# Réflexions sur l'apprentissage dans un univers multi-agents

Valérie Camps, Marie-Pierre Gleizes

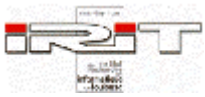
Systèmes Multi-Agents Coopératifs

# Réflexions sur l'apprentissage en univers multi-agents

**Journée du PRC GDR IA "Les systèmes multi-agents", Toulouse, Février 1996 et Journées du PRC GDR IA, Grenoble Mars 1997, Editions Hermès**

*Valérie CAMPS, Marie-Pierre GLEIZES*

*Valerie.Camps@univ-lr.fr, Marie-Pierre.Gleizes@irit.fr*



*Institut de Recherche en Informatique de Toulouse*

*118, Route de Narbonne*

*31062 Toulouse, Cedex - France*

# RÉFLEXIONS SUR L'APPRENTISSAGE EN UNIVERS MULTI-AGENTS

**Valérie Camps, Marie-Pierre Gleizes**
IRIT _ Université Paul Sabatier
118, route de Narbonne 31062 Toulouse cedex
{camps, gleizes}@irit.fr

**Résumé :**

La faculté d'apprentissage est nécessaire à tout système intelligent. L'apprentissage est d'ailleurs étudié dans le but d'améliorer le fonctionnement du système en lui permettant d'évoluer. Dans les systèmes d'Intelligence Artificielle Distribuée, le système dans son intégralité est capable d'exécuter des tâches dont la complexité est supérieure à la somme des tâches exécutées par les agents pris individuellement. Aussi l'apprentissage est plus complexe que dans les systèmes d'IA classiques car il faut prendre en compte le collectif. Apprendre dans les systèmes d'intelligence artificielle distribuée et dans les systèmes multi-agents peut être vu soit comme un apprentissage du collectif, soit comme un apprentissage individuel à partir du groupe.

La première partie de l'article est consacrée à présenter quelques méthodes ou modèles d'apprentissage mis en oeuvre en IAD et dans les systèmes multi-agents. Dans la seconde partie, sont définies les principales caractéristiques de l'activité d'apprentissage telles que : la distribution, le respect de l'autonomie des agents et l'utilisation de la coopération. Enfin, les réflexions en cours sur l'apprentissage au sein du projet LIAC sont exposées. Le papier se termine par la présentation d'un mécanisme permettant de mettre en oeuvre l'apprentissage au sein d'une organisation : la relaxation restreinte.

**Mots-clés :**

Intelligence artificielle distribuée, système multi-agent, autonomie, collectif, apprentissage.

## 1. L'APPRENTISSAGE DANS DES SYSTÈMES MULTI-AGENTS

L'apprentissage est une aptitude nécessaire pour amener un système à évoluer. De part le contexte d'utilisation des systèmes multi-agents, l'apprentissage semble être une faculté indispensable tant pour l'amélioration du fonctionnement du système que pour faciliter le développement d'application. C'est pourquoi de nombreux modèles commencent à être définis. Nous en présentons ci-dessous quelques uns des plus significatifs.

### 1.1 Le modèle de Sian

L'apprentissage dans le système de Sian (Sian, 91) est étudié pour que les agents améliorent leurs connaissances sur le domaine. L'originalité de cette méthode d'apprentissage réside dans le fait qu'elle utilise la coopération pour sa mise en oeuvre. Les agents confirment ou infirment, par des interactions coopératives, des hypothèses qu'ils ont déduites. Pour cela, les agents utilisent l'induction sur des exemples d'événements ayant eu lieu au sein de l'environnement qui est identique pour tous.

Le modèle présenté, appelé MALE (Multi-Agent Learning Environment), est écrit en PROLOG et est constitué :

- d'agents qui apprennent incrémentalement,
- d'opérateurs qui permettent aux agents d'exprimer leurs opinions,
- d'un tableau noir pour la communication entre les agents,
- d'une fonction d'évaluation des opinions individuelles,
- d'un mécanisme d'intégration des nouvelles connaissances.

