Aspects cognitifs des dialogues entre agents artificiels : l'approche par la cohérence cognitive

Philippe Pasquier

PHILIPPE PASQUIER

ASPECTS COGNITIFS DES DIALOGUES ENTRE AGENTS ARTIFICIELS

L'approche par la cohérence cognitive

Thèse présentée
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de doctorat en Informatique
pour l'obtention du grade de Philosophiæ Doctor, (Ph.D.).

FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL
QUÉBEC

Août 2005

Résumé court

Les cadres interactionnels actuels pour les communications entre agents (protocoles, stratégies de conversation, jeux de dialogue, . . .) garantissent la cohérence structurelle des conversations tenues. Pourtant, ce n'est pas tant l'habilité des agents à structurer leurs conversations qui nous intéresse que leurs aptitudes à tenir des conversations utiles quant à leurs objectifs individuels et collectifs. Pour traiter cette problématique, nous avons défini et implanté un modèle complet de la communication entre agents qui couvre les quatre dimensions classiques de la communication : syntaxe, structure, sémantique et pragmatique.

Au niveau syntaxique, nous proposons le langage DIAGAL[DIAlogue Game based Agent communication Language] qui se présente comme un ensemble de jeux de dialogue qui permettent la manipulation conjointe d'engagements sociaux. Du point de vue de la structuration des dialogues, les jeux de dialogue que nous proposons offrent une alternative à la rigidité des protocoles tout en capturant les aspects conventionnels de la communication, absents des approches considérant des actes de langages isolés. Dans notre approche, le niveau sémantique de la communication repose, quant à lui, sur les engagements sociaux qui capturent les interdépendances contractées par les agents lors des communications.

Dans ce contexte, notre contribution principale concerne les aspects cognitifs de la pragmatique. À cet effet, nous proposons une théorie cognitive de l'utilisation de ce cadre interactionnel basée sur la notion de cohérence cognitive et fondée sur des résultats non encore formalisés de sciences cognitives. Issue d'une unification de la théorie de la dissonance cognitive (une des théories majeures de psychologie cognitive) avec la théorie de cohérence cognitive (développée en philosophie de l'esprit), notre approche est formulée en termes d'éléments et de contraintes, notions familières en informatique. La théorie motivationnelle résultante est ensuite étendue afin de traiter la communication entre agents cognitifs. Sous les hypothèses de notre théorie, nous définissons alors une métrique de l'utilité des conversations entre agents.

Nous montrons comment cette théorie permet de résoudre en pratique de nombreux problèmes fondamentaux des aspects cognitifs de la pragmatique des communications entre agents. En particulier, nous proposons une première application de notre théorie pour l'utilisation automatique par des agents de type BDI [Beliefs, Desires and Intentions] des jeux de dialogue du langage DIAGAL. Ce faisant, nous introduisons un certain nombre d'outils techniques pour l'automatisation des communications entre agents tout en précisant quels sont nos apports théoriques pour les SMAs et plus généralement pour les sciences cognitives.

Résumé étendu

Les cadres interactionnels actuels pour les communications entre agents (protocoles, stratégies de conversation, jeux de dialogue, . . .) garantissent la cohérence structurelle des conversations tenues. Pourtant, ce n'est pas tant l'habilité des agents à structurer leurs conversations qui nous intéresse que leurs aptitudes à tenir des conversations utiles quant à leurs objectifs individuels et collectifs. Pour traiter cette problématique, nous avons défini et implanté un modèle complet de la communication agent qui couvre les quatre dimensions classiques de la communication : syntaxe, structure, sémantique et pragmatique.

Au niveau syntaxique, nous proposons le langage DIAGAL[DIAlogue Game based Agent communication Language] qui se présente comme un ensemble de jeux de dialogue qui permettent la manipulation conjointe d'engagements sociaux. Du point de vue de la structuration des dialogues, les jeux de dialogue que nous proposons offrent une alternative à la rigidité des protocoles tout en capturant les aspects conventionnels de la communication, absents des approches considérant des actes de langages isolés. Le niveau sémantique de la communication repose, quant à lui, sur les engagements sociaux qui capturent les interdépendances entre agents et entre les agents et leur environnement contractées lors des communications. Le langage DIAGAL a la bonne propriété d'être complet et adéquat par rapport à notre modèle de l'engagement social.

Dans ce contexte, notre contribution principale concerne les aspects cognitifs de la pragmatique. À cet effet, nous proposons une théorie cognitive de l'utilisation de ce cadre interactionnel basée sur la notion de cohérence cognitive et fondée sur des résultats non encore formalisés de sciences cognitives. Issue d'une unification de la théorie de la dissonance cognitive (une des théories majeures de psychologie cognitive) avec la théorie de cohérence cognitive (développée en philosophie de l'esprit), notre approche est formulée en termes d'éléments et de contraintes, notions familières en informatique. Basée sur le principe motivationnel de l'homéostasie de la cohérence cognitive, notre contribution s'articule autour d'une définition formelle du concept de cohérence cognitive et permet de répondre (même partiellement) à des questions trop rarement adressées dans leur généralité :

1. *Pourquoi un agent dialogue-t-il ?* Un agent dialogue pour essayer de réduire une incohérence cognitive qu'il ne parvient pas / veut pas / peut pas / réduire seul. On distingue les incohérences internes des incohérences externes, selon que les éléments impliqués dans l'incohérence sont exclusivement internes à un agent ou distribués entre plusieurs.

2. *Quand un agent doit-il prendre l'initiative d'un dialogue, à quel sujet et envers quels autres agents ?* Un agent s'engage dans un dialogue du fait d'une incohérence cognitive

qu'il ne peut résoudre seul. Il s'agit soit d'une incohérence interne qu'il ne parvient pas à résoudre seul ou une incohérence externe qui, de par sa nature distribuée, ne peut être réduite par l'action individuelle. Le sujet du dialogue concerne l'incohérence elle-même et les différents éléments de cognition qui y sont associés. Le(s) partenaire de dialogue est soit un agent jugé compétent et coopératif dans le cas d'une incohérence interne ou les autres agents impliqués dans le cas d'une incohérence externe.

3. *Par quel type de dialogue ou quelle unité de dialogue ?* Nous avons lié les différents types d'incohérence à la typologie des dialogues de Walton et Krabbe [1995], mais nous soutenons que notre approche est suffisamment générique pour pouvoir s'appliquer à n'importe quel cadre interactionnel basé sur les engagements sociaux. Dans cette thèse, nous avons utilisé les jeux de dialogues du langage DIAGAL comme cadre interactionnel.

4. *Comment définir et utiliser la notion d'utilité des dialogues ?* En suivant le principe de cohérence qui stipule que les agents préfèrent la cohérence à l'incohérence et la définition classique en micro-économie de la notion, la fonction d'utilité individuelle subjective d'un dialogue se définit directement comme l'éventuel gain de cohérence cognitive dû au dialogue moins le coût des différentes actions dialogiques. De plus, nous avons défini la notion d'utilité espérée en cas de succès du dialogue (les dialogues n'étant que des tentatives de résolution d'incohérence qui, par la même, peuvent échouer). Cette notion d'utilité espérée permet aux agents de choisir dynamiquement parmi les différentes unités de dialogue disponibles, leur permettant ainsi de structurer automatiquement leurs dialogues.

5. *Quand arrêter un dialogue ou le cas échéant comment le poursuivre ?* Un dialogue se termine lorsque l'incohérence qui est son objet est réduite. Sinon, il continue par structuration d'autres unités de dialogue suivant la chaîne de réduction de l'incohérence ou des incohérences apparues lors des tentatives antérieures (parfois, la réduction d'une incohérence en fait apparaître d'autres).

6. *Quels sont les impacts du dialogue sur les états mentaux des agents, aspects privés à ceux-ci ?* Dans le cas où le dialogue (vu comme tentative de réduction d'incohérence) échoue, ou lorsqu'un agent doit satisfaire un certain nombre d'engagements incohérents avec ses cognitions privées, il peut continuer à tenter de changer le monde extérieur pour rétablir la cohérence ou bien modifier ses propres cognitions pour restaurer la cohérence. Ce dernier cas est appelé changement d'attitude en référence au phénomène psychologique qu'il formalise.

7. *Quelle intensité donner aux forces illocutoires des actes de langages/dialogue utilisés ?* Le degré d'intensité des actes de discours/langage générés est influencé[1] par la

[1] Ce facteur n'est pas unique : rôles sociaux, hiérarchie sociale (on ne donne généralement pas d'ordre à un supérieur hiérarchique, ...) ont aussi leur influence.

magnitude de l'incohérence attaquée. Plus l'incohérence est importante, plus le degré d'intensité sera élevé. Ainsi, dans le cas d'un directif (qui vise à engager l'interlocuteur) l'agent pourra produire un conseil ou une suggestion si l'incohérence est faible, une requête ou une demande si elle est plus intense et un ordre ou une supplication si l'incohérence est importante.

8. *Quels sont les impacts du dialogue sur l'humeur des agents ?* Selon le principe de cohérence, la cohérence cognitive est une source de satisfaction tandis que l'incohérence est une source de mécontentement. On déduit simplement la réaction émotionnelle de l'agent de la dynamique de sa cohérence cognitive : la joie apparaît lors d'une réduction réussie, la tristesse lors d'une tentative de réduction échouée, la peur d'une future tentative de réduction importante, le stress et l'anxiété de la persistance de l'incohérence.

9. *Quels sont les impacts des dialogues sur les relations sociales entre agents ?* Dans notre formalisme, puisque les agents peuvent calculer et mémoriser l'utilité des dialogues, ils peuvent ajuster leurs relations sociales en fonction de l'utilité des dialogues passés. Par exemple, ils peuvent renforcer leurs relations envers les agents avec lesquels les dialogues ont été efficaces et utiles, ...

La théorie de la cohérence cognitive que nous avançons répond donc en théorie et en pratique à ces questions de manière unifiée. Concrètement, nous proposons une première application de notre théorie pour l'utilisation automatique par des agents de type BDI [Beliefs, Desires and Intentions] des jeux de dialogue du langage DIAGAL. Ce faisant, nous introduisons un certain nombre d'outils techniques pour l'automatisation des communications entre agents tout en précisant quels sont nos apports théoriques pour les SMAs et plus généralement pour les sciences cognitives.

Short abstract

Different approaches have investigated the syntax and semantics of agent communication languages. However, these approaches have not indicated how agents should dynamically use communications. Instead of filling this pragmatics gap, most approaches have mainly focused on the 'structure' of dialogues even though developers are more interested in agents' capabilities of having 'useful' automated conversations with respect to their goals rather than in their abilities to structure dialogues. We addressed this problem that requires re-inquiring the four main dimensions of (agent) communication : syntax, structure, semantics and pragmatics (the theory of the use of language).

At the syntactic level, we have developed an agent communication language called DIAGAL [DIAlogue Game based Agent communication Language] which consists of a set of dialogue games that allows for the grounded manipulation of social commitments. At the structural level, DIAGAL dialogue games offer a good compromise between the lack of flexibility of protocols while taking into account the conventional aspects of dialogue which were missing in isolated speech acts approaches. At the semantics level, we provide a social commitments model that captures the inter-dependencies contracted by the agents toward each other during dialogues.

In that context, our main contribution concerns the cognitive aspects of pragmatics. To this end, we develop in this thesis a motivational theory for the use of such a conventional and social agent communication framework. Our approach is based on cognitive science results that have not been formalized yet. Our theory consists of a formalisation and a unification of the cognitive dissonance theory (one of the major theories of cognitive psychology) with the coherence theory issuing from philosophy of mind. Our approach is formulated in terms of elements and constraints which allow making it computational. This theory allows formally defining and exploiting the notion of utility of dialogues.

We show in this thesis how this approach allows solving many theoretical and practical problems in agent communication. As a validation of this approach, we extend classical BDI [Beliefs, Desires and Intentions] agents to allow them to automatically use DIAGAL dialogue games. The resulting framework provides the necessary theoretical and practical elements for implementing our theory. In doing so, it brings in a general scheme for automatizing agents' communicational behaviour.

