
Une formalisation de la régulation par des normes d'une société constituée par des personnes et par des agents artificiels.

Jean Sallantin * — **Didier Ferrier** ** — **Julien Cotret** *** — **Dominique Luzeaux** **** — **Christophe Fagot*****

* *Laboratoire d'Informatique de Robotique et de Micro électronique de Montpellier, 161 rue Ada, 34392 MONTPELLIER CEDEX 5.*
jean.sallantin@lirmm.fr

** *Université Montpellier I, Rue de l'université, 34 000 Montpellier Cedex 5*
Didier.Ferrier@univ-montp1.fr

*** *Normind Cap Omega, Rond point Benjamin Franklin, CS39521 Montpellier cedex 2,*
cotret, fagot@normind.com.
**** *DGA/UMTER*
dominique.luzeaux@polytechnique.org

RÉSUMÉ. Depuis des millénaires, le Droit régit les sociétés humaines. Le Droit est désormais mis en demeure de réglementer les comportements exercés par des personnes agissant en utilisant les moyens de communication modernes ou assistées par des systèmes automatiques agissant au nom de personnes ou d'institutions. Cette situation crée le besoin de se donner des définitions acceptées par ces disciplines pour poser un modèle de régulation s'appliquant à des systèmes complexes composés de personnes et d'agents artificiels. Cet article propose de telles définitions et les illustre par la présentation de conditions pour une régulation d'agents artificiels personnels ayant la capacité juridique de gérer nos cartes de crédits et nos agendas.

ABSTRACT.

MOTS-CLÉS : machine autonome, système multi-agents, droit

KEYWORDS: autonomous machine

Introduction

La notion d'assistant logiciel personnel n'est pas encore là, mais il n'y a pas pour autant d'obstacle technique à concevoir et réaliser les successeurs de nos téléphones mobiles qui pourront se servir de notre carte de crédit, de notre agenda, de nos contacts, de notre signature et qui nous assisteront dans des tâches de la vie courante. Le web nous place déjà sous l'influence de machines, qui nous appartiennent ou appartiennent à d'autres comme Google et qui diffusent des informations. Ces machines ne disposent pas d'estimation du risque lié à leur gestion du secret et de la transparence de leur activité, de la preuve ou de la contingence des informations diffusées, et encore de la simplicité ou de la complexité de leur production.

Quand on parle de protection de l'image de la personne, de sa vie privée, des actes qu'elle peut accomplir, des responsabilités qu'elle peut assumer, on convoque le Droit. Et bien d'autres fonctions du Droit, comme celle d'assurer voire de fluidifier les échanges commerciaux, pourraient tirer parti d'assistants personnels autonomes. Le conseil économique et social européen promeut d'ailleurs une approche proactive du Droit ¹. Cette démarche met en avant des principes de subsidiarité, d'autorégulation, et de corégulation. Elle procède d'une meilleure prise en compte des besoins des utilisateurs du Droit avec en particulier le besoin de comprendre comment le Droit s'applique dans leur cas.

Eu égard à la notion d'assistant personnel autonome, il faut cependant reconnaître que l'idée d'autonomie n'est pas à ce jour définie formellement. De plus, il n'y a pas encore de discours fondateur sur cette notion, mais un ensemble de discours que l'on peut croiser et opposer dont font partie ceux des systèmes multi-agents et ceux du Droit. Par exemple, le Droit confère aux sujets capables la plénitude d'exercice de leurs droits. La capacité juridique étant définie par opposition à l'incapacité qui découle d'une inaptitude physique, intellectuelle, voire juridique (statut de la personne). Le sujet capable est donc reconnu dans ses fonctions de jugement, de décision, d'action, mais que veulent dire ces notions pour un agent logiciel autonome ?

L'objectif principal de cet article est de définir formellement des notions utilisées en Informatique ou en Droit pour décrire le comportement de systèmes pouvant être : un agent, un ensemble d'agents artificiels ou encore une société constituée par des personnes et des agents artificiels.

L'Informatique et le Droit se rejoignent d'une part par l'obligation de définir soigneusement leurs notions et d'autre part d'intervenir partout dans la société. Le défi de cet article est de parvenir à définir des notions utiles aux deux disciplines dans leur approfondissement théorique comme dans leur utilisation pratique. Un tel travail est justifié s'il approfondit la définition de notions utilisées par une discipline scientifique, il est fructueux s'il lève un coin d'ombre pour l'une des disciplines concernées et il

1. Avis du comité économique et social européen sur "l'approche proactive de la pratique du Droit : un pas en avant vers une meilleure réglementation au niveau de l'U.E." INT/415- CESE 1905/2008 ENJL/sg

est scientifique quand des scientifiques trouvent des situations concrètes invalidant le système de définition mis en place ; or les situations concrètes foisonnent dans les pratiques du droit et de l'informatique.

Pour écrire formellement ces définitions, nous avons choisi la théorie des catégories car elle sert d'une part en mathématique à définir des correspondances entre domaines des mathématiques et car d'autre part elle donne une base formelle à des raisonnements portés par des diagrammes. Nous pratiquons ainsi la démarche structurale utilisée en Science Humaine et Sociale dans bien des domaines.

Notre objectif est de définir un modèle de la régulation des systèmes que nous qualifions de systèmes mixtes (constitués d'agents artificiels et d'humains en interactions) nous présentons d'abord les Systèmes Multi Agents (SMA). Nous présentons ces définitions avec un petit problème d'école ayant sens pour les deux disciplines.

Dans la section suivante, nous allons appliquer notre démarche structurale à la définition de la notion de système autonome en partant de notions propres à l'informatique.