Le domaine d'application de MALE concerne le marché des matières premières. Le système est constitué de trois agents qui ont des connaissances concernant différents pays : l'Afrique, l'Amérique et l'Asie. La décomposition est issue de contraintes spatiales. Le but des agents est de prédire les prix des matières premières.

Le mécanisme d'apprentissage est essentiellement basé sur l'interaction entre agents. Pour apprendre, ils échangent des opinions et des jugements de valeurs sur des hypothèses. A chaque hypothèse est associée une valeur de confiance basée sur le nombre d'instances utilisées pour la générer. L'apprentissage a lieu quand tous les agents ont réussi à trouver un consensus sur l'hypothèse considérée.

Les trois phases du cycle d'apprentissage sont :
- la génération d'hypothèses initiales par un des agents. Différents critères peuvent être définis pour le choix des hypothèses initiales à prendre en compte pour l'apprentissage (évidence, échec pour l'explication, une force associée aux croyances...).
- la coopération pour arriver à une version de ces hypothèses acceptée par tous. Celle-ci est vue en termes de représentation des connaissances, d'évaluation d'hypothèses et d'unification de toutes les opinions reçues.
- l'intégration des informations ainsi apprises par les agents. A ce niveau, le problème de la révision des croyances se pose pour intégrer les nouvelles connaissances.

Cette méthode demande donc une homogénéité au niveau de la représentation des connaissances, l'utilisation d'un langage commun pour communiquer une opinion sur une hypothèse, et la définition de fonctions d'évaluation des hypothèses. Les agents communiquent au moyen d'un tableau noir et d'un langage composé de 9 opérateurs relatifs à

des hypothèses tels que :
- changer le statut d'une hypothèse de "proposée" en "convenue",
- n'exprimer aucune opinion sur une hypothèse.

Tous les agents sont dotés d'un mécanisme d'apprentissage identique. Un agent décide du choix des hypothèses sur lesquelles va porter l'apprentissage et sollicite les autres pour apprendre. Ainsi le mécanisme d'apprentissage est distribué et garantit l'autonomie des agents.

### 1.2 Le modèle de Sekaran et de Sen

Le modèle présenté dans (Sekaran, 94) utilise l'algorithme du Q-learning basé sur le schéma d'apprentissage par renforcement. Après s'être intéressés à l'apprentissage dans un domaine coopératif où deux agents travaillent ensemble sur une tâche commune sans partage explicite de connaissance ou d'information (Sen, 94), Sekaran et Sen étendent leur approche à un domaine non coopératif, où les agents ont des buts conflictuels. Ces agents ne partagent ni connaissance, ni résultat. Ils apprennent une stratégie à partir d'observations et l'améliorent avec l'expérience. L'originalité de ce travail est qu'il n'y a ni échange de connaissance explicite entre les agents ni dépendance de relations.
Cette méthode est mise en oeuvre dans le monde des blocs dans lequel deux agents autonomes ayant différentes capacités et aucune connaissance sur le domaine, doivent pousser un bloc d'une position initiale à une position finale, dans un espace euclidien. A chaque étape, chaque agent applique une force suivant un certain angle et le bloc est déplacé. La session est terminée dès que le bloc atteint l'une des deux positions finales ou lorsqu'il est hors du champ de jeu. Chaque agent apprend, par l'intermédiaire de tentatives répétées, une politique optimale pour atteindre la position voulue.
Le retour (feed-back) reçu après chaque action est une fonction inversement exponentielle de la distance entre la position courante du bloc et le chemin optimal associé à leurs buts respectifs. Sur la base de ces feed-back, les agents mettent à jour leur politique individuelle en utilisant l'algorithme du Q-learning. La convergence des matrices de stratégies (il n'y a pas de mise à jour significative des politiques individuelles sur un ensemble d'essais) est utilisée comme critère d'arrêt pour les expériences.

