentielles qui interdéfinissent les divers lieux. C'est ici qu'apparaît

¹Les principales définitions sont données en annexe, on trouvera les développements dans [3], [4], et [5].

la dimension structuraliste du modèle: l'agent, l'individu, ne sont pas des primitives, il sont spécifiés par un réseau de relations. On verra plus loin que l'identité individuelle n'a de sens que si la topologie de départ est séparée.

2.2. Structure du point de vue interne

Formellement, à l'ouvert U on associera $F(U)$ l'ensemble de tous les traits¹ qui opposent U aux ouverts non contenus dans U . C'est dans l'interprétation de l'opération de restriction qu'est introduite et postulée l'autonomie relative du système. Si V est un ouvert contenu dans U , les traits de $F(V)$ contiennent les restrictions à V des traits de $F(U)$, *la restriction à V du trait s de U est interprétée comme le trait s vu depuis V* . C'est là qu'est introduite la dynamique interne du système, même si elle n'est pas étudiée en tant que telle. Il est opportun à ce stade de citer Varela ([7]):

"Un système autopoïétique est organisé comme un réseau de processus de production de composants qui (a) régénèrent continuellement par leurs transformations et leurs interactions le réseau qui les a produits, et qui (b) constituent le système en tant qu'unité concrète dans l'espace où il existe, en spécifiant le domaine topologique où il se réalise comme réseau."

"Un observateur qui appréhende un système autopoïétique comme une unité dans un contexte qu'il perçoit et décrit comme son environnement peut distinguer, au sein du système, les perturbations qui proviennent de l'extérieur des perturbations qui proviennent de l'intérieur; mais, pour le système autopoïétique lui-même, elles sont intrinsèquement indistinguables."

Pour nous il en va de même pour les systèmes sociaux. Un système est effectivement produit par ses agents: si la structure produite s'analyse dans les termes d'une structure extérieure, la correspondance n'en est pas pour autant magique: les principes de production internes sont l'internalisation des contraintes externes. Ce processus d'internalisation doit être pris au sens fort pour échapper aux théories du reflet, il faut le penser comme un processus génétique sous peine de voir les agents comme des marionnettes agies par des

¹ Une relation binaire qui oppose deux ouverts se marquera par un trait pour chaque ouvert.

déterminismes transcendants.

Donc du point de vue interne tous les traits caractéristiques sont relatifs à la position occupée. L'ouvert V contenu dans U est déterminé par les relations qui l'opposent aux autres ouverts contenus dans U , mais aussi aux autres ouverts de l'espace par l'intermédiaire de U , ces dernières relations subissant un effet de diffraction ou un effet de perspective du à la position particulière de V .

3. LE FAISCEAU ENGENDRE ET L'INTERNALISATION

Le préfaisceau évoqué ci-dessus se donne donc comme structure déjà internalisée et intensionnelle. Il est naturel de pousser plus avant les processus d'internalisation et de spécification, et de poser la question: que se passe-t-il au niveau de l'individu?

Topologiquement l'individu est caractérisé comme intersection de tous les voisinages ouverts qui le contiennent¹. Quels sont alors les traits qui caractérisent cet individu x ? La réponse est très simple car la mécanique algébrique est déjà en marche, ce sont les germes en x des traits définissant les ouverts qui contiennent x .

Ces affirmations demandent quelques précisions techniques.

3.1. *Fibres et germes*

DEFINITION 3.1.1. Soit $\phi(x)$ l'ensemble des voisinages ouverts contenant x , et F un préfaisceau, la famille $(F(U), U)_{U \in \phi(x)}$ forme un système inductif, elle possède donc une limite inductive appelée *la fibre de F en x* , et notée F_x .

La fibre est obtenue comme le quotient de la réunion disjointe $(\coprod F(U))_{U \in \phi(x)}$ par la relation d'équivalence \mathcal{R} définie par:

Soit $U, V \in \phi(x)$, $s \in F(U)$ et $t \in F(V)$: $s \mathcal{R} t$ ssi il existe $W \in \phi(x)$, $W \subseteq U \cap V$, tel que $s|_W = t|_W$.

DEFINITION 3.1.2. La classe d'équivalence de s pour la relation \mathcal{R} s'appelle le germe de s en x et est notée s_x .

¹Si la topologie n'est pas séparée, il est donc impossible de distinguer des individus caractérisés par les mêmes traits.

Ainsi la fibre de F en x est l'ensemble des germes de traits en x , et le germe du trait s en x est *le trait s vu localement en x* . Nous verrons que ce passage au niveau local permet de reconstruire de manière canonique l'organisation globale, mais d'un point de vue interne.

EXEMPLE 3.1.3. On considère le faisceau Ω défini par:

$$\forall U \in \mathcal{O}(X), \Omega(U) \equiv \{V \in \mathcal{O}(X) \mid V \subseteq U\}$$

Ω est un faisceau particulièrement important dont nous reparlerons, il représente la structure topologique de X . Soit x un point de X que doit-on entendre par fibre de Ω en x ?

$$\text{si } x \in U, U_x = X_x = V_x \text{ où } V \in \phi(x)$$

$$\text{si } x \text{ est séparé de } U, U_x = \emptyset_x.$$

On voit donc que la vision de x sur la topologie est solipsiste. Tous les ouverts qui contiennent x sont identifiés et ceux qui sont séparés sont ignorés.

3.2. Faisceau engendré

DEFINITION 3.2.1. Un préfaisceau F est un faisceau si pour tout recouvrement d'ouverts $U = \cup_{i \in I} U_i$ on a les propriétés:

(F1) Soit s et t deux éléments de $F(U)$: $(s|_{U_i} = t|_{U_i}, \forall i \in I) \Rightarrow s = t$.

(F2) Pour toute famille $(s_i)_{i \in I}$, $s_i \in F(U_i)$, vérifiant: $s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j} \forall i, j \in I$, il existe un recollement; c'est à dire un élément s de $F(U)$ tel que: $s|_{U_i} = s_i, \forall i \in I$.

Pour tout préfaisceau F on peut effectuer une construction canonique qui lui associe un faisceau noté F^+ . Si F est déjà un faisceau le faisceau engendré est évidemment isomorphe à F . Cette construction qui est esquissée ci-dessous, revient à extensionnaliser la définition du préfaisceau F , puisque l'on va considérer toutes les structures au niveau des points x , puis récupérer les structures globales comme ensembles dont les éléments sont les structures locales.

Tout d'abord on considère la somme disjointe des fibres pour tous les points x de l'espace topologique de base X , soit: $(\coprod F_x)_{x \in X}$.

A tout ouvert U de X on va associer l'ensemble $F^+(U)$ des fonctions $f: U \rightarrow (\coprod F_x)_{x \in X}$ telles que:

$$(1) f(x) \in F_x$$

(2) Pour tout $a \in U$, il existe un voisinage V_a et $s \in F(V_a)$ tels que:
 $\forall x \in V_a, f(x) = s_x$.

Comme l'on munit $(\coprod F_X)_{X \in X}$ de la topologie la moins fine qui rende continues les fonctions $s^*: U \rightarrow (\coprod F_X)_{X \in X}$, $s \in F(U)$ qui à x associent s_x ; le faisceau F^+ est donc le faisceau des sections de fonctions continues de X dans $(\coprod F_X)_{X \in X}$.

DEFINITION 3.2.2. L'espace topologique $(\coprod F_X)_{X \in X}$ est appelé *l'espace étalé* associé au faisceau F^+ , il est identifié à F^+ .

Une base d'ouverts de $(\coprod F_X)_{X \in X}$ est: $s^*(U) = \{s_x, x \in U\}$, $s \in F(U)$, U ouvert de X .

F^+ étant un faisceau on montre aisément comment sont reconstruits des traits globaux. Soit U un ouvert de X , et $U = \cup_{i \in I} U_i$ un recouvrement ouvert de U , à ce recouvrement on peut associer dans $(\coprod F_X)_{X \in X}$ un ouvert $\cup_{i \in I} s^*_i(U_i)$ où $s_i \in F(U_i)$; en d'autres termes on se donne une collection de germes de traits, et la propriété de recollement du faisceau F^+ nous assure que cette collection donne un trait caractérisant U .

Si F est au départ un faisceau on n'obtient rien de plus que les traits existants, mais si F n'est qu'un préfaisceau, sont obtenus des traits nouveaux issus de regroupements inédits entre germes. Les traits étant des marques différentielles, on voit que le passage au niveau des individus produit de nouveaux systèmes de différences syncrétiques qui légitiment la structure topologique.

Pour résumer cette partie, nous avons vu comment s'internalise un système de différences, et comment il est reproduit à partir des caractéristiques individuellement perçues.

4. LE TOPOS $Sh(X)$

Il nous reste à penser la manière dont le système évolue. Classiquement on accorde aux agents une vision partielle du système et la faculté d'élaborer des stratégies ([2]). Nous allons voir comment ces capacités s'interprètent naturellement dans la structure d'un topos.

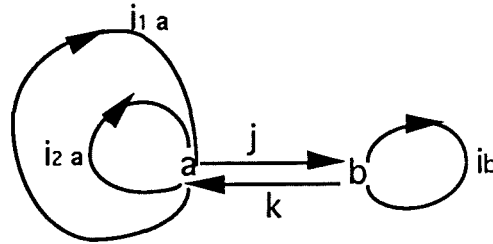
4.1. *Topoi*

Des définitions extrêmement techniques que l'on trouvera en annexe il faut retenir qu'un topos est une structure qui internalise l'ensemble des sous-objets d'un objet, et l'ensemble des transformations entre deux objets; c'est-à-dire qu'ils

sont représentés par des objets de la catégorie. Il possède un objet particulier Ω appelé *classificateur de sous-objets*, et les objets notés Ω^F et G^F , représentent respectivement l'ensemble des sous-objets de F et l'ensemble des morphismes de F dans G .

EXEMPLE 4.1.1. • La catégorie des ensembles est un topos, Ω est n'importe quel ensemble à deux éléments (on choisit donc $\{0, 1\}$), et B^A est l'ensemble des graphes des fonctions de A dans B .

• La catégorie des graphes finis est un topos, Ω est le graphe:



Ce qui nous importe avant tout est le théorème suivant:

THEOREME 4.1.2. La catégorie $Sh(X)$ des faisceaux sur un espace topologique X est un topos.

4.2 Le faisceau Ω^F

Pour la catégorie $Sh(X)$ le classificateur de sous-objets est le faisceau Ω mentionné dans l'exemple 3.1.3. Le faisceau Ω^F représente l'ensemble des transformations de F dans Ω , il représente donc l'ensemble des sous-objets, c'est à dire des sous-faisceaux de F .

Or le paragraphe 3 nous a montré que l'on pouvait considérer F comme un espace topologique, un espace étalé sur X . Alors les sous-faisceaux de F ne sont rien d'autre que les ouverts de F . Chaque sous-faisceau est une vision partielle de la structure, F en étant la vision totale. Nous avons vu que les ouverts élémentaires de F sont de la forme $s^*(U) = \{s_x, x \in U\}$, $s \in F(U)$, ce faisceau élémentaire est un point de vue possible sur le système, celui obtenu à partir d'un seul trait et depuis un seul ouvert. Ω^F apparaît donc comme l'ensemble des descriptions que le système peut faire de lui même des plus frustes aux plus complètes. Il est alors tentant d'interpréter Ω^F comme *le paysage* du système au sens de Ehresmann & Vanbremeersch ([2]) ou comme *domaine cognitif* du système au sens de Varéla ([7]):

"Le domaine d'interactions d'un système autopoïétique est son domaine *cognitif*. Autrement dit, le domaine cognitif d'un système autopoïétique est composé de l'ensemble des descriptions qu'il peut formuler. Le mode particulier de l'autopoïèse de tout système autopoïétique détermine donc son domaine cognitif et la diversité de son comportement."

Il est important de réaliser que Ω^F est donné avec F puisque formellement il suffit de connaître les germes des traits et la topologie de X.

4.3. *Les transformations*

Une transformation de l'état F en l'état G du système peut être vue soit comme un morphisme de faisceaux, soit comme une application continue entre deux espaces étalés. Au vu des descriptions du système dont ils disposent via Ω^F , les agents peuvent entreprendre de modifier ou de supprimer certains traits. C'est au paragraphe 5 que nous discuterons de la nature des descriptions formulables à un niveau d'organisation U.

Dans le processus d'autonomisation du champ artistique Bourdieu ([1]) montre comment les tenants de "l'art pour l'art" luttent pour substituer les critères purement artistiques de perception de l'œuvre aux critères liés au monde social.

"L'histoire du roman, au moins depuis Flaubert, peut aussi être décrite comme un long effort pour << tuer le romanesque >>, selon le mot d'Edmond de Goncourt, c'est-à-dire pour purifier le roman de tout ce qui semble le définir, l'intrigue, l'action, le héros: cela depuis Flaubert et le rêve du << livre sur rien >> ou les Goncourt et l'ambition d'un << roman sans péripéties, sans intrigue, sans bas amusements >> jusqu'au << Nouveau Roman >> et la dissolution du récit linéaire (...)." ¹

Il faut alors comprendre comment des changements nécessairement locaux produisent des effets globaux. Le modèle fournit une réponse très simple.

Soit f la transformation de l'état F en l'état G, f est continue donc l'image

¹Cf. P. Bourdieu, *Les règles de l'art. Genèse et structure du champ littéraire*. Editions du Seuil. 1992. p.335. On notera que les intentions exprimées par les acteurs des changements sont les signes d'un champ littéraire largement autonomisé puisqu'elles s'énoncent en termes purement esthétiques.

réciroque de tout ouvert de G est un ouvert de F . Considérons un ouvert élémentaire $t^*(U) = \{t_x, x \in U\}$, $t \in G(U)$; $f^{-1}(t^*(U))$ est nécessairement un ouvert de F c'est-à-dire une réunion d'ouverts élémentaires de F , ainsi:

$$f^{-1}(t^*(U)) = \cup_{i \in I} s^*_i(U_i).$$

Il faut donc que cette collection de germes ait été transformée en une collection de germes de G qui se recolle en t . Pour ce faire il faut que les transformations des germes initiaux aient été uniformes et compatibles sur les intersections $U_i \cap U_j$.

Ainsi l'on met en évidence des formes nécessaires d'action collective amenant à des changements locaux. Il est clair que plus le changement sera important, par le nombre de traits modifiés ou la taille de l'ouvert sur lequel portent les traits, plus seront importantes les formes de coordination.

4.4. Le faisceau G^F

Le faisceau G^F représente au sein du topos $\text{Sh}(X)$ l'ensemble des transformations entre les faisceaux F et G . Ce faisceau s'interprète naturellement comme l'ensemble des stratégies possibles pour atteindre G .

Les constructions des faisceaux G et G^F peuvent n'être que virtuelles, car les changements sont envisagés de divers points de vue dans le système, l'état résultant G_0 n'étant que très rarement conforme aux visées d'un groupe particulier¹. La différence entre état envisagé G et état final G_0 sera mesurée comme dans ([2]) par un morphisme de faisceaux (foncteur si l'on parle en termes catégoriques, fonction continue si l'on parle d'espaces étalés) entre G et G_0 .

5. INTERPRETATION LOGIQUE

Nous avons donc montré que tous les éléments qui participent à l'évolution de notre système sont des objets ou des morphismes du topos $\text{Sh}(X)$. Or à tout topos on peut associer son langage interne qui est une théorie intuitionniste des types ([5] et annexe), le topos devient alors un modèle de cette théorie. Il en va

¹ Ici nous suivons très exactement l'approche de [2] pour les systèmes hiérarchiques évolutifs.

donc de même pour le topos $\text{Sh}(X)$, et le théorème ci-dessous formalise la sémantique du langage interne de $\text{Sh}(X)$.

THEOREME.5.1. Soit F un faisceau sur X , U un ouvert de X , $a: h_U \rightarrow F$. Alors:

- (0) $U \Vdash a$ ssi $a^* = U$, dans le cas $F = \Omega$;
- (1) $U \Vdash b \in \beta$ ssi $\beta^*(U)(b^*) = U$, où $b: h_U \rightarrow F$ et $\beta: h_U \rightarrow \Omega^F$;
- (2) $U \Vdash \top$ toujours;
- (3) $U \Vdash \perp$ ssi $U = \emptyset$;
- (4) $U \Vdash \varphi(a) \wedge \psi(a)$ ssi $U \Vdash \varphi(a)$ et $U \Vdash \psi(a)$;
- (5) $U \Vdash \varphi(a) \vee \psi(a)$ ssi $U = V \cup W$ avec $V, W \in \mathcal{O}(X)$ tels que $V \Vdash \varphi(a|_V)$ et $W \Vdash \psi(a|_W)$;
- (6) $U \Vdash \varphi(a) \Rightarrow \psi(a)$ ssi pour tout $V \subseteq U$, si $V \Vdash \varphi(a|_V)$ alors $V \Vdash \psi(a|_V)$;
- (7) $U \Vdash \forall_{y \in G} \psi(y, a)$ ssi, pour tout $V \subseteq U$ et tout $b: h_V \rightarrow G$, $V \Vdash \psi(b, a|_V)$;
- (8) $U \Vdash \exists_{y \in G} \psi(y, a)$ ssi il existe un recouvrement ouvert $U = \cup_{i \in I} U_i$ et des flèches $b_i: h_{U_i} \rightarrow G$ tels que $U_i \Vdash \psi(b_i, a|_{U_i})$ pour tout $i \in I$.

Nous citons ce théorème pour référence car l'interprétation des huit conditions est essentielle à notre entreprise. Si l'intérêt du théorème est clair pour tous ceux qui s'intéressent à la logique intuitionniste et à ses modèles, il est moins évident d'en tirer un enseignement pour notre propos. Pour ce faire nous allons tenter une interprétation risquée, mais légitime de notre point de vue; on peut la rapprocher de l'interprétation faite par Van Dalen ([6]) de l'activité du mathématicien idéal au sens de Brouwer.

En 4.2. nous avons interprété le faisceau Ω^F comme l'ensemble des descriptions possibles du système pour lui-même. Il faut alors s'interroger sur le statut précis de cette notion de description pour les individus et surtout pour les groupements d'individus (les ouverts). Qu'est ce que l'on connaît du système au niveau U ? Une proposition de type $F \rightarrow \Omega$ dans le langage interne a pour sémantique dénotationnelle un sous-faisceau du faisceau F , c'est-à-dire une description partielle du système. On va donc identifier les prédications possibles à un niveau U avec les descriptions accessibles au niveau U .

Les huit conditions du théorème 5.1, définissent inductivement les prédications possibles sur le système pour un niveau U donné.

Les propositions atomiques sont caractérisées par les conditions (0) et (1).

- La condition (0) dit qu'au niveau U le système connaît sa topologie, c'est-

à-dire l'ensemble des ouverts contenus dans U.

- La condition (1) dit qu'au niveau U le système est capable de formuler une description de lui-même en termes de sous-faisceau de traits (pour le niveau U). En d'autres termes on peut formuler l'appartenance d'un trait à un sous-faisceau.

Les conditions (2) et (3) ne demandent aucune interprétation, les conditions suivantes montrent comment s'élaborent des descriptions de plus en plus complexes.

- La condition (4) dit que l'on peut former la conjonction de deux descriptions si les deux descriptions sont simultanément valides au niveau U.
- La condition (5) introduit les aspects proprement intuitionnistes: on peut au niveau U donner une description en termes disjonctifs si chacune des descriptions est valide sur un ouvert (suivant les règles de validité propres à cet ouvert) contenu dans U, les deux ouverts recouvrant U. On voit donc comment au niveau U il peut y avoir enrichissement¹ des descriptions, à partir de descriptions locales.
- La condition (6) s'analyse en termes identiques, la formation d'une description plus riche, signe d'une plus grande réflexivité, n'est possible que si elle s'appuie sur des faits vérifiés pour tous les ouverts contenus dans U.

Les deux dernières conditions extrêmement générales permettent d'exprimer des descriptions quantifiées d'un état G du système dépendantes de traits d'un état F du système. Elles mettent en évidence la faculté d'élaborer des stratégies et de prévoir (de son point de vue) l'état futur du système. Supposons que l'état G provienne de F par la transformation d'un trait a au niveau U (ce qui implique la transformation compatible de tous les $a|_V$ pour tout $V \subseteq U$), il est alors possible de formuler des descriptions de G, aussi complètes que celles de F (n'oublions pas que les règles de formation sont inductives).

6.CONCLUSION

¹Cet enrichissement du point de vue logique correspond de fait à une nécessaire synthèse entre des descriptions locales différentes, il peut être donc considéré comme un affaiblissement.

Le modèle présenté est donc complet, le système autonome évolue au sein du topos $\text{Sh}(X)$ au sein duquel sont construits tous les éléments nécessaires à cette évolution, descriptions internes possibles, langage de description interne, objectifs, et stratégies.

Bien sur il reste beaucoup à faire pour la description fine des liens qui unissent et opposent les constituants élémentaires du système, et pour celle des mécanismes de restriction et de recollement.

Dans le cadre de cet article nous espérons seulement avoir convaincu le lecteur de l'homogénéité et de l'économie de notre construction; cela tient au fait que le traitement mathématique n'est pas ad-hoc, il possède sa logique propre qui est spécifiquement mathématique; on ne peut nous reprocher que d'avoir forcé les interprétations.

ANNEXE

CATEGORIE. Une *catégorie* \mathfrak{C} est une collection de deux sortes d'entités des objets et des morphismes (ou flèches) entre ces objets vérifiant les conditions suivantes:

- deux morphismes $f: A \rightarrow B$ et $g: B \rightarrow C$ peuvent être composés en $gf: A \rightarrow C$
- pour tout objet A il existe un morphisme $1_A: A \rightarrow A$ tel que pour tout $f: A \rightarrow B$, on a: $f1_A = f = 1_B f$;
- pour tout $f: A \rightarrow B$, $g: B \rightarrow C$, $h: C \rightarrow D$ on a: $(hg)f = h(gf)$.

CATEGORIE CARTESIENNE FERMEE. Une *catégorie cartésienne fermée* est une catégorie \mathfrak{C} possédant des produits finis (elle a donc un objet terminal) et telle que, pour tout objet B de \mathfrak{C} , le foncteur $(-)\times B: \mathfrak{C} \rightarrow \mathfrak{C}$ a un adjoint droit, noté $(-)^B: \mathfrak{C} \rightarrow \mathfrak{C}$. Ce qui veut dire que pour tout objet A , B et C de \mathfrak{C} , il existe un isomorphisme: $\text{Hom}_{\mathfrak{C}}(A \times B, C) \cong \text{Hom}_{\mathfrak{C}}(A, C^B)$

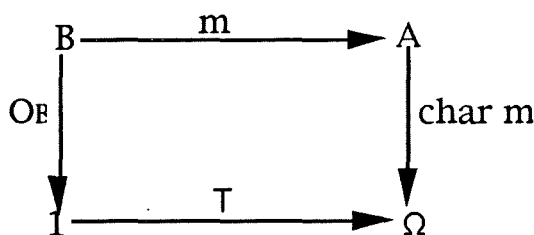
FONCTEUR. Un *foncteur* covariant (resp. contravariant) F entre les catégories \mathfrak{A} et \mathfrak{B} est un morphisme qui envoie les objets de \mathfrak{A} sur des objets de \mathfrak{B} et les flèches de \mathfrak{A} sur des flèches de \mathfrak{B} de telle sorte que: $f: A \rightarrow A'$ est envoyée sur $F(f): F(A) \rightarrow F(A')$ (resp. $F(f): F(A') \rightarrow F(A)$)

En outre il doit préserver l'identité et la composition, c'est à dire:

$$F(1_A) = 1_{F(A)}, \text{ et } F(gf) = F(g)F(f).$$

TOPOS. Un *topos (élémentaire)* \mathfrak{T} est une catégorie cartésienne fermée dont le foncteur de sous-objets est représentable. Ce qui s'exprime par l'existence d'un objet Ω , appelé le *classificateur de sous-objets*, et d'un isomorphisme naturel: $\text{Sub} \cong \text{Hom}(-, \Omega)$.

Plus précisément, cela signifie qu'il existe une flèche $\top: 1 \rightarrow \Omega$ (appelée *vrai*) telle que pour tout objet A , et tout monomorphisme $m: B \rightarrow A$, il existe une flèche unique *char m* telle que le diagramme ci-dessous soit un produit fibré.



L'objet Ω^A noté aussi PA peut-être interprété comme "l'ensemble" des sous-objets de A .

LANGAGE INTERNE D'UN TOPOS. Le langage interne $\mathfrak{L}(\mathcal{T})$ d'un topos \mathcal{T} a pour types les objets de \mathcal{T} .

Il possède comme termes de type A en les variables x_i de type A_i ($i = 1, \dots, n$) les polynômes $\varphi(x_1, \dots, x_n): 1 \rightarrow A$ en les indéterminées $x_i: 1 \rightarrow A_i$. En particulier:

- les variables de type A sont les flèches indéterminées $x: 1 \rightarrow A$;
- $*$ est $1 \rightarrow 1$;
- $\langle a, b \rangle$ est $\langle a, b \rangle: 1 \rightarrow A \times B$;
- $a = a'$ est la composée de $\langle a, a' \rangle: 1 \rightarrow A \times A$ et de $\delta_A: A \times A \rightarrow \Omega$ où $\delta_A \equiv \text{char} \langle 1_A, 1_A \rangle$;
- $a \in \alpha$ est la composée de $\langle \alpha, a \rangle: 1 \rightarrow PA \times A$ et de $\varepsilon_A: PA \times A \rightarrow \Omega$ où $\varepsilon_A \equiv \varepsilon_{\Omega, A}$;
- $\{x \in A \mid \varphi(x)\}$ est $\lambda_{x \in A} \varphi(x)$, l'unique flèche $\alpha: 1 \rightarrow PA$ telle que $x \in \alpha \bullet \bullet_x \varphi(x)$.

LEMME DE YONEDA. (Yoneda 1954) Soit \mathcal{C} une catégorie localement petite (i.e la classe des morphismes entre deux objets est un ensemble), et A un objet de \mathcal{C} . On note h_A le foncteur contravariant $\text{Hom}_{\mathcal{C}}(-, A)$.

Si $F: \mathcal{C}^{\text{op}} \rightarrow \text{Sets}$ est un foncteur contravariant, alors $\text{Nat}(h_A, F)$ la classe des transformations naturelles entre F et h_A (ici la classe des morphismes de faisceau) est en correspondance biunivoque avec $F(A)$.

NOTATION Soit $a: h_A \rightarrow F$ une transformation naturelle, on notera a^* l'objet de $F(A)$ qui lui est associé par le Lemme de Yoneda.