Ensuite, nous proposons et discutons la définition formelle du jeu dialectique entre les notions de secret et de transparence.

Enfin nous retournons à la pratique informatique en montrant comment réaliser en pratique une régulation par des normes d'un système mixte.

Les agents artificiels dans les systèmes multi-agents

Le domaine des Systèmes Multi-Agents (SMA) est un sous-domaine à part entière de l'Intelligence Artificielle. C'est avant tout un paradigme de programmation ayant pour but de concevoir, réguler ou simuler au mieux des systèmes complexes en se basant sur notre compréhension des organisations humaines (grandement influencé par les travaux fondateurs (March *et al.*, 1958) .

Définition 1 (Agent artificiel) *Un agent artificiel est, selon une des définitions les plus communément acceptées, un système informatique, situé dans un environnement, capable d'actions flexibles et autonomes dans cet environnement dans le but d'atteindre des objectifs fixés lors de sa conception. (Wooldridge et al., 2000)*

Chaque agent est une boîte noire dont on ne connaît pas les objectifs immédiats. Un des enjeux majeurs devient alors de fournir au système des capacités d'organisation et de régulation pour lui permettre de garder son équilibre et ainsi permettre à la communauté d'agents de travailler à atteindre leurs objectifs. Il est important de noter que les stratégies internes de chaque agent peuvent très bien inclure le recours à des actes tels que le mensonge, le vol ou la triche et que le système doit être capable de réguler (si nécessaire) ce type d'actions. Ces agents possèdent chacun leurs propre représentation du monde, leurs propres objectifs et définissent eux-mêmes des stratégies pour

essayer de les atteindre. C'est par ces interactions entre ces agents dans un environnement ouvert et hétérogène que naît la complexité. Le système formé par l'ensemble des agents est un système en équilibre dynamique (White, 1992)

Notre vision des SMA essaye volontairement de rester au plus proche du paradigme initial en évitant de prendre des hypothèses fortes, telles que la connaissance à priori du comportement interne des agents (on parle alors de systèmes ouverts).

On retrouve ici des conditions similaires aux problématiques humaines, à savoir la régulation d'agents autonomes visant à maintenir l'ensemble du système fonctionnel. Les normes en général et le droit en particulier font partie des outils dont dispose l'homme pour réguler les systèmes complexes que forment nos sociétés.

De nombreux travaux dans la communauté SMA montrent que les normes sont un moyen efficace de réguler de tels systèmes complexes. En utilisant la pression sociale (Carr *et al.*, 2008) ou en définissant des institutions explicitant les normes du système (Esteva *et al.*, 2000).

On voit que l'on peut transposer de façon naturelle les paradigmes du Droit à la régulation au monde des agents artificiels. Nous travaillons par ailleurs sur une plateforme logicielle multi-agents permettant justement de réguler les actions des agents par des institutions explicites en utilisant la logique de raisonnement du droit (Wright, 1951)

Pour illustrer la régulation d'agents artificiels par des normes et l'interaction potentielles de ceux-ci avec les normes du droit, nous allons prendre un exemple. Pour cela, nous nous inspirerons largement des travaux concernant les agents personnel assistant appelés "elfes" dans les travaux cités (Chalupsky, 2002).

Le but de l'expérimentation était de mettre en place des agents assistants dans un groupe de chercheurs. Les elfes étaient chargés de certaines tâches, jours et nuit pendant 7 mois. Chaque chercheur possédait son propre Elfe. Chaque elfe avait accès à plusieurs données concernant son possesseur : sa position géographique (via un appareil GPS) , ses préférences (culinaires et autres), son agenda. Ils avaient accès aussi à une certaines quantités de ressources comme les fax et le service de réservation des salles entre autres. Ils se chargeaient de tâches comme la réservation de salle, l'organisation de réunion en fonction des disponibilités de chacun et de leurs position géographique, mais aussi la commande automatique de repas. Ils transmettaient à leurs possesseurs des informations qui pouvaient leur être utiles, comme le retard d'un vol d'avion par exemple.

Au travers de son elfe, l'utilisateur choisissait de déléguer une partie de ses actions (la commande du déjeuner par exemple).

Cette expérimentation soulève plusieurs questions intéressantes.

L'assistant doit-il toujours demander l'avis de son instructeur, que se passe-t'il quand il ne demande pas son avis ? Que se passe-t'il quand il se trompe ? Comment le

maître peut-il réguler le comportement de l'assistant quand ce dernier a de nombreuses manières d'agir ?

Pour répondre à ces questions nous proposons un scénario extrapolant l'expérimentation de (Chalupsky, 2002) qui mettrait de tels assistants dans un contexte juridique réel.

Nos elfes s'occupent ici de l'achat via internet de produit de base utilisés dans une équipe de recherche en chimie. Ceux-ci gèrent les stocks, les demandes de chacun et s'occupent de trouver sur des sites de ventes en lignes des produits aux meilleurs prix. Ces elfes sont régulés par un ensemble de normes tels que des normes de qualités concernant les produits ou liées au cout de ceux-ci. Ces elfes ont la possibilité de demander leurs avis aux membres du laboratoire si cela s'avère nécessaire. Ils ont aussi de bonnes notions de la réglementation et savent quelles substances sont autorisées sur le sol français et quelles normes de qualités sont obligatoires. Confronté à l'achat d'un produit sur un site étranger l'agent se retrouve confronté à un produit qui enfreint un certain nombre de loi française. Sa capacité à évaluer une situation vis-à-vis de normes lui permet d'identifier clairement la situation comme illégale, il peut donc réagir en fonction de sa propre stratégie.