Le champ de force d'un agent est plus grand que celui d'un autre agent s'il est capable de vaincre l'autre agent et de pousser le bloc vers son but. Le nombre d'essais pour converger augmente lorsque le champ de force de l'agent le plus faible augmente (parce qu'il offre plus de résistance).
Lorsque les deux agents possèdent des capacités identiques, aucun d'eux n'est capable de pousser le bloc vers son but. Ils manoeuvrent alors pour pousser le bloc dans une position intermédiaire. Comme la force de l'agent le plus faible augmente, la position finale s'éloigne de la position de l'agent le plus fort. Ce phénomène est accentué lorsque le nombre d'options disponibles pour l'agent le plus faible augmente. Ceci parce que l'agent le plus fort converge vers une politique sous optimale ce qui peut être évité en choisissant un schéma d'actions probabilistes plutôt qu'une politique déterministe.

Les auteurs montrent que les agents peuvent utiliser des schémas d'apprentissage de renforcement pour atteindre leur but, qu'ils soient coopératifs ou conflictuels, et sans avoir de modèle l'un de l'autre.

### 1.3 Le système de Mataric

Le système multi-agent étudié (Mataric, 94) est composé d'un groupe de 4 robots mobiles réalisant une tâche de fourragement pour rechercher des palets dans une zone donnée et les ramener à la "maison". Les agents n'ont pas de modèle prédéfini du monde. Ils sont tous homogènes car construits à l'identique et n'emploient pas de modèle explicite des autres. La notion de groupe n'apporte aucun effet supplémentaire car le comportement collectif n'est ici que la somme des activités individuelles.

Les comportements sont des lois de contrôle guidées par le but dans un domaine donné et

pour atteindre un ensemble de buts. Les comportements sont plus généraux que les actions car ils ne détaillent pas les états. Le répertoire des comportements de base utiles pour le fourragement est simple et fixé comme suit :
- droite-ligne (déplacement en avant en évitant les collisions),
- dispersion (respect d'une distance minimum entre les robots),
- repos,
- retour-maison.
La combinaison de ces 4 comportements de base peut être intéressante pour répondre à une grande variété de situations et forment donc un ensemble intéressant à apprendre en fonction des conditions définies ci-dessous.

Les conditions sont des prédicats sur les capteurs qui correspondent à un sous-ensemble particulier de l'espace d'états. Chaque condition est définie comme une partie de l'état nécessaire et suffisant pour l'activation d'un comportement particulier.
Les conditions des prédicats sont : avoir-palet?, a-la-maison?, intrus-proche?, heure-nocturne?, prise?, relâchement?. Les conditions de prise et de relâchement sont câblées dans les robots. Dès qu'un robot détecte un palet entre ses doigts, il l'agrippe ; similairement, dès qu'il arrive à la maison avec un palet, il le relâche.

Apprendre à sélectionner des comportements est par définition un problème d'apprentissage par renforcement, car il est fondé sur une corrélation entre les comportements effectués et les résultats qu'il reçoit en retour. Les algorithmes d'apprentissage par renforcement tentent de maximiser les récompenses et de minimiser les punitions au cours du temps.
L'apprentissage par renforcement dans des domaines situés peut être formulé comme l'apprentissage des conditions nécessaires et suffisantes pour l'activation de chaque comportement du répertoire, tel que le comportement de l'agent maximise ses récompenses au cours du temps. La stratégie d'apprentissage par renforcement utilise le Q-learning et est strictement individualiste; les agents apprennent par eux-mêmes. Le mécanisme d'apprentissage est identique et distribué dans chacun des robots.
Les événements suivants produisent des renforcements positifs immédiats : saisir un palet, déposer un palet à la maison et dormir à la maison. Les événements suivants produisent au contraire des renforcements négatifs immédiats : déposer un palet hors de la maison, dormir hors de la maison.

Lorsque le groupe de robots augmente, il y a un accroissement des interférences entre les agents et un déclin de performance pour tous les algorithmes étudiés. Dans un scénario idéal, les autres agents devraient favoriser l'apprentissage individuel. Mais cela n'est possible que dans des sociétés où l'apprentissage est mutuel avec des règles sociales. Ces règles sociales impliquent des comportements altruistes. Un avantage central de l'apprentissage est la possibilité d'apprendre plusieurs types de comportements en parallèle.