REFERENCES

- [1] BOURDIEU P., *Les règles de l'art. Genèse et structure du champ littéraire*. Editions du Seuil. 1992.
- [2] EHRESMANN A. C., VANBREMEERSCH J.-P., "Un modèle pour des systèmes évolutifs

avec mémoire, basé sur la théorie des catégories", *Revue Internationale de Systémique*, Vol. 5, N°1, (1991), pp 5-25.

[3] GODEMENT R., *Théorie des faisceaux*, Paris, Hermann, 1964.

[4] GOLDBLATT R., *Topoi, the categorial analysis of logic*, North-Holland, 1979.

[5] LAMBEK J. & SCOTT P.J., *Introduction to higher order categorial logic.*, Cambridge, Cambridge University Press, 1986.

[6] VAN DALEN D., "Intuitionistic logic", *Handbook of Philosophical Logic, Volume III, Alternatives to Classical Logic*, D. Reidel Publishing Company, (1986), pp 225-339.

[7] VARELA F.J. *Autonomie et connaissance: essai sur le vivant*. Editions du Seuil.1989.

Transitions de phase et phénomènes critiques dans les systèmes collectifs

Eric Bonabeau

CNET Lannion B - OCM / TEP
Route de Trégastel
22301 Lannion Cédex

bonabeau@lannion.cnet.fr

Résumé : Les transitions de phase critiques, et plus généralement les phénomènes critiques hors équilibre, semblent jouer un rôle important dans les capacités de traitement de l'information (en un sens plus ou moins bien défini) des systèmes naturels, en particulier lorsque ce traitement est distribué ou collectif.

I Motivations cognitives et systèmes distribués

"Votre crédulité réside-t-elle dans quelque centre de crédulité de votre cerveau ? Un neuro-chirurgien pourrait-il vous ouvrir le crâne et effectuer quelque opération délicate permettant de réduire votre crédulité tout en vous laissant, pour le reste, parfaitement intact ? Si vous pensez que oui, vous êtes plutôt crédule, et vous devriez peut-être envisager cette opération."

D.Hofstadter, "Gödel, Escher, Bach"

Le paradigme informationnel en Sciences Cognitives

La plupart des modèles et des réalisations des Sciences Cognitives actuelles se fondent sur un paradigme général et fondamental, celui du traitement de l'information. Un système cognitif y est vu comme un appareil complexe acceptant un flux de données de l'environnement par l'intermédiaire de capteurs, puis faisant opérer sur ces données un ensemble de mécanismes permettant d'extraire une information signifiante et pertinente en vue d'une réaction adaptative à travers des effecteurs et un équilibrage de l'état interne. Un exemple frappant est constitué par les modèles connexionnistes où la notion de traitement de l'information est présente à tous les niveaux, aussi bien dans l'architecture des modèles que dans les modes d'apprentissage et d'exploitation. De manière plus générale, tous les modèles computationnels considèrent les processus cognitifs qu'ils tentent de simuler ou de reproduire comme des ensembles de fonctions de traitement de l'information qui interagissent. Il en est ainsi par exemple pour les modèles de l'intelligence artificielle classique, des réseaux sémantiques aux systèmes à base de règles.

Psychologie Cognitive et approche boîte noire

Le système traitant de l'information n'y est pas nécessairement vu comme une boîte noire, bien que ce soit là un moyen usuel et commode d'approcher la notion de traitement de l'information. Cette approche de type boîte noire pose un certain nombre de problèmes et n'est pas toujours explicative. Les physiiciens et les biologistes, pour ne citer qu'eux, offrent d'autres voies d'exploration. L'idée d'auto-organisation, qui remonte aux années cybernétiques, exprime une alternative possible : un système auto-organisé répond aux perturbations causées par des modifications enregistrées au niveau de ses couplages avec l'environnement dans lequel il est plongé. L'auto-organisation vise à lui permettre de répondre de manière pertinente ou encore viable. Ces descriptions grandiloquentes paraissent bien floues à un scientifique. Mais on voit apparaître depuis quelques années un domaine, celui de la science des systèmes complexes, qui même s'il n'est pas bien circonscrit, commence à produire des résultats conceptuellement nouveaux à partir d'une idée fédératrice simple : un grand nombre d'éléments simples interagissant de façon non-linéaire peuvent générer un comportement global complexe. En particulier, ce comportement complexe peut se traduire par une capacité universelle de traitement de l'information. Pour être plus concret, on peut se demander comment une colonie d'insectes peut s'adapter à son environnement, recruter des sources de nourriture (et ceci de façon quasi-optimale) ou construire des nids dont l'architecture est fort enchevêtrée, avec seulement une information locale, généralement bruitée et incomplète. On peut se demander également comment un réseau d'ordinateurs peut arriver à optimiser le temps d'utilisation des machines de manière distribuée, ou encore comment on peut prédire le comportement d'un ensemble d'agents économiques. Ce que l'on cherche à prédire, c'est la dynamique d'une grandeur agrégée ou macroscopique d'intérêt (le taux d'utilisation de machines, la variations des retours de bénéfices, des taux d'intérêts, etc...). On peut aussi chercher à contrôler lorsque l'on sait prédire.

Pour la psychologie cognitive, l'individu est considéré comme un système de traitement de l'information, qui "transforme les informations de nature physique en informations de nature mentale ou représentationnelle" (G.Tiberghien). Le but de la psychologie cognitive (aujourd'hui confondue avec la psychologie expérimentale), est alors de comprendre ce qui se passe dans "la boîte noire du psychisme", c'est à dire en particulier comprendre la nature des objets mentaux : représentations, catégories, concepts, symboles, mécanismes logiques, attitudes, états d'âme, stéréotypes... Les outils mentaux -et sensori moteurs- permettent la sélection et l'interprétation des informations reçues. Mais cette approche de type "boîte noire" possède quelques inconvénients majeurs :

- On ne sait pas, dans ce cadre, définir la notion d'information : qu'est ce que l'information ? La question se pose avec d'autant plus d'acuité que l'on associe au concept d'information une qualité de pertinence qui n'a aucun sens dans un système non "incarné", i.e non situé dans un environnement donné. Signification et action sont deux aspects très peu pris en compte.

- Le problème qui remet le plus en cause l'approche "boîte noire" est celui de l'existence implicite d'inputs et d'outputs. Ce que recherche en fait la psychologie cognitive, c'est une relation causale entre inputs et outputs du système cognitif. En dehors du fait qu'il est difficile de bien cerner les limites du système cognitif humain par exemple, il est également très risqué de parler d'inputs et d'outputs, notions qui

supposent un prédécoupage de l'espace phénoménologique en "observables". Or, ces observables, dont nous disons qu'elles constituent les inputs et les outputs d'un système, sont imposées par l'expérimentateur. Il n'est pas évident que ce découpage ait une pertinence optimale vis à vis du fonctionnement du système étudié. Même la notion de stimulus, qui correspond à un affaiblissement de celle d'input, présuppose que nous soyions capables de déterminer les divers stimuli auxquels est soumis un système.

- Enfin, il existe un troisième problème lié à l'approche classique de traitement de l'information, moins visible sans doute que les deux précédents mais également essentiel : le problème fonctionnaliste, qui consiste à voir un système implémenter une fonction. Le cas extrême de fonctionnalisme réside dans le paradigme informationnel poussé à son comble : le substrat est inessentiel, et seule la fonction de traitement de l'information compte. Mais, comme Putnam, fonctionnaliste repent, l'a bien indiqué, la structure précède la fonction, elle est une condition nécessaire à l'interprétation fonctionnelle, et l'interprétation d'une structure en terme de fonction reste et restera une interprétation : on ne saura distinguer si le substrat implémente "réellement, objectivement" la fonction, ou si tout se passe comme si il implémentait réellement cette fonction. C'est cela qu'il faut garder en mémoire : une structure n'est pas créée pour réaliser une fonction, la fonction est un effet de la structure lié à des paramètres environnementaux et à l'instanciation particulière de cette structure dans un système physiquement réalisé.

Ce texte n'est certainement exempt de reproches, mais s'attache cependant à éviter les défauts qui viennent d'être évoqués. En particulier, lorsque l'on emploie le terme information, c'est dans un sens qui peut sembler affaibli : il s'agit de structures, que l'on peut stocker, transmettre, transformer et ces trois composantes sont les opérations centrales de toute "computation", ou du moins de tout processus dont nous avons l'intuition qu'il compute quelque chose. D'autre part, l'idée de boîte noire est définitivement rejetée : le système doit être considéré comme plongé dans un environnement. Il est évident que la notion d'environnement, habituellement oubliée des méthodes réductionnistes classiques, amène à lier étroitement structure et fonction, puisque l'on ne peut plus dissocier une structure de la fonction qu'elle "implémente" (point de vue subjectif) dans un environnement donné. Cette constatation apparemment anodine, revêt en fait une grande importance, en particulier dans un cadre d'étude aussi abstrait que celui des phénomènes critiques : en effet, on voit certes des structures "complexes" apparaître au voisinage d'une transition de phase, mais les implications fonctionnelles de ces structures complexes sont loin d'être immédiates. D'ailleurs, plus fondamentalement, on peut se demander en quoi la "pression sélective" favorise l'apparition de structures complexes : on comprend bien intuitivement la complexification fonctionnelle, puisque c'est cette dernière qui est directement responsable des capacités d'un organisme (au sens classique aussi bien qu'au sens de super-organisme, de société, etc...) à se maintenir dans son environnement, alors que les structures sous-jacentes sont beaucoup moins directement impliquées. Une structure peut implémenter plusieurs fonctions et une fonction peut être réalisée par plusieurs structures.

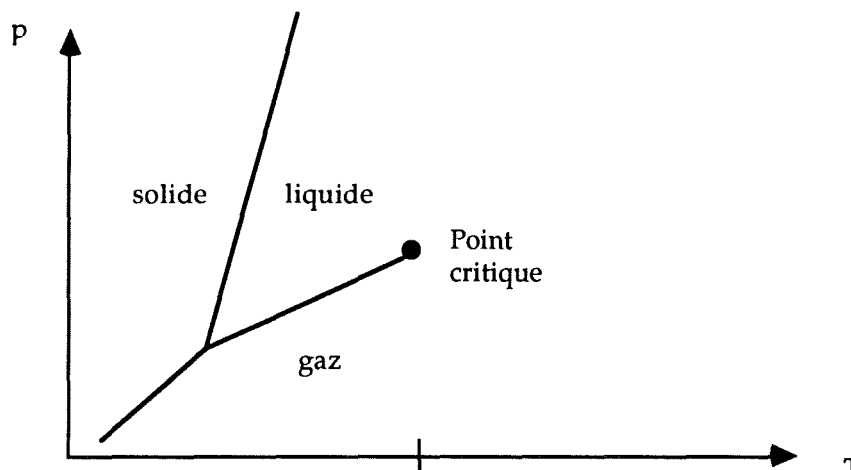
II Les transitions de phase et les phénomènes critiques : une introduction physique et mathématique

La notion de transition de phase a été jusqu'à récemment essentiellement physique. Ce n'est que depuis une quinzaine d'années, depuis la contribution de Feigenbaum sur les applications de l'intervalle, que l'idée d'appliquer les théories développées en physique des transitions de phase (critiques) aux systèmes dynamiques a germé. Il est donc naturel d'expliquer ce qu'est une transition de phase au sens où les physiciens l'entendent, avant de passer à l'analogie mathématique qui sera assez brève. Il faut noter dès maintenant que l'analogie est difficile formellement (puisque'elle n'est pas vraiment résolue) alors que la phénoménologie rend facile et enivrante l'application d'un vocabulaire physique à des systèmes dynamiques.

Classification des transitions de phase (1) : grandeurs thermodynamiques

Ehrenfest a proposé une classification des transitions de phase en prenant l'enthalpie libre G comme potentiel thermodynamique : $G = U - TS + pV$. Il définit :

- les transitions de phase du premier ordre comme celles pour lesquelles on observe des discontinuités dans des grandeurs thermodynamiques liées aux dérivées premières de G (comme l'entropie);
- les transitions de phase du second ordre pour lesquelles on observe des discontinuités dans les grandeurs liées aux dérivées secondes de G (comme C_p) alors que ses dérivées premières sont continues;
- de manière générale une transition de phase est d'ordre N si les $N-1$ premières dérivées de G sont continues alors que la dérivée d'ordre N est discontinue.

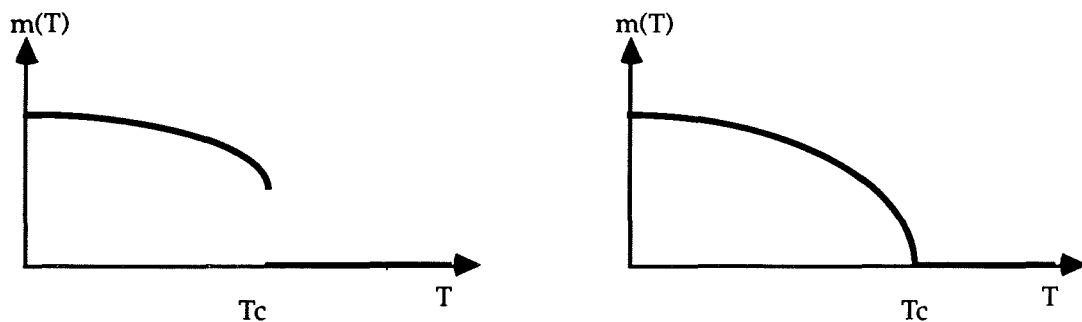


Classification des transitions de phase (2) : paramètres d'ordre

La classification de Landau, introduite pour rendre compte de transitions sans chaleur latente (pour certaines transitions de phase du second ordre selon Ehrenfest, on observe une divergence de la chaleur spécifique à la transition) est quant à elle reliée à la notion de paramètre d'ordre et de brisure de symétrie. Le paramètre d'ordre est une grandeur physique extensive, nulle dans la phase la plus symétrique, et non nulle dans la phase la moins symétrique : généralement, la phase symétrique est la plus désordonnée, tandis que la phase moins symétrique est plus ordonnée. Cette constatation est plutôt intuitive si l'on considère des propriétés d'invariance statistique : un gaz est invariant par n'importe quelle transformation isométrique, et le passage à la forme ordonnée fait perdre un grand nombre d'invariances. Si le paramètre d'ordre ne peut être généralement défini que pour des transitions sans chaleur latente, il n'est pas définissable pour des transitions entre variétés polymorphes ou liquide-gaz. De manière générale, on ne peut définir de paramètre d'ordre que lorsque les groupes de symétrie des deux phases obéissent à une propriété d'inclusion stricte : si les deux groupes de symétrie ne sont pas comparables ou s'ils sont égaux (comme c'est le cas pour la transition liquide-gaz), on n'a pas de paramètre d'ordre (bien que la densité pour la transition liquide-gaz puisse être considérée comme un paramètre d'ordre généralisé). Il est à noter que les transitions sans paramètre d'ordre sont toutes du premier ordre au sens d'Ehrenfest.

Pour ce qui est des transitions avec paramètre d'ordre, c'est à dire les transitions pour lesquelles le groupe de symétrie de la phase ordonnée est strictement inclus dans le groupe de symétrie de la phase désordonnée, on peut en distinguer deux types : celles où le paramètre varie continûment (premier ordre) et celles où il est discontinu (second ordre). Nous verrons dans le paragraphe suivant la notion de longueur de corrélation : la variation du paramètre d'ordre est liée à la variation de la longueur de corrélation. Si la longueur de corrélation tend vers l'infini lorsque l'on s'approche de la transition, cela signifie que la taille des îlots de fluctuations s'étend continûment pour atteindre le système tout entier, c'est à dire que le paramètre d'ordre varie continûment. Si par contre la longueur de corrélation ne diverge pas, cela signifie que le phénomène est en quelque sorte avorté et que le passage d'un groupe de symétrie à l'autre se fait de manière brutale : le paramètre d'ordre est discontinu.

D'autre part, un apport majeur de Landau a été également de caractériser les fonctions thermodynamiques comme fonctions du paramètre d'ordre : par exemple l'énergie libre est une fonctionnelle définie comme l'intégrale sur le système d'une "densité d'énergie libre" exprimée en fonction du paramètre d'ordre : $F(m) = \int f(m) dx$. Rappelons que $m = m(x)$ est une fonction extensive, à laquelle on peut conjuguer une variable intensive, généralement un champ extérieur (exemple de la magnétisation et du champ magnétique). Cela constitue une approximation qui lisse les fluctuations de l'énergie libre. Une autre approximation a consisté à ne faire dépendre f que des dérivées premières et secondes de m , et que f ne dépend de la variable H conjuguée de m que par le terme $-Hm$. Si $H=0$, la minimisation de F correspond alors à $\text{grad } m = 0$. Enfin, en insistant sur le caractère drastique des hypothèses de régularités nécessaires, Landau a supposé que F pouvait être développée en série de Taylor au voisinage de $m = 0 \rightarrow F(m) = F_0 + a m + A m^2 + B m^3 + C m^4$, avec les coefficients fonctions de P et T . Un raisonnement très simple sur des considérations de symétries et de minimisation de F , amène à la conclusion que a, A, B doivent être nuls : le point de transition est donc en m^4 . En fait, au voisinage du point de transition, on écrit : $F(m) = F_0 + A (T-T_c) m^2 + C m^4$, les termes impairs étant impossibles.



Phénomènes critiques, groupe de renormalisation, longueur de corrélation

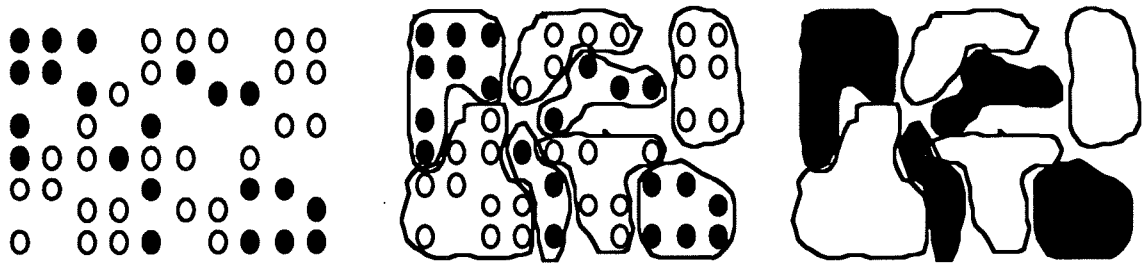
La théorie de Landau repose sur des hypothèses de régularité très fortes : or, au voisinage d'un point critique, ces hypothèses ne sont plus vérifiées, le potentiel thermodynamique n'est plus analytique. En fait, la théorie de Landau consiste à négliger les fluctuations à un endroit où elles deviennent précisément dominantes (défaut partagé par le champ moyen). L'explication de Landau est une explication de thermodynamique macroscopique et ne rend pas compte de la "raison microscopique" des transitions de phase (Petitot). On en vient là à la notion de longueur de corrélation, bien illustrée par les transitions magnétiques dans les verres de spins : lorsque l'on s'approche de la transition de phase, on a apparition d'îlots de spins prenant la même orientation par intercorrélations locales. On appelle longueur de corrélation la valeur moyenne l de la taille des ces îlots (on définit de la même manière le temps de corrélation comme le temps moyen de vie de tels îlots). Au voisinage d'un point critique, l diverge : les fluctuations deviennent macroscopiques et les corrélations locales se transforment en corrélations de longue portée. Cette insuffisance de la théorie de Landau s'est trouvée corroborée par l'expérience, en particulier par la détermination d'exposants critiques (reliés par des lois d'échelles) ne correspondant à ceux prédits par Landau. Pour répondre à cette insuffisance, on peut séparer trois étapes :

- L'hypothèse d'homogénéité de Widom : au voisinage d'un point critique, l'énergie libre est donnée par

$$F(T,m) = (T-T_c)^{2-a} f(m / (T-T_c)^b)$$

où f est une fonction homogène. On voit que l'énergie libre est alors composée d'une partie analytique et d'une partie non analytique.

- L'hypothèse de similarité de Kadanoff : on retrouve la forme postulée par Widom grâce à cette hypothèse d'invariance d'échelle (lois d'échelle entre les exposants critiques). Au voisinage de la criticalité, on a divergence en loi puissance des grandeurs thermodynamiques et la propriété qu'un hamiltonien local se transforme en hamiltonien global par l'ajustement de constantes de couplage renormalisées. L'idée de Kadanoff est illustrée par la figure suivante, qui indique comment on peut passer du local au global en "renormalisant", c'est à dire par exemple en regroupant des spins par blocs, et en changeant d'échelle pour traiter ces blocs de spins comme des spins individuels dont la valeur est la valeur moyenne du bloc.



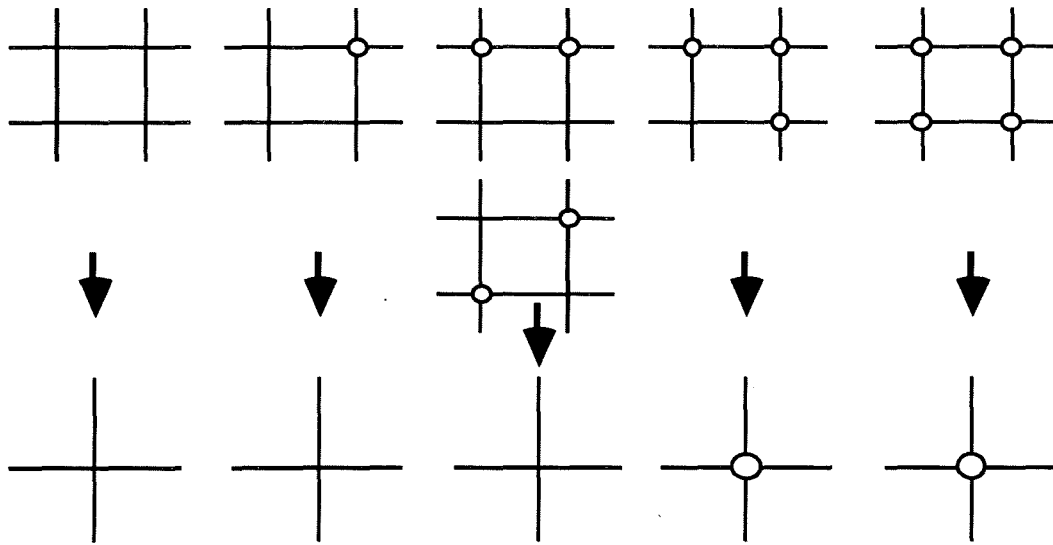
- Enfin, Wilson a formalisé de façon très rigoureuse les idées de Kadanoff, et a montré en particulier que l'on pouvait développer les exposants critiques en puissances de $\epsilon=4-d$ où d est la dimensionnalité du système.

Limite thermodynamique

La notion de limite thermodynamique est essentielle : tous les calculs qui sont faits et qui permettent d'établir des faits concernant les transitions de phase ne sont théoriquement valables que lorsque le nombre de particules tend vers l'infini. D'ailleurs, il a été observé qu'il n'y avait pas de transition de phase dans des verres de spins finis, ce qui est conforme aux calculs.

Un exemple simple et utile : la transition de percolation

La percolation permet de comprendre facilement certains concepts complexes. Imaginons un quadrillage carré sur lequel sont placés de façon aléatoire des atomes conducteurs (un site occupé par un atome conducteur est dit conducteur): la probabilité qu'un site du quadrillage soit occupé est p . On peut appliquer l'idée de renormalisation de la manière suivante : on se place au niveau supérieur, en prenant des blocs de quatre sites; si les quatre sites sont occupés, le bloc est conducteur, de même si trois sites sont occupés -> on dira donc que le bloc est "occupé" dans ces deux cas. Par contre, si deux sites ou moins de deux sites sont occupés, le bloc ne sera pas conducteur, ou ne sera conducteur que dans une direction-> on dira que le bloc n'est pas occupé. Lorsque $p=0$, la partie conductrice du treillis est vide. Lorsque l'on augmente p petit à petit, on observe l'apparition de zones conductrices. Enfin, lorsque p atteint une valeur critique p^* , un amas conducteur infini apparaît avec une probabilité 1. L'invariance d'échelle à la transition de phase (le treillis passe de non-conducteur à non conducteur) implique avec notre formation de blocs, que : $p^* = p^{*4} + 4 p^{*3} (1 - p^*)$, en effet, la probabilité qu'un bloc soit conducteur est donnée par $p^{*4} + 4 p^{*3} (1 - p^*)$, donc l'invariance de cette probabilité par changement d'échelle conduit à l'équation ci) dessus. La solution de cette équation est $p^*=0.76$ (on élimine les solutions triviales ou non physiques), à comparer au résultat expérimental $p^*=0.75$



Transitions de phase dans les systèmes dynamiques

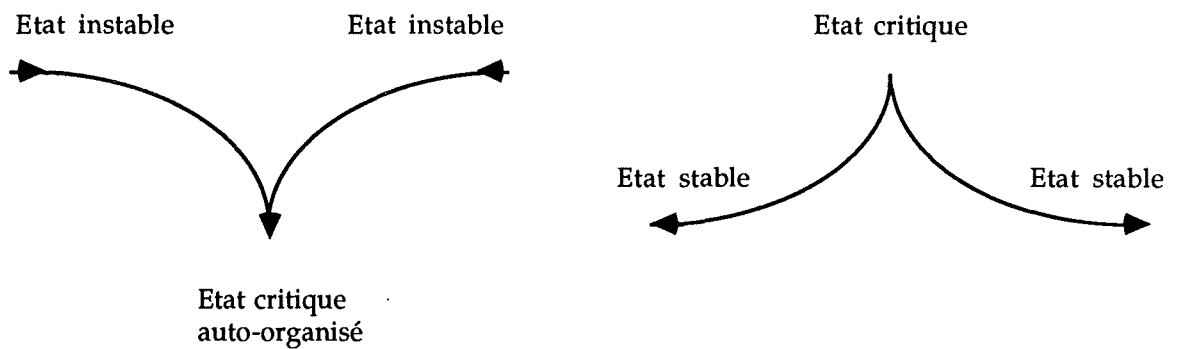
La phase qualifie en physique du solide un état de la matière, et un changement de phase correspond à un changement dans le comportement dynamique du système matériel considéré, bien que les problèmes liés à la dynamique des systèmes de physique du solide soient très mal compris. Même la transition de percolation peut revêtir une signification physique (par exemple état conducteur ou non conducteur). Si l'on veut appliquer la notion de transition de phase à un système dynamique, il faut donc également que l'on puisse avoir un changement qualitatif abrupt du comportement du système pour une certaine valeur dite critique d'un paramètre (qui joue un rôle équivalent à celui de la température en physique). Intuitivement, cela correspond pour un système dynamique à une bifurcation. Des systèmes aussi simples que des applications (unidimensionnelles) de l'intervalle peuvent subir une telle bifurcation dans leur comportement qualitatif. Par exemple, considérons la suite donnée par $X_{n+1} = k f(X_n)$, où f est simplement une fonction différentiable possédant un maximum v . Si on augmente progressivement la valeur de k depuis 0, on observe un changement de phase vers une dynamique chaotique lorsque k atteint une valeur critique. Cette route vers le chaos se fait par une série de doublements de périodes à des valeurs k_i du paramètre, jusqu'à ce que la période devienne infinie à la valeur critique. Si $f(v)-f(x) \sim |x-v|^z$ alors on a une universalité qui ne dépend que de z . Dans le cas où f est "quadratique", i.e $f(v)-f(x) \sim |x-v|^2$ avec $z=2$ l'invariance par changement d'échelle se traduit par un coefficient universel : $a=2.502907875\dots$, où a représente le rapport d'échelle entre deux doublements de périodes successifs, c'est à dire $a = k_i - k_{i-1} / (k_{i+1} - k_i)$. Cela correspond à une application directe du groupe de renormalisation. Le parallèle avec les transitions de phase est donc plus que superficiel.