Avant-propos

Quoi qu'on en dise, une thèse de doctorat est un travail collectif. Il y a les encadrants, au premier rand desquels on trouve le directeur de recherche. Dans mon cas, Brahim Chaibdraa m'a fait bénéficier de sa confiance et de son expertise à un point qui dépasse le simple professionnalisme. Il y a également les collaborateurs proches : l'équipe du DAMAS [Dialogue, Apprentissage et systèmes Multi-agentS][2], en particulier ceux qui ont travaillé avec moi sur les communications agents : David Bourget, Benjamin Rivalland, Nicolas Andrillon, Marc-André Labrie, Nicolas Maudet, Mathieu Bergeron et Roberto Flores ainsi que Jamal Bentahar et Bernard Moulin du laboratoire d'informatique cognitive. C'est finalement tout le personnel du département d'informatique et de génie logiciel de l'université Laval que je me doit de remercier.

Plus largement, cette thèse n'aurait pas eu la forme qui est la sienne sans les nombreuses discussions avec les chercheurs dont les domaines d'intérêts sont connexes aux miens. Aussi, sans pouvoir tous les nommer, il convient de remercier :

- les chercheurs de la communauté de recherche sur l'intelligence artificielle et les systèmes multiagents, en particulier : Munindar Singh, Frank Dignum, Leon Van der Torre, Jean-Paul Sansonnet, Amal El Fallah Seghrouchni, Pablo Noriega, Jean-Pierre Müller, Pierre Glize, Marie-pierre Gleizes, Catherine Tessier, Laurent Chaudron, Humbert Fiorino, François Legras, David Sadek, Jean-Pierre Briot, Olivier Boissier, Cosmin Carabelea, Marc-Philippe Huget, Jean-Louis Dessalles, Simon Parsons, Peter McBurney, ...
- les chercheurs en linguistique ou en traitement automatique du langage naturel : Christian Brassac, Jean Caelen, Luc Lamontagne, Chantal Enguehard, Thierry Lemeunier, Guillaume Chicoisne, Guy Lapalme, ...
- les chercheurs des différentes disciplines de sciences cognitives et de sciences sociales avec lesquels nous avons eu à interagir de manière continuelle : Frédéric Dehais

[2] http ://www.damas.ift.ulaval.ca/

(sciences cognitives), Guy Paquette (psychologie sociale), Daniel Vanderveken (philosophie du langage), Paul Thagard et Renée Bilodeau (philosophie de l'esprit), Paul Shultz et Mark Lepper (psychologie cognitive), Juliette Rouchier (sociologie et économie), ...

Lorsqu'on ne sait pas où on va, on peut à tout le moins se souvenir d'où on vient et c'est l'ensemble des professeurs qui m'ont formé qui devraient être remerciés ici, en particulier ceux qui m'ont enseigné l'informatique, l'intelligence artificielle et les sciences cognitives : Guy Mineau, Michel et Claudette Cayrol, Didier Dubois, Henri Prade, Jérome Lang, Robert Demolombe, Laurence Cholvy, Marie-Pierre Gleize, Luis Farinas del Cerro, Fabrice Évrard, Andreas Herzig, Laure Vieu, Frédéric Benhamou, Mohamed Quafafou, Philippe Lamarre, Sylvie Cazalens, Elie Milgrom, André Thaize, Pierre Wodon, Philippe Delsarte, Michel Sintzoff, Marc Lobelle, Peter Van Roy, Yves Willems, ...

Finalement, il y a les proches, la famille, les amis que nous ne mentionnerons pas ici à l'exception de Véronique sans qui le Canada m'aurait probablement semblé trop froid.

Le conflit est père de toute chose.

Fragment 53, Héraclite. (v. 470 av. J.-C.)

Table des matières

Liste des tableaux

Table des figures

Introduction

En 1997, Deep Blue battait le champion du monde d'échec en titre Garry Kasparov. Fin 1999, IBM annonçait pour 2005 Blue Gene, un système 1000 fois plus puissant que Deep Blue. John Koza [Koza et al., 1999], père des techniques de programmation génétique développe des systèmes dont les inventions sont brevetées qui annoncent l'aire de l'« human competitive artificial intelligence »[3]. Depuis que Ray Kurzweil, dans son ouvrage « The age of spiritual Machines », sous-titré « when computers exceed human intelligence » [Kurzweil, 1999], encensé par le Walt Street Journal, annonce des ordinateurs plus intelligents que l'humain pour 2020[4], l'intelligence artificielle a un impact considérable (et grandissant exponentiellement[5]) sur nos vies quotidiennes (culture incluse).

Quoi qu'on en pense, l'intelligence artificielle est l'un des paradigmes de la modernité. En effet, si la modernité désigne cette volonté occidentale de comprendre et de contrôler la nature. Cette réalité, partiellement consommée, s'accompagne du fait que dans la nature on trouve le vivant, et dans le vivant l'intelligence. La modernité, c'est donc entre autres de comprendre, au point de savoir la reproduire, l'intelligence naturelle. L'approche moderne de l'intelligence naturelle trouve son mythe dans l'intelligence artificielle .

Synergie de l'informatique et des sciences cognitives, l'intelligence artificielle est une science interface qui est à l'écoute des autres disciplines concernées par la cognition humaine ou animale que sont la linguistique, la psychologie, la philosophie de l'esprit et du langage, les neurosciences, la sociologie et l'économie. L'intelligence artificielle tente d'en formaliser les propos pour pouvoir simuler informatiquement les théories résultantes. Les fins de ce pro-

[3]Expression anglophone utilisée par Koza et que l'on pourrait traduire par « l'intelligence artificielle qui rivalise l'intelligence humaine ».

[4]Le titre de l'ouvrage de Kurzweil pourrait être traduit « L'age des machines spriritueles » et son sous titre « Lorsque les ordinateur deviennent plus intelligents que le humains ». Les idées de l'auteur, qui se base sur une extrapolation de l'évolution des travaux de recherche en cours et passés, sont également vulgarisées sur son site : http ://www.kurzweilai.net/.

[5]Dès 1965, Gordon Moore, alors président d'Intel, prévoyait la croissance exponentielle des capacités de calcul et des usages qui en seraient faits. La loi dite de Moore n'a pas été démentie depuis, difficile de croire qu'elle le sera dans un futur proche.

cessus de formalisation et de simulation sont multiples, puisqu'il s'agit à la fois (1) de valider (et plus souvent d'invalider) les théories simulées ainsi que de les enrichir des différentes réflexions soulevées par le modèle formel et (2) de faire progresser le front de connaissances de l'informatique autant d'un point de vue pratique que théorique. Dans le premier cas, l'apport ne se résume pas à la simulation d'une théorie de science cognitive puisque les concepts formalisés sont étendus, hybridés voir simplement ignorés au profit d'autres types de savoir. Il en résulte un retour fécond sur les sciences cognitives et de nombreux liens transdisciplinaires sont établis à ce carrefour de l'informatique, de la systémique (au sens large) et des sciences cognitives. Le second cas, quant à lui, offre une lecture de l'intelligence artificielle comme paradigme d'ingénierie du logiciel et s'il s'agit de distinguer entre science fondamentale et technologie applicative, l'intelligence artificielle se développe de front dans ces deux champs.

En cinquante ans d'existence, des avancées considérables ont été réalisées par l'intelligence artificielle sur la modélisation informatique du raisonnement et de la rationalité limitée, de l'apprentissage et de l'adaptation, de la vision et de l'audition et du traitement du langage naturel. Les techniques et algorithmes issus de ces recherches, qu'ils soient d'inspiration naturelle ou pas, sont appliqués dans tous les domaines touchés par l'informatisation. De ces techniques résultent des systèmes proactifs et autonomes, c'est-à-dire posant sur leur environnement des actions qui ne soient pas de simples réactions aux variations de celui-ci. L'intelligence artificielle produit des systèmes complexes, c'est à dire dont l'issue du processus décisionnel ne peut être complètement prédite même en connaissance du modèle comportemental utilisé[6], conçus pour exhiber des comportements qui seraient qualifiés d'intelligents s'ils étaient humains.

Classés dès 1995 parmi les technologies clés du 21ème siècle par le Scientific American [Maes, 1995] et faisant l'objet d'un nombre grandissant de conférences et de publications spécialisées, les agents intelligents et les systèmes multi-agents se sont imposés avec une importance croissante comme le paradigme de l'intelligence artificielle moderne. S'il y a de nombreuses manières de justifier ces notions dans leur généralité, il est bien délicat de les définir en particulier. C'est que de nombreuses architectures d'agent ou de systèmes multi-agents reposant sur des conceptions diverses ont étés produites dans des visées variées. Le plus simple est peut-être de rappeler, en informaticien, que la notion d'agent a été introduite pour venir compléter celle d'objet (au sens de la programmation orientée objet). En effet, il y a consensus sur le fait que les agents artificiels diffèrent des objets informatiques classiques par leur autonomie et leur pro-activité. Les techniques de programmation agent, complémentaires des techniques orientés objet progressent rapidement. Aussi, les avancées théoriques et pratiques à l'égard des systèmes multi-agents sont impressionnantes, comme en témoigne le dynamisme de la communauté de recherche qui y est associée.

[6]Du fait, entre autres, de la sensibilité aux conditions initiales, centrale dans les sciences de la complexité.

Parmi les nombreux modèles d'agents existant[7], et malgré le continuum qui les unis, on distingue généralement les agents réactifs, qui se contentent de réagir à leurs stimulus internes et externes, des agents dits cognitifs qui manipulent des représentations explicites de leur environnement, c'est-à-dire « raisonnnent ».

Les systèmes multi-agents sont constitués d'agents qui interagissent avec leur environnement. Cet environnement inclut, lorsqu'il ne s'y réduit pas, les autres agents. Dans ces systèmes, la communication inter-agent est la méthode la plus commune pour permettre aux agents de se coordonner. Dès lors, l'étude et la modélisation de la communication entre agents est un des aspects privilégiés de cette jeune discipline scientifique. On distingue deux grands modes de communication entre agents : la communication indirecte qui est une communication par signaux via l'environnement généralement utilisée pour les agents réactifs et la communication directe par envoi de message, habituellement associée aux agents cognitifs. Dans cette thèse, nous introduisons, après l'avoir motivé un modèle générique de la communication directe entre agents cognitifs[8].

Plan de la thèse

Le texte de cette thèse est divisé en deux parties, que trois annexes viennent compléter.

Dans une première partie, nous présentons un état de l'art de la modélisation des communications dialogiques entre agents cognitifs, duquel découle notre problématique. Après avoir introduit et défini les notions saillantes pour la modélisation dialogique en général (chapitre 1), nous allons délimiter, grâce à une revue de la littérature fournie (chapitre 2), un certain nombre de problématiques concernant les conversations entre agents (chapitre 3).

Dans une seconde partie, nous présentons nos contributions quant à ces problématiques. Dans les chapitres 4 et 5, un cadre interactionnel complet est introduit. Celui-ci repose sur un modèle de l'engagement social flexible et de son respect (chapitre 4), opérationnalisé par un langage de communication agent reposant sur la notion de jeu de dialogue (chapitre 5). Une théorie des aspects cognitifs de la pragmatique, au sens d'une théorie de l'usage du langage (adaptée au cadre interactionnel susmentionné) est ensuite développée (chapitre 6). On montre alors comment cette théorie peut être appliquée à l'automatisation de la communication entre agents cognitifs de type BDI [Beliefs, Desire, Intentions] (chapitre 7).

[7] On pourra consulter Boissier [2001], qui présente un état de l'art des modèles et architectures d'agent dans leur diversité.

[8] Cette thèse traitant d'agents cognitifs et de communication directe par envoi de message, c'est dans ce sens que nous utiliserons les termes agents et communication dans le reste du texte.

Finalement, les perspectives d'extension et de raffinement de ces contributions sont présentées (chapitre 8) avant que de conclure (chapitre 9). En outre, certains éléments complémentaires sont présentés en annexe afin de ne pas alourdir le texte. En particulier, la distinction entre sémantique linguistique et sémantique mathématique est discutée (annexe A) et les éléments de psychologie cognitive et sociale concernant les théories motivationnelles (annexe B) ainsi que les attitudes et le changement d'attitude (annexe C) sont introduits.