Quelle sera sa stratégie ? Il peut demander l'avis de ses maîtres qui vont lui demander d'acheter quand même le produit pour des raisons non connues de l'agent (comme une dispense temporaire ou une tolérance). Il peut aussi se retrouver dans une situation complexe où quelle que soit son action, il doit enfreindre un certain nombre de normes pour atteindre ses objectifs. Mais il est qualifié pour commander le produit sans demander l'avis d'un de ses maîtres. Et en cas de problème, on pourra donc facilement retrouver la source de l'infraction (car elle aura été explicite pour l'agent).

Mais allons plus loin, notre elfe devient donc capable d'identifier clairement les situations dans lesquelles il sera obligé de transgresser des réglementations. L'identification de cette situation et de son contexte permet donc à notre elfe de faire deux choses :

- Il peut mettre en place (ou on peut lui donner) des stratégies spécifiques et des comportements alternatifs dans de telles situations et capitaliser sur son expertise dans de telles situations.
- Il peut aussi établir une communication sémantique avec son maître en lui expliquant la situation, son contexte et les risques encourus. Il pourra ensuite au cours de cette conversation le guider pas à pas dans la recherche d'une solution en lui explicitant à chaque jugement les impacts vis-à-vis des institutions et ses conséquences.

Les scénarios présentés illustrent comment un agent artificiel pourrait agir en liant correctement ses systèmes d'action, de perception, de décision et de supervision. De tels systèmes sont techniquement réalisables, certainement utiles mais les risques liés à leur usage par la société ne sont pas encore maîtrisés. La section suivante définit un cadre permettant de qualifier l'autonomie de tels systèmes.

1. Définir des systèmes autonomes

Comme en Droit, où la capacité d'agir librement, d'exercer des droits et de répondre de ses actes est conférée à des personnes physiques ou morales, nous allons définir une notion d'autonomie qui s'applique aussi bien à un agent qu'à un système multi-agents.

La définition que nous proposons est de nature structurale car elle organise une relation interne entre des sous-systèmes constituant un système d'une manière purement algébrique. Comme les exemples qui vont suivre vont le montrer, nous souhaitons qu'il puisse être réalisé par des personnes mais aussi par des personnes assistées par des machines et par des elfes.

Les premières définitions ont pour but de définir la capacité d'autonomie comme venant de celle d'anticiper une impossibilité de décider ou l'inutilité d'un jugement. La dernière définition est plus abstraite car elle va caractériser une régulation de deux systèmes autonomes étant dans une relation dialectique.

Définition 2 (Boucle et Système APDJ) Soit quatre systèmes servant à définir les *ACTIONS*, les *PERCEPTIONS*, les *DECISIONS* et les *JUGEMENTS* d'un Agent ou d'un système multi-agents. Une boucle APDJ est définie par quatre relations qui traduisent respectivement les actions en perceptions, les perceptions en décisions, les décisions en jugements et les jugements en actions. Un agent ou un système multi-agent est un système APDJ quand ses actions sont coordonnées selon une boucle APDJ.

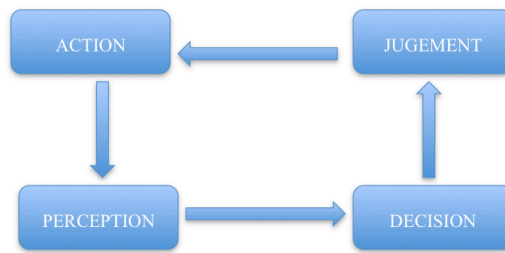


Figure 1. Les actions d'un agent sont contraintes selon le diagramme de gauche. Un agent autonome est capable d'anticiper une décision ou un jugement.

Considérons notre Elfe, il peut réaliser de nombreuses actions, prendre un rendez-vous, faire un achat sur le WEB. Il organise ses actions et il a une perception de ce qu'il fait. Notre agent va mémoriser ses séquences d'actions et enregistrer des indicateurs lui signalant des succès et des échecs de ses actions. Sa capacité de décision va lui permettre de proposer à sa supervision des plans d'actions de manière à atteindre ses objectifs.

Un système APDJ n'a pas la faculté d'anticiper l'effet de sa supervision, même si il est capable d'inventer des comportements adéquats. Pour rendre un système adéquat,

nous allons définir les notions d'anticipation d'une décision ou d'un jugement. Notre système sera adéquat s'il sait anticiper correctement une décision ou un jugement donc agir de telle manière que son action puisse être souvent en conformité avec la décision ou le jugement qui sont censés la justifier ².

Définition 3 (Système autonome) *Un système autonome est un système APDJ complété par deux relations qui traduisent respectivement les décisions en actions et les perceptions en jugements. Un agent ou un système multi-agent est dit autonome quand il est un système autonome.*

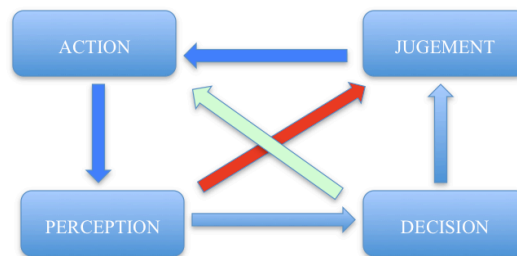


Figure 2. *Deux relations rendent au système la capacité d'anticiper une décision ou un jugement.*

Un système agit de manière autonome quand ses actions provoquent des perceptions qui provoquent des décisions qui construisent des plans d'actions. Un tel système anticipe alors l'effet des jugements sur ces comportements.

Mais quand ce système se trouve en situation d'incapacité de réaliser une action, il le signale à la supervision en charge des jugements de telle manière que cette dernière puisse modifier ses conditions d'action. Dans ce cas, il anticipe le résultat du jugement.