Les dynamiques symboliques : l'espace de phase est partitionné, les éléments de la partition sont étiquetés au moyen de symboles, et la dynamique du système (discrétisé dans le temps) est une suite d'états successifs auxquels on fait correspondre le label de la zone à laquelle l'état appartient. On peut donc ainsi caractériser le système.

Etats critiques auto-organisés

La plupart des systèmes naturels sont trivialement dissipatifs (flux d'énergie, d'entropie, de matière, provenant de l'extérieur du système). Cela pose un certain nombre de problèmes, en particulier l'impossibilité de traitement simple par la mécanique statistique de Gibbs-Boltzmann. C'est là qu'intervient la notion de criticalité auto-organisée : ce paradigme introduit par Bak, Tang & Wiesenfeld en 1987, prétend unifier deux phénomènes fortement présents dans la nature : les structures fractales et le bruit en $1/f$. Ces deux propriétés correspondent à des lois d'échelles spatiales et temporelles, usuellement trouvées à des points de transitions de phase critiques dans les systèmes à l'équilibre. L'absence d'échelle spatio-temporelle caractéristique permet de réduire la description du système à quelques paramètres pertinents qui le décrivent complètement, macroscopiquement. Les transitions de phase critiques correspondent également à une divergence des longueurs et temps de corrélation, indiquant une portée infinie dans la transmission cohérente de perturbations. Les phénomènes d'opalescence critique (décroissance en puissance des corrélations spatiales) et de ralentissement critique (absence de temps caractéristique) y sont intimement liés. La notion d'état critique auto-organisé reprend ces idées mais s'applique à des systèmes dissipatifs. Il se trouve qu'un point critique est instable dans un système à l'équilibre, et seul le ralentissement critique, correspondant à un gel de la dynamique, permet d'utiliser le terme "état" critique. Or, dans le modèle proposé par BTW, l'état critique est attractif : la dynamique d'un système dissipatif à N corps (N grand) amène ce système dans un état métastable caractérisé par l'absence d'échelle de temps et l'apparition de structures fractales spatiales.

Le fait que l'on puisse parler d'état critique est en soi un pari, puisque la notion de phénomène critique implique l'idée de dynamique. L'état critique correspond à une transition de phase, qui comme le rappelle Stéphane Roux dans sa thèse, est une séparation fugace et ponctuelle entre deux états stables. Kenneth Wilson voyait trois états qualitativement différents de la matière : l'état ordonné (relativement à un certain paramètre), l'état désordonné (les gaz et liquides et les solides amorphes), et l'état critique, entre deux phases. Le terme d'état se comprend en présence de deux phénomènes suivants, ralentissement critique et attractivité du point critique pour certaines dynamiques. Au voisinage d'un point critique, la dynamique des systèmes est généralement très ralentie, ce qui confère à ce point en apparence sans épaisseur un statut d'état. L'exemple d'empilement donné par Per Bak pour illustrer l'idée d'état critique auto-organisé n'est pas unique et semble être applicable à d'autres phénomènes naturels, même s'il n'est pas toujours possible de construire une dynamique qui converge vers un état critique. Le bruit de scintillement peut également avoir des origines diverses, et l'auto-organisation critique n'est qu'une cause parmi d'autres. Quant aux phénomènes qui semblent bien se réduire à ce type d'explication, on peut citer par exemple la formation des séismes, des turbulences, l'agrégation limitée par diffusion, l'évolution des populations, certains aspects de la trame économique et la percolation. Dans tous les cas, les modèles utilisés sont simples. Peut-être trop simples. Il n'empêche que leur pouvoir prédictif n'est pas négligeable. L'idée est la même dans tous les cas : pour reprendre l'exemple du tas de sable, si l'on ajoute à un tas de sable des grains de sable un à un, le tas atteint une taille critique pour laquelle des avalanches ont lieu; ces avalanches peuvent avoir n'importe quelle amplitude et de n'importe quelle durée, et le tas de sable est toujours ramené à sa configuration critique, soit par l'ajout de grains de sable, soit par le déclenchement d'avalanches.



On voit donc que tout l'intérêt des états critiques auto-organisés réside dans le caractère attractif de l'état métastable critique pour la dynamique du système. Cette propriété caractéristique d'un certain nombre (la généralité de ces modèles fait l'objet d'un débat acharné) de systèmes dissipatifs possédant un grand nombre de degrés de liberté implique en particulier la possibilité pour un tel système de se maintenir au voisinage d'une zone critique. Les transitions de phase critiques de la physique de l'équilibre correspondent à des points mathématiques, par essence fugaces : si les phénomènes critiques sont d'une telle importance pour l'étude des systèmes naturels, une certaine robustesse et une certaine stabilité (en un sens à préciser) sont souhaitables, et l'on ne peut attendre de la nature qu'elle ait permis aux systèmes naturels de stationner en un point sans épaisseur et qui plus est instable (une petite perturbation fera basculer le système dans l'une ou l'autre des phases).

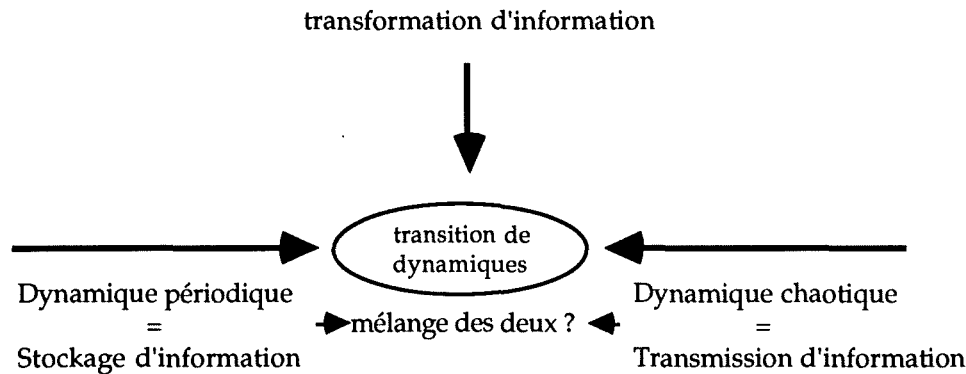
III Transitions de phase et phénomènes critiques: quel intérêt ?

Les transitions de phase du second ordre et les phénomènes critiques semblent présenter un intérêt immédiat, indépendamment de leur application spécifique aux systèmes collectifs traitant de l'information : en effet, de tels phénomènes ont pour propriété de générer des structures fractales dans l'espace et le temps, des lois d'échelles, et ce sont à ce jour presque les seuls phénomènes connus possédant cette propriété. Or on trouve les lois d'échelle partout dans la nature. La conclusion qui s'impose est que les phénomènes critiques sont sans doute très répandus. Dans la suite, on s'intéresse à des aspects plus proches du traitement de l'information.

a - Capacité de traitement de l'information aux transitions de phase

Le lien entre la capacité de traitement de l'information et les transitions de phase est relativement naturel et provient de la position intermédiaire par définition de la transition de phase entre deux milieux possédant des propriétés différentes. Prenons par exemple un système dynamique gouverné par un paramètre et présentant un transition vers le chaos (period-doubling, band-merging, crises, ...) : cette transition se fait à partir d'une dynamique d'attracteurs dont la période grandit (double par exemple) jusqu'à diverger pour une valeur critique du paramètre. Au-delà de la valeur critique, la dynamique du système est chaotique. Dans la dynamique périodique, les exposants de Lyapounov sont négatifs, ce qui induit un bon pouvoir de stockage de l'information, mais par contre aucune capacité de transmission de

cette information. De l'autre côté de la valeur critique, on a au contraire une très bonne transmission de l'information, puisque la moindre perturbation dans les inputs est amplifiée à l'infini dans le système, avec évidemment une absence de capacité de stockage.



On voit sur le schéma précédent que le problème est en quelque sorte "surdéterminé" : si on a bien rencontré de deux conditions nécessaires à la computation (stockage et transmission de structures) à la transition de phase, une troisième condition est nécessaire (le fait que ce groupe de conditions forme un ensemble de conditions suffisantes est un autre problème délicat : ces trois conditions doivent être remplies de manière cohérente), et concerne la transformation d'information. Il est évident que l'idée intuitive que l'on a de la computation implique de telles transformations d'information, qui doivent se traduire dans le cas d'un système physique par des modifications cohérentes de structures. L'indice qui laisse penser que cette condition peut être remplie au voisinage d'une transition de phase du second ordre est la divergence de la chaleur spécifique C_p . En effet, dans la formule

$$\Delta Q = C_p \Delta T$$

on voit que la chaleur spécifique correspond à l'énergie qu'il faut donner à un système pour augmenter sa température d'un degré. La divergence de la chaleur spécifique peut alors être interprétée de la façon suivante : l'énergie qui est fournie au système n'est pas utilisée pour faire augmenter sa température lorsque l'on se trouve au voisinage d'une transition de phase, mais pour transformer des structures en d'autres structures.

Au-delà de cette métaphore, on peut énoncer le résultat de Langton, qui s'inspirant du travail de Wolfram sur les automates cellulaires, et de ses propres observations (existence de structures se propageant et se réfléchissant sur des structures fixes), a démontré que les automates se situant au voisinage de la transition de phase avaient un pouvoir computationnel équivalent à celui d'une machine de Turing. Il est bien évident que cela ne signifie pas que les automates chaotiques n'aient pas également une telle puissance, mais seulement que la preuve constructive de Langton s'applique bien aux automates se situant près de la transition de phase. C'est cette preuve constructive qui est intéressante ici, puisqu'elle permet de montrer l'existence d'une puissance de calcul universel au voisinage de la transition : la variation de l'information mutuelle en fonction de l'entropie n'est pas en soi étonnante si l'on considère un automate cellulaire comme un système physique qui peut subir une transition de phase du second ordre, avec donc une divergence de la longueur de

corrélation (ce que traduit la divergence de l'information mutuelle d'une certaine manière). Ce qui est étonnant, c'est la concomitance entre cette transition de phase et la puissance de calcul universel.

Un autre indice présente également un intérêt certain vis à vis de la puissance de calcul : il s'agit d'une ressemblance phénoménologique entre la structure de l'espace des automates cellulaires et celle de l'espace des programmes lorsque le paramètre de description est lié au Halting problem : lorsque l'on s'approche de la transition de phase, les périodes transitoires sont de plus en plus longues, avant que l'automate ne prenne une dynamique bien "tranchée", périodique ou chaotique. Pour utiliser une image anthropomorphique, disons que l'on ne peut pas prédire si la période transitoire va s'arrêter. On voit bien là le parallèle avec le halting problem, puisqu'il existe des programmes pour lesquels on ne peut pas savoir s'ils s'arrêteront, lorsqu'on les fait débiter sur des inputs arbitraires. En comparant la classe des automates périodiques à celle des computations qui stoppent, et la classe des automates chaotiques à celles des computations qui ne stoppent jamais, on peut prolonger cette comparaison, en mettant en parallèle la classe des automates à périodes transitoires divergentes avec celle des computations dont on ne peut pas dire si elles s'arrêteront. Le parallèle est également troublant, entre la notion d'indécidabilité et celle de ralentissement critique (phénomène observé lors des phénomènes critiques). En fait, les physiciens qui s'intéressent aux liens entre computation et systèmes physiques font de l'existence de périodes transitoires infinies une condition nécessaire pour qu'un système physique soit doté d'une capacité universelle de traitement de l'information. C'est ainsi que les relaxations lentes (relaxations vitreuses ...) prennent un intérêt tout particulier.

b - Transitions de phase et organisation

On peut réconcilier la théorie d'Atlan concernant le contenu informationnel avec des notions plus éclairantes : en effet, pour Atlan, en partant de la relation $R = 1 - [H(x,y)/H(x)H(y)]$, on en déduit que $H(x,y) = (1-R)[H(x)+H(y)]$, ce qui signifie que lorsque R diminue, sous l'effet du bruit, $H(x,y)$ augmente. Il est évident que si l'on appelle $H(x,y)$ le contenu informationnel de (x,y) , on en déduit que le contenu informationnel de (x,y) augmente sous l'effet du bruit (principe d'organisation par le bruit). Mais on peut donner le nom que l'on veut à $H(x,y)$, il n'en demeure pas moins que $H(x,y)$ mesure le désordre de (x,y) , ou en terme d'information, l'information moyenne qu'apporte une mesure sur le système, ou l'incertitude que l'on a sur le comportement du système : bref, le théorème que l'on a alors s'énonce par "le désordre augmente sous l'effet du bruit", ce qui apparaît comme une trivialité. Si l'on traduit $H(x,y)$ en terme de taux de diversité dans les comportements possibles du système -ce qui d'une certaine manière est le cas-, alors on aboutit cette fois à l'énoncé "le taux de diversité" du système augmente sous l'effet du bruit. Cette manière de présenter la chose possède l'avantage d'être à la fois non-triviale et vraie dans l'acception courante du terme "diversité". Il est évident que lorsqu'un système est dans une zone de dynamique chaotique, il explore un grand nombre d'états : son taux de diversité est donc élevé. Mais cela ne veut pas pour autant dire qu'il s'agit là d'un avantage pour le système, comme l'a pressenti intuitivement Atlan en disant que les comportements viables du système se situent entre le cristal et la fumée, au centre de l'échelle entropique. Pourtant ces équations simples sur $H(x,y)$ ne confirment pas cette intuition. C'est la notion de transition de phase qui permet de prendre conscience de l'intérêt du concept de bruit organisateur : il n'y a qu'à la transition de phase que le bruit est utile. Pourquoi ? Parce que, comme nous l'avons vu, la chaleur spécifique y diverge, montrant que l'énergie ΔQ n'y est pas

utilisée pour augmenter la température, mais pour modifier des structures. Les dynamiques ordonnées ont de bonnes capacités de stockage de l'information, mais les perturbations sont étouffées, l'information n'est pas transmise : le bruit ne peut être organisateur. Par contre, si les dynamiques chaotiques correspondent à un grand pouvoir de traitement de l'information -en ce sens que la positivité d'un ou de plusieurs exposants de Lyapounov entraîne la propagation et l'amplification à l'infini de la moindre perturbation-, elles ne permettent pas le stockage de données. Enfin, aucun de ces deux types de dynamiques ne permet la transformation du bruit en énergie structurante. Bref, seuls les voisinages de transitions de phase permettent l'organisation par le bruit. Le compromis réalisé à la transition de phase est clair : on a d'un côté ce qu'Atlan qualifie de conditions nécessaires pour l'existence d'un processus d'auto-organisation, à savoir une certaine redondance initiale ("réservoir pour l'organisation") et une certaine stabilité vis à vis des perturbations (sans laquelle le système serait détruit par la moindre perturbation); et de l'autre côté, on a la diversité nécessaire à la richesse exploratoire du système, et à la création de significations.

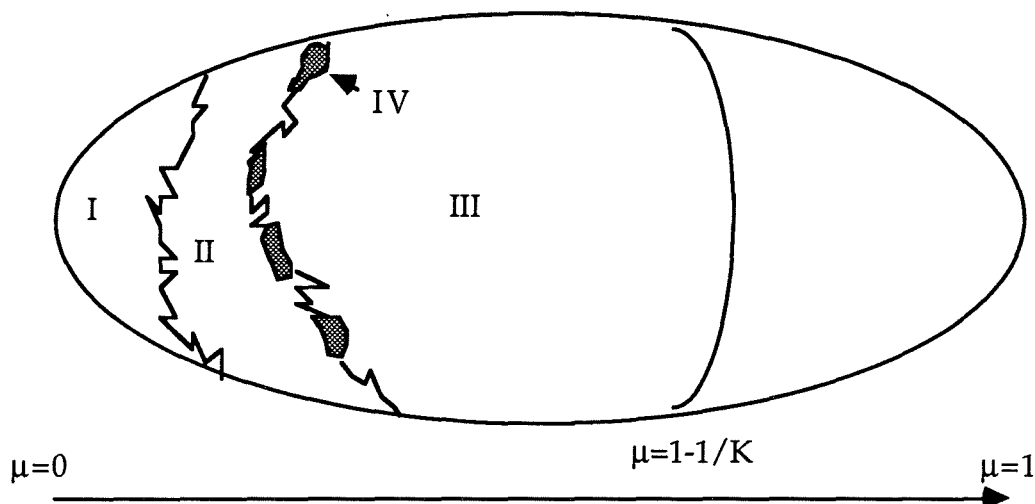
c - Classification des automates cellulaires

Les automates cellulaires ont été très étudiés depuis leur introduction par John von Neumann il y a quarante ans - qui voulait prouver l'existence d'un ordinateur universel auto-reproducteur. S'ils ont captés l'attention d'un grand nombre de chercheurs c'est sans aucun doute pour la richesse et la diversité de leur phénoménologie (il faut ajouter pour ces dernières années l'essor des modèles sur réseau), figurées par l'exemple frappant et célèbre du jeu de la vie de Conway. Parmi toutes les règles qu'il est possible d'imposer à un automate cellulaire, un petit nombre seulement -pour une topologie et un ensemble d'états fixés et isotropes - aboutissent à une dynamique complexe possédant la richesse du jeu de la vie. Ce n'est que depuis quelques années (1989-1990) que l'on sait caractériser par une grandeur mathématique "l'intérêt" dynamique d'une règle. Ce paramètre, qui ordonne partiellement l'espace des automates cellulaires, a permis d'observer une transition de phase lorsqu'on le fait varier (observation faite en mesurant l'information mutuelle en fonction du paramètre qui se trouve être une fonction croissante de l'entropie).

Stephen Wolfram a proposé une classification des comportements des automates cellulaires (discrets, déterministes). Cette classification comporte quatre groupes de comportements : la classe I regroupe les dynamiques possédant un point fixe, la classe II regroupe les dynamiques à orbites périodiques, la classe III regroupe les dynamiques chaotiques -il est à noter que stricto sensu toutes les orbites d'un automate déterministe à nombre d'états fini sont périodiques : la notion de chaos est à prendre ici au sens large, c'est à dire que lorsque la taille des orbites devient suffisamment grande, le comportement est dit chaotique -, et la classe IV regroupe des dynamiques que Wolfram qualifie de 'complexes', sans vraiment en comprendre les structures. Les dynamiques de la classe IV sont caractérisées par de très longues périodes transitoires et des patterns spatio-temporels complexes possédant des structures oscillantes et se propageant. Chris Langton a introduit un paramètre permettant d'élucider au moins partiellement le rapport entre le comportement observé et la nature de la règle de transition dans le cas des automates cellulaires à une dimension. Il a également observé que pour la classe IV, l'entropie moyenne est relativement faible, le taux de diffusion est quasiment nul, et l'information mutuelle présente un pic (information mutuelle temporelle au sens de Shannon). C'est justement l'information mutuelle qui semble constituer le cœur des systèmes

organisés : il est évident qu'un système structuré, organisé, présente de la cohésion et de la diversité, ce qui est possible lorsqu'un compromis est établi entre l'ordre et le chaos, lorsque les éléments du système sont corrélés sans l'être trop.

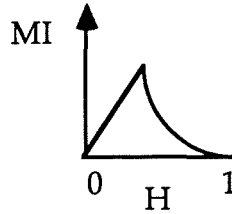
Dans son partitionnement de l'espace des automates cellulaires, Chris Langton a remarqué qu'un tout petit nombre de règles aboutit à une dynamique 'complexe'. Le schéma suivant (pour lequel l'axe vertical correspond à une grandeur qui n'a pas encore été déterminée) indique la structure approximative de cet espace.



Le paramètre μ utilisé par Langton est le pourcentage d'entrées dans la table d'une règle, qui n'ont pas pour sortie correspondante un état nul prédéfini. Si l'on considère un automate cellulaire à une dimension, à k états et possédant une topologie à voisinages de rayon r ($= 2r+1$ voisins), si S_0 est l'état 'nul' fixé, et si enfin M entrées de la table d'une règle (parmi les $T=[k]^{2r+1}$ entrées) ont pour sortie correspondante un état différent de S_0 , alors le paramètre μ a pour valeur pour cette règle $\mu=M/T$. Il est évident que cette grandeur n'est pas bi-univoque, et que d'autre part, on a une symétrie de l'espace des automates cellulaires à $\mu=1-1/K$. L'entropie moyenne, mesurée en fonction de μ , présente une discontinuité pour certaines valeurs critiques, indiquant une transition vers le chaos : il s'agit là d'une transition qui correspond en physique statistique à une transition de phase du premier ordre. Dans un petit nombre de cas, cependant, on a une variation plus régulière de l'entropie, qui présente un saut continu, comme c'est le cas pour des transitions de phase du second ordre. L'interprétation de ce résultat est donnée par le schéma précédent, qui indique que des chemins permettent de passer de la région II à la région III en passant par une région critique. Ce passage n'est cependant pas obligatoire, ce qui rend la comparaison avec la phénoménologie des transitions de phase encore plus aigüe : on peut passer de l'état solide à l'état gazeux sans passer par la phase liquide, mais on peut également passer par la phase liquide (il est à noter que si les gaz et les liquides sont appelés fluides, les gaz -et les solides- sont aujourd'hui mieux modélisés que les liquides).

L'information mutuelle a également été mesurée. L'information mutuelle permet d'indiquer le degré de corrélation de deux variables A et B : en théorie, $IM(A,B)=IM(B,A)=H(A)+H(B)-H(A,B)$, où H désigne l'entropie, et $H(A,B)$ désigne l'entropie jointe. En pratique, la mesure de la corrélation est donnée par $M=\sum_{i,j} P_{ij} \log [P_{ij}/P_i P_j]$. L'entropie, comme l'information mutuelle, peut être

directionnelle, c'est à dire spatiale ou temporelle, ou spatio-temporelle. L'information mutuelle spatiale bloc à bloc et l'information mutuelle temporelle d'un instant à l'autre sur un même site, donnent le même type de résultat sous la forme d'une courbe entropie/information mutuelle :



Si l'on mesure l'information mutuelle en fonction de μ , ou plutôt en fonction de $\Delta\mu = \mu - \mu_c$, où μ_c est la valeur critique de μ pour un chemin donné vers le chaos, on a $M(|\Delta\mu|) \sim e^{-c|\Delta\mu|}$. Si l'on compare l'information mutuelle (une différence d'entropies) à la chaleur spécifique ($=\partial H/\partial T$) en physique statistique -ce qui a déjà été fait pour des verres de spins bidimensionnels-, on peut déterminer l'ordre de la transition en étudiant le comportement de l'information mutuelle. On peut monter d'autre part, en utilisant la théorie du champ moyen, que $\mu_c = 1/2 - 1/2 [1 - 2/(r+1)]^{1/2}$. Donc, plus le voisinage est grand, plus la transition se fait à faible μ , i.e à faible entropie. Autre constatation, la probabilité que la transition se fasse en passant par la phase complexe est très petite. On retrouve des aspects familiers et classiques des transitions de phase du second ordre en se souvenant que les corrélations spatiale aussi bien que temporelle décroissent très lentement à la transition de phase : or l'information mutuelle constitue une mesure de corrélation dont on peut montrer qu'elle est plus performante que les mesures classiques de corrélation (a- l'information mutuelle capture également la partie non-linéaire de la dynamique, et b- l'information mutuelle peut s'appliquer à des dynamiques symboliques, alors que les corrélations classiques dépendent fortement des valeurs). Donc, l'accroissement du taux d'information mutuelle (constaté dans le temps comme dans l'espace) rend très intéressant le parallèle avec les transitions de phase.

Chris Langton a mis en évidence des structures fixes, oscillantes ou se propageant dans les automates de la classe IV. Ces structures (murs de réflexion, solitons,...) supportent un pouvoir de computation universel comme le montre la procédure constructive de Langton. Pour être exact, il n'a montré le pouvoir de computation universelle que pour un petit nombre de règles de la classe IV, mais l'intuition d'une extension de cette propriété à l'ensemble de la classe IV semble être bonne. D'autre part, si une procédure constructive positive permet de conclure en faveur du pouvoir de computation universelle, il n'en est pas de même pour une procédure constructive qui échoue. Donc, on ne peut pas conclure sur le pouvoir de computation des automates des autres classes, d'autant que la facilité que l'on a à montrer le pouvoir universel pour ceux de la classe IV peut parfaitement être lié à la méthode employée. A ce propos d'ailleurs, la méthode employée est celle de Von Neumann, Wolfram, etc... qui procède par reconstruction de fonctions logiques. Il n'est pas certain -c'est un euphémisme- que la manière de "computer" de la nature passe par ces voies détournées. Toutefois, montrer l'accroissement soudain et drastique d'une capacité à traiter de l'information conduit à une vision fonctionnelle des transitions de phase, dans laquelle les structures complexes (complexité dans l'espace et dans le temps) qui apparaissent supportent effectivement une fonction complexe, celle de traitement de l'information au sens -mal- défini dans l'introduction.

d - Caractérisation des transitions de phase dynamiques par des hiérarchies d'automates

I) Complexité de KCS : La complexité de Kolmogorov-Chaitin-Solomonoff (KCS) est une alternative à certaines des limitations de la théorie classique de l'information, en ce qu'elle permet de mesurer (théoriquement) la complexité de données indépendamment d'un ensemble de tous les messages. Si D est la donnée dont on veut mesurer la complexité, alors KCS(D) est la représentation minimale de D sur une machine de Turing universelle (MTU), c'est à dire la taille du plus petit couple (Programme+Données) nécessaire à la reproduction de D au moyen d'une MTU. Mais KCS n'est pas non plus une grandeur adaptée à la mesure de la complexité telle que nous en avons l'intuition : en effet KCS est maximale lorsque la donnée est parfaitement aléatoire, comme la chaîne de bits produite par un pile ou face parfait. [La description MTU la plus courte d'une telle chaîne de 0 et de 1 consiste à faire PRINT (Data), c'est à dire à recopier totalement la chaîne.] Or, un pile ou face ne nous paraît pas complexe (un indice : la chaîne générée n'est pas structurée), et d'autre part est facile à modéliser mathématiquement. Enfin, KCS n'est pas discriminante : sur les 2^n chaînes de bits de longueur n, il y en a $2^{n-k+O(1)}$ qui vérifient l'inégalité $KCS \leq \text{Max}(n) - k$ [où Max est la complexité maximale pour une chaîne de taille n : on a $\text{Max}(n) = n + \log(n) + O(1)$]. On a donc exponentiellement peu de chaînes qui ne soient pas de complexité maximale.