### 1.4 La méthode de Shaw

Shaw (Shaw, 89) a travaillé sur l'apprentissage des compétences et des connaissances des agents ; sa méthode est destinée à des sociétés d'agents cognitifs, qui communiquent par envoi de messages. Il considère les systèmes de DAI comme des organisations adaptatives, capables de s'enrichir en apprenant à partir d'une expérience passée. Ainsi, les agents sont dotés de mécanismes qui leur permettent de s'améliorer périodiquement et progressivement.

La méthode est basée sur deux processus : un processus d'offre de résolution de tâches et un

processus de transformation génétique.
Le processus d'offre est une extension du réseau contractuel de Smith et Davis (Smith, 80) dans lequel les agents travaillent par partage de tâches en respectant le cycle suivant :

- décomposition d'une tâche en sous-tâches,
- allocation des sous-tâches aux agents,
- résolution des sous-tâches par les agents choisis,
- intégration des solutions partielles.

L'extension consiste à récompenser les agents sélectionnés pour résoudre une tâche au cours du processus de résolution. Cette récompense permet d'augmenter la "force" de l'agent. La force quantifie donc le service rendu par l'agent à la collectivité, elle reflète sa capacité à résoudre des tâches. Elle est utilisée ensuite pour déterminer quels sont les caractéristiques importantes des agents.
Grâce au processus de détection des caractéristiques désirables, des algorithmes génétiques de mutation ou de croisement sont utilisés pour remplacer les agents "faibles" par des agents "forts". La performance globale du système est ainsi augmentée. Pour cela, chaque agent possède une représentation génétique de ses caractéristiques et de ses capacités.
Ce sont donc les agents qui interviennent le plus dans le processus de résolution qui sont choisis pour avoir leurs capacités dupliquées car ils possèdent les forces les plus grandes. La coopération est le critère utilisé pour guider l'apprentissage au sein de ce système. Shaw l'exprime ainsi : "The better agents that have been successfully completing more tasks would increasingly be favored by the bidding process..., the less useful agents are forced to adapt and change their configuration by mimicking some of the features of the successful agents."

L'architecture des systèmes à base de classifieur a été appliquée de manière distribuée au sein de chacun des agents pour permettre l'apprentissage. Par contre, les algorithmes génétiques sont centralisés car ils nécessitent la connaissance des forces assignées à tous les agents du système.

Cette méthode a été mise en oeuvre dans le cadre d'un système de fabrication. Chaque agent est un intervenant au niveau de la fabrication et les chromosomes représentent tout ce que l'agent sait faire.
Dans ce système, l'apprentissage concerne les connaissances des agents puisque les agents qui vont être modifiés intègrent les caractéristiques des agents les plus sollicités. Cette modification entraîne aussi une réorganisation du système c'est-à-dire que les interactions vont se produire de manière différente.

### 1.5 La méthode de Weiβ

Weiβ (Weiβ, 93) propose une méthode destinée à des sociétés d'agents réactifs qui communiquent constamment avec leur environnement et entre eux par envoi de messages. Cette méthode s'inscrit dans le cadre d'un apprentissage collectif. Les agents possèdent une vision incomplète de leur environnement. Ainsi, un agent a seulement une information locale sur l'état de l'environnement, et cette information peut différer de celle que possède un autre agent. Il n'y a pas de hiérarchie entre les agents.

La méthode d'apprentissage distribuée préconisée vise à une meilleure coordination entre les agents de la société. Pour cela, deux algorithmes d'apprentissage, ACE (Action Estimation) et AGE (Action Group Estimation), tous deux issus de l'algorithme du "bucket brigade" ont été développés. Le traitement réalisé par ces deux algorithmes peut être décrit de manière

simplifiée comme l'exécution répétée du cycle suivant :

- la détermination de l'action. Dans une première étape chaque agent détermine, suivant ce qu'il connaît de l'état de l'environnement, l'ensemble des actions qu'il peut effectuer ;
- la compétition. Les agents concourent pour le droit de devenir actif. Cette compétition englobe le calcul et l'annonce de l'offre ainsi que la sélection des actions qui sont actuellement effectuées ;
- la répartition des crédits. Dans cette dernière étape, les agents attribuent à l'aide de la méthode du "bucket brigade", un crédit à chacun des autres agents selon l'estimation de leurs actions.