Première partie

État de l’art, problématique et objectifs

Cette thèse se divise en deux parties. Cette première partie présente notre état de l'art[9] du vaste domaine de la communication entre agents artificiels cognitifs. Cet état de l'art, sans avoir la prétention d'être complet, introduit les éléments nécessaires pour comprendre notre problématique. C'est pourquoi, de manière un peu inhabituelle, il se trouve placé avant celle-ci. Cela donne également la possibilité à des lecteurs qui ne sont pas du domaine d'aborder la lecture de cette thèse avec un minimum de pré-requis puisque l'essentiel de ceux-ci sont présentés dans cet état de l'art.

Ainsi, dans les deux premiers chapitres, on présente, par l'intermédiaire de leurs fondements théoriques, ce qui nous a semblé être les principaux acquis du domaine. Le premier chapitre introduit la communication entre agents cognitifs et présente les généralités du domaine. On y introduit la théorie des actes de langage qui sert de fondement aux formalismes de représentation des énoncés et on y discute les aspects syntaxiques des langages de communication entre agents qui en ont découlé. On y présente également quelques caractéristiques fondamentales de la communication dialogique qui permettent d'entrevoir les difficultés pour passer d'une modélisation de l'énoncé à une modélisation du dialogue. Le second chapitre présente, en insistant sur leurs fondements théoriques, les deux grandes familles de modélisation des dialogues entre agents actuelles. On y distingue les approches intentionnelles qui reposent sur la modélisation des états mentaux privés des agents et les approches conventionnelles et sociales qui mettent l'accent sur la dimension sociale du dialogue.

Cet état de l'art critique motive notre intérêt pour les aspects cognitifs de la pragmatique des communications. Dans le troisième et dernier chapitre de cette partie, nous présentons donc notre problématique de recherche, avec les objectifs qui en découlent. La seconde partie de cette thèse présente et discute nos contributions quant à ces objectifs.

[9] Cet état de l'art est basé sur celui que nous avons publié dans la revue internationale de sciences cognitives, In Cognito [Pasquier et Chaib-draa, 2004b]. Il étend et met à jour les états de l'art publiés par Maudet et Chaib-draa [Maudet, 2001; Maudet et Chaib-draa, 2002].

Chapitre 1

Communication entre agents cognitifs : généralités

1.1 Introduction

Comme l'a argumenté Craig [1993], le grand nombre de théories relatives à la communication reflète la diversité des idées sur le sujet. Dès lors, si on ne peut trouver une théorie unificatrice des théories de la communication, il faut composer avec la multiplicité des approches. Selon Littlejohn [2002], le but des recherches dans les domaines de la communication ne devrait plus être la recherche d'un hypothétique modèle standard qui rendrait le champ statique et « mort ».

À l'inverse, dans le champ des SMAs [Systèmes Multi-Agents] à base d'agents cognitifs, le besoin d'un modèle de communication standard se fait sentir et les efforts se multiplient dans ce sens. Les technologies agents et multi-agents permettent de concevoir et de développer des applications complexes. La caractéristique fondamentale de celles-ci dans le paradigme actuel de l'informatique répartie est l'habileté des agents à communiquer entre eux de manière utile à leurs objectifs tant individuels que collectifs.

Cependant, faire le point sur l'étude des communications entre agents cognitifs tout en rendant compte de ses fondements est une tâche délicate. En effet, les recherches dans ce domaine empruntent (comme c'est l'habitude dans le domaine des SMAs) des notions à de nombreux domaines des sciences cognitives. On trouvera pèle-mêle : philosophie du langage, philosophie de l'esprit (épistémologie), socio-linguistique, linguistique, psychologie sociale, sociologie, dialectique, intelligence artificielle et intelligence artificielle distribuée. La stupé-

fiante diversité des approches ainsi que la complexité des cadres conceptuels développés, si elles rendent compte de l'importance et de l'envergure du domaine, rendent également toute tentative de synthèse du domaine hasardeuse et partielle.

Une première critique serait d'ailleurs de constater que malgré leur intérêt scientifique théorique fondamental indéniable, ces études se sont bien souvent éloignées de la réalité informatique. On ne présente généralement pas d'algorithme, on n'analyse pas les approches en termes de complexité computationnelle et on discute peu de l'implémentation des idées présentées. La frontière entre l'étude du dialogue pour ce qu'il est chez les humains et sa modélisation dans un cadre explicitement multi-agents est souvent confuse. Voyons pourquoi cette critique serait bien malvenue en rappelant pourquoi la communication humaine est la métaphore privilégiée des travaux sur les communications entre agents artificiels cognitifs.

1.1.1 Discussion : communication inter-humaine et inter-agents.

Dans les systèmes multi-agents la communication est un point clé. Mais est-ce que, pour autant, les modèles de communication des agents doivent être basés sur ceux issus de la recherche sur le langage naturel et les communications entre humains ? Il y a évidemment plusieurs courants au sein des recherches concernant les SMAs. Nous nous limiterons ici aux SMAs cognitifs, basés sur le paradigme de représentation symbolique des connaissances et de formalisation du raisonnement issu du courant cognitiviste des sciences cognitives. L'interaction dans ce type de système requiert des techniques de communication plus sophistiquées que les solutions traditionnelles de communication entre modules logiciels : passage de données ou appel de procédures à distance. Dans un système multi-agent hétérogène et ouvert, il faut prendre en compte :

- *l'hétérogénéité des agents* : les messages doivent être mutuellement compréhensibles alors que les points de vues des agents ne sont pas forcément mutuellement consistants ;
- *l'échange de savoir* : un agent rationnel doit pouvoir manipuler des croyances sur les autres et en particulier sur leurs comportements, croyances et intentions. En effet, un tel agent doit être capable d'identifier et d'expliciter les conflits qui font obstacle à la résolution de ses problèmes. Pour ce faire, il doit pouvoir exprimer ces différents types de connaissances et non simplement transmettre de simples données.
- *le contrôle local* : les agents doivent être autonomes. C'est-à-dire que leur comportement ne doit pas dépendre d'un planificateur central ni d'interactions pré-définies. L'agent doit être capable de développer sa propre stratégie de communication dynamiquement.

– *la structure organisationnelle* : Pour éviter l'explosion combinatoire de la quantité de communications au sein du système, il est commun d'avoir recours à une structure organisationnelle qui distribue les rôles ainsi que les relations hiérarchiques et les comportements attendus qui leur sont associés.

Dès lors que les humains parviennent à intégrer ces dimensions, il n'est pas inutile de prendre exemple sur les modèles de la communication humaine pour élaborer ceux des SMAs. L'observation des conversations humaines est la base de l'élaboration de protocoles sophistiqués. Les humains ont développé des techniques d'interaction très perfectionnées qui s'accommodent de leur rationalité limitée. Puisque les agents artificiels sont eux-aussi limités dans leur rationalité[1], on peut s'en inspirer.

Il faut également garder à l'esprit que le langage naturel a le plus grand pouvoir d'expression ! Ceci n'est pas sans conséquence lorsque l'on souhaite l'interopérabilité des différentes architectures d'agents. Pouvoir garantir qu'un langage de communication agent, à l'instar du langage naturel, permet de tout dire et sous certaines conditions d'être compris de tous est un aspect majeur. Doter un cadre de communication inter-agent de ce type de pouvoir expressif supprime la tentation pour un développeur de spécialiser son système de communication, ce qui rendrait ses agents incompréhensibles pour ceux créés par d'autres développeurs. Aussi, une approche générique comme celles inspirées du langage naturel permet plus facilement de s'accommoder de l'hétérogénéité des systèmes développés.

En outre, le langage naturel comme étalon et source d'inspiration commune donne un cadre unificateur à des recherches souvent difficiles à comparer autrement. Notons, que les aller-retour que cela implique entre les modèles de communication inter-agents et les modèles de communication humaine participent du projet de l'intelligence artificielle et plus généralement des sciences cognitives.

Finalement, les chercheurs qui travaillent sur le dialogue et les conversations avec des outils informatiques poursuivent (indépendamment ou non) les objectifs suivants :

– élaboration d'outils de traitement automatique du langage naturel (linguistique informatique et informatique linguistique) [Kayser, 2001] ;

[1] Les modèles d'agents artificiels rencontrent, au même titre que les autres logiciels, les limites de décidabilité et de calculabilité révélées par l'informatique théorique suite aux débats sur la calculabilité intuitive qui ont occupé nombre de mathématiciens prestigieux de la première moitié du 20ème siècle. On pourra consulter [Wolper, 1991] pour une présentation de ces résultats, [Simon, 1957] pour une justification de l'impossibilité de la rationalité parfaite et [Russell et Norvig, 2003, p.972-973] pour une discussion sur les conséquences sur la modélisation des agents.

– élaboration de systèmes de dialogue homme-machine en langage naturel [Caelen, 1996] ;

– conception de systèmes d'aide au dialogue, de médiatisation (collecticiel) [Lamontagne et Lapalme, 2004] ;

– élaboration de la composante interactionnelle de systèmes multi-agents ;

– modélisation linguistique, cognitive, physiologique ou neuro-biologique avec simulation et validation informatique.

Cette simple liste permet de bien comprendre que ces chercheurs ont un certain nombre d'ambitions qui si elles ne sont pas toujours incompatibles sont bel et bien différentes. Certaines recherches visent l'universalité des résultats et souhaitent embrasser le genre humain, d'autres étudient des types de dialogues particuliers (dialogue orienté tâche, supervisée, argumentation, ...) ou des types de systèmes particuliers (SMA, IHM [Interfaces Homme-Machines], simulations de systèmes sociaux ou biologiques, ...). De ce champ d'une grande diversité, nous tenterons donc d'extraire les éléments pertinents pour les communications dialogiques entre agents. Pour cela, il convient de préciser le type d'agents dit « cognitif » dont il est question dans les SMAs.

1.1.2 Hypothèses : caractéristiques des agents cognitifs

Au fur et à mesure que le domaine des SMAs se développe, les modèles d'agents cognitifs se complexifient[2]. Ainsi, on prête aux agents intelligents actuels un certain nombre de caractéristiques et de compétences dites cognitives du fait de leur nature anthropomorphique. Voyons ici les hypothèses qui sont faites concernant les agents cognitifs, capables d'utiliser des « formes » dialogiques :

– *capacités d'action et de perception* : ces capacités, présentes dans tous les modèles d'agents permettent aux agents de percevoir et d'agir sur leur environnement. On retrouve ici la boucle d'interaction avec l'environnement introduite par la cybernétique et élaborée depuis.

– *capacités cognitives* : représentation explicite des connaissances (incluant généralement le maintien d'un modèle des autres aussi complet que possible) et raisonnement formel, apprentissage (ou au moins actualisation des connaissances : mise à jour et révision des connaissances), planification garantissant la proactivité et l'autonomie.

[2]Le lecteur néophite pourra consulter [Weiss et Co, 2001, chapitre 8], [Wooldridge, 2001b, chapitres 3 et 4] ou [Boissier, 2001] pour un survol des modèles d'agent cognitifs.

– *capacités sociales* : gestion des engagements (cela suppose le raisonnement temporel), capacités de communication (manipulation d'un cadre interactionnel) ;

Tous ces éléments sont assemblés en une théorie comportementale qui abstrait un modèle de fonctionnement interne de l'agent (formalisation du raisonnement en dehors de toute instanciation) basé sur un modèle cognitif (représentation et structuration des connaissances). Les théories comportementales de ce type doivent fournir pour un agent des éléments concernant par exemple : sa stratégie de raisonnement, son modèle déductif/inductif, sa théorie de l'action et de la causalité, ses méthodes de planification et de satisfaction de buts, son système de dynamique de croyance et de révision de croyance, ses capacités, . . . [3]

Finalement, la façon de communiquer d'un agent doit être compatible avec son fonctionnement interne. On peut même dire que la capacité de communication d'un agent fait partie intégrante de son modèle cognitif et comportemental. Pour un agent donné, une conversation est un processus dynamique qui met en jeu l'essentiel de ses ressources cognitives.