On parle souvent de complexité algorithmique au sujet de KCS. C'est globalement vrai, à ceci près que la complexité algorithmique stricto sensu est définie comme le taux de croissance de KCS lorsque la taille de la donnée tend vers l'infini :

$$\text{Alg} = \lim_{t \rightarrow \infty} \text{KCS}(X_t) / t$$

La pertinence de cette grandeur dépend de la nature de la source qui génère la donnée : il est nécessaire que cette dernière soit stationnaire, alors que KCS est valable dans tous les cas. Une décroissance de KCS à partir d'une certaine taille peut indiquer un changement dans le comportement de la source, aussi bien que l'apparition d'une symétrie de longue portée. Si l'on définit l'entropie par blocs

$$H_n = \sum_{\{S\}} P_n(S_1, \dots, S_n) \log P_n(S_1, \dots, S_n)$$

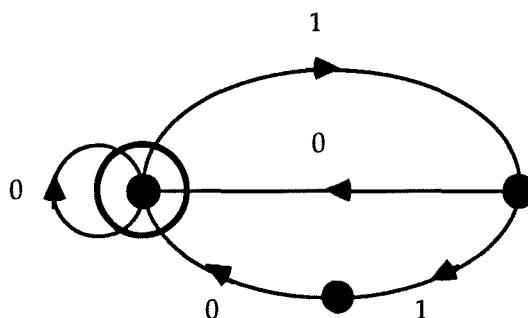
où les S_i sont les symboles composant la donnée, alors on a la propriété suivante : $h = \lim_{n \rightarrow \infty} H_n/n$, et h et Alg sont essentiellement identiques. La différence entre les deux vient de ce que Alg mesure l'information nécessaire pour spécifier les n premiers bits, tandis que h mesure l'information nécessaire en moyenne pour spécifier n'importe quelle série de n bits, n tendant dans les deux cas vers l'infini. Dans le cas d'une source stationnaire, cette différence n'existe pas.

II) Une grandeur plus générale : Cette grandeur KCS, qui, bien que non calculable, mesure le degré d'aléatoire de données, est l'instanciation d'une grandeur relative. En effet, on peut généraliser la notion de complexité KCS en introduisant la complexité conditionnelle par rapport à un ensemble de symétries. Si S est une symétrie, où un moyen de détecter des régularités dans la donnée D, alors on définit $C(D/S)$ comme l'information contenue dans les classes d'équivalence induite par S sur D + la quantité d'information contenue dans D et qui n'est pas explicable par S. Pour rendre les choses plus claires, disons que $C(D/S)$ est la plus petite représentation de D sur une machine computant la symétrie S. En utilisant la notation $C(D/S)$, on peut définir KCS : $\text{KCS}(D) = C(D/\{\text{MTU}\})$ où $\{\text{MTU}\}$ désigne

l'ensemble de toutes les symétries dont une MTU peut tirer parti (toutes les symétries computables par une MTU). Cette vision des choses a amené Crutchfield à poser comme hypothèse que la complexité "naturelle" est bien cernée par $C(D)=C(D/\{MBT\})$, où $\{MBT\}$ désigne l'ensemble des symétries computables par une MBT, une "Machine de Bernoulli-Turing" qui est une MTU augmentée d'un registre aléatoire : les symétries déterministes ne sont plus les seules computables, il faut également tenir compte des symétries statistiques. L'observation qui a débouché sur cette définition vient de la facilité que l'on a à modéliser à la fois les processus prédictibles (Pt, une horloge) et les processus parfaitement aléatoires (Bt, un pile ou face parfait). [Il faut noter qu'à l'égard des mesures de complexité faites sur des dynamiques symboliques de systèmes dynamiques, chaos et aléatoire ont le même statut.] Pour une MBT, ces deux types de processus sont effectivement les plus simples à reproduire, ce qui place le maximum de la complexité C au centre de l'échelle entropique, une situation qui correspond mieux à notre intuition. D'autre part, l'une des faiblesses de la complexité KCS, le fait que KCS mesure la difficulté de reproduire une donnée de manière exacte -en tenant compte même des détails "non significatifs", sans importance-, est éliminée du fait de l'acceptation de symétries statistiques comme éléments de régularité dans la donnée.

Etant donné que la complexité $C(D/\{MBT\})$ n'est pas plus calculable que KCS -oserai-je dire qu'elle est encore moins ?-, Crutchfield a proposé une méthode constructive permettant d'approximer cette complexité. Pour comprendre l'idée de Crutchfield, il faut tout d'abord introduire la notion de complexité des grammaires.

III) Complexité des grammaires : L'idée de définir la complexité d'une grammaire est apparemment simple. Du moins pour ce qui est des langages "simples" de la hiérarchie de Chomsky. Rappelons qu'un ensemble de chaînes composées sur un alphabet fini (chaînes de bits par exemple) est appelé un langage formel. La hiérarchie de Chomsky distingue plusieurs niveaux de complexité dans les langages qui peuvent être réguliers, context-free, context-sensitive ou récursivement énumérables (ils se différencient par la généralité des règles de leurs grammaires et par la difficulté qu'il y a à tester l'appartenance d'une chaîne donnée au langage). Les langages de programmation sont des langages réguliers : par définition, ils sont représentés par des graphes déterministes orientés finis. Dans un tel graphe, les liens sont étiquetés par des symboles de l'alphabet, et chaque symbole apparaît au plus une fois sur l'ensemble des liens quittant une node donnée. La node de départ est unique. Une chaîne de caractères appartenant au langage correspond alors à un chemin unique sur le graphe, au contraire d'une chaîne incorrecte.



Par exemple, dans le graphe ci-dessus, l'ensemble des chaînes de caractères permises est composé de toutes les chaînes de caractères (0 et 1) sauf celles qui contiennent trois 1 consécutifs.

On peut alors définir la complexité d'un langage (ou plutôt de sa grammaire) comme la taille de l'automate minimal qui lui est associé. On pense donc au nombre de nodes, auquel on ajoute le nombre de liens, ce qui correspond intuitivement à la difficulté d'écrire les règles de grammaire. Pourtant, la mesure généralement utilisée ne tient pas compte des liens : Wolfram en particulier a introduit

$$RLC = \log n$$

où RLC désigne "Regular Language Complexity" et n le nombre de nodes du graphe. En fait, RLC se comprend comme la complexité d'évaluation du graphe lorsque l'on veut vérifier la validité d'une chaîne. En effet, pour une telle évaluation, il suffit de mémoriser le numéro de la node courante puis d'utiliser une table de look-up pour déterminer la prochaine node à choisir. S'il y a équiprobabilité des nodes, le temps moyen nécessaire pour choisir une node parmi n est $\log n$.

Si par contre les nodes ne sont pas équiréparties, c'est à dire si l'on dispose d'une distribution de probabilités stationnaire $\{P_k\}$ indiquant la probabilité d'être à la node k à n'importe quel instant, cette distribution apporte une information permettant de réduire le temps moyen d'évaluation, donc l'information nécessaire sera

$$SC = - \sum_{k=1}^n P_k \log P_k$$

où SC désigne "Set Complexity", la complexité par rapport à un ensemble de nodes affectées de probabilités. On voit que SC est une version réduite de RLC, puisque SC atteint son maximum, RLC, lorsque les nodes sont équiprobables. RLC et SC sont toutes deux finies pour des langages réguliers. Mais, alors que RLC n'est évidemment pas définie pour des classes supérieures de la hiérarchie de Chomsky, il n'en va pas nécessairement de même pour SC : les langages context-free et context-sensitive peuvent être représentés par des graphes infinis, et la probabilité qu'une node soit atteinte à n'importe quel instant diminue de manière drastique avec l'éloignement de la node à la node initiale; cette décroissance peut se révéler si rapide que SC est finie.

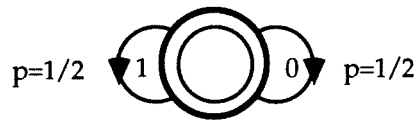
IV) Complexité de Crutchfield-Young : Avec les outils du paragraphe précédent, on peut comprendre très simplement l'esprit de la complexité de Crutchfield, qui est une mesure constructive qui se veut une approximation de $C(\cdot / MBT)$: il est évident que la notion de MBT est abstraite, et qu'une MBT est irréalisable (une MTU est également irréalisable "en laboratoire" stricto sensu, puisqu'elle suppose une bande infinie). Mais on peut, afin d'approximer une mesure théorique fondée sur un outil inexistant, réduire l'outil théorique et choisir un outil pratique évidemment bien plus faible. Crutchfield a choisi comme outil de substitution des automates probabilistes : en effet, un système dynamique auquel est associé une dynamique symbolique peut être vu comme un producteur de langages qui sont représentés par des graphes. Si l'on varie le paramètre d'ordre du système (par exemple le paramètre r dans la fonction logistique $r X (1-X)$), le langage devient de plus en plus diversifié au fur et à mesure que l'on s'approche du chaos, et la taille des automates (RLC) mesure cette diversification. L'idée d'utiliser des automates probabilistes au lieu d'automates déterministes correspond à une volonté de rendre l'aléatoire simple. Par exemple, une chaîne parfaitement aléatoire sera représentée par le graphe

Il existe un certain nombre de modèles théoriques ou computationnels de systèmes distribués.

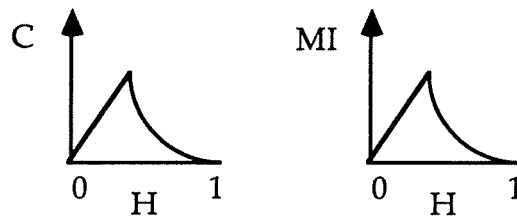
Références

- [Atlan 72] - H. Atlan, "L'organisation biologique et la théorie de l'information", Ed. Hermann, Paris, 1972
- [Atlan 74] - H. Atlan, "On a formal definition of organization", Journal of theoretical Biology, 45, pp.295-304, 1974.
- [Atlan 79] - H. Atlan, "Entre le cristal et la fumée", Seuil, Paris, 1979.
- [Atlan 85] - H. Atlan, "Auto-organisation fonctionnelle et création de signification", Cognitiva Cesta, Paris, June 1985.
- [Atlan 87] - H. Atlan, "Self-Creation of Meaning", Physica Scripta, 36, 563-576, 1987.
- [Auger 90] - P. Auger, "Dynamics and thermodynamics in hierarchically organized systems", ISFR International Series on Systems Science and Engineering, vol.V, Pergamon Press, 1990.
- [Bak & al. 88] - P. Bak, C. Tang, K. Wiesenfeld, "Self-organized criticality", Physical Review A 38, 364-374, 1988.
- [Bennett 82] - C.H. Bennett, "The Thermodynamics of Computation - A Review", International Journal of Theoretical Physics, 21, 905-940, 1982.
- [Bennett 87] - C.H. Bennett, "Dissipation, Information, Computational Complexity and the Definition of Organization", in 'Emerging Syntheses in Science', ed. D. Pines, Addison-Wesley, 1987.
- [Bennett 88] - C.H. Bennett, "Logical Depth and Physical Complexity", in 'The Universal Turing Machine, A Half-Century Survey', ed. R. Herken, Oxford University Press, 227-258, 1988.
- [Bennett 89] - C.H. Bennett, "Computational Measures of Physical Complexity", Lectures in the Sciences of Complexity, Ed. D. Stein, Addison-Wesley, 1989.
- [Bennett 90] - C.H. Bennett, "How to Define Complexity in Physics, and Why", in 'Complexity, Entropy and the Physics of Information', ed. W.H. Zurek, SFI, Addison-Wesley, 1990.
- [Bourgine & al. 92] - P. Bourguin, F. Varela, "Towards a practice of autonomous systems", Proceedings of ECAL 91, MIT Press 1992.
- [Campbell 89] - D. Campbell, "Introduction to Nonlinear Phenomena", Lectures in the Sciences of Complexity, Ed. D. Stein, Addison-Wesley, 1989.
- [Chaitin 75] - G. Chaitin, "Randomness and Mathematical Proof", Scientific American 232, 47-52, 1975.
- [Chaitin 79] - G. Chaitin, "Toward a Mathematical Definition of Life", in 'The Maximum Entropy Formalism', MIT Press, 477-498, 1979.
- [Chaté & al. 90] - H. Chaté, P. Manneville, "Criticality in cellular automata", Physica D 45, North-Holland, 1990.
- [Chauvet 87] - G. Chauvet, "Traité de Physiologie théorique", Masson, 1987.
- [Crutchfield & Young 90] - J. Crutchfield, K. Young, "Computation at the onset of chaos", Complexity, Entropy and the physics of Information, SFI studies in the Sciences of Complexity, vol.III, ed. W. Zurek, Addison-Wesley, 1990.
- [Eubank & al. 90] - S. Eubank, J.D. Farmer, "An Introduction to Chaos and Randomness", Lectures in the Sciences of Complexity, vol.II, Ed. E. Jen, Addison-Wesley, 1990.
- [Farmer 90] - J.D. Farmer, "A Rosetta Stone for Connectionism", Physica D 42, North-Holland, 1990.
- [Fogelman 91] - F. Fogelman Soulié, "Les théories de la complexité", Seuil, Paris, 1991.
- [Grassberger 86] - P. Grassberger, "Toward a Quantitative Theory of Self-Generated Complexity", International Journal of Theoretical Physics, 25, (9), 907-938, 1986.
- [Grassberger 89] - P. Grassberger, "Problems in Quantifying Self-Organized Complexity", Helvetica Physica Acta, 62, 498-511, 1989.
- [Hogg 90] - T. Hogg, "The Dynamics of Complex Computational Systems", in 'Complexity, Entropy and the Physics of Information', ed. W.H. Zurek, SFI, Addison-Wesley, 1990.
- [Huberman & al. 87] - B. Huberman, T. Hogg, "Phase transitions in artificial intelligence systems", Artificial Intelligence, 33, 155-171, 1987.
- [Huberman 90] - B. Huberman, "The Performance of Cooperative Processes", Physica D 42, North-Holland, 1990.
- [Kampis 91] - G. Kampis, "Emergent Computations, Life and Cognition", in World Futures, vol.31, pp 33-48, Gordon and Breach Science Publishers S.A, 1991.
- [Kauffman 89 a] - S. Kauffman, "Adaptation on Rugged Fitness Landscapes", Lectures in the Sciences of Complexity, Ed. D. Stein, Addison-Wesley, 1989.
- [Kauffman 89 b] - S. Kauffman, "Principles of Adaptation in Complex Systems", Lectures in the Sciences of Complexity, Ed. D. Stein, Addison-Wesley, 1989.

- [Kauffman 90] - S.Kauffman, "Requirements for Evolvability in Complex Systems: Orderly Dynamics and Frozen Components", in 'Complexity, Entropy and the Physics of Information', ed. W.H.Zurek, SFI, Addison-Wesley, 1990.
- [Kephart & al. 90] - J.Kephart, T.Hogg, B.Huberman, "Collective behavior of predictive agents", Physica D 42, North-Holland, 1990.
- [Koppel 88] - M.Koppel, "Structure", in 'The Universal Turing Machine, A Half-Century Survey', ed. R.Herken, Oxford University Press, 435-452, 1988.
- [Küppers 91] - B.-O.Küppers, "Information and the origin of life", MIT Press, 1991.
- [Langton 86] - C.Langton, "Studying Artificial Life with Cellular Automata", Physica D 22, North-Holland, 120-149, 1986.
- [Langton 89] - C.Langton, ed., "Artificial Life", Addison-Wesley, 1989.
- [Langton 91] - C.Langton, "Computation at the edge of chaos : phase transitions and emergent computation", PhD Dissertation, University of Michigan, 1991.
- [Langton 90] - C.Langton, "Computations at the edge of chaos", Physica D 42, S.Forrest, ed., North-Holland, 1990.
- [Langton 91] - C.Langton, C.Taylor, D.Farmer, S.Rasmussen, eds., "Artificial Life II", Addison-Wesley, 1991.
- [Li 90] - W.Li, N.Packard, C.Langton, "Transition phenomena in cellular automata rule space", Physica D 45, North-Holland, 1990.
- [Li 91] - W.Li, "On the Relationship between Complexity and Entropy for Markov Chains and Regular Languages", Complex Systems, 5, 381-399, 1991.
- [Lloyd 90] - S.Lloyd, "Valuable Information", in 'Complexity, Entropy and the Physics of Information', ed. W.H.Zurek, SFI, Addison-Wesley, 1990.
- [Mc Intosh 90] - H.Mc Intosh, "Wolfram's class IV and a good life", Physica D 45, North-Holland, 1990.
- [Mc Lennan 90] - B.Mc Lennan, "Evolution of communication in a population of simple machines", TR-UTK-CS-90.
- [Millonas 92] - M. Millonas, "Swarms, phase transitions and collective intelligence", Draft submitted to the proceedings of AL III.
- [Opper & al.] - M.Opper, S.Diedrich, "Phase Transition and If noise in Game Dynamical Model", Phys. Rev. Lett, vol. 69, 10, 1616-1619, 1992.
- [Perelson 90] - A.Perelson, "Theoretical Immunology", Lectures in the Sciences of Complexity, vol.II, Ed. E.Jen, Addison-Wesley, 1990.
- [Rasmussen & al. 90] - S.Rasmussen, C.Knudsen, R.Feldberg, M.Hindsholm, "The Coreworld: emergence and evolution of cooperative structures in a computational chemistry", Physica D 42, North-Holland, 1990.
- [Verschure 91] - P.Verschure, "Chaos-based Learning", Complex Systems, 5, 359-370, 1991.
- [Weisbuch 89] - G.Weisbuch, "Dynamique des systèmes complexes", Interéditions / CNRS, 1989.
- [Winfree] - A.Winfree, "The Geometry of Biological Time", Biomathematics vol.8, Springer Verlag.
- [Wootters & al. 90] - W.Wootters, C.Langton, "Is there a sharp phase transition for deterministic cellular automata ?", Physica D 45, North-Holland, 1990.



et la complexité de la chaîne est donc nulle. Crutchfield et Young, en appliquant cette méthode à la dynamique symbolique d'une application de l'intervalle (application logistique bien connue pour son diagramme de bifurcation) $X_{n+1} = r X_n (1 - X_n)$, a obtenu une figure fort similaire à celle de Langton mesurant l'information mutuelle des automates cellulaires. Dans la figure ci-dessous, C désigne la complexité de Crutchfield, tandis que MI désigne l'information mutuelle. H est l'entropie du système (entropie métrique dans la figure de gauche, et entropie de Shannon dans la figure de droite).



C étant calculée comme fonction du nombre de nodes de l'automate fini stochastique permettant de rendre compte de la dynamique du système, un saut qualitatif à la transition de phase (i.e une divergence du nombre de nodes) correspond à un saut qualitatif dans la grammaire nécessaire pour rendre compte du langage symbolique généré par le système..

e - Evolution et co-évolution au bord du chaos

1 - L'idée de coévolution au bord du chaos a été avancée par Kauffman, en terme de dynamiques subcritique, critique et surcritique. Cette idée est simple, et se fonde sur l'autre idée que l'évolution simultanée d'espèces qui modifient mutuellement leurs fitness amènent à une complexification fonctionnelle, en particulier dans le cas de parasites et de leurs hôtes : afin de résister à un parasite, un organisme augmente certaines de ses capacités, mais le parasite, pour répondre à cette augmentation va lui-même augmenter ses capacités de résistance aux réponses, puis l'organisme va réagir à cette réponse, etc... Celà se traduit par une procréation accrue ou au contraire diminuée de l'organisme et du parasite. Les deux éléments se maintiennent l'un l'autre dans une zone critique, où des avalanches peuvent se produire dans un sens comme dans l'autre. Celà fait penser à certains modèles d'épidémiologie ou aux équilibres ponctués (de loin).

2 - Evolution vers le bord du chaos : Packard a montré qu'une population d'automates cellulaires évolue vers une limite composée exclusivement d'automates "complexes", si leur fitness est mesurée par leur performance dans la résolution d'un problème difficile (dans ce cas il s'agit d'automates cellulaires monodimensionnels à deux états et trois voisins et la tâche à accomplir consiste à décider si le vecteur d'entrée contient plus de 0 que de 1 : si le nombre de 0 est

supérieur, l'automate doit finalement converger vers un état nul, et si le nombre de 1 est supérieur, l'automate doit converger vers un état complètement actif; en cas d'égalité, une période transitoire infinie n'est pas exclue).

f - Transitions dans les systèmes coopératifs lorsque l'on varie leur connectivité

Il s'agit là d'un point de vue un peu différent sur les transitions de phase : lorsque l'on varie certains coefficients liés à la connectivité ou la coopérativité des éléments en interactions, on observe un changement qualitatif dans le comportement du groupe, et dans ses performances (bifurcations, très similaires aux structures dissipatives comme la convection de Bénard).

1 - Huberman et Hogg ont proposé le modèle suivant (transition de percolation dirigée : considérons un grand ensemble de processus interagissant et possédant seulement des interconnexions locales, sous la forme d'arbres dont les nœuds (les processus) sont liées avec une probabilité z . On s'intéresse au nombre moyen de processus qui peuvent être activés à partir d'un processus donné. La valeur critique de z est 1 : pour $z < 1$, l'activation est localisée finie, alors que pour $z > 1$, l'activation est infinie. En effet, soit $N(u)$ le nombre de sites à une distance u d'un site donné, comme le chemin qui va d'une node à une autre dans un arbre est unique, le nombre de sites à une distance u qui peuvent être atteints à partir du site initial est $N(u) = z^u$. Donc, la distance moyenne d'activation est donnée par $1/\langle N \rangle \sum u N(u)$ ou $\langle N \rangle$ est le nombre moyen de sites dans l'arbres $\langle N \rangle = \sum N(u) = 1/(1-z)$. Donc, $\langle u \rangle = z/(1-z)$, et donc l'activation moyenne se propage à l'infinie lorsque $z > 1$. On peut appliquer ce modèle à des recherches heuristiques dans des arbres (de profondeur d et de probabilité de branchement b). Il faut noter qu'on a évidemment des effets de taille finie dans les systèmes réels. Une connaissance parfaite nécessite de visiter seulement d nœuds, alors qu'une absence de connaissance implique en moyenne la visite de la moitié des nœuds, ce qui est exponentiel en d . On peut supposer que l'on a une situation intermédiaire, où l'on dispose à chaque branchement d'une heuristique permettant d'éliminer les mauvaises branches avec une probabilité $1-p$ (la connaissance parfaite correspond à $p=0$ et l'absence de connaissance à $p=1$). Donc la probabilité effective de branchement est donnée par $z=bp$. Le nombre moyen de nœuds visités est $N = d+1/2 (b-1) p \sum N(s)$, où $N(s)$ est le nombre de nœuds d'un sous-arbre de profondeur d . Si d est très grand, le nombre de nœuds visités à chaque étape est donné par $n=N/d = (2-z-p) / [2(1-z)]$. On observe donc une transition dans la performance de la recherche.

2 - Millonas a montré comment la dynamique d'une société de fourmis artificielles vivant sur un graphe peut subir plusieurs types de transitions de phase dans un sens très généralisé. D'autre part, le caractère "autocatalytique" du modèle lui confère certaines propriétés d'autoorganisation.

3 - Opper et Diedrich ont montré que dans un modèle d'espèces en interactions sous une "pression de coopération" u , on observe une transition vers une dynamique instable lorsque u atteint une valeur critique. Cette transition également par le fait que le système composé d'espèces manifeste un bruit en $1/f$ en réponse à des perturbations extérieures.

g - Quelques modèles théoriques : social impact, jeu de la vie, économie, synergétique, système immunitaire

Etats critiques auto-organisés

P. Bourgin
Labo de Vie et d'Intelligence Artificielle
CEMAGREF
Parc de Tourvoie
92185 Antony
Tel: 40 96 61 79

Pour comprendre les systèmes naturels, qu'il n'a pas créés, l'homme est bel et bien contraint de s'en faire des modèles. Paradoxalement, son entreprise de modélisation a rencontré un certain succès dans deux classes extrêmes de systèmes : les systèmes déterministes, fortement couplés mais de faible dimension sont bien traités par la mécanique analytique, alors que la mécanique statistique a permis de modéliser des systèmes stochastiques, de très grandes dimensions mais faiblement couplés.

Mais comment modéliser des systèmes de grande dimension, dont les éléments interagissent de façon forte et non-linéaire ? Comment d'immenses assemblées de neurones, à la dynamique locale, construisent des réponses adaptées à des environnements complexes ? Comment d'immenses collections de lymphocytes parviennent à opérer cette distinction essentielle entre le soi et le non-soi ? Comment des insectes presque myopes parviennent à trouver leur nourriture, ou à effectuer de redoutables tâches de construction ? Comment des agents économiques, avec leur rationalité limitée, parviennent à faire émerger les structures complexes des économies modernes ? Comment un nombre immense d'espèces peuvent coévoluer au sein de l'écosphère ? Comment comprendre tous ces phénomènes d'intelligence collective, où un système de grande dimension, sans coordination centralisée, parvient à résoudre les problèmes essentiels à sa viabilité ?

Les développements récents de la théorie du chaos ne semblent pas apporter une réponse aux questions qui précèdent. Les systèmes chaotiques, tels qu'ils sont actuellement étudiés, sont des systèmes dissipatifs, de faible dimension, déterministes mais pouvant présenter toutes les caractéristiques de l'aléatoire. Certes, les systèmes mentionnés ci-dessus sont dissipatifs, comme parcouru par des flux d'énergie, de matière, ..venant de l'extérieur. Mais ils sont de grande dimension. De plus, ils ne nécessitent aucun réglage fin de leur paramètres pour se maintenir dans leur domaine de viabilité.

Récemment, un nouveau paradigme a été proposé pour les grands systèmes interactifs, celui des états critiques auto-organisés, avec comme signature spatiale une structure fractale et comme signature temporelle un bruit de scintillation en $1/f$ [Per Bak et al., 87]. Par auto-organisé, on entend simplement que le système n'a pas besoin d'un réglage fin des paramètres pour évoluer vers un état critique : l'état critique est un attracteur de la dynamique du système, robuste en fonction des conditions initiales, des variations de ses paramètres et des aléas. Comme dans tout état critique, il n'y a plus d'échelle caractéristique de temps et d'espace et l'autosimilarité spatio-temporelle se traduit par des structures fractales et un bruit en $1/f^1$.