Les deux algorithmes ACE et AGE proposés se différencient par leur manière de mettre en oeuvre la seconde étape, à savoir, la compétition.
Dans la méthode ACE, les agents concourent pour pouvoir effectuer des actions individuelles. Chaque agent fait une offre pour chacune des actions qu'il peut effectuer et l'annonce à tous les autres agents qui sélectionnent alors l'offre la plus élevée parmi toutes les offres proposées. L'agent qui possède cette dernière est autorisé à exécuter l'action choisie.
Dans la méthode AGE, les agents concourent pour effectuer des groupes d'actions et non plus des actions individuelles. Pour cela, un agent estime la pertinence d'un but ou d'une action non seulement sur la dépendance entre sa connaissance et l'état courant de l'environnement mais aussi sur la dépendance des contextes des activités possibles.

Une telle séquence nécessite que les agents connaissent les autres, et qu'ils possèdent le même algorithme d'apprentissage notamment pour le calcul des offres. Il est essentiel que les agents soient rationnels, et non égoïstes c'est-à-dire qu'il ne faut pas que les agents insistent pour exécuter une action incompatible avec les actions précédemment exécutées. Le choix d'une offre parmi toutes celles proposées induit également un contrôle centralisé ce qui supprime l'autonomie de l'agent pour l'apprentissage.

Ces deux algorithmes ont été étudiés expérimentalement dans le monde des blocs. Les agents doivent à partir d'une configuration initiale aboutir à une configuration finale. Chaque agent est responsable d'un cube. Tout en ayant une vue limitée de l'état de l'environnement, les agents parviennent à agencer les cubes pour atteindre la configuration désirée. Cette application montre que les deux algorithmes sont capables d'étudier des séquences stables d'ensembles d'actions et que l'algorithme AGE fournit de meilleurs résultats que l'algorithme ACE.

## 2 DES ORIENTATIONS POUR L'APPRENTISSAGE

### 2.1 Analyse des modèles existants

Dans ces quelques modèles, les auteurs montrent que l'activité d'apprentissage collectif permet au système d'être plus performant (du point de vue de son fonctionnement) qu'un système dans lequel l'apprentissage est réalisé de manière classique.
L'étude de ces méthodes fait apparaître deux niveaux d'apprentissage : les connaissances des agents s'enrichissent soit au niveau du domaine des compétences, soit au niveau des interactions entre les agents. L'apprentissage classique en IA permet certes d'enrichir les connaissances relatives au domaine, mais le fait que l'agent interagisse avec d'autres lui permet d'apprendre plus sur le domaine que s'il est isolé. C'est le cas dans la méthode utilisée

par Sian (Sian, 91) et par Sekaran (Sekaran, 94). Les performances associées au fonctionnement du système, sont aussi améliorées par le fait que les agents se coordonnent ou interagissent mieux comme avec la méthode de Weiβ (Weiβ, 93).

Dans la plupart des systèmes étudiés, les agents sont autonomes pour la résolution, mais dans les méthodes d'apprentissage présentées par Weiβ et Shaw une synchronisation est requise pour apprendre. Ceci peut nuire à l'autonomie de l'agent. Ces réalisations montrent aussi que tous les agents ont tendance à utiliser le même algorithme d'apprentissage.