Cette description, même succincte, des agents cognitifs actuels, met en évidence leur anthropomorphisme. À ce niveau d'abstraction, les concepts utilisés sont issus des sciences cognitives dans leur pluralité et n'ont de rapport avec l'informatique que par leur expression formelle. Ceci vient conforter notre analyse sur les liens entre la communication langagière humaine et la communication directe entre agents artificiels de la section précédente.

Il convient toutefois de tempérer cette position, car si les modèles de conversation pour agents artificiels sont des contributions aux modèles de conversations dans leur généralité, ils ne couvrent pas l'étendue souhaitée pour les modèles du dialogue entre humains. En particulier, les étapes d'énonciation (mise en forme du message formel) et de formalisation (formalisation du message naturel) des messages et les aspects ayant trait à la multimodalité (intégration multimodale et répartition multimodale) ne sont généralement pas traités dans les travaux sur les communications entre agents artificiels[4]. En définitive, la nature formelle

[3] De telles théories peuvent être composées, par exemple :

– d'éléments sur le savoir et l'action [Moore, 1990b] ;

– d'une architecture BDI [Belief, Desire and Intention] [Rao et Georgeff, 1995] ;

– de savoir-faire et gestion du temps [Singh, 1994] ;

– d'une théorie de l'intention [Cohen et Levesque, 1990a] ;

– d'éléments de théories économiques, théorie des jeux [Werner, 1992].

[4] Ils le sont par contre dans les systèmes concernant la communication homme-machine comme pour les agents interfaces ou les agents amenés à évoluer au sein de communautés mixtes (voir à ce propos les travaux de Chicoisne [2002] ou Lemeunier [2000, 2003]).

des agents artificiels permet d'éviter d'avoir à traiter toutes les ambiguïtés associées au langage naturel. En définitive, dans un cadre multi-agents, une grande partie de la complexité (et par là même de la richesse) des communications humaines disparaît.

À l'inverse, certaines exigences, liées à leur finalité informatique, contraignent les modèles de communication entre agents artificiels à ne pas être que de simples cadres d'analyse (comme il est commun d'en trouver en linguistique, en dialectique, ...) mais à fournir des outils et langages formels, réifiables en structures de données et procédures effectives (au sens de l'informatique théorique).

Malgré cela, vouloir rendre compte des modèles de communication entre agents par leurs fondements nous amènera à considérer des modèles de la communication dialogique plus généraux. Ce type de réflexion, intégrant différents domaines des sciences cognitives, nous semble seul permettre une vision globale et une réflexion féconde.

Le reste de ce chapitre est organisé de la manière suivante. La prochaine section (section 1.2) introduit quelques généralités sur la modélisation des énoncés linguistiques et la formalisation des langages de communication agent subséquente, tandis que les sections suivantes présentent des généralités sur le dialogue et sa modélisation (section 1.3) et motivent le passage de l'énoncé au dialogue (section 1.4) rendant ainsi compte de l'évolution historique du domaine. Le chapitre suivant, quant à lui, introduit et discute plus avant les modèles de dialogue entre agents cognitifs actuels.

1.2 Des actes de langage aux langages de communication agent (ACLs)

1.2.1 Théories des actes de langage : « Quand dire c'est faire. »

Historiquement, les unités conversationnelles ont été étudiées avant le cadre conversationnel dans lequel elles doivent prendre place. La *théorie des actes de langage* (et ses nombreuses variantes) est en ce domaine la référence incontournable. La théorie des actes de langage [Wittgenstein, 1953; Grice, 1957; Austin, 1962; Searle, 1969; Vanderveken, 1990] est issue de la philosophie du langage. Initialement pensée pour le langage naturel, sa nature formelle la rend utilisable pour les modèles computationnels. L'idée maîtresse de cette théorie est qu'une instance d'utilisation de la langue est une action comme les autres : « dire c'est faire » [Austin, 1962]. Pour chaque acte de langage « primitif », on distingue quatre composantes qui peuvent être vues comme quatre actes :

– *énonciation* : le locuteur fournit l'énoncé dans le contexte par transmission ou prononciation du message ; c'est le niveau physique.

– *acte locutoire ou locution*[5] : l'interlocuteur (ou les interlocuteurs, le cas échéant) a perçu l'énonciation. Il lui faut interpréter le sens de l'énoncé en termes de son contenu propositionnel. Si le contenu propositionnel interprété est celui que le locuteur voulait transmettre, on dira que l'aspect locutoire de l'acte est accompli avec succès. Par exemple : « il pleut », « it's raining » et « es regnet » sont trois énoncés différents qui correspondent à un seul acte locutoire de contenu propositionnel : il pleut. Dans la suite, on notera p ce contenu propositionnel qui correspond à ce qui est dit.

– *acte illocutoire ou illocution* : cet acte traduit les intentions du locuteur envers son (ses) interlocuteur(s). Une fois que l'interlocuteur a perçu et interprété le sens propositionnel de l'énoncé, il doit inférer ce que le locuteur a voulu exprimer par cet énoncé. « On va être riche » peut être interprétée selon le contexte comme une information, une prédiction, une promesse ou une blague ironique et cela change tout. Si l'interlocuteur saisit le sens que le locuteur a voulu donner à son énoncé on dira que l'acte illocutoire a réussi.

– *acte perlocutoire ou perlocution* : un tel acte porte sur les effets du message sur le destinataire : action, modification de croyance, modification de ses attitudes propositionnelles (AP). L'effet perlocutoire concerne la réaction du destinataire, les effets de son interprétation sémantique du message.

Le terme « acte de langage » est souvent employé pour désigner un acte illocutoire. Notons qu'il existe différentes variantes de cette théorie, certaines étant de nature « inférentielles » [Bach et Harnish, 1979] et d'autres plutôt « analytiques » [Searle, 1979].

Force illocutoire et contenu propositionnel

Dans leurs analyses des différents types syntaxiques de phrases du langage naturel, certains linguistes ont précisé les aspects illocutoires. Les actes illocutoires (qui sont la transmission du sens d'une phrase dans un contexte donné) consistent en l'application d'une force illocutoire F sur un contenu propositionnel p, ce que l'on notera $F(p)$ dans le reste de ce

[5] Certains auteurs, plus fidèles à la formulation d'Austin [1962] préfèrent présenter l'énonciation comme partie intégrante de l'acte locutoire qui est alors décomposé en actes phonétique (production de sons), phatique (production de mots appartenant à un vocabulaire et organisés conformément à une grammaire) et rhétique (production de l'acte phatique ayant un sens particulier pouvant faire intervenir des références).

texte[6]. Par exemple, les énoncés « allons-nous en ! » et « nous allons partir » ont le même contenu propositionnel (le locuteur et son entourage vont s'en aller) mais des forces illocutoires différentes (respectivement une force illocutoire d'ordre et une force illocutoire d'assertion pour le futur). D'après Searle et Vanderveken [1985], chaque force illocutoire peut être divisée en six composantes : un but illocutoire, un mode d'accomplissement de ce but, des contraintes sur le contenu propositionnel, des conditions préparatoires, des conditions de sincérité et un degré d'intensité de ces mêmes conditions de sincérité. Détaillons cela.

Le but illocutoire - Le but illocutoire relie la proposition énoncée au monde réel. Pour Searle, il existe cinq utilisations possibles du langage qui sont caractérisées par les cinq buts illocutoires (on parle aussi des cinq types d'actes de langage) :

1. *assertif/représentatif* : le locuteur exprime un contenu propositionnel qui se réfère au monde passé, actuel ou futur tel qu'il se le représente. Exemples d'actes illocutoires assertifs : affirmation, assertion, conjecture, rappel, accusation, témoignage, prédiction, ...
2. *directif* : le locuteur donne une directive représentée par le contenu propositionnel au(x) destinataire(s). Exemples d'actes illocutoires directifs : ordre, demande, prière, invitation, conseil, recommandation, ...
3. *commissif/promissif/engageant* : le locuteur s'engage (vis-à-vis du destinataire) à accomplir l'action représentée par le contenu propositionnel. Exemples d'actes illocutoires promissifs : promesse, menace, renonciation, acceptation, vœu, serment, ...
4. *expressif* : le contenu propositionnel concerne l'humeur mentale et l'affect du locuteur. Exemples d'actes illocutoires expressifs : déclaration d'amour, félicitation, remerciement, insulte, ...
5. *déclaratif* : le locuteur accomplit l'action représentée par le contenu propositionnel du simple fait de sa locution. Exemples d'actes illocutoires déclaratifs : excommunication, nomination, ratification, leg, ajournement, bénédiction, ...

Le but illocutoire est la principale composante de la force illocutoire car il indique le lien du contenu propositionnel avec le monde. Un locuteur qui accomplit un acte illocutoire peut avoir toutes sortes de buts perlocutoires et d'autres intentions. Par exemple, il peut vouloir amuser, convaincre, embarrasser ou choquer. Mais dans tous les cas, il a au moins l'intention

[6] Cette distinction introduite par Austin est justifiée par Searle qui observe que les deux dimensions F et p peuvent être niées indépendamment. Par exemple, la phrase « Je promets de venir » peut être niée de deux manières : « Je promets de ne pas venir » ou « Je ne promets pas de venir » selon que c'est le contenu propositionnel ou la force illocutoire qui est niée.

d'accomplir le but illocutoire de son acte concernant son contenu propositionnel. Une des justifications de la complétude de cette classification est que les cinq buts illocutoires couvrent les différentes directions d'ajustement possibles entre l'utilisation de la langue et le monde. En effet, d'un point de vue logique, il n'y a que quatre directions d'ajustement possibles pour un acte de langage :

1. *la direction d'ajustement des mots aux choses* : l'énoncé offre une représentation des choses (peut-être fausse). En cas de satisfaction d'un acte de langage ayant cette direction d'ajustement, le contenu propositionnel correspond à un état de choses existant indépendamment de l'énonciation dans le monde. Par exemple, les actes de langage ayant un but illocutoire assertif ont la direction d'ajustement des mots aux choses. En effet, ils ont pour but de représenter comment les choses sont dans le monde. Un énoncé comme : « Le ciel est bleu. » sera doté de cette direction d'ajustement.

2. *la direction d'ajustement des choses aux mots* : l'énoncé propose un processus de changement des choses. En cas de satisfaction d'un acte de langage ayant cette direction d'ajustement, le monde est transformé de façon à satisfaire son contenu propositionnel. Les actes de langage ayant un but illocutoire directif ou commissif ont la direction d'ajustement des choses aux mots. Leur but est que le monde soit transformé (respectivement par l'interlocuteur ou le locuteur) de sorte qu'il corresponde à leur contenu propositionnel. Un énoncé comme : « Fermes la porte derrière toi !. » sera doté de cette direction d'ajustement.

3. *la double direction d'ajustement* : l'énoncé est un changement des choses. En cas de satisfaction de l'acte illocutoire ayant la double direction d'ajustement, le monde s'ajuste au contenu propositionnel et cet ajustement consiste en l'énonciation elle-même. Les actes de langage dont le but illocutoire est déclaratif ont la double direction d'ajustement. Un énoncé comme : « Les États-Unies d'Amérique déclarent la guerre à l'Irak. » sera doté de cette direction d'ajustement.

4. *la direction d'ajustement vide* : l'énoncé et les choses sont indépendants. Les actes de langage dont le but illocutoire est expressif ont la direction d'ajustement vide. Ce type d'acte de langage est supposé toujours satisfait puisqu'il ne réfère pas aux états de choses du monde (extérieur) mais à l'état mental du locuteur. Un énoncé comme : « Je ne me sens pas très bien ! » sera doté de cette direction d'ajustement.

Le mode d'accomplissement - La plupart des buts illocutoires peuvent être atteints de différentes façons. Ainsi, un acte de langage directif peut être réalisé de différentes manières : autoritaire, douce, supplicative, ... Le mode d'accomplissement spécifie comment le but illocutoire doit être atteint. En français, ceci est exprimé par des adverbes comme : obligatoirement, éventuellement, peut-être, ...