¹ en fait, il s'agit d'un bruit en $f^{-\alpha}$, avec α voisin de 1 et en tout cas compris entre 0 (un bruit blanc) et 2 (le mouvement brownien)[Per Bak et al., 88]. Le fait que α soit plus ou moins voisin de 1 fait l'objet de nombreux débats.

Les structures fractales [Mandelbrot, 1982] abondent dans la nature mais on comprend encore mal la dynamique de leur formation [Kadanoff, 1986]. Les bruits de scintillation sont eux aussi très fréquents dans la nature. On comprend alors tout l'intérêt porté aux signatures spatiales et temporelles des états critiques auto-organisés. La divergence des longueurs et des temps de corrélation n'a pas seulement pour conséquence une autosimilarité spatio-temporelle : elle conduit aussi à une portée infinie dans la transmission cohérente des perturbations et l'observation d' "avalanches" de toute taille et de toute durée au sein du système.

Ces diverses caractéristiques des états critiques auto-organisés (structures fractales, bruit de scintillation, avalanches de toute taille et de toute durée) en font un paradigme particulièrement important pour l'étude des grands systèmes interactifs, dans des domaines aussi variés que les tremblements de terre, la turbulence, les phénomènes de transport, les fluctuations économiques, les phénomènes de coévolution en écologie.

I. Le paradigme du tas de sable

Le tas de sable peut être considéré comme un paradigme d'état critique auto-organisé que l'on peut aisément observer dans des expériences ou des simulations : les tas sous-critiques et les tas sur-critiques tendent naturellement vers l'état critique, avec leur pente critique caractéristique. Le tas est instable en de nombreux endroits, mais l'état critique reste un attracteur global. Lorsqu'on le perturbe au voisinage de l'état critique, on peut observer non seulement le caractère attracteur de cet état critique mais aussi l'invariance de propriétés statistiques, telles que la distribution des avalanches en taille et durée. Si on mouille le sable ou si on met des paravalanches, les mêmes phénomènes d'état critique et d'avalanches résistent aux modifications mineures des règles gouvernant le système.

Dans la suite du paragraphe, on s'intéresse seulement aux simulations et à leur interprétation. Ces simulations sont réalisées à l'aide d'automates cellulaires. La question de l'état critique auto-organisé comme attracteur robuste est discutée à partir du modèle en une dimension (1D) ; l'examen des propriétés statistiques des avalanches est ensuite examinée en dimension supérieure (2D et 3D).

I.1 Modèle 1D et protocoles

Une pile de sable 1D est définie par sa hauteur $h(n)$ sur un intervalle $[0, N]$. Cette hauteur peut s'interpréter comme l'énergie potentielle du grain le plus haut placé. Sa dynamique, comme automate cellulaire, s'exprime à partir de la différence² :

$$z(n) = h(n) - h(n+1)$$

I.1.1. dynamique de l'automate cellulaire :

Cette dynamique comporte deux règles. La première donne l'évolution locale lorsqu'on perturbe le tas en ajoutant un grain. La deuxième représente le basculement d'un grain de sa position à la position suivante, lorsque la pente dépasse la pente critique :

- 1- ajout en position n :

$$z(n-1) < z(n-1) - 1,$$

$$z(n) < z(n) + 1$$
- 2- diffusion en position n : si $z(n) > z_c$ alors

²Si on interprète $h(n)$ comme l'énergie potentielle du grain le plus haut placé, $z(n)$ est une différence d'énergie potentielle.

$$\begin{aligned} z(n-1) &<- z(n-1)+1 \\ z(n) &<- z(n)-2 \\ z(n+1) &<- z(n+1)+1 \end{aligned}$$

La deuxième règle est en fait une diffusion lorsqu'une instabilité locale apparaît dans le système. L'instabilité locale provoque une dissipation d'énergie, puisque le grain qui a basculé a perdu de l'énergie potentielle.

Un tel système possède un très grand nombre d'états métastables : tous les états tels que $z(n) \in [0, z_c]$ sont métastables ; il y a donc z_c^N configurations métastables. Ces configurations évoluent par addition de grains vers l'état critique, qui est globalement attractif. Mais il faut tout de suite remarquer que cet état critique est en même temps le plus instable des états métastables : l'ajout d'un grain dans n'importe quelle position introduit une instabilité.

L'existence d'un grand nombre d'états métastables dans le bassin d'attraction d'un état critique semble être une caractéristique générale des états critiques auto-organisés. Les perturbations font passer d'un état métastable à un autre état métastable du même bassin, de sorte que l'état critique reste attracteur.

I.1.2. protocole d'expérimentation numérique :

Pour les équations différentielles, il ne suffit pas de fixer la dynamique ; on sait que les conditions aux limites, les conditions initiales sont essentielles. Dans le cas des automates cellulaires, il faut aussi se donner des conditions initiales et des conditions aux limites ; mais il faut en outre fixer la distribution de probabilité des perturbations, puisqu'on s'intéresse à l'évolution stochastique du système sous l'effet de perturbations.

Le protocole d'expérimentation numérique doit être soigneusement défini, si l'on veut mettre en évidence la faible dépendance de l'état critique auto-organisé par rapport aux perturbations, aux conditions initiales et aux conditions aux limites. On a explicité ci-dessous trois hypothèses pour chaque classe de conditions :

conditions initiales :

- i1. $\forall n \ z(n) = 0$ (il n'y a pas de sable)
- i2. $\forall n \ z(n) > z_c$ (l'état est surcritique en tout point)
- i3. $z(n)$ quelconque

conditions aux limites :

11. mur en $n = 0$, échappement en $n = N$
 - $z(0) = 0$ (c.a.d $h(0)=h(1)$)
 - si $z(N) > z_c$ alors
 - $z(N-1) <- z(N-1)+1$
 - $z(N) <- z(N)-1$
12. mur en $n = 0$, mur en $n = N$
13. échappement en $n = 0$, échappement en $n = N$

perturbations :

- p1. le sable est ajouté avec une distribution spatiale uniforme
- p2. le sable est ajouté à gauche
- p1. le sable est ajouté avec une distribution spatiale non uniforme

En croisant ces diverses hypothèses, il a 27 protocoles possibles. Tous n'ont pas été étudiés. Mais tous ceux qui l'ont été ont conduit au même résultat : ces systèmes de transport évoluent jusqu'à un point critique où le transport est devenu stationnaire. Le modèle du tas de sable peut, en effet, servir à bien d'autres interprétations, comme le

transport d'un fluide sous l'effet de différence de pression, comme un flot d'électrons sous l'effet d'une différence de voltage, etc..

I.2. Lois puissances et lois d'échelle dans les modèles 2D et 3D

Le système devient stable précisément au point critique où les perturbations se propagent à longue portée. Il n'y a plus d'échelle caractéristique de temps et d'espace et on observe des avalanches de toute taille et de toute durée. La taille d'une avalanche est le nombre de grains qui ont basculé entre deux ajouts successifs de grains ; la durée d'une avalanche est le nombre de pas nécessaires pour obtenir un nouvel état métastable. Cela revient à supposer que la dynamique du système (donc des avalanches) est beaucoup plus rapide que la dynamique des perturbations.

Plus précisément, au voisinage de l'état critique, les distributions en taille et en durée obéissent à des lois puissances, dont la décroissance est lente à l'infini. Les lois puissances pour les tailles et durées obtenues expérimentalement en 2D et en 3D sont les suivantes [Per Bak et al, 88]

	tailles $s^{-\tau}$	durées $T^{-\alpha}$	bruit en $f^{-\beta}$
2D(grille 50*50)	1	0.43	1.57
3D(grille 20*20*20)	1.37	0.92	1.08

tableau 1 : lois puissance en taille et durée des avanches

On montre qu'une superposition d'impulsion d'une quantité physique avec une distribution de durée $D(T) \approx T^{-\alpha}$ conduit à un spectre de puissance³ en $S(f) \approx f^{-2+\alpha}$ [Per Bak et al, 88], c.a.d à un bruit de scintillation en $f^{-\beta}$ avec la relation :

$$\beta = 2 - \alpha \quad (1)$$

Enfin la taille et la durée de vie d'une avalanche sont également liées par une loi puissance:

$$s \approx T^{1+\gamma}$$

On montre que cette loi puissance n'est pas indépendante des autres lois-puissance, et on obtient la nouvelle relation :

$$\alpha = 2 - \beta = (\gamma + 1)\tau - 2\gamma \quad (2)$$

Les relations (1) et (2) sont les lois d'échelle "universelles" pour la classe des systèmes de transport modélisables par l'automate cellulaire du tas de sable. Les coefficients α , β , γ , τ dépendent uniquement de la dimension du substrat physique sur lequel s'effectue le transport.

II. Le jeu de la vie

³La fonction de dissipation de l'énergie correspond au basculement d'un grain :

$$f(x,t) \text{ (dissipation d'énergie à la position } x \text{ au temps } t)$$

$$F(t) = \int f(x,t) dx \text{ (la dissipation instantanée d'énergie)}$$

Le spectre de puissance de la dissipation d'énergie est alors :

$$S(f) = \int \langle F(t_0+t) F(t_0) \rangle \exp(2\pi i f t) dt, \text{ où } \langle \dots \rangle \text{ est la moyenne temporelle sur } t_0.$$

Avec ces notations, la taille d'une avalanche s'écrit :

$$s = \int_0^T F(t) dt \text{ (dissipation intégrée au cours de la durée de l'avalanche)}$$

Le jeu de la vie [Conway, cf Gardner, 70] est une abstraction d'un système écologique, où les individus meurent lorsqu'il y a une trop faible ou une trop forte densité de population autour d'eux et se reproduisent lorsque la densité est moyenne.

Ce modèle va nous fournir une nouvelle classe d'universalité pour les états critiques auto-organisés. Il présente aussi l'intérêt théorique de ne pas satisfaire à une loi de conservation locale : avec le tas de sable, il y avait une conservation locale du nombre de grains ; avec le jeu de la vie, ce n'est plus le cas.

Le protocole de l'expérience est le suivant :

a- dynamique :

- si une case vide est entourée par deux cases occupées, un nouvel organisme apparaît
- si une case occupée est entourée par plus de trois cases occupées, l'organisme meurt
- dans les autres cas, son état ne change pas.

b- conditions initiales : les cases occupées sont tirées au hasard

c- conditions aux limites : périodiques, au sens où il y a un raccordement "torique" des bords haut et bas, des bords droit et gauche.

d- perturbations et avalanches :

chaque fois que le système atteint un état stationnaire⁴, on ajoute un organisme au hasard. La durée de l'avalanche est le temps mis par le système pour retrouver un état stationnaire. La taille de l'avalanche est le nombre de naissances et de décès pendant la durée de l'avalanche.

Après 40 000 perturbations sur une grille 150*150, Per Bak et al. [89] obtiennent les lois puissances suivantes pour la taille et la durée :

	tailles	durées
	$s^{-\tau}$	$T^{-\alpha}$
2D(grille 150*150)	1.4	1.6

tableau 2 : lois puissance dans le jeu de la vie

La répartition des cases occupées présente une structure fractale : le nombre moyen de cases occupées dans un rayon r suit la loi puissance $D(r) \approx r^{d-1}$, avec la dimension fractale $d = 1.7$.

Pour d'autres protocoles de perturbation (individus supplémentaires avant la stabilisation, en des sites spécifiques et non au hasard), on obtient les mêmes valeurs pour les exposants τ , α , d . Ces exposants ne dépendent pas du détail des protocoles de perturbations, ou encore des conditions initiales. Par contre, ils dépendent de la dimension de l'espace dans lequel est plongé le système : ils ne sont plus les mêmes pour le jeu de la vie en dimension 3.

Le jeu de la vie est un système coévolutif très sommaire. S. Kaufmann et S. Johnsen [91] ont proposé un modèle de coévolution beaucoup plus sophistiqué, où la contribution de certains gènes à la fitness dépendent de façon non linéaire d'autres gènes. Ils observent que ces processus coévolutifs conduisent à des équilibres de Nash : dans ces

⁴Un état stationnaire est un état périodique simple, qui comporte des groupes locaux stables (de période 1) et des groupes locaux cycliques de période 2. Les groupes de période plus grande sont excessivement rares sur une grille de 150*150.

équilibres, des avalanches coévolutionnaires de toutes tailles se produisent, correspondant à des extinctions plus ou moins importantes d'espèces. Un tel modèle permet de donner une interprétation aux phénomènes d'équilibres ponctués : selon Eldredge et Gould [72], les fossiles révèlent une évolution non pas graduelle mais des équilibres ponctués, où de longues périodes stationnaires sont interrompues par de courtes périodes d'évolution rapide. Il n'est plus alors nécessaire d'invoquer une catastrophe d'origine externe pour expliquer un phénomène majeur d'extinction : la possibilité d'une grande avalanche est contenue dans la dynamique même de l'évolution.

III. Fluctuations économiques

Per Bak et al.[89] ont proposé un modèle de production où des sociétés fabricant différents produits sont positionnées sur un réseau à deux dimensions : la trame économique. Chaque société s'approvisionne en matières premières auprès de deux sociétés amont et fabrique des produits qu'elle essaie de vendre à deux sociétés aval. Le réseau a la forme d'un treillis et fonctionne à partir de la demande externe.

dynamique : chaque demande unitaire induit de proche en proche dans la trame économique une production $y(n)$ selon les règles suivantes :

- 1- Si l'entreprise dispose d'une unité de stock, elle la vend
- 2- Sinon, elle commande une unité de produit intermédiaire à chacune des deux sociétés amont, produit 2 unités : elle en vend une et stocke l'autre.

perturbations : N demandes unitaires tirées au hasard

Les simulations révèlent que, si la demande pour chacun de ces produits varie faiblement et de façon aléatoire, de nombreuses sociétés connaîtront des avalanches dans leurs ventes et leurs productions. La trame économique ainsi modélisée tend vers un état critique auto-organisé, dans lequel les fluctuations économiques sont intrinsèques.

On sait résoudre exactement ce modèle, à partir d'une méthode développée par Dhar et Ramaswamy [92]. Si y désigne la production totale induite par N demandes unitaires tirées au hasard, la distribution de y suit la loi suivante :

$$Q(y) = 1/N^{1/\tau} G(y/N^{1/\tau})$$

où $G(x)$ est une loi de Pareto-Levy [Mandelbrot, 63]

Une telle distribution n'a ni moyenne ni écart-type fini. Le théorème central limite classique ne s'applique plus, avec comme résultat une distribution gaussienne.

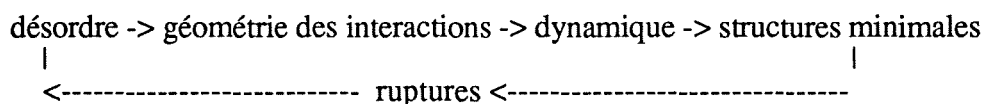
IV. Comment étudier les grands systèmes interactifs ?

modéliser ...

On a pu croire que l'étude analytique séparée des différents éléments d'un grand système puis de leurs interactions permettrait de prévoir le comportement global du système. Le paradigme des états critiques auto-organisés repose sur la démarche inverse, consistant à modéliser l'ensemble des éléments dans leurs interactions.

Il est certes impossible d'avoir des modèles de grands systèmes interactifs à la fois totalement réalistes et maniables du point de vue théorique. On est conduit à proposer des modèles abstraits et idéalisés, sous la seule condition qu'ils reproduisent des caractéristiques émergentes des systèmes observés. Cette attitude a été particulièrement féconde en physique, où elle s'est accompagnée d'un énorme effort de mathématisation et de théorisation. La vie artificielle en tentant d'abstraire les phénomènes dynamiques fondamentaux du vivant pour en reproduire les propriétés émergentes s'est engagée sur la même voie.

Les travaux consacrés aux états critiques auto-organisés concernent pour beaucoup le paradigme du tas de sable, qui a soulevé beaucoup de critiques. On pourrait s'attarder sur ces critiques, et peut-être conclure sur la mauvaise représentativité du paradigme du tas de sable. Plutôt que de focaliser l'attention sur le seul exemple du tas de sable, il est préférable de discuter la démarche de modélisation, en prenant pour point de départ la proposition de Stéphane Roux [90] : un milieu désordonné y est modélisé comme un réseau, dans un espace de dimension d , dont les noeuds présentent ou non des interactions locales : on obtient une géométrie particulière des interactions. Sur cette géométrie, on place une dynamique et on observe l'apparition d'un état critique auto-organisé, avec sa signature spatiale de structures "fractales" ; ces structures sont "minimales", puisque l'état critique est attracteur :



Remarquons l'importance du point de vue géométrique, au niveau de la géométrie des interactions, qui s'appuie sur la théorie de la percolation, comme au niveau des structures minimales, qui renvoie à la géométrie fractale. Les structures minimales évoluent à leur tour, sous l'effet de perturbations externes ou, plus naturellement, sous l'effet de la dynamique interne.

Cette démarche s'est avérée féconde pour plusieurs dizaines de modèles de systèmes physiques. Pour chaque modèle, on s'intéresse à diverses observables et on obtient, par des arguments théoriques ou des simulations, des lois puissances au voisinage de l'état critique. Les deux points essentiels sont alors les suivants, que le tableau 1 précisait dans le cas du tas de sable :

- a. Les exposants de ces lois puissances ne dépendent pas de la géométrie particulière des interactions, mais seulement de la dimension de l'espace dans lequel est plongé le réseau
- b. Les exposants ne sont pas indépendants entre eux : il existe des lois d'échelle entre ces exposants.

Ces deux points tout à fait remarquables ont conduit les physiciens à parler de classe d'universalité. En fait, on s'est vite aperçu qu'il y avait de nombreuses classes d'universalité ! Ces classes correspondent à la diversité des dynamiques. Une classe d'universalité est une classe de systèmes dynamiques, qui présentent les mêmes exposants et donc les mêmes lois d'échelle.

simuler ...et prédire ...

On cherche généralement à prédire la dynamique d'une grandeur agrégée. Pour des systèmes déterministes non chaotiques, telle que la terre en orbite autour du soleil, l'incertitude est constante dans le temps. Pour les systèmes déterministes chaotiques, une petite incertitude initiale augmente exponentiellement avec le temps. Il y a une échelle de temps au delà de laquelle il est impossible de faire des prévisions.

Pour les systèmes critiques auto-organisés, Per Bak conjecture un chaos faible [91] : une petite incertitude initiale augmenterait comme une loi puissance au cours du temps. Les prévisions à moyen terme sont raisonnables même si leur précision décroît avec le temps. Mais la question de savoir si l'incertitude des prévisions similaires, économiques et météorologiques augmentent avec le temps suivant une loi puissance ou exponentielle reste ainsi ouverte.

théoriser...

On assiste à une unification théorique des phénomènes critiques [Petitot, 86], sur laquelle il est possible de prendre appui pour la théorisation des états critiques auto-organisés. Plusieurs théories contribuent à cette unification.

La théorie de la percolation offre un moyen qualitatif de parler de criticalité dans un réseau d'interactions : c'est une géométrie de la contagion, dans laquelle les arêtes du réseau laissent ou non passer les interactions. Un amas infini d'interactions apparaît lorsque la probabilité d'avoir une arête passante dépasse un certain seuil de percolation. Il faut dépasser ce seuil de percolation pour que des fluctuations locales se propagent presque partout dans un système désordonné.

La théorie des catastrophes fournit une modélisation qualitative aussi bien au niveau des effets macroscopiques des fluctuations qu'au niveau des phénomènes microscopiques de brisure de symétries.

La théorie de la renormalisation et des fractals rend compte des phénomènes d'invariance d'échelle que l'on rencontre au voisinage des points critiques. Le point essentiel de la renormalisation est de permettre une fantastique réduction du nombre des variables dans un grand système interactif. La renormalisation reste certes une opération très difficile à expliciter. Mais le principe même d'une réduction drastique du nombre des variables due à la criticalité est un fait tout à fait remarquable. On aurait alors les implications récurrentes suivantes pour le passage de sens d'un niveau d'organisation au suivant :

- un système à l'état critique auto-organisé pourrait être modélisé par un petit nombre de paramètres, même s'il s'agit d'un grand système interactif.
- si le niveau d'organisation suivant évolue vers un état critique auto-organisé, la même règle s'applique

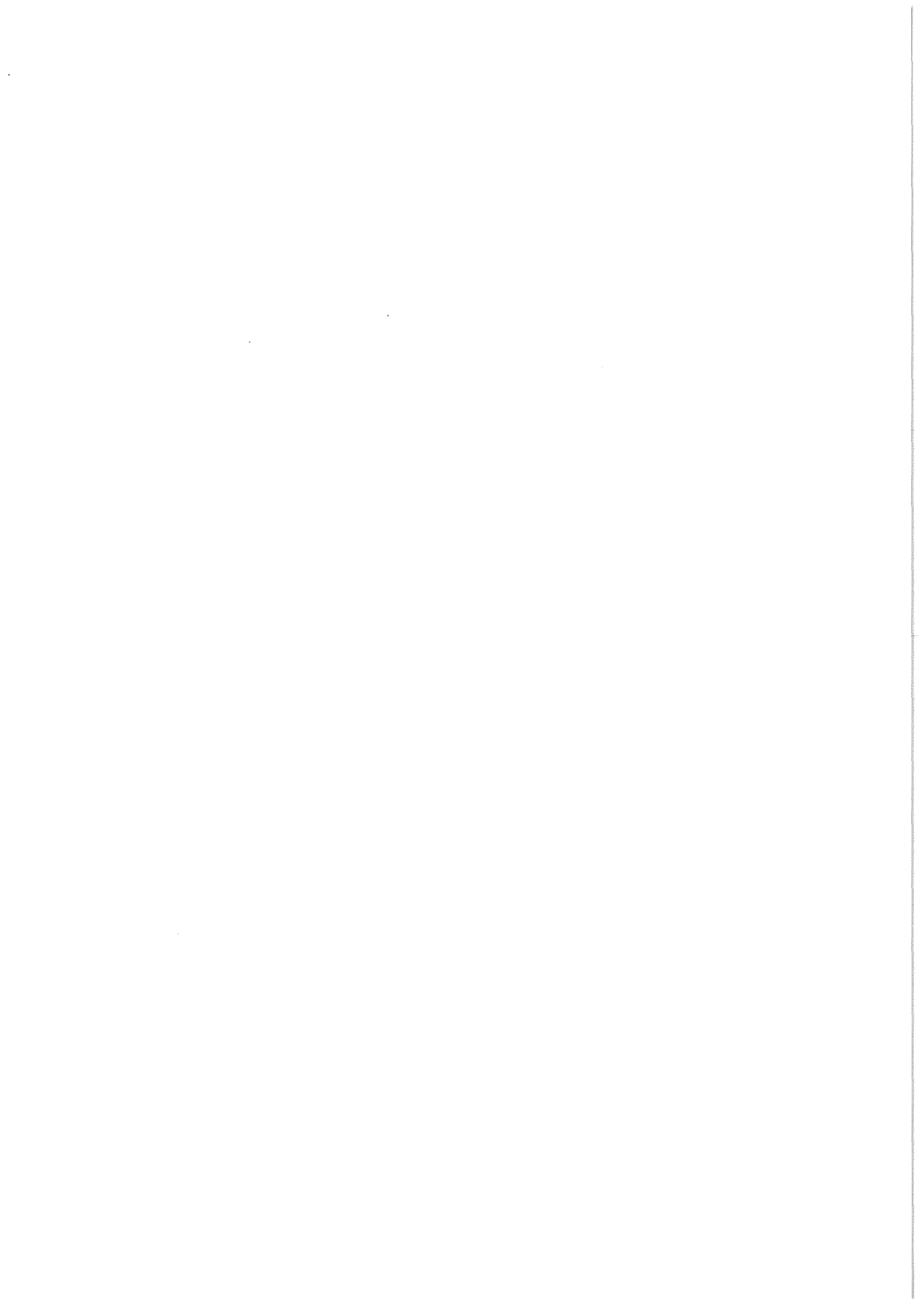
Dans la modélisation d'un grand système interactif, une pratique courante consiste à "simplifier" la description de ces éléments, sans pour autant perdre une certaine reproduction des comportements émergents de ces systèmes. L'idée de criticalité permettrait de fonder de telles pratiques modélisatrices.

Il reste la question centrale de la robustesse de l'état critique auto-organisé et son caractère attracteur qui sont des idées neuves. Un état critique est généralement vu comme extrêmement fragile, réalisable pour une valeur unique de certaines variables comme la température. Mais le beau théorème de Pirogov et Sinai montre que toute perturbation du modèle peut être exactement compensée afin de rester dans l'état critique [par ex. Plum, 86].

Une classe des grands systèmes interactifs, avec leurs éléments en interaction forte à courte portée, pourrait ainsi évoluer continuellement vers un état critique sans jamais l'atteindre car de petits événements y déclenchent des réactions en chaîne à longue portée pouvant conduire à des avalanches de toutes tailles et de toute durée. Telle est la question théorique qui reste posée par nos observations de cette classe de grands systèmes interactifs. Une réponse à cette question viendrait combler un grave déficit dans nos capacités de modélisation et apporterait une compréhension profonde d'une large classe de phénomènes physiques, biologiques et d'intelligence collective.

Bibliographie :

- Bak P., Chao Tang and Kurt Wiesenfeld, 1988, self-organized criticality, in Physical Review A vol 38, n° 1, july 88.
- Bak P., Kan Chen & Michael Creutz, 89, Self-organized criticality in the 'Game of Life', Nature, vol 342, 14 dec 1989.
- Bak P. et Kan Chen, 1991, Les systèmes critiques auto-organisés, in Pour la Science n° 161, mars 91, pp.52-62.
- Bak P., Kan Chen, Jose A. Scheinkman & M. Woodford, 92, Self Organized criticality and fluctuations in economics, Santa Fe Institute Report.
- Dhar D & R. Ramaswamy, 1989, Phys. Rev Lett. 63, 1659.
- Gardner, M, 1970, Game of Life, Scientific American 223 (4),120-124.
- Eldredge N. et S.J. Gould, 1972, Punctuated equilibria : an alternative to pyletic gradualism, Models in paleobiology, ed. T.J.M. Schopf, 82-115. Freeman Cooper, 1972.
- Grimmet G., 1989, Percolation, Springer Verlag.
- Kadanoff L.P., Phys. Today 39(2), 6 (1986)
- Langton C.G., 1990, Life at the edge of chaos, Artificial Life II, ed. C.G. Langton, C. Taylor, J.D. Farmer, S. Rasmussen, Addison Wesley.
- Lindgren K., 1990, Evolutionary phenomena in simple dynamics, Artificial Life II, ed. C.G. Langton, C. Taylor, J.D. Farmer, S. Rasmussen, Addison Wesley.
- Mandelbrot B., 1963, the variation of certain speculative prices, J. Business of the U. Chicago, 37, 393.
- Mandelbrot B., 1982, The fractal geometry of nature, Freeman, San Francisco.
- Petitot J., 1986, Epistémologie des phénomènes critiques, C.A.M.S P.024, EHESS, Paris.
- Prum P, 1986, Processus sur un réseau et mesures de Gibbs, Masson.
- Roux S., 1990, Structures et désordres, Thèse de l'ENPC, Paris.