A partir de cet existant, on peut dégager les principales contraintes auxquelles doit répondre l'activité d'apprentissage.
Tout d'abord, il est essentiel que le processus d'apprentissage du système multi-agent soit implémenté de manière **distribuée** c'est-à-dire qu'il soit localisé chez les agents et qu'il opère à partir des informations connues localement par l'agent qui l'abrite. De plus, chaque agent doit pouvoir utiliser un processus d'apprentissage adéquat c'est-à-dire lié à la manière dont est implémenté l'agent. Le mécanisme d'apprentissage peut différer d'un agent à l'autre, ceci est d'autant plus réalisable qu'il est distribué. Comme nous l'avons vu dans les travaux exposés précédemment, les agents peuvent apprendre par observation des autres mais aussi en communiquant. Cette contrainte de distribution est en accord avec le fait que la distribution à la fois des données et du contrôle est de plus en plus présente dans les systèmes multi-agents actuels. Les agents n'ont donc qu'une vue partielle des autres agents et de l'environnement. Il peut sembler, au premier abord, que les méthodes d'apprentissage mises en oeuvre dans un tel contexte soient moins performantes que celles mises en oeuvre dans le cadre de sociétés où tous les agents ont une vue globale du système, pourtant elles sont excellentes du fait de l'incomplétude des connaissances.

Ensuite, l'activité d'apprentissage ne doit pas nuire à l'autonomie des agents. Cette autonomie leur permet d'avoir un fonctionnement asynchrone dans leurs activités de raisonnement, d'action sur l'environnement, de communication. Pour conserver cette notion d'asynchronisme au niveau de l'apprentissage, il faut que l'agent décide, en fonction de son état et de ses connaissances locales, de l'instant où il doit apprendre. Le processus d'apprentissage doit donc être **asynchrone**.

Finalement, il nous semble fondamental que la **coopération** guide l'apprentissage. En effet, un des concepts clés des systèmes multi-agents est la coopération. Elle est utilisée au cours de l'apprentissage soit pour déterminer quelle connaissance relative au domaine apprendre (Sian), soit pour déterminer quelles caractéristiques posséder (Shaw)...

### 2.2 Le principe du système LIAC

Les hypothèses de travail pour l'apprentissage dans LIAC (Langage d'Intelligence Artificielle Collective) sont induites par les propriétés des systèmes multi-agents. Les caractéristiques tant des agents que du système contraignent de manière forte la réalisation de l'activité d'apprentissage. Le système multi-agent doit favoriser l'aspect distribué en son sein et être ouvert. De plus, la conception d'un système multi-agent doit conserver l'autonomie des agents et le principe de localité. Un des points fondamentaux est la coopération entre les agents. C'est pourquoi l'apprentissage doit respecter et étendre ces principes.

Les agents de LIAC sont composés de connaissances telles que ses compétences, de croyances sur les autres et d'attitudes sociales. D'une manière générale les **attitudes sociales**

peuvent être définies à différents degrés : la sincérité ou le mensonge, l'altruisme ou l'égoïsme... Elles conditionnent fortement le comportement d'un individu au sein d'un collectif**.** Diverses études ont été réalisées notamment par Goldman et Rosenschein (Goldman, 94) et Sekaran et Sen (Sekaran, 95), (Sen, 95) pour analyser l'influence des attitudes sociales des agents sur le comportement collectif. Les résultats obtenus montrent que le système fonctionne mieux quand les agents coopèrent. C'est pourquoi nous avons défini la coopération comme attitude sociale exclusive des agents de LIAC. Les agents ont pour objectif de détecter et de tenter d'éliminer les situations non coopératives. Une implantation de ces attitudes a été testée sur le domaine du "tileworld" (Piquemal, 96). Les comparaisons effectuées entre des agents ayant différentes attitudes sociales montrent que les meilleures performances sont atteintes avec la coopération.

**Structure d'un agent générique**

Les connaissances que le système doit apprendre ne sont pas celles liées aux compétences sur le domaine d'application. Cet apprentissage peut en outre être réalisé avec les méthodes classiques ou à partir de l'observation des autres agents. Notre intérêt est centré sur l'apprentissage collectif de l'organisation et concerne les croyances. Il permet à un système d'être plus performant dans le sens où la coordination et/ou la coopération s'améliore(nt) au cours de sa vie. Ceci est réalisé en recherchant une meilleure organisation c'est-à-dire que l'agent soit au bon endroit, au bon moment, dans le chemin du raisonnement.