Cette représentation diagrammatique est très compacte mais elle a le mérite de donner naturellement lieu à une formalisation via la théorie des catégories (Luzeaux, 1998) (Luzeaux, 2009).

Une des pierres angulaires de la théorie des catégories est d'adopter de manière systématique un point de vue relationnel : tout peut être défini comme une flèche (i.e. intuitivement une relation ou une transformation, ou plus précisément un morphisme) entre objets, et d'ailleurs les objets eux-mêmes peuvent être définis en utilisant simplement des flèches (en fait, la simple différence conceptuelle a priori entre une flèche et un objet est que les flèches peuvent être composées et définissent donc un « depuis » et un « vers » ; mais un objet peut alors aussi être défini comme une flèche dont le

2. Le paradoxe de Libet : les techniques d'enregistrement de l'activité cérébrale humaine montre que bien souvent l'arrêt d'une action précède de quelques secondes le jugement d'arrêter l'action.

« depuis » et le « vers » sont confondus). On se donne donc des objets et des flèches, i.e. des morphismes, entre eux. ³

L'ensemble des flèches suivantes constitue le graphe de régulation du système. En composant les flèches par transitivité on définit les cycles de régulation du système. Deux nouveaux cycles apparaissent :

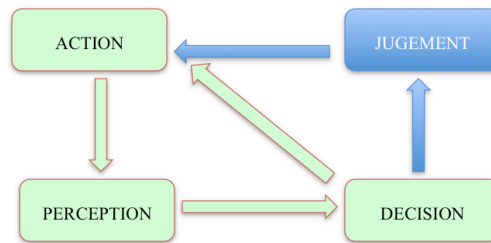


Figure 3. Le cycle donné par le triangle (perception-décision-action) définit la capacité du système. Pour les systèmes informatiques il peut être produit par des calculs d'optimisation.

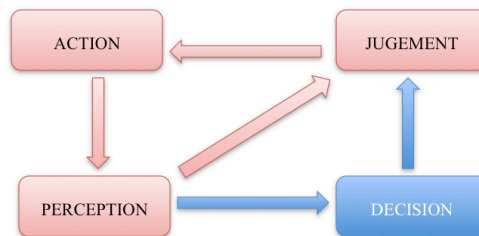


Figure 4. Le cycle donné par le triangle (perception-jugement-action) définit la réaction aux incapacités . Pour les systèmes informatiques, il est réalisé par des systèmes de raisonnement symbolique.

Définition 4 (Système supervisé APDJ) Dans un système supervisé APDJ, les opérations des sous systèmes Actions, Perception et Décision sont réalisées par l'apprenti

3. Si l'on applique récursivement les définitions, on peut considérer une catégorie particulière dont les objets sont eux-mêmes des catégories ; dans ce cas les morphismes entre ces objets composites sont appelés foncteurs. Si l'on explicite l'action d'un foncteur, on s'aperçoit qu'un foncteur entre catégories peut être défini comme associant aux objets de la première catégorie des objets de la seconde, et aux morphismes de la première des morphismes de la seconde, de telle manière qu'un morphisme entre deux objets de la première catégorie est envoyé sur un morphisme entre les images des objets. Ainsi un foncteur est une transformation qui conserve d'une certaine manière la structure de base entre les catégories considérées. Si l'on prend encore plus de recul, on peut considérer la catégorie dont les objets sont des foncteurs ; alors les morphismes entre foncteurs sont les transformations naturelles.

ou par l'instructeur. L'instructeur seul supervise la prise de jugement. Un agent assistant est un système supervisé autonome.

Nous considérons que notre elf est un agent assistant, dans la mesure où il sait anticiper les décisions et les jugements de son instructeur. Les flèches diagonales de la figure 1 signalent cette capacité de prendre un jugement en fonction de la perception en bonne correspondance avec la décision. L'introduction du superviseur rend plus lisible la notion d'acquisition d'une autonomie. L'elf est autonome quand il anticipe correctement les volontés de son superviseur. Un elf autonome ne dérange pas sans raison son instructeur. Quand l'elf se trompe, l'instructeur dispose d'un exemple nouveau à corriger ou à sanctionner pour faire évoluer la décision ou le jugement de son agent.

Nous allons une dernière notion qui est conférée à un système autonome stabilisé

Définition 5 (Stabilité d'un système autonome) *Considérons :*

- une première catégorie incluant comme objets les systèmes « action » et « perception », et comme morphismes les flèches entre ces derniers, c'est-à-dire du point de vue interprétatif, la traduction par les différents acteurs des actions en faits constatés.
- une seconde catégorie, dont les objets sont le « décision » et la « supervision », et les morphismes les flèches entre ces derniers, ce qui traduit à présent le travail d'inférence et d'échange propre à l'analyse rationnelle,
- un couple de foncteurs entre ces deux catégories dont on pose qu'il définit une adjonction (propriété mathématique entre les deux catégories qui traduit le fait qu'il y a une correspondance à un certain niveau d'abstraction entre les opérations faites dans chaque catégorie).