Cognition without Neurones: Perception, Learning and Memory in the Immune System.

I. Introduction.

It is widely accepted as self-evident that neurones (or neuro-mimetic automata in the case of computer simulations) are a necessary and sufficient basis for cognitive phenomena, and hence that neurophysiology is unconditionally a constitutive element of the cognitive sciences. In this paper, I shall question that assumption by providing a possible counter-example. The science of immunology is characterized by the fact that immunologists spontaneously attribute "cognitive" capacities to the immune system: recognition, learning, memory and self/non-self discrimination. In order to examine the validity of the suggestion that the immune system may be "cognitive", I shall propose some theoretical criteria for defining what we mean by "cognition". These criteria can then be used to elaborate certain requirements that the neurosciences should satisfy in order to constitute a fully articulated part of cognitive science.

II. The immune system.

All vertebrate animals possess an "immune system" composed primarily of a special class of cells, the *lymphocytes*, which circulate freely in the blood and lymph. The interactions of lymphocytes, both amongst themselves and with other cells and molecules, are mediated primarily by a special class of protein molecules, the *immunoglobulins*, which exist in two forms, either bound in the cell-membrane of the lymphocytes, or as free molecules secreted into the body fluids. The membrane-bound immunoglobulins function as receptors: when they are "occupied" by other molecules, the lymphocyte is activated, either to *secretion* of large numbers of immunoglobulin molecules, or to *proliferation* forming a clone of lymphocytes with a common ancestor. It is possible to distinguish two different regions in each immunoglobulin molecule: the "constant" or "framework" region which is virtually the same in all immunoglobulins, and which ensures that they all have the same overall shape and structure; and the "variable" region, which is different from clone to clone¹ and which forms the "combining site" responsible for the specific interactions of the immunoglobulin with other molecules. The diversity of the variable regions is generated by a remarkable process of somatic DNA rearrangement which,

¹ The variable region is identical in all the cells within a given clone.

as far as is known, occurs only in lymphocytes. This diversity is so great that no two clones have identical combining sites; at the same time, in a population of 10^8 lymphocytes¹, there exists at least one clone which will interact specifically, via its immunoglobulin receptors, with any molecular shape whatsoever (of the appropriate size). Immunologists describe this phenomenon by saying that the "repertoire" of the immune system is "complete" (Ref).

In classical immunology, the immune system is activated primarily by external "antigens", typically bacteria and viruses. When an antigen is introduced into the system, the specific lymphocyte clones that interact² with it will be stimulated, both to proliferate³, and to secrete large quantities of the corresponding immunoglobulins. Thus, in the absence of "network" interactions between the various lymphocyte clones themselves (see below), the introduction of an antigen leads to a strong "immune response" which triggers the destruction of the antigen. This mode of operation of the immune system is known as the "clonal selection theory".

III. A network model of the immune system.

In recent years, immunological theory has been enriched by a new approach to understanding the organization of the system. The initial impetus was given by Jerne (1974), who pointed out that if the immune repertoire is indeed complete, then the immunoglobulin receptors of one clone should interact with the immunoglobulins produced by certain other clones. This has led to a concept of the immune system as an autonomous, self-activating network. In order to gain some idea as to how such a network might work, we have built a mathematical model of the immune system (refs). Even in its simplest form, to be described below, this model is a highly non-linear dynamic system; we have therefore performed computer simulations to study its behaviour.

The basic constituents of this model are a set of lymphocyte clones, numbered $i=1, \dots, n$. Throughout the life of an individual animal, new clones are continuously generated (in higher vertebrates, in the bone marrow) and presented as candidates for recruitment into the population of activated lymphocytes. In order to represent the interactions between clones, we

¹ This is the number of lymphocytes in a mouse.

² Immunologists speak of "recognition"; I will examine in the discussion whether the use of this cognitive metaphor can be justified.

³ Hence the term "clonal selection", in analogy with the natural selection of neo-darwinian evolutionary theory

have used the "shape-space" concept (ref) according to which the "shape" of the combining site of each immunoglobulin can be defined by m stereochemical parameters. The values of these parameters can be taken as coordinates in an m -dimensional "space", so that each molecular shape corresponds to a point in "shape-space". In practice, we have found it useful to consider that each point in shape-space corresponds not to a single shape, but to a pair of complementary shapes (distinguished as "black" and "white") having high-affinity interactions between them. This makes it possible to calculate the affinity m_{ij} between any two shapes i and j as:

$$m_{ij} = \exp(-d_{ij}^2) \quad \text{Equation [1]}$$

(where d_{ij} is the distance in shape-space between a black/white point i and a white/black point j).

In the simple form of the model presented here, there is no proliferation; clones are either present at unit concentration, or absent. The net effect of all the other clones in the network on any given clone i is then defined by the "field" h_i , where

$$h_i = \sum m_{ij} \quad \text{Equation [2]}$$

(summation over $j=1..n$).

On this basis, the model postulates that the criteria for the recruitment of a new clone and for the maintenance of a pre-existing clone are the same: if the field experienced by a clone is within a "window" between a lower threshold and an upper threshold, the clone is recruited/maintained; outside the window, the clone is not recruited/eliminated from the population.

The simulations to be shown here employ a 2-dimensional shape-space, not because this is a particularly realistic assumption (to the extent that the shape-space concept is valid at all, the dimensionality is certainly higher), but because this lends itself to a graphic visualisation of the results which contributes greatly to their interpretation. Each simulation commences with a single "black" clone at the centre of the shape-space. "Black" and "white" clones are generated at random points in the shape-space and proposed as candidates for recruitment into the network. After each recruitment, the entire population is re-evaluated and clones with a field outside the "window" are eliminated. The result of this process is illustrated in Figure 1; schematically, it comprises two phases.

Initially, the population grows until the whole of the available shape-space is occupied by black and white clones, but in a rather disorganized

manner (Figure 1a and 1b). During this phase, a phenomenon of "local collapse" is frequently observed, due to the following mechanism: (i) the presence of a "black" (or white) clone creates a field permitting the recruitment of a neighbouring "white" (or black) clone; (ii) this recruitment often has the effect of increasing the field for the first clone above the upper threshold, so that it is eliminated; and (iii) this elimination abolishes the field for the second clone, so that it too is eliminated. In other words, the process displays a form of "self-organized criticality" characterized by numerous local "avalanches". The initial phase comes to an end when this perpetual self-reorganization leads to the emergence of certain configurations in shape-space which appear to be quasi-stable (Figure 1c and 1d). The general nature of these configurations is even clearer if the "edge effects" are eliminated by joining opposite edges of the 2-dimensional square, so that the shape-space has the topology of a torus (Figure 2).

The quasi-stable configurations in question are based on the formation of linear "chains" of black and white clones. Each black chain runs parallel to a facing white chain (and vice versa). The region "behind" a black (or white) chain is bounded by a second black chain which runs parallel to its own facing white chain, and so on. The quasi-stability of this configuration can be explained as follows. In the region between a pair of facing black and white chains, the field is above the upper threshold, too high for clones of either colour to be recruited. In the region "behind" a chain, on the other hand, the field is too low for clones of the same colour to be recruited. In other words, the configuration of a pair of parallel facing black and white chains is stable because each chain creates a field (at a distance corresponding to the interval between them) within the window for recruitment/maintenance of the other. Finally, there is one remaining form of recruitment which could potentially disrupt this overall configuration: that of white clones in the back-to-back region bounded by black chains (or vice versa). Sporadically, this can and does occur. However, the isolated white clone immediately creates a field leading to the recruitment of neighbouring black clones; because of the asymmetry due to the surrounding chains of black clones, the field for the white clone will systematically exceed the upper threshold, leading to its elimination by a

process basically similar to that involved in the phenomenon of "local collapse" discussed above¹.

In conclusion, we can say that certain characteristic configurations emerge precisely because of their relative stability. This process is somewhat akin to natural selection, but with two notable differences compared to the "clonal selection" of the classical theory. Firstly, the entities selected are not individual clones, but entities at a higher level of organization, i.e. global configurations of organized sets of clones. Secondly, the operative agent of selection is not a fixed, pre-given, external entity; rather, the constraints are endogenously generated by the process itself.

Having thus established some of the major self-organizing characteristics of the network, the next question concerns the interaction of the network with antigens. We have modelled antigens as molecular shapes whose presence is independent of the field they experience. The results of simulations in which the ontogenesis of the network occurs in the presence of such antigens are shown in Figure 3; the antigens are represented by square symbols in order to distinguish them from the circular symbols representing lymphocyte clones.

What we see is that the network retains its self-organizing capacity, giving rise to the same sort of emergent configurations as previously. The novel feature is that the black and white antigens appear to be systematically incorporated into chains of the same colour (Figures 3a and 3b). In other words, the previously unconstrained disposition of the chains is now constrained by the coupling of the network with the antigens. Figure 3c shows that this constrained disposition is maintained even if the constraining antigens are removed². That this is a dynamic feature, rather than the result of static inertia, is shown in Figure 3d: the introduction of discordant antigens (black antigen in a white chain, white in a black) results in a realignment of the chains such that they incorporate antigens of like colour.

¹ Connoisseurs of the Japanese game of Go will recognize striking - and aesthetically pleasing - analogies, both in the actual forms of the emergent configurations and in the strategic reasons underlying them.

² I shall examine whether this can be considered as an example of "memory" in the discussion.

IV. Discussion: is the immune system "cognitive" ?

We are now in a position to turn to the central question of this paper: is there any sense in which it might be meaningful to say that the immune system is "cognitive" ? In order to discuss this question with any clarity, we require some sort of definition of the term "cognitive". As a working definition, I propose the following: the minimal requirement for an entity to be "cognitive" is that it should be capable of both perception and action; moreover, its actions should be guided by its perceptions in such a way that a meaningful constraint is satisfied.

Firstly, then, does the immune system "perceive" its environment? Yes, if by "perceive" we mean that the interactions between the system and its environment reliably lead to characteristic modifications in the internal state of the system. To be more explicit, we need to take into account the fact that the immune system has two distinct modes of operation: the "classical" mode of clonal selection, and the "network" mode modelled in the simulations described above. In the first mode, the system perceives an antigen by mounting an immune response, i.e. by producing high concentrations of immunoglobulins with strong affinity for the antigen. In the second mode, coupling with an antigen leads to realignment of the chains of "black" and "white" clones such that they incorporate the antigen in question.

Secondly, is the immune system capable of "actions" guided by its perceptions? Again, the answer is yes, on condition that we accept to describe its operation in terms that are inhabitual in classical immunology. By mechanisms that it is not necessary to describe in detail here, an immune response triggers the destruction of the antigen that provoked it. In the case of the network mode, a positive answer depends on a speculative hypothesis: it has been suggested that the incorporation of antigens into "chains" may create functional connections between the set of antigens thus incorporated (ref).

Thirdly, does the way in which the actions are guided by the perceptions conform to some non-trivial, meaningful constraint? With this question, we are at the heart of the matter. The classical mode of operation is appropriate in the case of external antigens such as bacteria and viruses: their destruction, triggered by an immune response, defends the organism against infectious disease. The network mode of operation is appropriate in the case of "self-antigens", i.e. molecules which are a part of the organism itself. A network response to bacteria and viruses would fail to protect the

organism from infection; an immune response to self-antigens would cause auto-immune disease. Thus, the immune system may be regarded as "cognitive" precisely to the extent that it systematically enters into the appropriate mode of operation. As a matter of empirical fact, this does seem to be the case; but immunology will only fully qualify as a "cognitive science" if the mechanisms by which this comes about can be elucidated. This is the issue conventionally known as "self/non-self discrimination"; in view of its central importance, it is worth examining the matter more closely.

To begin with, it is to be noted that the two modes of operation are mutually exclusive: if the system responds in the network mode to a given antigen, that effectively precludes an immune response to the antigen in question. This can be clearly seen by considering the fate of a lymphocyte freshly emerged from the bone-marrow. If the lymphocyte is recruited into the network, it will be activated and become part of what has been called the "Central Immune System" or CIS (ref). If it is not recruited, it will remain as a small resting cell for two or three days before dying. The rate of bone-marrow production is extremely high (it is sufficient to reconstitute the entire immune system in a week), so that at any single point in time these inactivated, moribund cells make up fully 80% of the total lymphocyte population; they constitute what has been called the "Peripheral Immune System" or PIS. It is to be noted that although normally moribund, the cells in the PIS will be activated if they encounter an antigen having high affinity with their immunoglobulin receptors; in this case, since they do not possess a network organization, they will give rise to a classical immune response.

It is important to appreciate the relationship between the CIS and the PIS. It is the CIS which is positively defined: it is composed of cells integrated into a self-activating network. The PIS is negatively defined: it is simply composed of the residue of cells which are not recruited into the network. Since the "repertoire" of lymphocytes produced by the bone-marrow is "complete", it follows that the repertoire of the PIS is "complete minus the repertoire of CIS".

We are now in a position to specify more precisely the conditions under which the distribution of roles between the network mode of operation (instanciated by the CIS) and the classical mode of operation (instanciated by the PIS) will be "appropriate" in the sense defined above. The essential requirements are that the repertoire of the CIS should

reliably include all the self-antigens of the body; but that it should not extend too far beyond this, otherwise the repertoire of the PIS will be correspondingly reduced and may not cover the totality of pathogenic micro-organisms whose function it is to destroy. The first requirement is probably met by the timing of developmental events in ontogeny: the CIS is the first to develop, in the pre-natal embryo, at a time when only self-antigens are present. The second requirement is more problematic, and is in fact a major focus of our current research. The problem is that in the very simple model described above, the "network" expands to occupy the whole of the available shape-space; this would reduce the repertoire of the PIS to nothing. This problem is, however, clearly posed, and it should be possible to solve it by an appropriate extension of the model¹.

Summing up, then, it seems that it is both plausible and fruitful to envisage the hypothesis that the immune system is indeed "cognitive" in the sense defined here. This signifies that the employment of cognitive metaphors is not just loose heuristic talk, but may correspond to a scientific reality. As we have seen, in immunology a central issue is the positive constitution of a self-identity; the capacity for self/non-self discrimination is a secondary consequence. It is interesting to note that since, in an outbreeding population, no two individuals are identical, this self-identity is clearly *learnt*. Turning the cognitive metaphor around, our analysis of the immune system suggests that learning is an epigenetic phenomenon, fundamentally inseparable from the ontogenetic organization of the system as a whole. In this perspective, it may be that "memory" is basically nothing other than a reflection of the fact that the current state of a cognitive system arises from the total history of its ontogenetic constitution (ref Rosenfield).

V. Conclusions.

What are the potentially generalisable lessons to be gleaned from this investigation of whether it is scientifically reasonable to attribute cognitive faculties to the immune system? One point that emerges very clearly is that the question cannot be answered by analysis of the individual components of the system. Cells and molecules may or may not be parts of a cognitive

¹ In technical terms, one promising possibility involves including interactions between the immunoglobulin-producing B-lymphocytes (the only type considered here) and the T-lymphocytes which recognize digested protein fragments "presented" on cell surfaces by the molecules of the "Major Histocompatibility Complex".

system, but in and of themselves there is nothing cognitive about them. Nor is it sufficient to analyse the low-level interactions between these components. Thus, the stereo-chemically specific interactions of immunoglobulins, both with each other and with other molecules, are basic to the dynamic functioning of the system; but again, as such, they are not cognitive. The same goes for the activation of a cell by the interaction between its membrane-bound receptor molecules and particular ligands. It only becomes legitimate to speak of such interactions as instances of "recognition" if they are considered in the context of a global system which is indeed cognitive. Even the entire immune system is not, in itself, cognitive; the decisive questions turn on the quality of the interactions between the system and its environment. Thus, the minimal level of hierarchical organization at which it is possible to speak of cognition involves perceptually-guided actions; the key question is whether the coupling between action and perception satisfies some meaningful constraint.

If it is possible to give an affirmative answer to this last question - and I have suggested that this may be the case for the immune system - then (on the definition used here) it is legitimate to speak of cognition. Now, however, a fascinating question arises: what is the "object" of this cognition? In the case of the immune system, we have seen that the answer is: the incorporation of self-antigens into the dynamic configurations of lymphocyte clones which are constitutive of the immunological self-identity; and discrimination between these and external non-self antigens which evoke destructive immune responses. What is remarkable is that this "object" can only be specified by reference to categories constituted by the functioning of the immune system itself. Thus, the immune system is capable of perception and learning; it is undoubtedly adaptive, in the sense that it makes a major contribution to the survival of the organism of which it is part; but it does not perceive, learn or adapt to anything pre-existing or definable in the realm of purely objective reality. It perceives, learns, and adapts (under strong constraints) to aspects of reality that have only been brought into existence through the participation of its own cognitive activity. To sum up in a phrase: the object and subject of immune cognition are inseparable.

Are these considerations generalisable? In particular, are they applicable to neurophysiology? Before deciding, let us see what would be the implications.

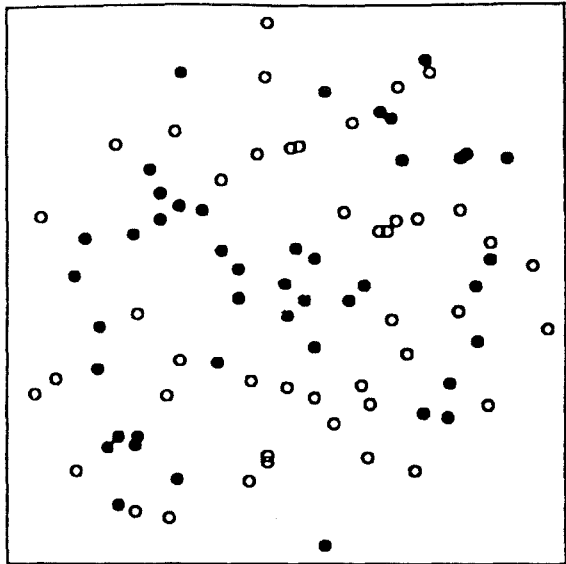
To begin with, this perspective would suggest that neurotransmitters and voltage-dependant ionic channels in neuronal membranes are not, in themselves, cognitive. Nor are neurones and synapses; nor even is the brain, considered as such in isolation. In order to be cognitive, the nervous system must be considered in the functional context of an animal accomplishing perceptually-guided actions in an environment. In this respect, it is surely significant that neurones first appear in evolution in the role of linking sensory inputs to locomotory actions. An interesting feature here is that such actions will, via the animal's movement in its environment, give rise to correlatable variations in its sensory perceptions. The significance of a central nervous system with a connexionist architecture may be that it makes it possible for the animal to identify such correlations, and thus to develop conceptual categories based on its active involvement in the world.

This brief evocation may be sufficient to show that the perspective put forward here, although at first sight rather negative concerning much current research in neurophysiology, can actually serve to integrate the neurosciences into the mainstream of cognitive science by focussing on what is cognitively significant in the underlying molecular and cellular mechanisms. It may therefore be acceptable to generalize the conclusions suggested by the more detailed analysis of the immune system which forms the central theme of this paper. Cognitive perception, learning and adaptation are anything but arbitrary: they are strongly constrained by a "reality principle" which, in the case of living organisms, relates to their viability (ref) and their capacity to maintain their autopoietic organization (ref). The objects of cognition are, however, neither pre-existent nor definable in terms of a totally independent and purely objective reality. Thus, to interpret the function of the nervous system in terms of "feature detectors" or "grandmother neurones" is to go badly astray. The objects of cognition, with or without neurones, are brought into existence by the coupled perceptions and actions of the cognitive system itself.

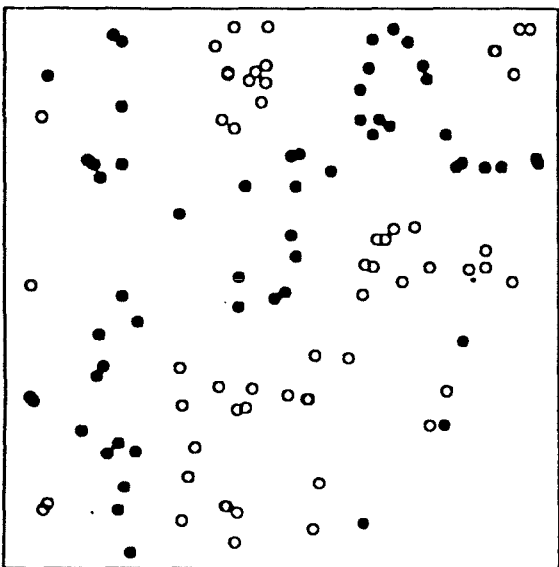
Acknowledgements.



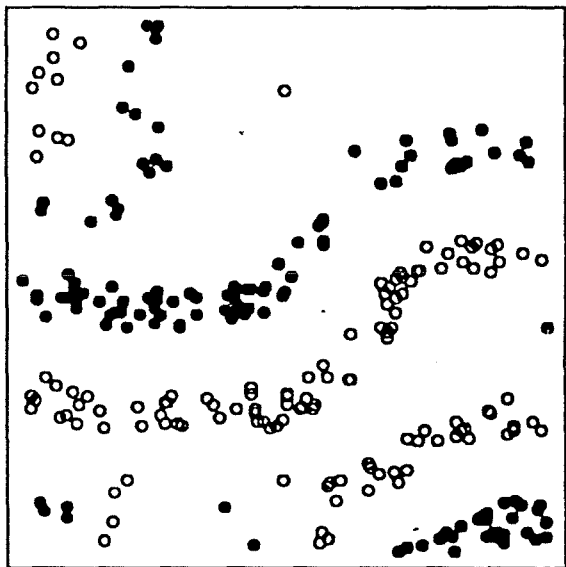
a



b



c



d

Figure ~~1~~ 1

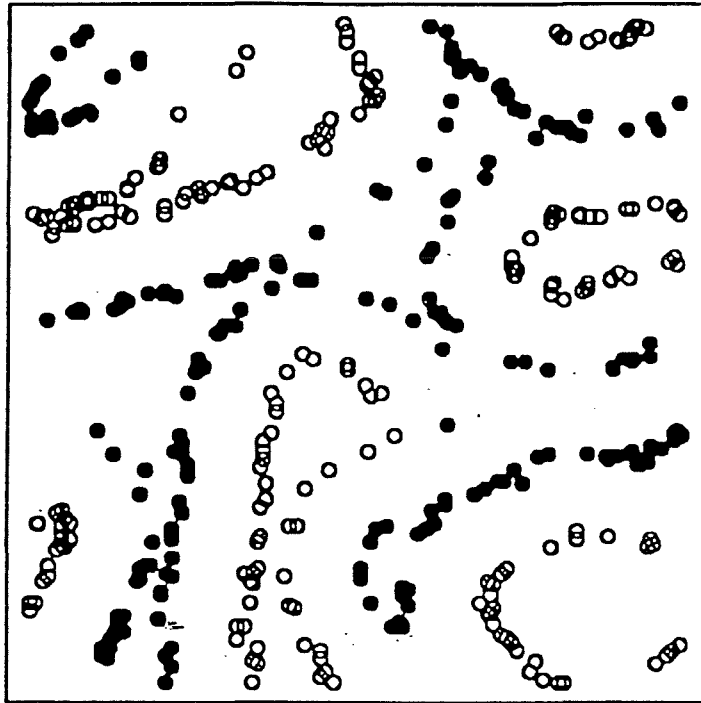
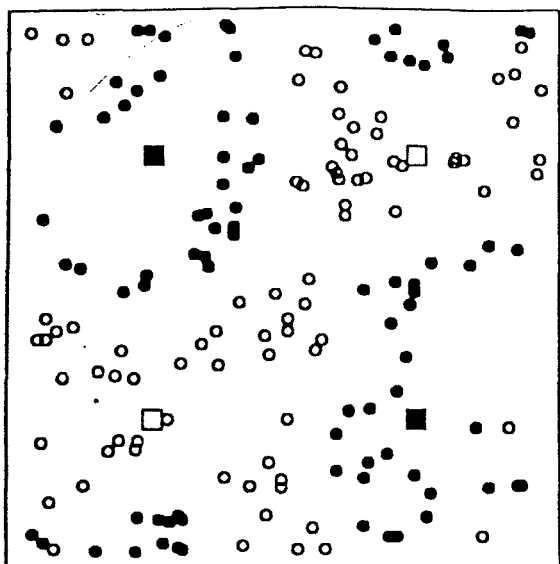
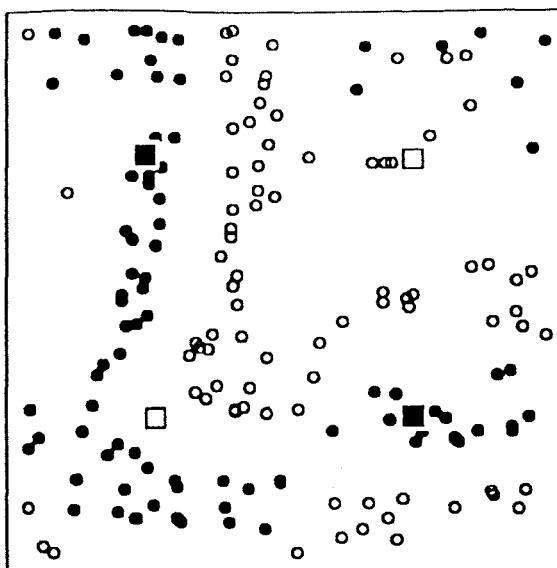