Comment tendre vers une organisation optimale? Il faut supprimer la propagation d'information inutile du point de vue du récepteur. Par exemple, il faut supprimer le nombre d'agents intermédiaires pour la réponse à une requête ; il faut également supprimer l'envoi de résultat à un agent qui n'en fera rien. Ceci permettra d'atteindre plus rapidement la solution, car diminuer le nombre de communications entre les agents et diminuer un fonctionnement non rentable des agents conduit à une utilisation optimale des ressources.

L'agent doit être doté de capacités à détecter une organisation non optimale, par exemple lorsqu'il reçoit une information dont il ne comprend pas le sens. Mais aussi il doit pouvoir participer à une modification de l'organisation à laquelle il appartient pour atteindre l'organisation optimale. Pour cela l'agent doit être rationnel et altruiste. L'organisation optimale peut être définie par "tout est bien utilisé". Pour arriver à une organisation optimale,

il faut que le système s'auto-organise, que les agents aient des aptitudes à s'auto-organiser. Ainsi nous pensons que **l'apprentissage dans un univers multi-agent correspond à l'auto-organisation du système**.

Étudier l'apprentissage en intelligence artificielle collective comme la recherche exclusive d'une organisation optimale peut sembler réducteur dans le sens où :
- chaque agent possède un corpus de compétences fixé initialement et invariable,
- le corpus de compétences est complet dans le système i.e. il existe une organisation optimale qui peut fournir les comportements souhaités du système.

C'est une démarche scientifique rigoureuse qui nous y a conduit, car apprendre simultanément deux choses (l'organisation d'une part et la compétence d'autre part) c'est courir le risque de la confusion pour décider de leurs intérêts respectifs. Mais bien entendu aucune organisation optimale ne sera trouvée si le corpus de compétences des agents est incomplet, ce qui est en général le cas dans les systèmes. Nous répondons à cela par une **organisation du système en strates où chaque agent est lui-même un système multi-agent** composé d'agents de granularité plus faible comme Minsky l'a proposé dans sa société de l'esprit. Ainsi, la compétence d'un agent de strate élevée change si son organisation interne est modifiée ce qui nous fait revenir aux deux niveaux d'apprentissage précédemment évoqués.

### 2.3 La méthode de base pour l'auto-organisation dans LIAC

La méthode permettant l'auto-organisation que nous présentons ici est basée sur la modification des connaissances sociales entre agents. Elle permet la rediffusion de connaissances à des voisins particuliers pour satisfaire ou fournir un résultat à des agents pertinents. Elle entraîne une auto-organisation de la société basée sur la découverte d'une nouvelle connaissance sociale entre agents.

Notre méthode repose sur le paradigme de la relaxation défini par Lesser (Lesser, 80) dont le fonctionnement est le suivant : lorsqu'un agent a besoin de faire connaître une information, il va la communiquer à tous ses voisins. Si ces derniers sont intéressés par cette information ils la mémorisent et à leur tour la font passer à leurs voisins. Si en revanche ils la jugent inutile ou erronée, ils la détruisent. Notre paradigme, appelé relaxation restreinte, diffère de ce dernier car il consiste à ne diffuser une information qu'aux agents susceptibles de l'utiliser. Dans de telles conditions, un agent qui vient d'inférer un résultat ne le communique qu'aux agents susceptibles de l'utiliser et un agent qui recherche une information pour continuer son raisonnement ne sollicite que les agents susceptibles de la posséder.

Un agent juge de la pertinence d'un autre agent, donc sélectionne les agents qui lui semblent pertinents pour la recherche, à l'aide d'un appariement entre ses connaissances sociales et l'information cherchée ou devant être transmise. Le nombre ainsi obtenu pour chaque agent est ensuite comparé à un seuil que nous nous sommes fixé au préalable. S'il lui est supérieur l'agent est dit pertinent auquel cas l'agent initiateur lui envoie un message appelé message d'appel d'offre.