Les contraintes sur le contenu propositionnel - Certaines forces illocutoires imposent des conditions sur l'ensemble des contenus propositionnels qui pourraient leur être associés. Ainsi, les actes de langage dont le but illocutoire est commissif ou directif doivent avoir un contenu propositionnel représentant des actions futures respectivement du locuteur ou des interlocuteurs. Par exemple, lorsque le locuteur énonce une promesse, celle-ci doit avoir pour contenu propositionnel une action future du locuteur : « je ne le ferai plus, c'est promis ! ».

Les conditions préparatoires - Lors de la performance d'un acte illocutoire, le locuteur a généralement des croyances sur le contexte de son énonciation. Par exemple, un locuteur qui donne un conseil croit généralement que cela peut aider l'autre partie. Les conditions préparatoires d'une force illocutoire déterminent quelles doivent être les croyances du locuteur pour qu'il puisse accomplir un acte de langage ayant cette force.

Les conditions de sincérité - Dans l'accomplissement d'un acte illocutoire, le locuteur transmet un contenu propositionnel avec un but illocutoire, mais aussi des informations concernant ses états mentaux. Cela signifie que si la communication est une « extérieurisation », elle renseigne aussi sur l'état intérieur du locuteur. Par exemple, lorsqu'un locuteur énonce une demande, il a un but illocutoire directif auquel peut être associé un désir, un regret, une inquiétude, ... De tels états mentaux sont des attitudes propositionnelles $m(p)$ où p est le contenu propositionnel et m est un mode psychologique (croire, espérer, désirer, regretter, ...).

Comme tout locuteur peut mentir et transmettre des états mentaux qui ne sont pas réellement les siens, on peut distinguer les actes illocutoires sincères (le locuteur a les états mentaux qu'il exprime) des insincères. Les conditions de sincérité indiquent quels états mentaux devraient être présents chez le locuteur lorsqu'il produit un acte illocutoire sincère. Par exemple, une condition de sincérité des actes illocutoires assertifs est que le locuteur doit croire le contenu propositionnel. De même, une condition de sincérité des actes illocutoires directifs est que le locuteur doit désirer que le contenu propositionnel soit accompli. Ou encore, une condition de sincérité des actes illocutoires commissifs est que le locuteur doit avoir l'intention de réaliser le contenu propositionnel.

Le degré d'intensité - Évidemment, les attitudes propositionnelles associées aux conditions de sincérité le sont avec une intensité qui dépend de la force illocutoire. Ainsi, une supplication dénote un désir plus grand qu'une simple demande. L'intensité d'une force illocutoire est donc indiquée par un degré.

Actes illocutoires complexes

Il est important de noter que la notion de force illocutoire est itérable. On peut, par exemple, affirmer qu'une demande a été faite, proposer une demande ou demander une proposition, ... Tous ces actes prennent la forme $F(F(p))$ ou plus généralement $F(\ldots F(p)\ldots)$ et il n'y a pas de limitation à ce type d'imbrication. En outre, certains actes illocutoires utilisant des connecteurs comme « et » ou « mais » ne sont pas de la forme $F(p)$. Il s'agit d'actes illocutoires dits complexes qui sont généralement de la forme $F_1(p_1) \wedge F_2(p_2)$. Par exemple, S_0 est à la fois une assertion et une question.

S_0 *: Il est six heures, non ?*

Un certain nombre de connecteurs illocutoires ont été proposés par Searle et Vanderveken [1985] dans le cadre de leur logique illocutoire.

Actes de langage indirects

Au niveau linguistique, la théorie des actes de langage classique prévoit qu'il est possible de déterminer la force illocutoire d'un acte à partir de sa forme linguistique. Les marqueurs lexicaux ou syntaxiques doivent permettre cette opération. C'est ce que l'on appelle « l'hypothèse de force littérale ». La réalité n'est pas si simple et dans de nombreux cas, l'acte locutoire ne suffit pas à déterminer l'acte illocutoire correspondant. Seule l'utilisation « d'expressions performatives explicites » de la forme « je te <verbe performatif> que ... » donne le type de l'acte réalisé sans ambiguïté, encore que l'on puisse trouver des cas discutables. Dans les dialogues réels, de nombreux actes de langage ne sont pas du type indiqué par leur forme linguistique (que nous nommerons forme littérale ou acte de surface). C'est ce que l'on appelle les actes de langage indirects. Par exemple, l'énoncé S_1 suivant s'interprète généralement comme une requête à laquelle on répond en donnant une salière et non comme une question fermée (à laquelle on répond par oui ou non).

S_1 *: Peux-tu me passer le sel ?*

Bien des difficultés subsistent quant à l'étude de ce type d'actes, et ce, même si les travaux de Grice [1969] sur la notion d'implicature (section 2.2.1) fournissent des indications sur le type d'indirection réalisée pour passer de la question littérale à la requête indirecte.

Les conditions de succès des actes illocutoires

Les conditions de succès d'un acte illocutoire sont l'ensemble des conditions qui doivent être réunies dans le contexte de l'énonciation pour que le locuteur réussisse à accomplir cet acte. Les conditions de succès d'un acte de langage sont déterminées de façon univoque par la force illocutoire et le contenu propositionnel de l'acte de langage. Un acte illocutoire $F(p)$ est accompli avec succès si et seulement si le locuteur réalise le but illocutoire de la force F sur la proposition p avec le mode d'accomplissement, les conditions préparatoires, les conditions de sincérité, le degré de puissance de F et que p vérifie les conditions sur le contenu propositionnel de F. Par exemple, la condition de succès de la requête S_2 est que le locuteur doit espérer que le récepteur accomplisse les actions représentées par le contenu propositionnel à savoir manger sa soupe.

S_2 : Mange ta soupe !

De même, la condition de succès d'une promesse est que le locuteur doit être prêt à s'engager à accomplir les actions représentées par le contenu propositionnel dans le futur.

Les conditions de satisfaction des actes illocutoires

La plupart des actes de langage concernent le monde via leur contenu propositionnel. Même si un acte illocutoire est accompli avec ses conditions de succès remplies, il peut ne pas être satisfait quant à son rapport au monde. Un acte de langage est satisfait si son contenu propositionnel est rendu vrai selon la direction d'ajustement au monde propre à son but illocutoire. Les conditions de satisfaction d'un acte illocutoire sont les conditions qui doivent être réunies pour que l'acte de langage soit entièrement satisfait (c'est-à-dire que chacune de ses trois composantes soit satisfaite). Par exemple, la condition de satisfaction d'une promesse est que le locuteur accomplisse ce qu'il a promis. De même, une demande (comme S_2) n'est satisfaite que si l'interlocuteur accomplit les actions qui sont représentées par le contenu propositionnel[7].

1.2.2 Les langages de communication agent (ACLs)

Dans les SMAs cognitifs, l'hypothèse la plus répandue est que la communication inter-agent sera plus fructueuse si elle se fait par l'intermédiaire d'un langage de communication

[7] Dans la théorie des actes de langage, la notion de satisfaction est une généralisation de la notion de vérité.

explicite. La propriété essentielle qui rend le langage utile, c'est que le sens de ses signes soit partagé. Ceci est vrai pour les langues vivantes mais aussi pour tous les signes codés comme les coups de sifflet d'un arbitre de football. Ce que l'évolution a permis pour les langages humains, la standardisation le tente pour les agents artificiels [Finin et al., 1999], on utilise alors au minimum :

- *un dictionnaire de vocabulaire/signes commun(s)* : l'ontologie des services (dont le sens est au moins partiellement commun aux agents) qui permet une interprétation commune des contenus propositionnels des actes de langage.
- *des actes de langage* utilisant cette ontologie qui servent de briques de bases de la communication et correspondent aux attitudes propositionnelles transmises entre les agents.

Les langages de communication agent, notés ACL [Agent Communication Language] dans le reste du texte, prennent donc place dans une couche logiquement supérieure à celle des protocoles de transfert de données informatiques (TCP/IP, HTTP, IIOP, ...) et adressent le niveau intentionnel et social des agents. Ils se différencient donc des mécanismes de la théorie de l'information[8] non seulement par leur structure et leur syntaxe plus complexe, mais aussi par leurs spécifications génériques et précises. Leur puissance expressive, héritée de la théorie des actes de langage, les rend suffisamment génériques pour envisager satisfaire les besoins des multiples applications des SMAs, allant de l'ingénieurie de systèmes de résolution de problèmes à la simulation sociale. Les sections suivantes présentent la palette des ACLs actuels, dont les deux plus connus et « standards » sont KQML et FIPA-ACL.

KQML [Knowledge Query and Manipulation Language]

Apparu avant FIPA-ACL, KQML [Finin et Fritzson, 1994] fourni un ensemble d'actes de langage standards et utiles. Proposé en 1993 par le consortium DARPA-KSE [Knowledge Sharing Effort], ce langage est structuré selon trois niveaux enchâssés [Labrou et al., 1995] :

1. *la couche de communication* : renseigne la communication (identité du récepteur, de l'émetteur et nature de la communication). Elle est minimale car KQML ne prend pas en charge le transport lui-même (TCP/IP, SMTP, IIOP ou autres).

[8] La théorie de l'information, développée par Shannon et Weaver [1975], sert de base théorique à la communication dans les réseaux informatiques.

2. *la couche message* : donne des indications sur le contenu du message, en particulier le langage et l'ontologie utilisés pour le contenu ainsi que le type d'acte de langage attaché au contenu. C'est la couche centrale de KQML qui définit le type d'interaction que des agents-KQML pourront avoir.

3. *la couche de contenu* : contenu du message exprimé en KIF [Knowledge Interchange Format], Prolog, KQML ou autre. Notons que KQML ne traite pas cette couche, si ce n'est pour savoir où le contenu commence et se termine. Comme le contenu du message est opaque, c'est à la couche message de le renseigner.

Les primitives de KQML sont appelées performatifs même si elles ne sont pas des actes performatifs au sens linguistique du terme. Cette appellation qui prête à confusion a d'ailleurs été critiquée. Ces performatifs sont divisés en trois catégories :

1. 7 performatifs de régulation de conversation traitent quelques cas particuliers (`sorry`, `error`) et permettent quelques variantes de la conversation (`standby`, `ready`, `next`, `rest`, `discard`) ;

2. 17 performatifs de discours permettent l'échange d'informations et de connaissances (`ask-if`, `tell`, `deny`, `stream-all`, ...) ;

3. 11 performatifs d'assistance et de réseau pour étendre la conversation à plus de deux agents (`forward`, `broker-all`, ...).

KQML est issu d'un projet de la DARPA, le KSE [Knowledge Sharing Initiative], initialement prévu comme moyen d'échange d'informations entre programmes à base de connaissances. Cependant, sa structure orientée message et la généricité de ses primitives lui permettent d'être utilisé comme ACL. KQML a été l'ACL le plus utilisé (et implémenté) dans la communauté SMA pendant les années 90. Aussi, puisque le développement de KQML n'a pas été centralisé, plusieurs variantes incompatibles sont nées. À la fin de années 90, l'ACL développé par la FIPA [2004] [Fondation for Intelligent Physical Agents] a progressivement détrôné celui-ci.

FIPA-ACL

C'est le seul effort réellement organisé pour créer un ACL standard. Comme FIPA-ACL est défini par une corporation de chercheurs [FIPA, 2000], sa mise à jour est lente mais chaque version est scrupuleusement vérifiée. Il a été conçu pour palier aux faiblesses des différentes

versions de KQML. FIPA-ACL[9] diffère de KQML en ce qu'il a été directement doté d'une sémantique. En effet, la version originale de KQML ne décrivait que la syntaxe de ses messages et rien n'était dit sur leur sens précis (indépendamment qu'ils correspondaient grossièrement à différents types d'actes de langage). Ce n'est que plus tard, qu'une sémantique a été proposée pour KQML. Les aspects sémantiques des langages de communication agents seront présentés en détail section 2.2.3. FIPA-ACL est né d'ARCOL.

ARCOL [ARtimis Communication Language]

ARTIMIS est une plateforme générique pour agents communicants développée par France Télécom [Sadek et al., 1997]. Dans ARTIMIS, un agent peut communiquer avec un humain aussi bien qu'avec un autre agent. Les faits de communication des agents sont modélisés comme des actions rationnelles via ARCOL. Une expression ARCOL est écrite en SL [Semantic Language] pour le message et utilise SCL [Semantic Content Language] pour le contenu.