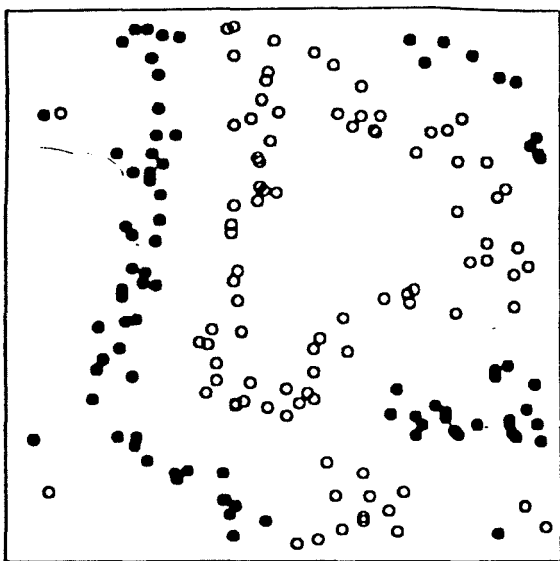
Figure 2



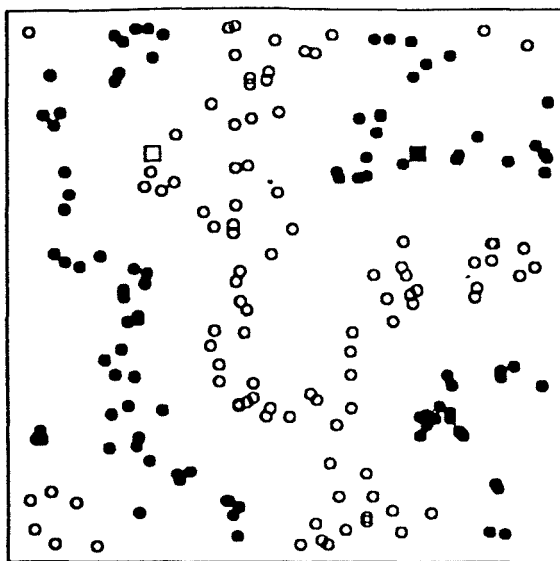
a



b



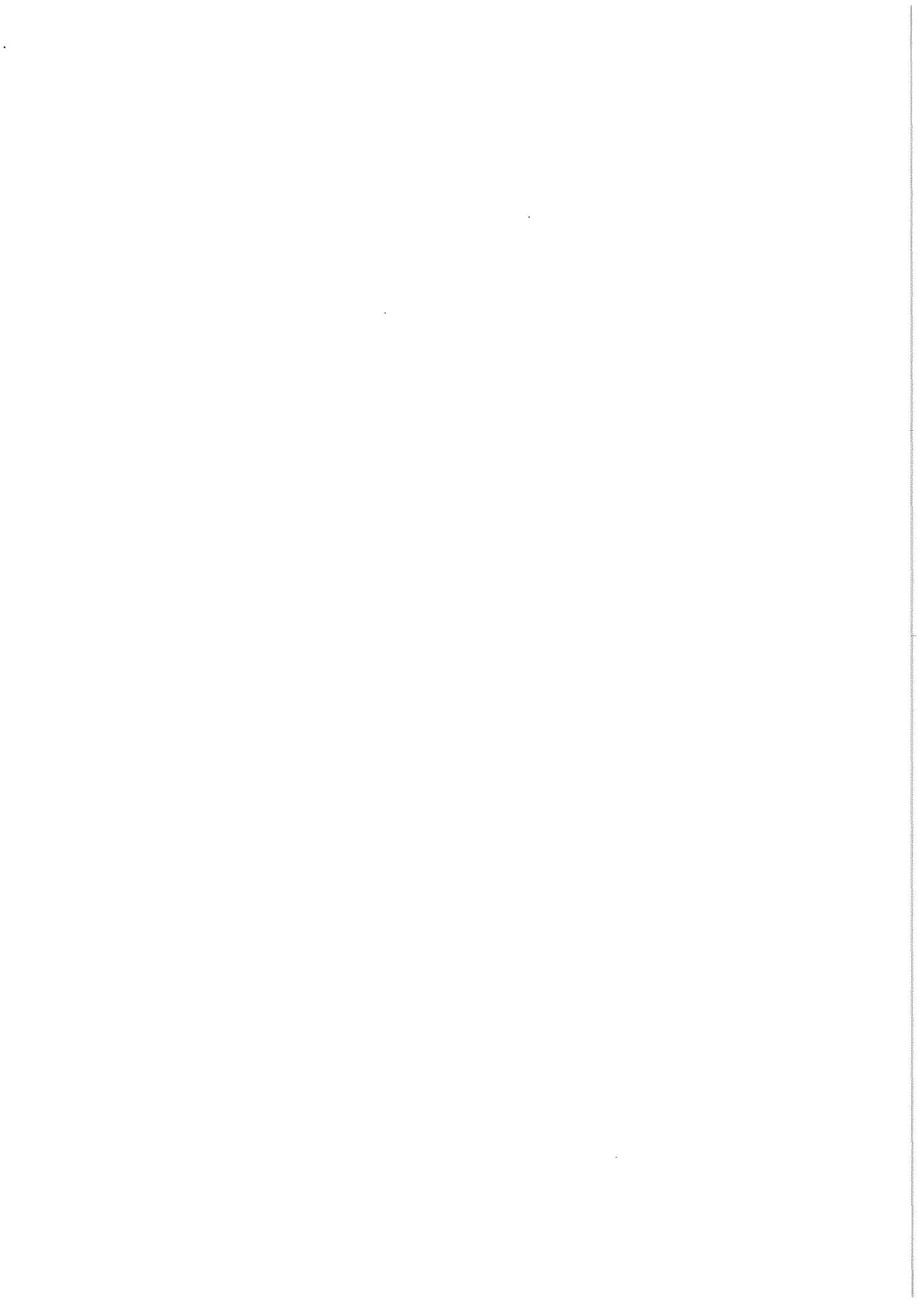
a



b

Figure 3

~~Figure 3~~



ADAPTATION AND REGULARITY EXTRACTION : A CONSTRUCTIVIST APPROACH

Guillaume DEFFUANT, Etienne MONNERET

Laboratoire d'Intelligence Artificielle
CEMAGREF

BP 121, 92185 ANTONY FRANCE

Tel: 40 96 61 21

Fax: 40 96 60 80

e-mail: def@cemagref.fr, etienne@cemagref.fr

Abstract

This paper proposes a general framework for adaptive systems which is inspired by Maturana and Varela's autopoiesis and Piaget's theory of adaptation. It appears that geometric descriptions of the system and its environment in the same topological space are particularly appropriate in this approach. The considered systems are therefore made of simple geometric components which can be modified by actions of the environment. Processes of component regeneration (creation, destruction and link-up) allow this structure to evolve and to become more likely to fit different environments. The result is an evolving geometric form. This framework is used in the particular case of regularity extraction. In this case however, the component regeneration must be itself adaptive. A new model built according to the previous principles, the self-building strings, is described as an illustration.

Topic : Ontogeny, learning and evolution.

1.Introduction

The autopoiesis framework [Varela et al. 74] is founded on an abstraction of the basic mechanisms observed in the biological cell. It emphasizes the living systems capacity for producing themselves, and creating their own identity. This self-production occurs thanks to regeneration processes of basic components (creation, destruction and link-up), which allows the constitution of the system as unity defining a topological domain.

This approach can easily be related to the general problem of adaptation and leads to a distinction between passive and active adaptation. The former concerns the components plasticity to the perturbations coming from the environment, the latter the components regeneration processes. It has many common points with Piaget's conception of biological adaptation [Piaget 37].

Moreover, this allows to design a general framework for adaptation, in which the main tendencies of adaptive systems can take place (connectionist networks, genetic algorithms, swarm intelligence, immune algorithms, swarm intelligence...). This is the first contribution of this paper.

This has also deep implications concerning the relations between the system and its environment. From a practical point of view, the system must be described in the same topological space as the elements of its environment. This can be achieved by the use of a geometric description. Few models in the literature take this point into account. However, it seems very important to us because it allows to design more easily efficient component regeneration processes.

With such a description, it seems that powerful regularity extractions are achievable. The question of regularity extraction arises when the system evolves in a changing environment. If the component regeneration remains the same, there is no improvement in the system's structural changes in order to fit periodically encountered situations. This is possible with a two level adaptation, which involves two types of components. Level2-components are responsible for a control on the regeneration processes of level1-components, and they exhibit active and passive adaptation at their level.

We give an example of a new adaptive model, the self-building strings, which follows the principles of our general framework. The basic geometric components of this model are linear segments. The first level adaptation allows to fit a noisy form given by points. We intend to build a second level of adaptation.

First of all, the precise definition of autopoiesis is recalled, and the question of the relations between the environment and the system is addressed. Then the general framework for adaptative systems is proposed. The self-building strings are described as an illustration of our general framework. Several strategies allowing to implement the second level adaptation, leading to powerful pattern recognition, are finally drawn.

2.Autopoiesis and adaptation

2.1.Definition of an autopoietic system

Thanks to interdependent processes of component regeneration, autopoietic systems produce their own identity : they distinguish themselves from their background. The precise definition from Varela [Varela 79], originally inspired by the biological cell, is the following :

An autopoietic system is organized (defined as a unity) as a network of processes of production (transformation and destruction) of components that produces components that :

- (1) through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them; and
- (2) constitute it (the machine) as a concrete unity in the space in which they exist by specifying the topological domain of its realization as such a network.

The two points of the definition are very different. The first one refers to properties which are expressed in commonly used scientific words (components, processes of generation, destruction) and can be easily the object of a consensus in a scientific community. It expresses that the system have to be self-maintained : the new components regenerate the transformation processes.

The second point is more subtle. It refers to the notion of a distinguishable unity, which needs more explanation. It is clearly linked to the emergence theory, because this unity can be seen as a higher level coming from lower interaction processes. The constitution of the topological domain where the processes exist, is then a retroaction of this higher level on the lower one.

In this paper, we focus on the first property of autopoiesis which can be more clearly related to adaptation capacities. The continuous component regeneration gives the system a structural plasticity, and allows it to fit different environments.

2.2. Interactions with the environment

It has to be noticed that the autopoiesis framework gives a new perspective for the interactions between the system and its environment. This point is very important because it has deep implications for cognitive science and epistemology.

The interactions between the system and its environment are generally described through an input/output framework. It is especially the case of connectionist models. The system and its environment are not considered in the same topological space, they are represented side by side, as in figure 1a. The system involves specific components which receive the inputs from the environment, and then propagate it. The system processes this information, and gives the corresponding output.

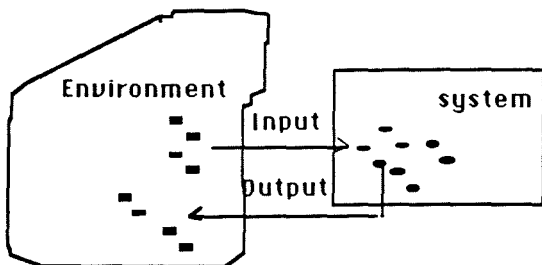


fig. 1a : the input/output framework

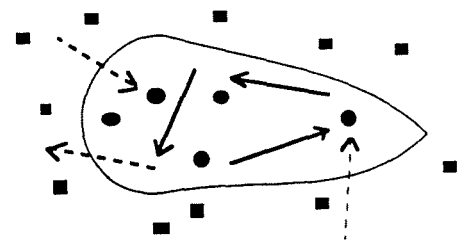


fig. 1b : the autopoiesis framework

In the autopoiesis definition, the environment cannot be represented this way [Varela 1979, 1991]. The system specifies its topological domain in the space where its components exist. The system must therefore be represented in the same space as its environment. Furthermore, its components are submitted to the perturbations coming from this environment. This is schematized on figure 1b.

3.A general framework for adaptive systems

In this section, autopoiesis is related to the adaptation capacity, and a general framework for adaptive systems is drawn. A second level adaptation is defined in order to design powerful regularity extraction tools.

3.1. Cooperation between active and passive adaptation

Autopoiesis can be easily related to general capacities of adaptation. The concurrence between the perturbations coming from the environment and the component regeneration suggests to distinguish two types of adaptation (passive and active adaptation) which must cooperate.

3.1.1. Passive adaptation

Passive adaptation is a plasticity of the system components under the environment perturbations. Geometrically, this plasticity is expressed by deformations and movements of the components under attractive or repulsive forces. The laws governing this plasticity are similar to physical ones, and it is assumed that they cannot be changed by the system. This can be compared to Piaget's accommodation.

Many examples of passive adaptation can be found in the literature. All connectionist networks, (backpropagation [Le Cun 85, Rumelhart et al. 86], LVQ [Kohonen 82]), with a fixed architecture can be classified in this category. In this case, a training example attracts the system in the direction corresponding to a gradient descent.

3.1.2. Active adaptation

Active adaptation corresponds to the components regeneration in the autopoiesis framework (it can be compared to Piaget's assimilation). A complete active adaptivity can only be achieved through the following processes :

-Component creation : this process is very important in order to give the system the possibility to fit complex environments. This process is similar to a chemical reaction allowing the production of new elements (in the biological cell), or to individual reproduction (in a system made of living components). Component creation is widely used as an adaptive strategy. In genetic algorithms [Holland 75, Goldberg 89], new elements of a population are continuously created. Many recent connectionist architectures function also by new cell recruitment [Carpenter & Grossberg 87, Marchand & al. 90, Nadal & Mézard 90, Fahlman & al 90, Deffuant 90], and so does immune network [Perelson 90, Bersini & Varela 90].

-Component destruction : in order to regenerate itself, the system must eliminate inefficient or useless components. This increases the system's possibilities for fitting its environment. This process has to be related to the chemical reaction of component destruction (in the biological cell), or individual death (in a system made of living components). The components destruction is essential in genetic (selective pressure) or in immune systems. Its importance has also been related to the overfitting problem in connectionist networks [Hinton 86, Scarlett & Zee 88, Sietsma & Dow 88].

-Component link-up : the possibility to link-up components is crucial in order to get cooperative effects among the components. Moreover, this allows to build the system as a unity (cf the second condition of autopoiesis). Different definitions can be adopted for links. In a weak sense, it can be a possibility of communication, or of mutual influence between components. This is the case in Brooks's subsumption robotic architecture [Brooks 91], in genetic or immune algorithms or even in swarm intelligence [Deneubourg & Goss 1989]. However, it can be argued that such links cannot provide a real unity in the system. The use of geometric components allows to define it in a stronger sense : a link occurs between two components when they share one (or several) subcomponents.

In order to get wide adaptation possibilities, we argue that a system must realize a cooperation between passive and active adaptations. It must be noticed that very few artificial systems realize this cooperation. A reason is that active adaptation modifies the system's topology, which has usually no clear interpretation in the space of the environment. It is therefore very difficult to establish efficient strategies of structural modifications which increase the fitness possibilities. This is another argument for describing the system and its environment in the same topological space [cf perceptron

membranes, Deffuant 92].

3.2. Regularity extraction and second level adaptation

When the rules governing the component regeneration processes remain the same in a changing environment, the system's efficiency does not increase when the same situations are periodically encountered. In our framework, improvements could only occur when the active adaptation is itself adaptive in our general sense. The component regeneration is then adaptive, and the system is said adaptive at the second level.

Such a system must involve two types of components, one for each level. The components of the second level are responsible for transformations in the dynamics of creation, destruction or link-up of those of the first one. They are submitted to perturbations of their environment (passive adaptation) and they are continuously regenerated through processes of creation, destruction and link-up. The immune system clearly exhibits such an organization, and interactions between hormonal and neural substances could follow this schema, and even adaptation of higher levels.

The second level of adaptation is particularly interesting because it can be related to pattern recognition. If level2 components organize themselves in order to accelerate the building of first level structures fitting situations which occur frequently (or are important), this building can become practically instantaneous, and can be interpreted as the recognition of a pattern.

In the following section, our general framework for adaptation is illustrated by the description of a new model : the self building strings.

4. The self-building strings

4.1. First level adaptation

In this section, an example of practical system inspired by our framework is given : the self building strings. The description of elementary components and their dynamics are given.

4.1.1. Basic components

The system and its environment are described in the same topological space (here the Euclidian plane). The environment is constituted of points distributed in the plane in order to draw a form. The system's components are segments given by their two extremities. Two components are linked when they share one extremity. The emergent form is therefore a piecewise string in the plane, which fits the points of its environment. The string can be deformed by the attraction of the environment points, or thanks to the component regeneration.

4.1.2. Passive adaptation

The passive adaptation is given by the components reactions under the environment's perturbations. In this application, the segments are deplaced or deformed according to the same simple laws of attraction deriving from a potential energy E_p .

For one segment, E_p is given by a Gaussian centered on the segment and equal to zero on it (fig. 2a). Moreover, given one data point M , its projection P on the line supported the segment, and the nearest extremity E of the segment from M (fig. 2b), the potential energy is given by :

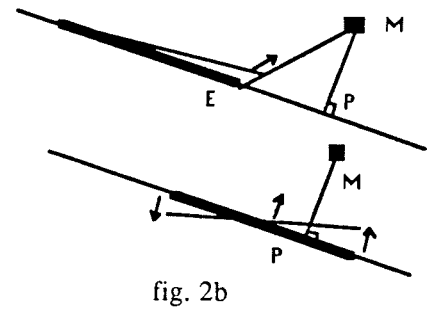
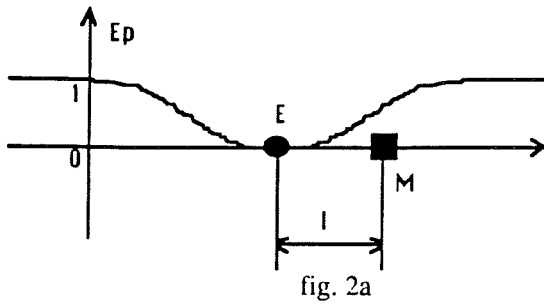
$$E_p(l) = (1 - e^{-l^2/d^2})^2$$

If P is on the segment (fig. 2b bottom)

$$l = \sqrt{(x_m - x_p)^2 + (y_m - y_p)^2}$$

otherwise (fig. 2b top)

$$l = \sqrt{(x_m - x_e)^2 + (y_m - y_e)^2}$$



The derivative of the potential energy is given by :

$$dE_p(l) = -k_1 \frac{l}{d} (1 - e^{-l^2/d^2})^2 dl,$$

where d is the typical distance of the points having an action on the segment, and k_1 is a constant. The potential decreasing is obtained by small movements in the direction of descendent gradient : if $E(x_e, y_e)$ is an extremity of the segment, we apply at each step :

$$\Delta x_e = -k_2 \frac{dE_p}{dx_e} \qquad \Delta y_e = -k_2 \frac{dE_p}{dy_e}$$

where k_2 is a constant. When the origin is chosen in the middle of the segment, and in the case where P in on the segment, this gives little rotations around the origin and little translations in the normal direction of the segment (fig. 2b bottom). In the case where P is outside the segment, it gives little extensions of the segment (fig. 2b top).

Data points are acting only on the nearest segments. Let l_m be the minimal distance of the segments to one given point. The point is acting on all segments that are at a distance less or equal to $l_m + d_1$ (fig. 3).

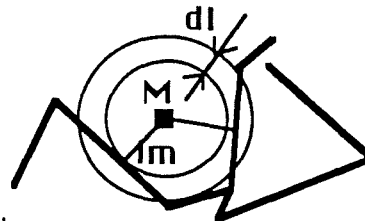


fig. 3

4.1.3.Active adaptation

The active adaptation corresponds to the components regeneration (creation, destruction, linkage). These procedures are applied periodically, while the system is submitted to the perturbations coming from its environment. The cooperation between passive and active adaptation occurs especially in the destruction process, which takes into account the number of perturbations received by each component.

-Creation : a segment is randomly chosen with a given probability, and is cut in the middle in order to produce two segments. This is the basic mean of new component creation.

-Link-up : when the distance between the extremities of two segments is less than a threshold, these two points are merged, and the two segments are linked by this common extremity. When two segments are linked, the deformations of one have consequences on the other. This allows cooperative and competitive effects.

-Destruction : the utility of a component is evaluated thanks to the frequency of perturbations it receives. When this frequency is below a threshold, the component is judged useless, and it is removed. Moreover, when a segment size is below the linkage threshold, i.e. when the segment is reduced to a simple point, it is also eliminated. These destructions can be interpreted as natural decay of elements which are not surrounded by a sufficiently favorable environment.

Figure 5 illustrates the cooperation between passive and active adaptation.

4.2.Second level adaptation

At this step, the system's adaptation capacities can be compared to those of a plant. It grows on a favorable ground. However, it will not grow or reorganize faster if it encounters periodically the same ground configurations. In order to achieve such this type of behavior, a second level adaptation is necessary : the component regeneration must be itself adaptive. We describe now how this second level adaptation could be implemented.

4.2.1.level2 active adaptation

In this approach, the second level basic components influence level1 active adaptation : they favor level1 component creation, linking or destruction.

Level2 component creation : Level2 components must correspond to the regularities of level1 functioning. Stabilized level1 configurations which are present for a long time must trigger level2 component creation. This requires a stabilization test which can be achieved thanks to a distance on level1 configurations. Level2 components can be the global level1 configuration which is duplicated, or only stabilized parts of it.

Level2 component link-up : when level2 patterns exhibit subcomponents (segments) which are close

to each other, these components are merged and become common to both patterns. This allows to simplify level2 structure, and to create paths between its components. These can lead to associations between the different parts. Furthermore, these links create a topological network of level2 components, in which links correspond to topological similarities (cf figure 4).

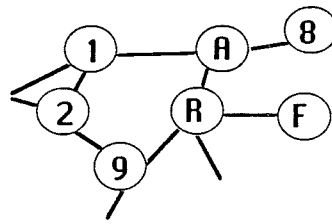


Figure 4

Topological network of level2 components. The links represent topological similarities.

Level 2 component destruction : an activation is associated to level2 components, which corresponds pattern's frequency of occurrence. When this activation decreases below a threshold, the pattern is eliminated.

4.2.2. Passive adaptation

The level2 passive adaptation concerns deformations of level2 components under the action of their environment. As level2 correspond to regularities, these deformations will occur much more slowly than at level1. Furthermore, level2 components must be perturbed only by stabilized level1 configurations. The nearest (if it is not too far) level2 component of the stabilized level1 configuration will be slightly attracted and modified by this level1 configuration.

4.2.2. Action of level2 components on level1 active adaptation

Level2 components correspond to regularities extracted, they must therefore favor the construction of level1 configurations already encountered, according the current state of the level1 configuration and of the environment.

This can be achieved by giving level2 components an activation corresponding to the proximity of their segments to the points of the environment, which are far from level1 components. The most activated level2 pattern will produce level1 components, which will accelerate the development of the level1 configuration. This is illustrated by Figure 6.

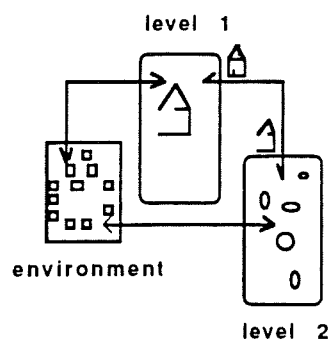


Fig. 6

level2 adaptation : level2 components produce level1 components according to previously encountered configurations. This allows to accelerate the form construction.

These ideas are only at their beginning and many different strategies are certainly possible for the level2 dynamics. Self-building strings seem particularly appropriate for their exploration because of

the simplicity of its components.

5. Conclusion

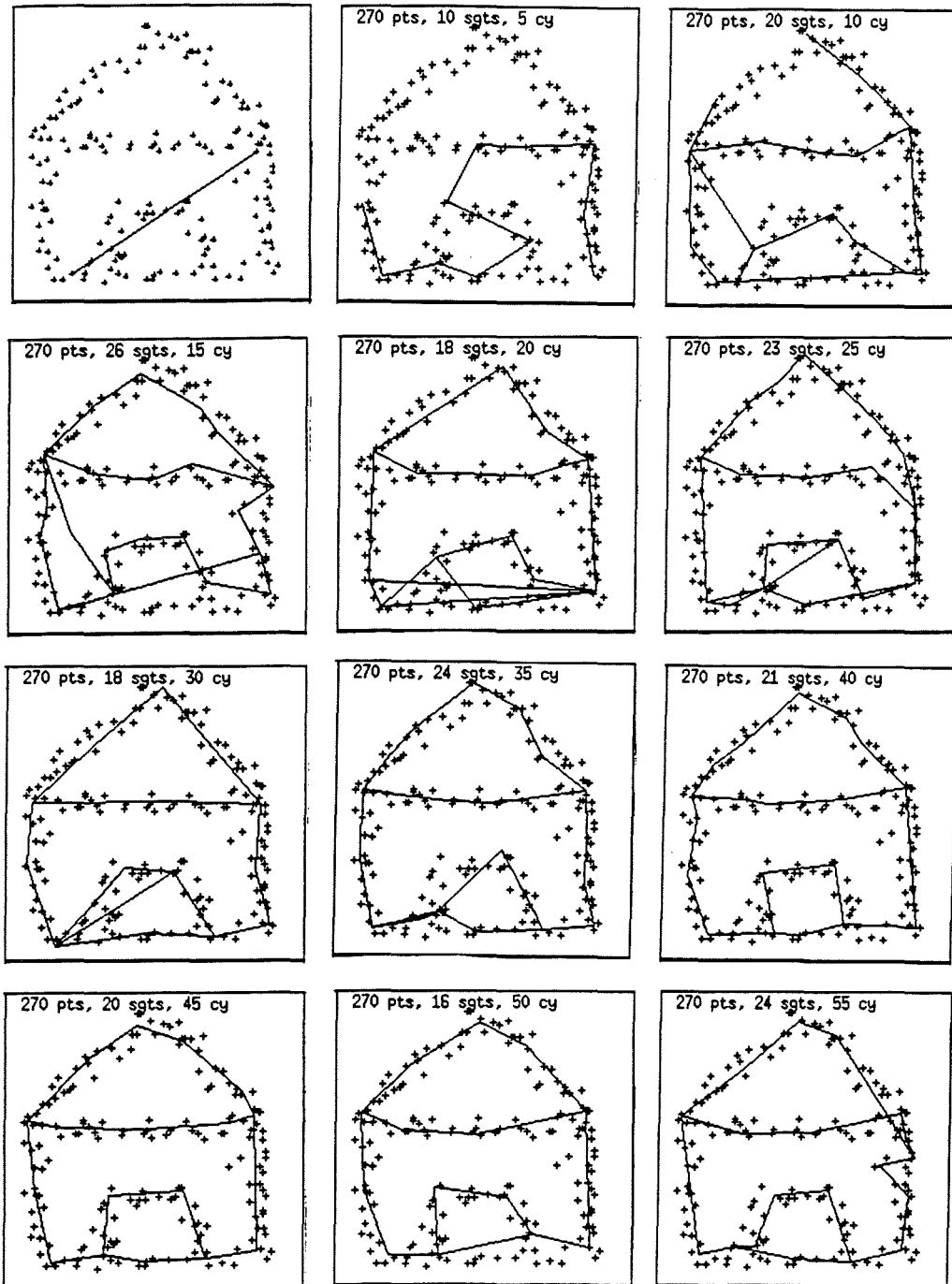
We have proposed a general framework for adaptive systems which draws a new perspective for the understanding of existing models, and for the design of new ones. This framework is founded on the deep scientific and philosophical works of Piaget and Varela.