Pour savoir quand la recherche se termine et surtout éviter les bouclages de l'algorithme nous attribuons une durée de vie au message envoyé par l'agent demandeur qui décroît au fur et à mesure de sa diffusion dans la société. Il est donc réémis qu'un nombre maximal de fois (que l'on s'est fixé). Malgré cette caractéristique, il se peut que l'agent demandeur attende très longtemps la réponse à son problème, en particulier si l'agent recevant l'appel d'offre a un nombre important de messages à considérer avent de le traiter. Pour éviter ce problème, l'agent bloqué limite le temps de recherche c'est-à-dire son temps d'attente. Si au delà du temps de recherche qu'il a autorisé, il n'a pas reçu l'information manquante, il considère cette dernière comme inconnue et il doit alors demander l'aide de l'utilisateur pour pouvoir

poursuivre son raisonnement. Dans le cas d'une communication de résultat il continue son raisonnement sans douter de sa serviabilité. Enfin, nous avons mis au point une méthode locale à l'agent demandeur lui permettant de faire le parallèle entre les deux notions de temps introduites de sorte à ce que durant la période de recherche allouée, le message soit relaxé autant de fois que désiré.
Nous avons jugé du bien fondé de cette méthode à partir d'un simulateur reproduisant le plus fidèlement possible les principes énoncés précédemment. Les résultats sont mis en évidence dans (Camps, 95). Outre les résultats obtenus, cette méthode nous conforte dans l'idée qu'il est possible d'apprendre autrement qu'à l'aide d'offres et de récompenses. De plus, elle présente l'avantage d'être insérable dans tout système multi-agent existant, car elle est totalement indépendante de la sémantique véhiculée dans les messages et donc du domaine d'application.

## 4. RÉFÉRENCES


CAMPS Valérie, GLEIZES Marie-Pierre
Principes et évaluation d'une méthode d'auto-organisation
Journées francophones IAD & SMA, Saint Baldoph Mars 1995 (337 - 348)

GOLDMAN Claudia V., ROSENSCHEIN Jeffrey S.
Emergent Coordination through the Use of Cooperative State-Changing Rule
AAAI 1994 (408-413)

LESSER V.R., ERMAN L.D.
Distributed interpretation: a model and experiment
IEEE Transactions on computers Vol. C-29 n°12 Dec. 1980, (1144-1163)

MATARIC Maja J.
Interaction and Intelligent Behavior
PHD of Philosophy Massachussetts Institute of Technology May 1994

PIQUEMAL-BALUARD C., CAMPS V., GLEIZES M-P., GLIZE P.
Properties of Individual Cooperative Attitude for Collective Learning
Porceedings of the seventh European Workshop on MAAMAW Eindhoven, The Netherlands, January 22-25 1996

SEKARAN Mahendra, SEN Sandip
Multi-Agent Learning In Non Cooperative Domains
ECAI 94

SEKARAN Mahendra, SEN Sandip
To help or not to help
Seventeenth Annual Cognitive Sciences Conference
July 22-25 1995 Pitsburg Pennsylvannia

SEN Sandip, SEKARAN Mahendra, and HALE Jones
Learning to coordinate without sharing information
Proceedings of the twelfth national conference on Artificial Intelligence, August 1994.

SEN Sandip, SEKARAN Mahendra
Using reciprocity to adapt to others

IJCAI 1995

SHAW M. J. and WHINSTON A.B.
Learning And Adaptation In Dai
in Gasser and Huhns 1989

SIAN Sati Singh
Adaptation Based On Cooperative Learning In Mas
in Proceedings of the sevcond workshop on Modelling Autonomous Agents in Multi-Agent World Editors Y. Demazeau & J-P Müller Editions North Holland 1991

SMITH R.G. and DAVIS R.
Frameworks for Cooperation in Distributed Problem Solving IEEE Transactions on Systems, Man, and Cybernetics Vol. SMC-11 n°1 January 1981 (61-70)

WEIß Gerhard
Learning To Coordinate Actions In Multi-Agent Systems
in Proceedings of the International Joint Conference on Artificial Intelligence (311-316) August 1993