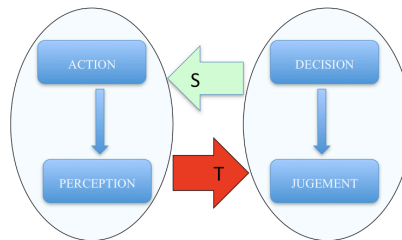


Figure 5. Deux foncteurs S et T mettent en relation la perception de l'action avec sa supervision

La troisième condition est une hypothèse forte mathématique, mais a priori pas tant du point de vue de la modélisation. Cette propriété traduit simplement le fait qu'un acteur sait utiliser, après traduction dans son propre corpus, les décisions et les

supervisions, et que réciproquement les autres savent aussi utiliser dans leur corpus, après nouvelle traduction, sa perception de ce que l'acteur a réussi à faire des savoirs d'autrui dans son propre corpus : ce n'est qu'une formulation de la nécessaire réflexivité (au sens de l'identification partielle possible entre un objet et la représentation de la représentation de cet objet).

L'adjonction donne alors un sens mathématique au cycle de régulation et elle permet de montrer qu'il se referme sur lui-même.

Illustrons la notion de stabilité pour des systèmes supervisés ^{45 6}.

Définition 6 (Qualification et habilitation d'agents) *Un système supervisé est stabilisé quand il est gouverné par les boucles de gestion de ses capacités et incapacités. Un agent assistant intervenant dans un système stabilisé est dit qualifié. Un agent instructeur dans un système stabilisé est dit habilité.*

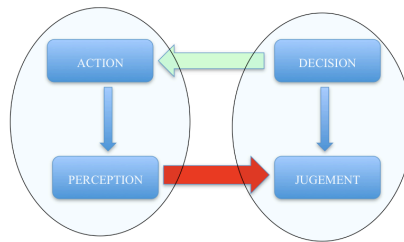


Figure 6. Dans un tel système l'instructeur donne des consignes et sanctionne son apprenti qui l'informe de ce qu'il perçoit de ses actes.

Exemple 1 (la conduite accompagnée) *L'apprenti doit tout d'abord passer avec succès son code. Pour cela, il doit faire un minimum de fautes de jugement face à des photos représentant des situations de conduite. Avant de passer son examen pratique, l'apprenti va subir pendant 3000Km l'instruction de ses parents. Cet apprentissage a comme objectif de leur forger des réflexes associant action perception et décision. L'apprentissage est un succès quand l'apprenti anticipe la décision et le jugement de son instructeur.*

L'apprenti dans la conduite accompagnée est un agent assistant. Le système APDJ est composé de l'apprenti et de son instructeur. Seul l'apprenti agit. Seul l'instructeur supervise. L'apprenti comme l'instructeur perçoit "il y a un stop" décision "tu devrais

4. Cette notion de stabilité évoque les travaux de René Thom car elle est structurelle.
 5. Elle évoque les travaux de Marvin Minsky car elle est réalisable par une société de l'esprit.
 6. Elle évoque le livre de Douglas Ofstadter, Gödel, Escher, Bach. Les Brins d'une Guirlande Eternelle car elle se définit comme étant une structure se refermant sur elle-même.

freiner" l'instructeur supervise le jugement "freine bon sang !". Il est attendu que l'apprenti acquiert la capacité de planifier de manière autonome ses actions. Il est aussi souhaité que l'apprenti prévienne quand il se sent incapable d'agir. S'il le fait à bon escient il est qualifié. Le système est stabilisé quand l'apprenti a intégré les consignes de son instructeur.

Exemple 2 (les juristes : juniors et senior) *Un juriste senior guide l'activité de modélisation de deux juristes juniors. Le senior et les juniors forment le système multi-agents. Les deux juniors forment le système assistant. Les juniors déterminent les clauses, les variables et les contraintes décrivant une classe de contrats. Le senior supervise la décision des juniors et valide un modèle. Pour acquérir leur autonomie les juniors soumettent au senior des contrats formellement corrects mais qu'ils perçoivent comme stupides. Le senior leur fait corriger le modèle.*

Lors des expérimentations que nous avons menées en relation avec le centre de Droit de l'entreprise, ce processus converge en moins de 10 interactions vers un modèle ne produisant plus que de contrats corrects mais réalisés de manière maladroite.

Cette section a défini la notion d'autonomie de manière structurale en caractérisant un système autonome par les relations entretenues entre les sous systèmes qui le composent.

Dans une première partie, les définitions ont précisé les flots d'information entre les composants d'un système autonome de manière à ce que l'on puisse aussi bien définir un agent, un système composé d'un couple d'agents, ou une société d'agents.

La définition de système autonome stabilisé se dégage de la notion d'anticipation pour apporter une forte propriété de stabilité. Nous avons montré comment cette définition s'applique à des systèmes composé d'un instructeur et d'un apprenti. Dans la section suivante, nous allons regarder comment selon la même démarche définir des notions habituelles en Droit mais absentes des systèmes multi-agents.

2. La régulation juridiques de systèmes socio-économique

Nous partons de deux idées : l'idée intuitive que le secret est nécessaire à la menée des affaires humaines et l'idée évidente que la transparence est nécessaire à leur supervision. Le Droit n'oppose pas le secret et la transparence et il constate que ces notions interviennent ensemble pour réguler le fonctionnement des systèmes socio-économiques.

Nous voulons définir cette régulation comme se ramenant au contrôle du double jeu du secret et de la transparence. Nous voulons également que cette régulation se réalise par la définition de normes servant à superviser le comportement des agents ⁷.

7. Un tel système est rationnel au sens de la rationalité économique car chaque agent pour agir prend en considération ce que les autres pensent de son comportement. Et cela est suffisant :

Cette section est consacré à une étude en droit de la notion de stabilité d'un système social. Cette notion de stabilité particularise la définition 5.

Définition 7 (Stabilité d'un système social) .

Posons deux systèmes :

Un premier système est celui de l'activité matérielle d'un ensemble d'agents. Un second système est celui de l'activité sociale de cet ensemble d'agents. Deux opérations adjointes servent à traduire chaque système dans le second. Comme les deux systèmes se traduisent l'un dans l'autre, le système de l'activité du système d'agents tient compte des décisions et des supervisions faites sur son comportement.