This framework leads especially to a description of the system and its environment in the same topological space. The use of geometric components for the system and its environment is therefore particularly appropriate. The distinction between active adaptation (corresponding to component regeneration) and passive adaptation (corresponding to the component plasticity under the environment perturbations) is also very important. In our view, rich adaptation occurs when the two types of adaptation cooperate efficiently.

The second level adaptation, in which the component regeneration of a first level is itself adaptive gives the possibility of powerful regularity extractions. Our model, the self-buiding strings, confirms the interest of this approach.

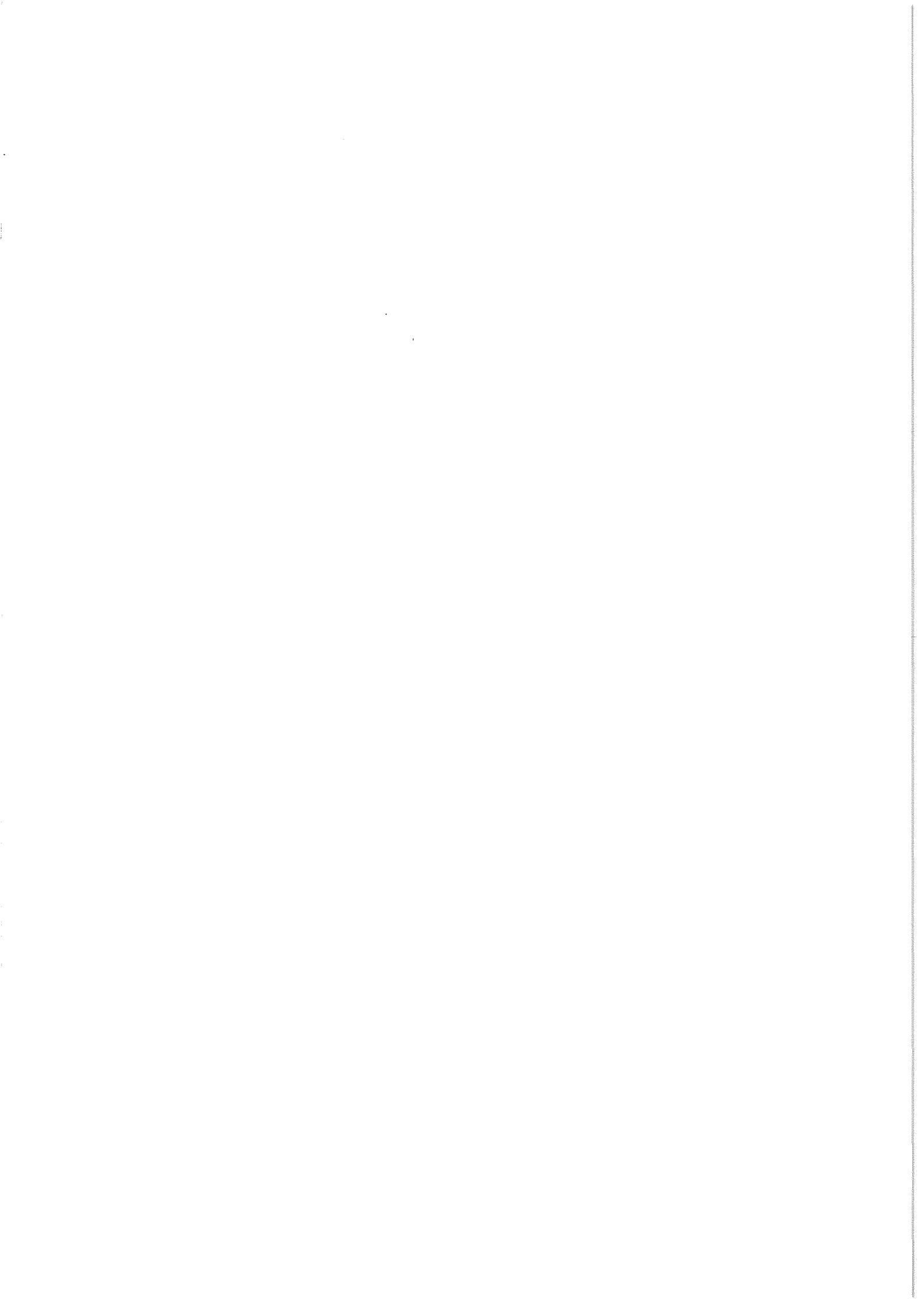
This framework corresponds to a new conception of cognitive systems in which the importance of a body, a topological form, is particularly stressed. Intelligence becomes therefore the capacity to build an adequate form with respect to the constraints of the environment.

Figure 5 : these pictures are showing the adaptation of a self-building string, starting with one segment between two randomly chosen data points. Each cycle every data point (statistically) send a perturbation to the self-building string. The system is pictured every five cycles.



References

- Bersini H. and Varela F. 1990, "Hints for Adaptive Problem Solving Gleaned from Immune Networks", in *Proceedings of Parallel Problem Solving from Nature*, 1st Workshop PPSN, Springer-Verlag.
- Brooks R., 1991, "New Approaches to Robotics", *Science* 253, pp. 1227-1232.
- Carpenter G.A. & Grossberg S., 1987, "A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine", in *Computer Vision, Graphics and Image Processing* 37, pp 54-115.
- Deffuant G., 1992, "Perceptron membrane : a new self-building connectionist network", to appear in *Neural Computation*.
- Deneubourg J.L. & Goss S., 1989, "Collective pattern and decision making", *Ethology, Ecology and Evolution*. 1, pp 295-311.
- Fahlman S.E. and Lebiere C., 1990, "The Cascade-Correlation Learning Architecture", in *Advances in Neural Information Processing Systems II* (Denver 1988), ed. D.S. Touretzky, pp 524-532, San-Mateo: Morgan Kaufmann.
- Frean M., 1990, "The upstart Algorithm : A Method For Constructing and Training FeedForward Neural Networks", *Neural Computation* 2, pp 198-209.
- Golberg D.E., 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading MA.
- Hinton G.E., 1986, "Learning Distributed Representations of Concepts", in *Proceedings of the Eighth Annual Conference of the cognitive Science Society* (Amherst 1986), pp 1-12, Hillsdale: Erlbaum.
- Holland J.H., 1975, *Adaptation in Natural and Artificial Systems*, Univ. of Michigan Press, Ann Arbor.
- Kohonen T., 1982, "Self-Organized Formation of Topologically Correct Feature Maps", in *Biological Cybernetics* 43, pp 59-69.
- Le Cun Y., 1985, "A learning scheme for asymmetric threshold networks", in *Cognitiva*, (CESTA-AFCET Ed.).
- Marchand M., Golea M. and Rujan P., 1990, "A Convergence Theorem for Sequential Learning in Two-Layer Perceptrons", *Europhysics letters* 11, pp 487-492.
- Mézard M & Nadal J.P., 1989, "Learning in Feedforward Layered Networks: The Tiling Algorithm", *Journal of Physics* 21, 1087-1092.
- Perelson A.S. and Forrest S., 1990, "Genetic algorithms and the Immune System", in *Proceedings of Parallel Problem Solving from Nature*, 1st Workshop PPSN, Springer-Verlag.
- Piaget J., 1936, *La naissance de l'Intelligence chez l'enfant*, Delachaux et Niestlé.
- Rumelhart D.E., Hinton G. and Williams R., 1986, Learning internal representations by error propagation. In J.L. McClelland D. E. Rumelhart and the PDP research group, editors, *Parallel distributed processing : Exploration in the microstructures of cognition*.
- Scalettar R., et Zee A., 1988, Emergence of Grandmother Memory in Feed Forward Networks : Learning with Noise and Forgetfulness. In *Connectionist Models and their Implications : Readings from Cognitive Science*, eds. D. Waltz and J.A. Feldman, pp 309-332. Norwood: Albex.
- Sietsma J. and Dow R.J.F., 1988, "Neural Net Pruning-Why and How", in *IEEE International Conference on Neural Networks*, San Diego, Vol I, pp 325-333, New-York: IEEE.
- Varela F., Maturana H. and Uribe R., 1974, "Autopoiesis : The Organization of Living Systems, its Characterization and a Model", in *Biosystems* 5, 187.
- Varela F., 1979, *Principles of Biological Autonomy*, Elsevier/North-Holland, New York.
- Varela F., 1991, "Organism : A meshwork of selfless selves", in *Organism and the Origin of Self*, Kluwer Associates, Dordrecht, A. Tauber (Ed.), pp. 79-107.



Simulation de systèmes auto-organisés

Renaud Cazoulat

LAIAC, Université de Caen 14032 Caen Cedex

Tel : 31 45 56 16

Email : Renaud.Cazoulat@univ-caen.fr

1) Introduction

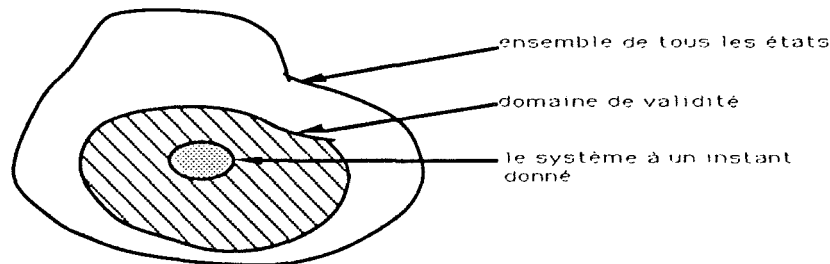
Il existe de nombreux exemples de groupes composés d'individus qui coopèrent entre eux. De cette coopération va émerger des propriétés et des comportements qu'un individu isolé est incapable de produire. Les colonies de fourmis sont l'exemple typique de cette coopération.

Afin de mieux comprendre les principes de coopérations dans ces groupes il paraît intéressant de les simuler par des systèmes artificiels. Collectivement aussi bien qu'individuellement, la simulation est basée sur un paradigme qui semble prometteur : l'**auto-organisation**.

2) Que va-t-on appeler système auto-organisé ?

C'est un système qui se caractérise par sa capacité à conserver son **identité**. Pour cela il doit répondre aux perturbations éventuelles par des actions visant à ramener le système vers un état d'équilibre.

La propriété qui en découle est le maintien de la cohérence interne du système au cours de son évolution. Le système va évoluer dans un sous ensemble d'états valides. Si le système sort de ce domaine de validité, il ne peut plus conserver son identité et donc son existence est menacée.



L'évolution du système est définie par un couple: **l'organisation** et la **structure**.

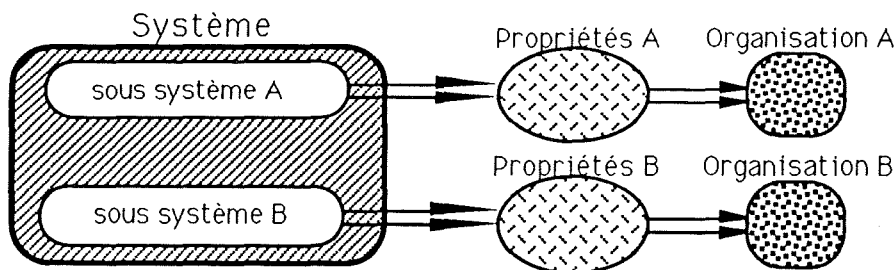
La structure décrit précisément tous les constituants du système. Ceux-ci peuvent être de plusieurs types: éléments détecteurs, moteurs, constructeurs. Ils interagissent entre eux afin de permettre au système de rester dans son domaine de validité. La manière dont ce fait cette interaction est définie par l'organisation.

L'organisation est l'ensemble de toutes les règles de fonctionnement et d'interaction des constituants. Elle est la composante principale du système. En effet deux systèmes de structure identique vont avoir des comportements différents si leur organisation est différente. Les règles de l'organisation sont en général peu complexes mais la dynamique induite par ces règles provoque l'**émergence** de nouvelles propriétés intrinsèques au système.

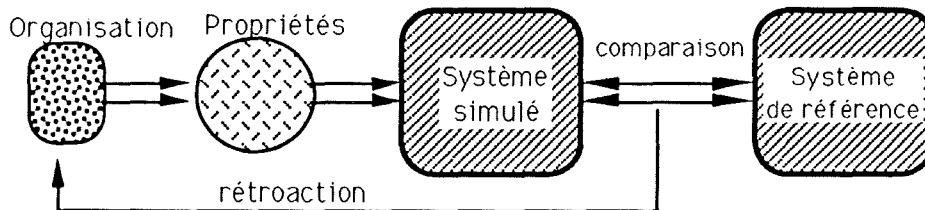
3) Intérêt de simuler un système auto-organisé.

L'étude de la propriété d'émergence est une tâche complexe. La première approche est l'**analyse descendante**. Elle consiste à regarder de manière globale un système évoluer, extraire les propriétés émergentes, puis à essayer d'en déduire les règles d'organisation de ce système. Malheureusement la déduction des règles en fonction des propriétés est une opération non triviale car il est difficile d'inférer des règles à partir d'un nombre important de propriétés émergentes.

De plus, des organisations différentes peuvent faire émerger des propriétés similaires. Autrement dit, il peut exister plusieurs explications du fonctionnement d'un système. Ces explications sont généralement incomplètes et ne portent que sur un aspect du système



Une alternative est l'**analyse ascendante**. L'optique est radicalement opposée. Pour étudier un système donné, on va bâtir des hypothèses sur son fonctionnement et construire un modèle de ce système, avec sa structure et son organisation. La simulation de ce modèle permet de comparer les propriétés qui en émergent avec celle du système de référence et suivant la validité des résultats obtenus, on va modifier les hypothèses et donc la structure et/ou l'organisation du modèle de façon à obtenir le maximum de similitude avec le système de référence.



Cette approche permet de contrôler complètement le fonctionnement du système simulé et donc de proposer une explication du système de référence qui, même si elle n'est pas **strictement** exacte, reste une explication valable et complète dans l'ensemble des explications possibles de ce système, tant que des différences ne se sont pas manifestées.

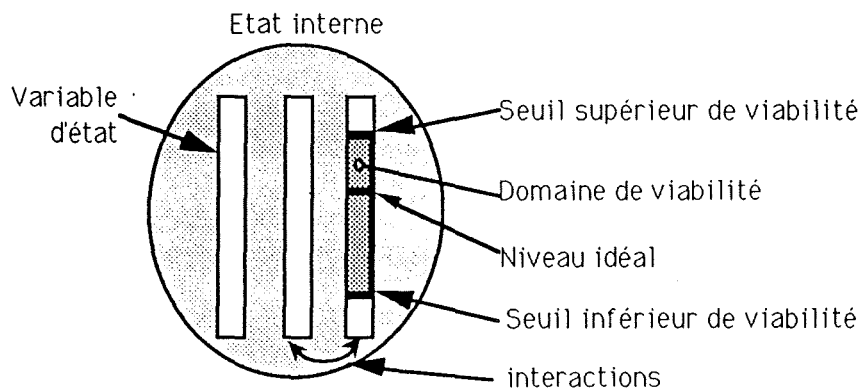
4) Application: simulation de vie artificielle.

On peut considérer une unité vivante comme un système auto-organisé. Le domaine de validité est aisément définissable: c'est l'ensemble des états internes qui permet à l'unité de rester vivante. Il peut être appelé **domaine de viabilité**. Il faut donc avant tout définir l'état interne.

4.1) L'état interne.

Il est composé de **variables d'états**. L'ensemble de toutes les valeurs des variables d'états à définir, à un instant donné, un des états possibles. Ces valeurs vont

devoir être comprises dans un **intervalle de viabilité** pour que le système reste dans son domaine de viabilité. Les bornes de cet intervalle sont les **seuils de viabilité**. Une valeur particulière est le **niveau idéal** que la variable doit atteindre pour que le système soit dans l'état le plus stable.



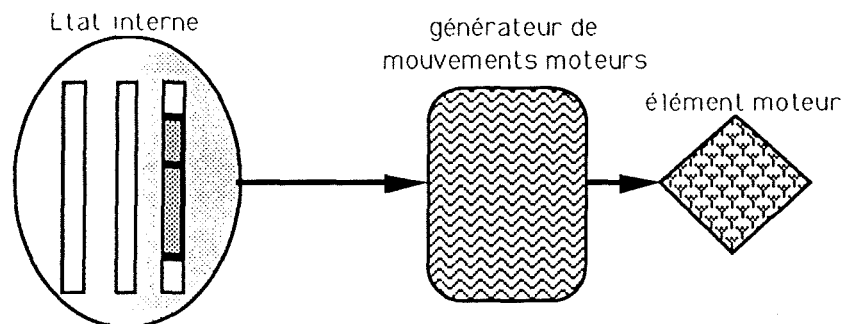
Les seuils de viabilité constituent les niveaux d'alertes du système. En effet en dehors de cet intervalle, l'intégrité du système risque d'être compromise. Ces variables peuvent représenter une donnée particulière mais elles peuvent également être inter-dépendantes, une modification d'une variable entraînant la modification d'une autre.

Dans un système vivant, ces variables correspondent en général à des indicateurs de bon fonctionnement. Par exemple le taux de glucose dans le sang est un de ces indicateurs. Si ce taux varie dans des proportions importantes, et ainsi compromet la survie du système, il va sortir du domaine de viabilité et atteindre les niveaux d'alertes. Le système pourra donc prendre des mesures d'urgence en conséquence.

4.2) L'expressivité du système.

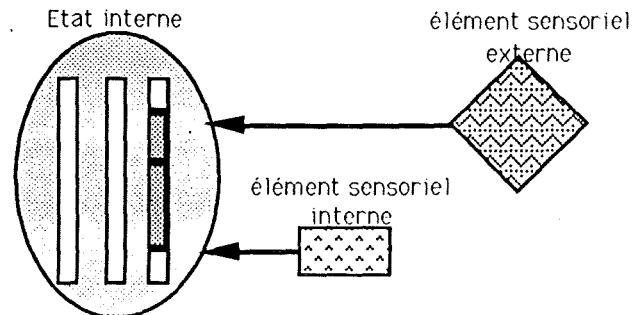
Le système doit en permanence rétablir son équilibre interne. L'équilibre interne est atteint quand toutes les variables internes sont proches de leur valeur idéale ou au moins dans leur domaine de validité. Le système va répondre aux perturbations par des réactions pour garantir son équilibre : c'est l'homéostasie. Ces réactions vont se faire au travers des systèmes moteurs.

Dans un système simulé, l'activité motrice va être contrôlée par un générateur aléatoire de mouvements. Ce générateur va tirer au sort un mouvement parmi tous les mouvements possibles. Le générateur dépend directement de l'état interne qui peut ainsi moduler l'activité motrice.



Par exemple l'activité motrice pourra être ralenti si l'état interne s'approche des limites du domaine de viabilité mais si un seuil d'alerte est atteint l'activité motrice sera au contraire stimulée.

Pour avoir connaissance des perturbations qu'il subit le système doit posséder des capteurs sensoriels. Deux types de capteurs se présentent : ceux qui réagissent aux perturbations extérieures et ceux qui réagissent à un changement d'état physiologique. Ces capteurs sensoriels vont influencer directement les variables de l'état interne et créer ainsi un processus d'attention (ou d'alerte) **extéroceptif**.

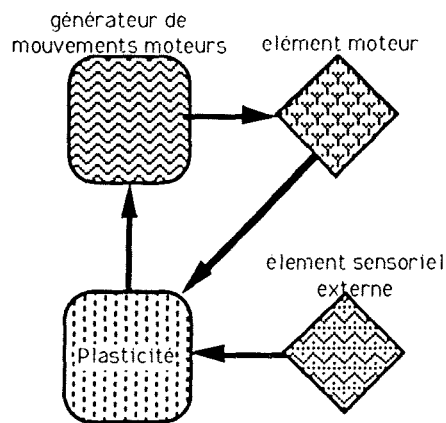


4.3) La plasticité du système.

Le système ainsi défini va fonctionner suivant la schéma d'influence:

élément sensoriel -> état interne -> élément moteur.

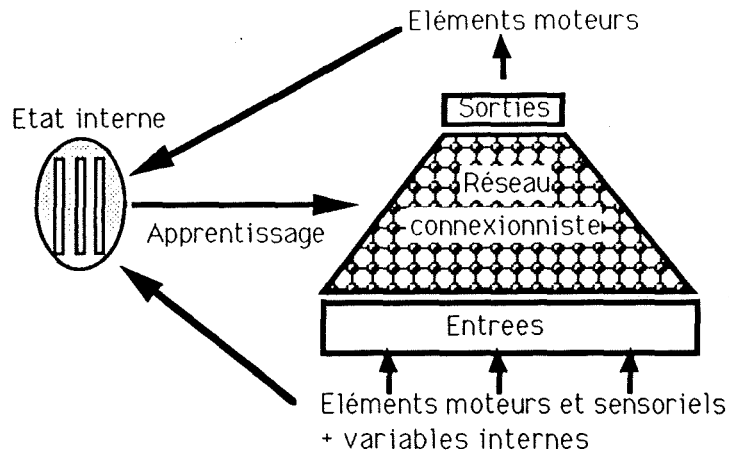
Pour réaliser une boucle complète et rompre la rigidité de ce schéma, il faut introduire un lien entre élément moteur et élément sensoriel. Ce lien proprioceptif est interne au système et permet une interaction directe entre éléments moteurs et sensoriels indépendamment de l'état interne. Ceci permet de créer une boucle **sensori-motrice** dans laquelle le mouvement moteur influe sur la perception sensorielle pour induire un nouveau mouvement moteur.



Un exemple de cette boucle sensori-motrice est le mouvement des yeux : lorsque l'on bouge la tête le point fixé par les yeux va se déplacer. Les yeux vont devoir bouger également pour contrebalancer le mouvement de la tête et continuer à fixer le même point.

Mais le système est complètement câblé. Il produira toujours des actions identiques pour des perturbations identiques. Pour lui ajouter la possibilité d'apprentissage et d'évolution, le système peut être doté d'un système d'apprentissage basé sur un réseau connexionniste. Ce réseau est le lien entre la boucle sensori-motrice et les variables internes.

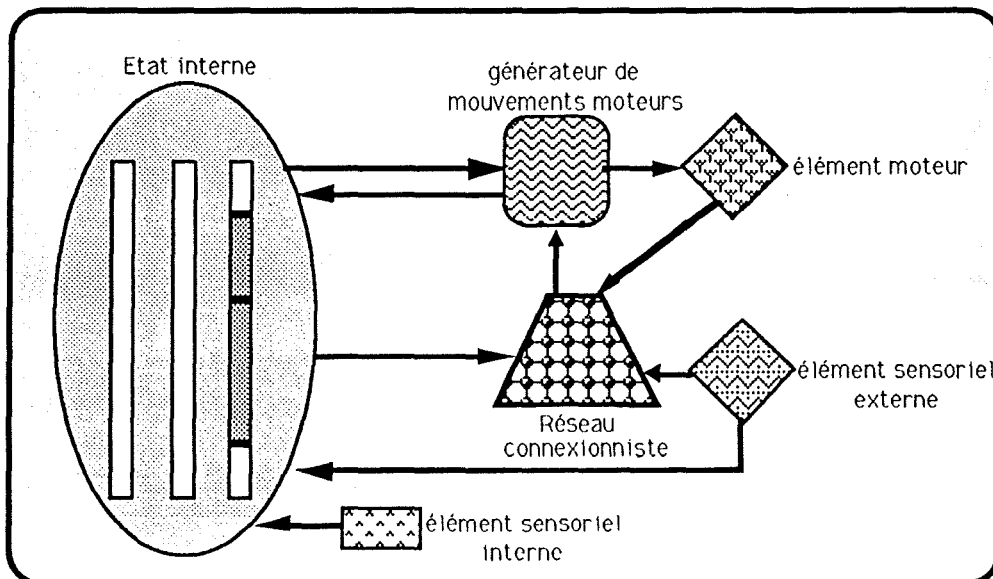
L'état interne, les éléments récepteurs et moteurs vont engendrer via le réseau des mouvements moteurs qui eux-même vont entraîner une modification de l'état interne et de la réponse des éléments récepteurs. Ces modifications vont d'une part déclencher l'apprentissage du réseau et d'autre part engendrer de nouveaux mouvements moteurs. Les interactions réalisent donc une boucle **auto-alimentée**. Le système est constamment en activité. Les perturbations s'inscrivent de manière naturelle dans cette boucle, ce ne sont pas elles qui déclenchent l'activité du système



L'apprentissage du réseau est basé sur la punition et la récompense. Toute action qui rapprochera l'état interne de l'état stable sera validée par le renforcement de la réponse du réseau. Réciproquement toute action qui éloignera l'état interne d'un état stable entraînera l'apprentissage de la réponse contraire.

Les possibilités induites avec l'introduction du réseau connexionniste sont primordiales pour obtenir l'émergence d'un nombre plus important de propriétés. De plus le système acquiert une individualité lors de son ontogenèse par son apprentissage qui sera différent de celui de tout autre système. C'est donc le moyen de différencier un système d'un autre tout en conservant l'identité de structure et d'organisation.

La structure et l'organisation d'un système auto-organisé sont récapitulés par le schéma suivant:



5) L'environnement d'un système.

Simuler un système implique de simuler également l'environnement dans lequel il va évoluer. En se basant sur un premier essai de simulation d'un environnement possédant des caractéristiques empruntées au monde des grenouilles, on peut décrire les caractéristiques fixes qu'un environnement doit posséder:

- des données matérielles, qui vont correspondre à des conditions physiques, par exemple des obstacles, les limites d'un étang.

- des lois déterministes, qui régissent des événements stables comme la présence d'un prédateur non mobile.

Cet environnement va également contenir des événements autonomes qui vont répondre à des lois de probabilités, comme la présence de proies, de prédateurs mobiles et surtout d'autres systèmes.

Suivant la situation du système dans son environnement, il va être soumis à des stimulations qui vont engendrer des actions motrices qui vont à leur tour changer la situation du système. La boucle de rétro-action est conservée.

Lorsque l'état interne du système va sortir du domaine de viabilité, la conséquence est la disparition du système de son environnement, c'est à dire la mort. Si l'on considère l'ensemble des systèmes comme étant lui-même un système auto-organisé, pour pallier à la disparition d'un des constituants, il faut introduire la possibilité de création d'un nouveau constituant : la naissance. La naissance peut être spontanée ou produite par la **reproduction** des constituants. Cette reproduction permet de conserver les systèmes qui vont réagir de manière adéquate aux conditions de l'environnement et ainsi réaliser un apprentissage génétique.

6) Conclusion.

Cette représentation d'un système artificiel permet de le simuler sur ordinateur. Cette simulation informatique est un moyen pratique de vérifier les hypothèses de coopérations entre les constituants d'un système. De plus, si l'on considère l'ensemble composé de l'environnement et des systèmes comme un "méta-système" (une sorte d'écosystème), la simulation permet également de dégager des règles de coopérations entre les systèmes eux-mêmes.