ARCOL définit quatre primitives mutuellement exclusives et composables : `Inform`, `Request`, `Confirmation` et `Inform Referent`. Une des limites d'ARCOL est qu'il présuppose la sincérité des agents, ce qui est un obstacle à l'objectif d'ouverture des ACLs.

ICL [InterAgent Communication Language]

OAA [Open Agent Architecture] est une plateforme multi-agent pour environnements distribués développée par SRI International [1999]. ICL en est le langage de communication agent. ICL est un langage déclaratif logique vu comme intergiciel (*middle-ware*) pour l'implémentation de mécanismes de coopérations entre les serveurs de services, les clients et les agents facilitateurs du cadre OAA. ICL utilise trois types de primitives : `Solve`, `Do` et `Post`. Le contenu des messages est exprimé en Prolog. ICL est cependant restreint à l'architecture OAA et à sa nature procédurale qui interdit les échanges bidirectionnels.

[9] L'ACL de FIPA s'appelle en fait ACL, mais pour éviter toutes confusions nous le nommerons FIPA-ACL dans le reste du texte.

AOP [Agent Oriented Programming]

À l'instar de ICL pour OAA, d'autres plateformes de développement orienté agent inclues un ACL. Shoham [1990a] a introduit le paradigme de programmation orientée agent (AOP) en empruntant à l'intelligence artificielle, à la théorie des actes de langages et à la programmation orientée objets. AOP ne dispose que de trois primitives communicationnelles composables : `Inform, Request, Unrequest.`

MAC [Mobile Agent Communication]

Un agent mobile est un programme qui peut migrer de façon autonome dans un réseau de machines hétérogènes. Du fait de son jeune âge, ce type d'agents relève plus du génie logiciel que de l'intelligence artificielle. De tels agents sont considérés comme de simples processus et non comme des agents cognitifs disposant d'attitudes mentales et évoluant dans une organisation sociale. Ainsi, les mécanismes de communication pour de tels agents ne sont pas à proprement parlé des ACLs et se limitent généralement à de simples mécanismes de passages de messages. Plusieurs alternatives, plus primitives que les ACLs, peuvent être utilisées : RMI [Remote Method Invocation] , CORBA [Common Object Request Broker Architecture]. Un exemple d'agents mobiles communicants par RPC [Remote Procedure Call] ou passage de message sont les Aglets [AGile appLETS] : des objets Java développés par le centre de recherche IBM de Tokyo.

Autres ACLs

On compte un certain nombre d'autres ACLs :

- COOL : un ACL conçu par l'Université Technique de Berlin et l'Enterprise Integration Laboratory à l'Université de Toronto.
- LOGOS : un ACL développé par la NASA avec des éléments spécifiques au contrôle aérien.
- PLACA : langage dérivé de AGENT-0 et KQML pour intégrer les intentions et des éléments de planification.
- APRIL et MAIL : langages développés dans le cadre du projet ESPRIT : IMAGINE.

1.2.3 Limitations des ACLs

Suite à cette présentation des principaux ACLs, développés à partir de la théorie des actes de langages, il convient d'indiquer les limitations de ceux-ci. Puisque les sémantiques de ces ACLs n'ont été définies que plus tard et que nous en traiterons au chapitre suivant, nous nous contenterons d'indiquer ici, brièvement, les limitations syntaxiques des ACLS. Celles-ci tombent dans deux catégories : le problème de couverture des ACLs et les problèmes concernant les ontologies.

Problème de couverture des ACLs

De nombreuses discussions sont encore en cours sur la couverture des différents types d'actes de communication par les primitives définies dans l'un ou l'autre des ACLs précédents. En philosophie du langage, on trouve les catégories suivantes, dont les cinq premières sont celles définies dans la théorie des actes de langage[10] : représentatif ou assertif, directif (ordre ou demande), commissif (promesse), expressif, déclaratif, permissif (autorisation) et prohibitif (interdiction). Or, force est de constater que les ACLs ne couvrent généralement pas toute cette panoplie. Par exemple, FIPA-ACL et KQML ne définissent que des primitives assertives et directives. Cela a pour conséquence que certaines attitudes propositionnelles (comme le désir, la crainte, ...) des agents cognitifs ne sont pas exprimables dans ces ACLs populaires [Chaib-draa et Dignum, 2002].

Problèmes concernant les ontologies

Le langage de la couche de contenu, ainsi que le vocabulaire qui y est utilisé doivent être compris et donc partagés entre les agents. Pour ce faire, un champ indiquant l'ontologie à utiliser est prévu dans KQML (originellement pour une ontologie d'Otonlingua) comme dans FIPA-ACL. Pour être utile, une ontologie relative à un domaine doit être complète et donc relativement volumineuse. Le problème de savoir quand et comment un agent doit intégrer une (nouvelle) ontologie reste ouvert.

[10] Les partisans de la théorie des actes de langage réduisent les autres catégories à des sous types de celles-ci

Conclusion

Nous détaillerons les sémantiques proposées pour KQML et FIPA-ACL en section 2.2.3, car elles participent des approches intentionnelles des dialogues agents traitées en section 2.2. Les modèles de dialogues sont en effet rendus nécessaires par le fait que chaque acte de langage est à interpréter via son contexte et plus particulièrement en rapport aux autres actes de langage qui l'entourent temporellement. L'acte de langage prend son sens comme élément d'un contexte et plus spécifiquement d'une conversation, d'un dialogue. La section suivante introduit des généralités sur les notions de dialogue et de conversation.

1.3 Généralités sur le dialogue

1.3.1 Définition

Dans les sociétés humaines, un acte de langage est rarement utilisé seul. Dans la communication utilisant le langage naturel, un système de signes parmi d'autres, on distingue le discours (un locuteur), du dialogue et de la conversation (au moins deux intervenants)[11]. Ce qui différencie le dialogue de la simple communication (le discours), c'est la recherche d'une *inter-compréhension*. Cette recherche impose aux interlocuteurs de s'assurer qu'ils se comprennent pour co-construire des interprétations communes. Par exemple, un dialogue de sourds n'est pas un dialogue. Si l'un des interlocuteurs n'intègre jamais la vision de l'autre dans la sienne, la co-construction est bloquée. Cette recherche d'inter-compréhension peut être synchrone (dialogue oral[12]) ou asynchrone (courrier, courriel, forum de discussion).

La co-construction d'une interprétation commune ne signifie pas que les interlocuteurs doivent être d'accord. Ils peuvent avoir l'interprétation commune de leur désaccord. Ajoutons que cette co-construction semble réussir plus ou moins selon un grand nombre de facteurs inhérents aux interlocuteurs (proximité culturelle et sociale, historique de leur relation, ...). La recherche d'une inter-compréhension optimale est facilitée par :

– *la recherche d'un langage commun* : sans langage commun, le dialogue est difficile. Notons que cette recherche se poursuit au cours du ou des dialogues par la construction de références communes [Clark et Wilkes-Gibbs, 1986] ;

[11] Si l'on souhaite préciser le nombre d'intervenants, on peut recourir à des termes plus précis : dilogue, trilogue, ...

[12] Éventuellement porté par écrit, comme les pièces de théâtre ou les exemples de ce texte.

- *la capacité à détecter les ambiguïtés et les incohérences dans le discours d'autrui* ;
- *la production et la recherche de retours (feedbacks) de compréhension* : pour s'assurer de leur compréhension réciproque, les interlocuteurs fournissent et cherchent chez l'autre des indices de compréhension. Ces indices peuvent être positifs (manifestations d'une bonne compréhension : signes d'acquiescement verbaux ou non-verbaux) ou négatifs (demande de clarification et d'explication), explicites ou implicites.
- *les échanges correctifs* : la qualité de l'inter-compréhension repose sur des interprétations. Comme il peut toujours y avoir erreur d'interprétation, la capacité à corriger sa compréhension ou celle d'autrui est fondamentale.
- *la capacité à méta-communiquer* : l'évitement et la résolution de conflits de compréhension peuvent nécessiter de dialoguer sur l'activité de dialogue en cours.

Toutes ces caractéristiques sont à la base d'une propriété plus générale du dialogue : *l'interactivité*. Il y a interactivité lorsque les actions des uns sont influencées par celles des autres. Cette propriété a une implication majeure : la non-prédictibilité globale de la structure du dialogue. En effet, contrairement aux phrases ou même aux discours, les conversations ne sont généralement pas l'expression de structures régulières, mais plutôt de co-constructions complexes. Néanmoins, on peut tout de même différencier les conversations orientées, cadrées ou contraintes, d'autres plus libres. Par exemple, des conversations orientées telles que les enchères, les entretiens d'embauche, les interrogatoires, les dialogues patient-docteur lors d'une consultation, ou encore les échanges entre élève et professeur sont plus structurés que des conversations libres entre amis. On dira que ces conversations contraintes, qui ont lieu au sein d'une activité précise, en lien avec une tâche particulière ou menée dans l'attente de certains résultats sont *orientées structure* et les conversations libres *orientées processus*. Le tableau 1 récapitule les différences entre ces deux types de conversation en reprenant les caractéristiques majeures du dialogue contraint et du dialogue libre indiquées par Searle [1992b].

1.3.2 La conversation comme activité commune

Il existe de nombreux travaux sur la sémantique des actes de langage (pour une synthèse, on pourra consulter [Pasquier, 2001b] ou [Labrou, 1996]). Mais peu d'études formelles traitent de la sémantique des dialogues. On pourrait penser que la sémantique d'une conversation est obtenue par composition des sémantiques des actes de langage qui la compose. Il n'en est rien. Illustrons cela à l'aide d'un exemple. Le graphe de conversation pour l'action

Orienté processus	**Orienté structure**
Peu de contraintes sur le type de contribution	Beaucoup de contraintes sur le type de contribution
De multiples buts essentiellement locaux	Peu de buts globaux
Ordre d'intervention assez libre	Ordre d'intervention assez fixe
Rôle des participants pas très bien défini	Rôle des participants bien défini
Contribution dépendante des contributions passées	Contribution rarement dépendante des contributions passées
Principe d'organisation local	Principe d'organisation global

TAB. 1.1 – Dialogue orienté structure vs. orienté processus, selon Chaib-draa et Vongkasem [2000].

de Winograd et Flores [1986] montre qu'une simple demande d'action est un processus complexe composé de plusieurs actes de langage impliquant le requérant, mais aussi l'agent à qui la demande est faite (la figure 1 présente ce protocole). Une telle demande est une action commune complexe et la sémantique des actes de langage qui composent le graphe ne saurait rendre compte de la sémantique de la conversation dans sa généralité (en particulier sa structure).

Cet exemple, parmi d'autres, montre que la communication entre agents se fait par conversation, séquence d'actes de langage dont la somme des sens isolés ne rend pas compte de la signification. C'est pourquoi, de nombreux chercheurs rendent aux conversations leur *dimension sociale* et tentent d'analyser le dialogue du niveau conversationnel vers le niveau des actes de langage plutôt que le contraire. Cohen et Levesque [1990b] ont proposé leur point de vue avec les buts persistants et les croyances mutuelles. Grosz et Kraus [1996] ont proposé les plans partagés, Singh [1994] l'engagement social et Traum [1997] étudie les croyances mutuelles.