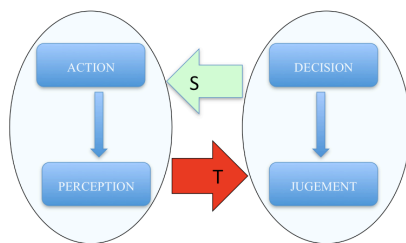


Figure 7. Les opérations de secret et de transparence établissent la cohérence entre les domaines de l'activité matérielle et socio-économique des agents

Dans la fin de cette section, nous allons discuter le caractère universel de cette dialectique entre secret et transparence dans l'activité humaine.

Tout d'abord les notion de secret et de transparence ne s'opposent pas.⁸

Dans un cadre éthique, les comportements privilégiant le secret peuvent être qualifiés de dissimulation et déloyauté. Ce qui s'oppose ainsi au secret est la libération de l'accès à l'information (sa révélation, ou sa diffusion).^{9 10} Nous pouvons de la même manière poser que la transparence s'oppose à l'opacité.

Posons que le secret requiert de l'opacité et que cette opacité vient de ce que le possesseur du secret doit le protéger par certaines actions comme la rétention d'infor-

il n'a pas a prendre en considération des raisonnements en abîme faisant intervenir ce que les autres pensent de ce qu'il pense de ce qu'ils pensent.

8. Colloque international Secret et Justice -*Secret et Justice, - le secret entre éthique et technique ?*, 3-4-5 dec 1998.L'espace juridique,éd. 2000

9. L'exigence légale va souvent au delà de l'impératif de transparence en imposant non seulement l'accès à l'information,mais aussi sa Diffusion, la délivrance de certaines informations DF 111

10. dans certains cas extrême comme le trafic de stupéfiants blanchiment de capitaux la délation auprès des autorités est une obligation

mation ou le mensonge.^{11 12} . S'il y a secret, il y a opacité (au sens de la restriction d'informations ou du masquage de ces dernières). Dans un jeu social, il n'est en principe possible de demander aux autres agents s'ils ont la preuve d'une information que si cette dernière n'est pas confidentielle. Au contraire, les informations non confidentielles peuvent être utilisées par d'autres agents et faire progresser la résolution du problème.

De la même manière, la transparence requiert des actions de diffusion de l'information qui peuvent être la levée du secret, le laisser connaître, le faire savoir.

La relation entre le secret et la transparence est dialectique : il faut une certaine transparence pour assurer le secret et la transparence implique un partage du secret.^{13 14 15 16} La transparence peut assurer le secret. Il peut se protéger en restant méconnu car la transparence fait croire en son absence. La mise en évidence permet de dissimuler.^{17 18}

Cette section discute la définition structurale de notions utiles en Droit pour la régulation de systèmes autonomes et sociaux. Ces notions n'ont pas leur correspondant en SMA. Leur prise en compte légitime une contrainte forte posée sur les système SMA qui est celle d'une adjonction entre une définition de l'activité matérielle du système et une définition de l'activité sociale du système.

Si on observe la figure 2, on voit la place très spécifique du système de supervision qui est celui où aboutissent les flèches traduisant les décisions en supervisions et celles qui traduisent les perceptions en supervisions via le foncteur de transparence.

Dans la section suivante, nous allons présenter un outil logiciel industriel développé pour superviser par de normes le fonctionnement d'une société d'agents.

3. Définir de manière opérationnelle une supervision par normes

Selon (Kelsen, 1996), le Droit positif se définit comme la science des Normes. De manière universelle, les systèmes de normes habilent des institutions à commander

11. le secret prospère dans l'ombre là où n'est pas imposée la transparence DF 112

12. le droit de retenir l'information vient de ce que l'information est source de pouvoir et d'autre part qu'elle a un coût et un prix... le détenteur a le droit et le devoir de ne pas le divulguer ; à la limite il a le droit de mentir, le mensonge étant une arme économique efficace. . . . il faut que la transparence n'excède pas ce qui est nécessaire au bon fonctionnement du marché D.Schmidt transparence et marchés financiers et boursiers p169 DF 112

13. obligation de discrétion des membre du comité d'entreprise C.trav. L 432-7

14. le marché doit être transparent ce qui est antinomique à l'idée de secret. Cependant, il est interdit d'échanger des informations sur les prix pratiqués et leurs évolutions.

15. le délit de bavardage oblige ceux qui sont au courant d'information privilégiées à les taire

16. Tout le monde doit avoir accès aux mêmes informations qu'elles soient justes ou fausses.

17. désigné par D.Ferrier comme l'effet bikini : ce qui est révélé est suggestif, ce qui reste caché est vital

18. Le Secret et la Transparence s'associent à la bonne conduite des affaires sans pour autant y introduire une progression morale

ou gouverner d'autres institutions. Il y a en Droit, pour Kelsen, une pyramide de pouvoirs institutionnels qui se trouve en correspondance avec une pyramide de normes. Les notions de qualification et d'habilitation de la section précédente relèvent du domaine juridique car elles font intervenir une institution supérieure qui va les attribuer aux institutions inférieures.

Cette vision est contestée car, dans le monde moderne, selon (Ost *et al.*, 2002), on est passé d'une pyramide des pouvoirs à un réseau des pouvoirs. De plus, comme l'ont montrées les sections précédentes, pour conférer au système son autonomie, interviennent également des notions d'auto-régulation et de co-régulation.