Chaib-draa et Vongkasem [2000] ont exploré les idées de Searle et Clark qui voient la conversation comme une *activité commune* des agents qui y participent. Pour Clark [1996], une activité commune est une activité qui implique au moins deux participants, chacun jouant un rôle au sein de cette activité. Les participants cherchent à accomplir certains buts collectifs (en outre de leurs éventuels buts individuels/privés). Les participants peuvent utiliser des procédures spécifiques pour atteindre leurs buts, mais doivent s'accorder sur le début et la fin de l'activité. L'exécution des procédures constituantes de l'activité peut être simultanée ou séquencée. Cette activité commune est un ensemble d'actions dont la plupart sont des *actions*

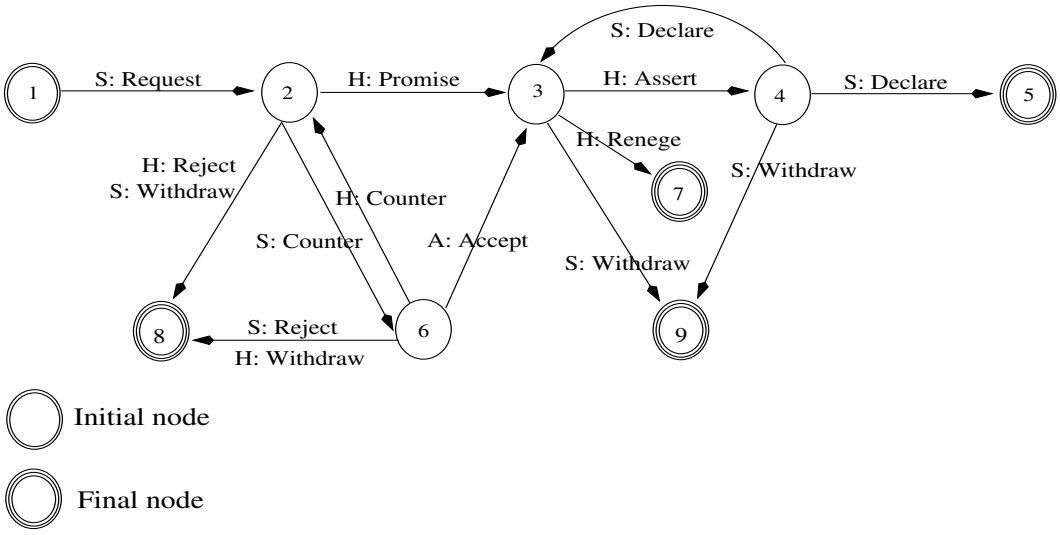

FIG. 1.1 – Protocole de conversation pour l'action par Winograd et Flores [1986].

communes. Une action commune est plus que la somme des actions individuelles des participants. La coordination, notamment, est en plus. Cela soulève le problème de la coordination d'une activité commune. Clark [1996] distingue quatre niveaux dans la communication langagière. Chacun de ces niveaux nécessite que les agents se coordonnent :

1. *niveau comportemental / attentionnel* (le locuteur exécute/l'interlocuteur prête attention) : l'émetteur d'un message doit d'abord s'assurer de l'attention du récepteur.
2. *niveau du signal* (le locuteur présente/l'interlocuteur identifie) : l'émetteur d'un message doit s'assurer de la bonne réception de celui-ci par le récepteur.
3. *le niveau du sens du signal* (le locuteur signifie/l'interlocuteur comprend) : le locuteur doit s'assurer de la bonne compréhension du message par l'interlocuteur. Les participants doivent avoir la même interprétation de l'information.
4. *le niveau de l'activité* (le locuteur propose/l'interlocuteur considère) : les participants doivent avoir identifié le projet, vouloir le réaliser, chacun doit être capable d'accomplir sa part de l'activité commune et enfin tous les participants doivent avoir les croyances mutuelles communes des points précédents et de celui-ci.

Ces niveaux ont une propriété de causalité descendante. En effet, proposer un projet (niveau 4) nécessite de s'assurer de la bonne compréhension de l'interlocuteur (niveau 3), ce qui implique de présenter un signal et de s'assurer de sa bonne réception (niveau 2), ce qui

requiert l'exécution d'un comportement pour attirer l'attention (niveau 1). Ainsi, valider un niveau signifie que tous les niveaux inférieurs sont validés.

Comme exemple d'activité commune, citons le duo en musique et plus proche de l'activité linguistique : les dialogues de recherche d'information comme les paires question-réponse. Idéalement, tous les types de dialogues possèdent les caractéristiques d'une activité commune.

1.3.3 Typologie des dialogues

Une typologie exhaustive des dialogues est illusoire puisqu'il existe une infinité de types de dialogues [Wittgenstein, 1953]. Mais, en se restreignant aux dialogues à but discursif, Vanderveken [1999] propose une classification basée sur les directions d'ajustement :

1. *dialogue à but descriptif* : les interlocuteurs dialoguent pour décrire l'état du monde ou se mettre d'accord sur cette description. C'est le cas des dialogues de persuasion, d'investigation ou de recherche d'information, des enquêtes, des examens, ...
2. *dialogue à but délibératif* : les interlocuteurs dialoguent pour s'engager ou engager les autres à réaliser certaines actions. C'est le cas des dialogues de négociation ou de délibération, ...
3. *dialogue à but déclaratif* : les interlocuteurs dialoguent dans le but de réaliser une déclaration commune. C'est le cas pour les dialogues tenus par des assemblées ou par un jury, ...
4. *dialogue à but expressif* : les interlocuteurs dialoguent pour exprimer leurs attitudes. C'est le cas des hommages, des bravos, des félicitations, ...

Dans le même esprit que pour la théorie des actes de langage, ces quatre types de dialogue sont raffinables selon leurs composantes : mode d'atteinte du but discursif, conditions thématiques, conditions d'arrière-plan, conditions de sincérité.

Une autre classification, issue des travaux en dialectique formelle est celle proposée par Walton et Krabbe [1995]. Ils définissent cinq types de dialogues principaux qui sont caractérisés par le but global et commun du dialogue et les buts privés de chacun des interlocuteurs :

1. *dialogue de persuasion* : le but global est de résoudre un conflit, le but privé de chacun des intervenants est de convaincre son ou ses interlocuteur(s) ;

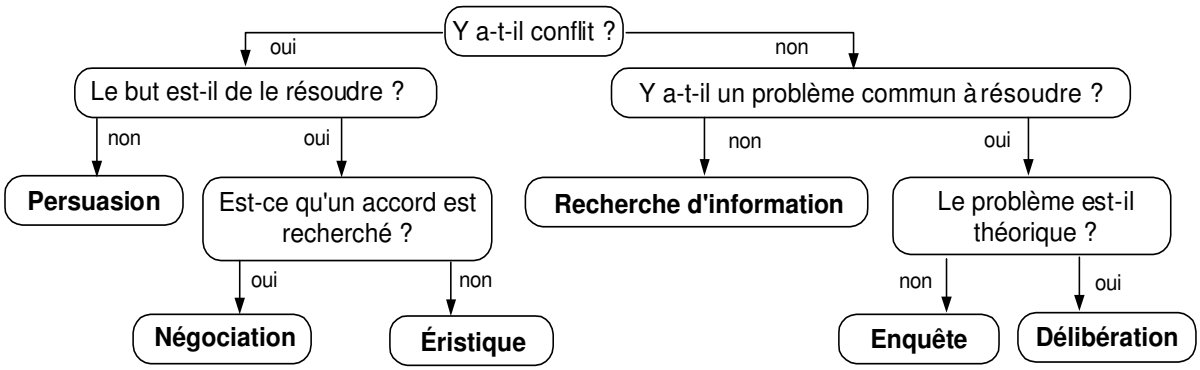

FIG. 1.2 – Situation initiale et type de dialogue d'après Walton et Krabbe [1995].

2. *négociation* : le but global est de résoudre un conflit en atteignant un accord, chacun des agents ayant pour but privé de maximiser sa satisfaction ;

3. *dialogue d'investigation (ou enquête)* : le but global est d'établir la véracité d'un fait, les buts individuels visent à participer à ce processus ;

4. *dialogue de délibération* : le but global est de s'entendre sur un plan, chacun travaillant à l'influencer à son avantage ;

5. *dialogue de recherche d'informations* : le but global est de trouver de l'information. C'est le seul type de dialogue qui ne soit pas symétrique, seul un des agents a pour but privé d'obtenir de l'information.

Walton et Krabbe ont également systématisé les conditions initiales de ces différents types de dialogues comme indiqué par la figure 1.2. On note que certains types de dialogues mentionnés par Vanderveken sont absents de la classification de Walton et Krabbe. Il est, par exemple, difficile de voir dans quelle catégorie de Walton et Krabbe les dialogues à but expressif pourraient se retrouver.

Ces classifications ne signifient pas que les dialogues soient d'un seul et même type tout au long de leur déroulement. Les dialogues observables dans la réalité seraient plutôt des compositions, parfois complexes, de ces types de base. On parle alors de *dialogues de types complexes*.

En outre, le choix du type de dialogue dépend du contexte. Bunt [1996] distingue les aspects statiques des aspects dynamiques du contexte. Pour les aspects statiques, il identifie :

– *le contexte physique et perceptuel* :

  - *co-présence* : c'est la situation de dialogue élémentaire (parce que la plus commune), le « face à face », dialogue oral, à un endroit donné et à un moment précis.
  - *non co-présence* : le dialogue est médiatisé (téléphone, courriel, courrier) par l'oral ou l'écrit.

- *le contexte social* : les rôles sociaux des interlocuteurs sont porteurs de droits et d'interdictions liés à l'activité dialogique (par exemple : professeur-élève, policier-témoin, patient-médecin, directeur-employé, client-vendeur, . . .). Généralement, l'interlocuteur qui a l'initiative est celui qui guide la discussion. Dans le cas général, on parle de dialogues d'initiatives mixtes.

- *le contexte cognitif* : ce contexte contient les différentes attitudes mentales des interlocuteurs. Il est donc essentiellement dynamique. Les aspects statiques qui peuvent être isolés en début de dialogue sont : l'identité des interlocuteurs, les capacités qu'on leur prête (on attribue certaines capacités, différentes, à un enfant ou à une machine, . . .) et le but principal de la conversation qui explique l'entrée en dialogue.

Notons que tous les travaux présentés dans ce document ne considèrent que les dialogues oraux médiatisés (cas de non co-présence), en laissant de côté la prosodie, les gestes, les mimiques et autres postures propres aux humains. Cette restriction est une conséquence des simplifications décrites en section 1.1.2 et tient à la nature des agents artificiels.

1.4 De l'énoncé au dialogue

Historiquement, les unités conversationnelles – les énoncés – ont été étudiées avant le cadre conversationnel, dialogique, dans lequel elles doivent prendre place. Pour ce qui est de la représentation des énoncés, c'est la théorie des actes de langage (section 1.2.1) qui est communément admise et utilisée en intelligence artificielle. Les langages de communications agents (section 1.2.2) n'en sont que des instanciations dans le cadre des systèmes multi-agents. Lorsqu'il a fallu les utiliser pour des conversations, de nombreux problèmes sont apparus. Il en a résulté une critique de la théorie des actes de langages (section 1.4.1) et plusieurs propositions visant à dépasser ces limitations ont été avancées (présentées sections 1.4.2, 1.4.3 et 1.4.4).

1.4.1 Critique des actes de langage

Les problèmes de la théorie des actes de langage apparaissent lorsque l'on considère les caractéristiques générales du dialogue telles que nous les avons présentées à la section précédente (section 1.3). On distingue trois sources de problèmes distinctes :

1. *le caractère non monologique du dialogue* : l'interlocuteur est plus qu'un simple locuteur/auditeur, les interlocuteurs sont engagés dans une activité commune (voir section 1.3.2) ;
2. *l'importance du contexte dialogique* : chaque acte de langage est à interpréter via son contexte et en particulier via les actes de langages qui l'entourent temporellement pour former un type de dialogue (voir section 1.3.3).
3. *le caractère multidimensionnel du dialogue* : en plus d'être utilisé pour effectuer une tâche, le dialogue est également utilisé pour se contrôler lui-même (voir les niveaux de dialogues, section 1.3.2).

Trois types de solution ont été proposés pour prendre en compte les spécificités du dialogue dans la théorie des actes de langage. Chacune de ces solutions se concentre sur l'une de ces dimensions problématiques et on présentera dans l'ordre : (1) la dialogisation des actes de langage (section 1.4.2), (2) les approches contextuelles (section1.4.3) et (3) les actes multi-niveaux et les actes de dialogue (section 1.4.4).

1.4.2 Dialogisation des actes de langage

De nombreux auteurs ont reproché aux actes de langage leur caractère monologique. Pourtant, il y a chez Austin [1962] la notion d'*uptake*. Dans l'exemple suivant, A formule un acte de langage potentiellement indirect (question fermée ou requête) et c'est la réponse (l'*uptake*) de B, à savoir B', B'' ou B''', qui détermine le sens (la force illocutoire, notamment) de l'acte de A.