La supervision par des normes est ainsi un composant essentiel du système car, comme le montre la figure 7, le système de jugement sert à établir l'accord entre le comportement physique d'un système et sa supervision par des décisions conduisant à des jugements.

Le système INTEGRE® est un environnement logiciel fournissant un environnement de régulation du comportement d'un agent ou d'une communauté d'agents, par inférence et contraposées, acceptant les comportements contradictoires et alternatifs. Il permet aussi d'accompagner la prise de jugement d'un agent en lui donnant de façon explicite un retour sur les conséquences de ses actions. Ces normes constituent une expertise stratégique utilisable par les agents dans leurs prises de jugements. Les comportements, normes et expertises sont exprimées sous la forme de préconisations et contre-indications, à l'aide de concepts définis dans des ontologies. L'objectif est également de fournir un système de construction interactive de système de normes, en définissant l'ensemble des concepts et des comportements nécessaires au bon fonctionnement du logiciel qui embarquera les connaissances modélisées.

3.1. Le vocabulaire

L'ontologie d'un domaine permet de hiérarchiser et composer l'ensemble du vocabulaire permettant de décrire les objets, actions et acteurs du domaine considéré. Le vocabulaire est divisé en 2 catégories : les concepts qui représentent les entités ayant une essence propre dans le domaine modélisé (par exemple, une « tâche », un « acteur », une « phase », une « localisation »), ou les informations permettant de décrire ces entités (par exemple, un « poids », une « durée », un « paragraphe ») ; les relations qui représentent les liens sémantiques que l'on peut établir entre différents concepts (par exemple, « est décrit par », « a subit », « est localisé »).

Le vocabulaire s'organise de manière hiérarchique. Au sein d'une hiérarchie, un terme (ou un mot) est le fils d'un terme « père » s'il représente une vision plus spécifique de ce père. Le vocabulaire étant scindé en deux ensembles (concepts et relations), ces termes sont donc répartis dans deux hiérarchies. Chacune de ces hiérarchies possède comme racine un terme générique dit « Universel ». Dans ces hiérarchies, il ne faut pas confondre la spécialisation d'un terme et la composition d'un terme. On peut, par exemple, dire que le concept du domaine de la planification « Phase » est une

spécialisation du concept « Tâche », mais on ne peut en faire de même entre « Phase » et « Ressource », car une ressource est l'une des informations permettant de décrire une phase.

Un domaine peut être modélisé en faisant appel à plusieurs ontologies. Il est ainsi possible de créer autant d'ontologies que de type de domaines connexes auxquels sont confrontés les agents et les personnes (juridique, industriel, humanitaire ...).

3.2. Les normes, règles métier, de bonne pratique, contraintes, workflows, savoir-faire ...

Les préconisations et contre-indications d'un domaine permettent de définir les normes, règles métier, contraintes, règles de bonne pratique, de fonctionnement, workflows, savoir-faire ... qui existent entre les termes de ce domaine, ou qui portent sur chaque objet du domaine. Ces préconisations et contre-indications permettent à la fois :

- de déduire de nouvelles informations concernant un objet en fonction des informations existantes ou de la description d'un contexte,
- de déconseiller la présence conjointe ou simultanées de certaines informations concernant un objet.

D'un point de vue descriptif, ces préconisations et contre-indications sont constituées de conditions portant sur les attributs des objets du domaine (par exemple, la date de début d'une tâche est postérieure au 1er Janvier 2008). Elles sont structurées par des conjonctions et/ou des disjonctions de conditions. Lors d'une phase d'exploitation (par opposition à la modélisation) d'un modèle, chacune des conditions peut être validée (vrai), invalidée (faux) ou indéterminée.

Les préconisations et contre-indications produisent un certain nombre d'informations de cohérence sous la forme de modalités accompagnant les déductions réalisées. Les 4 jugements utilisées dans ce cadre sont :

- le conseil,
- la contre-indication,
- la contradiction,
- l'incomplétude.

Par exemple, si une contrainte de type « fin-début » a été définie entre deux phases d'une opération financière, dès que la date de fin de la première phase est connue, la date de début de la deuxième phase peut alors être conseillée. Si le responsable de l'opération décide cependant d'avancer la seconde phase avant la fin de la première, la contrainte n'est plus respectée ; il y a donc contradiction entre la théorie (i.e. la contrainte) et la décision de l'expert. Il est à noter que cette contradiction n'empêche nullement la réalisation de la deuxième phase, ni la poursuite du processus d'aide à la décision pour l'opération concernée.

Il est à noter que, fréquemment, il s'avère que lors de la constitution d'une règle l'utilisateur constate qu'il manque du vocabulaire dans ses ontologies. C'est d'ailleurs la principale raison d'ajouter du vocabulaire : permettre de décrire de nouvelles règles ou compléter des règles existantes. Dans cette situation, il est important que l'utilisateur ne sorte pas de son contexte de travail. En d'autres termes, il ne faut pas fermer brutalement le contexte de description de la règle étudiée pour aller compléter le vocabulaire, puis revenir dans le contexte de description de la règle. Ces changements de contexte de travail impliquent des changements de posture cognitive qui nuisent à la qualité et à la pertinence des règles produites. L'utilisateur doit au contraire simplement compléter son vocabulaire en direct depuis le contexte de production des règles. A l'issue de son travail sur la règle, il retrouvera ces améliorations du vocabulaire qu'il pourra alors structurer ou décrire plus en détails, mais cette fois dans un contexte de structuration du vocabulaire.

3.3. Les domaines et institutions

Un domaine représente un usage. Il est défini par un ensemble d'ontologies et un ensemble d'institutions qui sont des ensembles cohérents de préconisations et contre-indications. Un des intérêts de la notion de domaine réside dans le fait qu'un domaine est consistant, c'est-à-dire qu'il contient toutes les informations nécessaires à son fonctionnement. Les institutions quant à elle permettent généralement de représenter différents points de vues réglementaires ou comportementaux au regard du domaine décrit.

3.4. Les environnements

Enfin, un environnement d'exécution sert à définir un cadre d'exécution. C'est dans un environnement que sont créés des objets et que s'appliquent les préconisations et contre-indications. Un environnement d'exécution est donc nécessaire que l'on soit dans une phase de validation ou dans une phase d'utilisation des connaissances modélisées. Car finalement, que l'utilisateur soit en phase de validation ou en phase d'usage réel des connaissances (directement ou via un agent le représentant au sein d'un système), seule sa posture d'esprit et son délai de prise de décision changent, pas ses demandes d'aide à la décision par les connaissances modélisées. Un environnement INTEGRE® est donc la perception qu'à un agent d'un monde et des des actions réalisées dans ce monde. Cet environnement INTEGRE® fourni à l'agent une aide au jugement et à la décision basée sur les systèmes normatifs que l'agent est supposé suivre et respecté. Mais les agents comme les hommes peuvent avoir la liberté et l'autonomie suffisante pour déroger à la règle, et transgresser les normes. Les conséquences de ces transgressions, volontaires ou contraintes, sont alors la mise en oeuvre de nouvelles normes.

Conclusion

Le Droit depuis longtemps et l'Informatique désormais servent à gouverner le jeu social. Il est important de réfléchir à leur synergie et de prendre en considération leur nécessité de disposer de pratiques efficaces.

Pour cela nous avons défini formellement des notions utilisées en Informatique et en Droit pour décrire le comportement de systèmes pouvant être : un agent autonome, un ensemble d'agents autonomes ou encore une société constituée par des personnes et des agents autonomes.

Ce travail a été facilité par le fait que ces disciplines se rejoignent sur l'obligation qu'elles ont à définir soigneusement leurs notions dont l'analyse montre qu'elles se complètent harmonieusement.

L'ambition de cet article est de définir pour ces deux disciplines les notions de "systèmes autonomes" et de "stabilité de système autonome" qui prennent sens en participant à leurs approfondissements théoriques.

Pour écrire formellement ces définitions, nous avons choisi la théorie des catégories qui rejoint la démarche structurale utilisée dans bien des domaines des Sciences Humaines et Sociales et donne une base formelle à ses raisonnements. En effet cette théorie sert en mathématiques à établir des ponts entre ses domaines et elle donne une base formelle à des raisonnements portés par des représentations diagrammatiques.

Le contrôle juridique est ici proactif et a posteriori car il ne s'agit pas de gripper le système en multipliant les réglementations, mais au contraire d'en libérer l'autonomie en lui permettant d'anticiper sur les dysfonctionnements et sur l'incohérence d'ensemble de normes.

La définition de la stabilité de systèmes autonomes composés d'assistants logiciels profite ici de la prise en compte du jeu dialectique entre secret et transparence qui est un régulateur du jeu social et en particulier de celui des affaires. Il illustre une convergence certaine entre les démarches informatique et juridique de la régulation des systèmes complexes.

Désormais rien de théorique et de pratique empêchent de réaliser une nouvelle génération de téléphones mobiles nous assistant dans nos affaires.

REMERCIEMENT Ce travail est financé par l' ANR intermed

4. Bibliographie

Carr H., Pitt J., Artikis A., « Peer Pressure as a Driver of Adaptation in Agent Societies », *ESAW 2008 St. Etienne France*, Engineering Societies in the agents World ESAW, Springer Verlag, p. 191-207, 2008.

Chalupsky H., « Electric Elves : Agent Technology for Supporting Human Organizations », *AI Magazine*, 2002.

- Esteva M., Rodríguez-Aguilar J., Arcos J., Sierra C., Garcia P., « Institutionalizing Open Multi-Agent Systems », , *Institutionalizing Open Multi-Agent Systems, icmas, pp.0381, Fourth International Conference on Multi-Agent Systems (ICMAS'00, IEEE Computer Society, 2000.*
- Gruber T. R., « A translation approach to portable ontology specifications », *Knowledge acquisition*, vol. 5, p. 199-220, 1993.
- Kelsen H., *Théories générales des normes*, PUF Presses Universitaires de France, 1996.
- Luzeaux D., « Category theory applied to digital systems theory. », *In 4th World Multiconference on Systemics, Cybernetics, and Informatics*, , Orlando, FL, USA,, 1998.
- Luzeaux D., « Vers une nouvelle théorie abstraite des systèmes en vue de leur ingénierie. », *Journées Internationales de l'Association Française pour l'ingénierie système*, Paris France, 2009.
- March J., Simon H., *Organisations*, John Wiley New York, 1958.
- Ost F., van de Kerchove M., *De la pyramide au réseau : pour une théorie dialectique du Droit*, faculté saint louis de Bruxelles, Bruxelles, 2002.
- Sowa J., *Conceptual Structures : Information Processing in Mind and Machine*, The Systems Programming Series, Addison-Wesley, 1984.
- White H. C., *Identity and control : a structural theory of social action*, Princeton University Press, 1992.
- Wooldridge M., Jennings N. R., Kinny D., « The Gaia Methodology For Agent-Oriented Analysis And Design », *Journal of Autonomous Agents and Multi-Agent Systems*, vol. 3, p. 285-287, 2000.
- Wright G. H. V., « Deontic logic », *Mind*, vol. 60, p. 1-15, 1951.