A : Vous allez à Berlin ?
B' : Oui
B'' : Oui, montez !
B''' : Vous pouvez monter.

Type d'acte communicatif	**Fonction expressive**	**Fonction évocatrice**
Assertif	Croyance (jugement)	Croyance partagée
Question	Désir d'information	Information désirée
Requête	Désir que X	X

TAB. 1.2 – Fonctions de quelques actes communicatifs selon Allwood [1994].

Malheureusement, cette notion d'*uptake* n'a pas été formalisée dans la théorie des actes de langage classique. Plusieurs extensions de la théorie des actes de langage ont donc été proposées en vue d'utiliser les actes de langage dans un cadre dialogique.

Trognon et Brassac [1992] dialogisent la logique illocutoire et en font une logique interlocutoire qui prend en compte la notion d'*uptake* présentée ci-dessus. Leur approche revient à considérer que la réponse de B (en reprenant l'exemple de la section précédente) « est une offre d'interprétation adressée à A qui l'examine pour voir si elle est une interprétation acceptable ». Leur principe de l'enchaînement conversationnel se décompose en trois moments [Brassac, 1994] : (1) A produit un énoncé, (2) B produit un énoncé qui (entre autres) propose une interprétation de celui de A et (3) A valide, invalide ou module l'interprétation de B (tout en proposant une interprétation de l'énoncé de B, chaque énoncé étant susceptible d'être une composition de (1), (2) et (3)). C'est ainsi que, pour ces auteurs, se résout le problème de l'interprétation, c'est-à-dire le problème de l'inaccessibilité des intentions des agents communiquants.

D'autres extensions de la théorie des actes de langage ont introduit la notion d'effet perlocutoire attendu (la notion d'effet perlocutoire a été introduite section 1.2.1). Très tôt, Allwood [1976] a distingué la fonction expressive (qui donne des informations sur le locuteur et le contexte) de la fonction évocatrice (qui indique les changements escomptés sur l'interlocuteur ou sur l'environnement) des actes communicatifs. Le tableau 1.2 indique ces deux fonctions pour différents types d'actes communicatifs. Sadek [1991a] a ensuite introduit la notion d'effet rationnel pour rendre compte de cette notion d'effet perlocutoire attendu. C'est cette approche qui a été retenue par la suite pour fonder les sémantiques mentalistes des ACLs, nous y reviendrons lorsque nous présenterons les approches intentionnelles à la section 2.2.

1.4.3 Théorie contextuelle des actes de langage

La théorie contextuelle des actes de langage est une variante célèbre de la théorie des actes de langage qui considère qu'un acte de langage n'est jamais produit ni perçu de manière complètement isolée. Cette version contextuelle de la théorie des actes de langage est

issue d'une remise en cause de l'hypothèse de force littérale qui affirme (pour les actes de langage directs) que la force illocutoire d'un énoncé peut-être déduite de sa forme syntaxique ([Searle et Vanderveken, 1985] proposent d'ailleurs une classification des verbes performatifs de l'anglais qui indique pour chacun la force illocutoire dont il est l'expression). Bunt et Black [2000] indiquent que la force illocutoire d'un acte est essentiellement pragmatique, c'est-à-dire que son attribution relève essentiellement de facteurs contextuels. Par exemple, l'énoncé S_3 suivant sera interprété différemment selon que l'on se trouve dans une école, au commissariat de police, dans la rue avec des amis ou dans le bureau de son directeur :

S_3 *: Qu'avez-vous fait hier ?*

L'idée est donc de se désintéresser de l'énoncé de l'acte dans sa forme pour se focaliser sur le contexte dans lequel il est survenu et sur lequel il va agir en retour. Bunt [1996] définit un acte de langage comme l'application d'une fonction communicative à un contenu propositionnel. Cette fonction communicative spécifie quelle mise à jour du contexte doit être effectuée pour prendre en compte l'information transmise. L'idée est qu'un acte de langage est dans ses effets une fonction du contexte vers le contexte. Tout dépend donc de la définition que l'on donne à la notion de contexte. Pour Bunt, le contexte n'est que la somme des contextes cognitifs des interlocuteurs, ce qui nous amène aux approches intentionnelles que nous présenterons au chapitre 2 (voir section 2.2).

Pour d'autres [Stalnaker, 1978; Levinson, 1979], il s'agit d'exclure les états mentaux des interlocuteurs du contexte. Le contexte est alors directement lié à l'activité dialogique. Il est constitué de ce qui a été accepté lors du dialogue. Par exemple, un acte assertif de contenu propositionnel p ajoute l'engagement propositionnel correspondant au contexte. D'autres actes permettent de retirer certains éléments du contexte. On se dirige alors vers les approches dites conventionnelles (voir section 2.3). Finalement, d'autres éléments influant sur la production des actes de langages relèvent du contexte. Le contexte incluant entre autres :

- *certaines conventions et hypothèses* : on parle en français et pas en anglais, on parle travail et on ne joue pas aux charades, ...
- *certains gestes et inflections de voix* : froncer les sourcils, montrer du doigt ou faire des mimiques expressives, ...
- *des éléments pertinents de l'histoire de la conversation* : pour fixer les références pronominales, ...
- *des faits ambiants pertinents* : « on se voit dans une heure. », signifie qu'il faut avoir une idée de l'heure qu'il est au moment de l'énonciation, ...

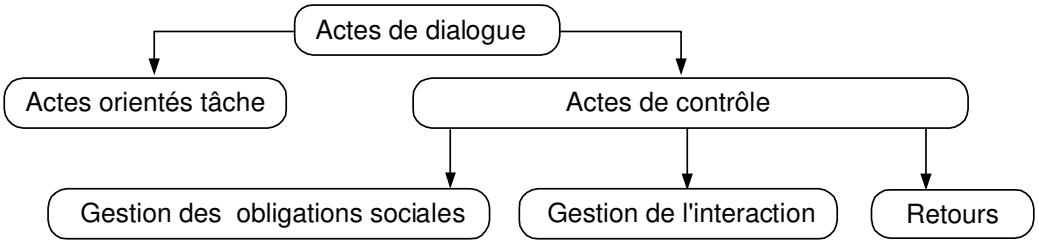

FIG. 1.3 – Typologie des actes de dialogue selon Bunt [2000].

Dans tous les cas, ces aspects contextuels ne sont pas pris en charge par la version analytique, Searlienne, de la théorie des actes de langage[13].

1.4.4 Les actes de dialogue et les actes multi-niveaux

Bunt [1996] a proposé une théorie des *actes de dialogue*. Selon Bunt et les partisans de l'approche contextuelle, on distingue les actes qui vont de l'avant (nommés « forward-looking » et qui sont plutôt initiatifs) des actes qui suivent (traduit de « backward-looking » et qui sont plutôt réactifs). De la même manière, les actes qui font progresser vers l'accomplissement de la tâche sous-jacente (TO-acts, pour « task-oriented acts ») sont distingués des actes de contrôle du dialogue (DC-acts, pour « dialogue control acts »). Ces derniers sont de trois types : les signaux de retour (*feedbacks*), les actes des gestions des obligations sociales et les actes de gestion de l'interaction (tour de parole, structuration du dialogue, contact entre les interlocuteurs). La figure 1.3 présente cette typologie des *actes de dialogue*. L'un des problèmes soulevés par cette approche est que l'étude des signaux de retour montre qu'ils sont généralement ambigus, essentiellement à cause des différents niveaux de l'interaction. En particulier, ces différents actes de dialogue ne sont pas mutuellement exclusifs comme le suggère la typologie de Bunt. Ils interviennent parallèlement sur les différents niveaux du dialogue (voir section 1.3.2). L'approche par les actes multi-niveaux, proposée par Traum et Hinkelman, est une tentative de prise en compte de cette structure en couches.

Traum et Hinkelman [1992] ont proposé des *actes de conversation multi-niveaux*. On y retrouve des actes de langage noyaux qui sont les actes de langage classiques, mais qui ne prennent leur plein effet qu'après établissement[14]. Les actes noyaux sont considérés comme des actions communes qui sont ajoutées au fond commun plutôt que comme de simples ac-

[13] Encore une fois, ses défenseurs, diront que ce n'est pas son objectif et repousseront cet aspect vers la théorie de l'action ou la pragmatique.

[14] Terme traduit de l'anglais *grounding*, la notion d'établissement désigne le processus par lequel une information est mise en commun. Cette mise en commun est généralement modélisée comme un processus en deux

Niveau de discours	Type d'actes	Exemples
SUB-UU	tour de parole	prendre-tour, garder-tour, donner-tour, assigner-tour
UU	établissement	initier, continuer, valider, réparer, demande-réparer, demande-valider, accepter, rejeter
Unité de discours (DU)	actes-noyau	informer, demander, s'engager
Multiple DUs	argumentation	élaborer, résumer, clarifier, question-réponse, convaincre, trouver un plan

TAB. 1.3 – Actes conversationnels multi-niveaux selon Traum et Poesio [1997].

tions du locuteur. Cette théorie des actes conversationnels suppose en outre trois autres types d'actes de langage : les actes de tour de parole, les actes d'établissement et les actes d'argumentation, plus complexes, car ils mettent en jeu plus d'un acte noyau. Les actes de conversation multi-niveaux sont récapitulés dans le tableau 1.3. Ils sont répartis à différents niveaux du discours : sous-énonciation (SUB-UU), énonciation (UU), unité de discours (DU) et discours (Multiple DUs). Cette approche des actes multi-niveaux réifie donc certaines des idées de Clark concernant les niveaux de la communication (section 1.3.2). Le cadre formel de Traum et Poesio [1997], issu des travaux sur le projet TRAINS de l'université de Rochester entend unifier la théorie des actes de langage et la DRT [Discours Representation Theory]. Nous ne présenterons pas cette approche plus avant puisqu'elle considère le langage naturel dans toute sa complexité et propose un traitement linguistique complet dont nous n'avons pas besoin dans le cadre des communications agents.

1.5 Conclusion

En conclusion de ces généralités et en gardant à l'esprit les réflexions de l'introduction de ce chapitre, nous pouvons poser quelques objectifs communs aux modèles de dialogues que nous allons considérer dans la suite de cet état de l'art. En effet, nous savons maintenant que les modèles considérés doivent fournir les outils par lesquels des agents artificiels vont dialoguer. Cette interaction ne peut pas, dans le cas général, être réduite à un simple message (one-shot communication) mais doit bien être constituée de séquences cohérentes de messages sur un même sujet ou au service d'une tâche ou d'un but commun.

étapes : présentation de l'information par le locuteur et acceptation (ou rejet, ou demande de clarification, ...) par l'interlocuteur. Nous reviendrons plus longuement sur cette notion d'établissement à la section 2.3.3.

En outre, en se basant sur les apports de la théorie des actes de langage, ces envois de messages ne sont pas vus comme une simple transmission d'informations, mais comme des actions agissant sur le monde (c'est-à-dire les états mentaux des interlocuteurs, leurs attitudes sociales, le contexte ou l'état du dialogue, selon les modèles). Le chapitre suivant poursuit cet état de l'art en synthétisant les propositions faites jusqu'alors pour la modélisation des dialogues entre agents.

Chapitre 2

Dialogues entre agents : des approches intentionnelles aux approches conventionnelles.

2.1 Introduction

Les propositions de modèles de structuration des dialogues entre agents faites jusqu'alors peuvent être regroupées en deux familles : les approches intentionnelles, parfois qualifiées d'approches mentalistes et les approches conventionnelles et sociales. Nous présenterons d'abord *les approches intentionnelles* (section 2.2), qui sont les plus anciennes et pour lesquelles les enchaînements d'énoncés résultent des intentions et plus généralement des états mentaux des interlocuteurs supposés sincères et coopératifs.

Nous présenterons ensuite *les approches conventionnelles et sociales* (section 2.3) qui sont l'hybridation des approches conventionnelles et des approches sociales. Les approches conventionnelles mettent l'accent sur les contraintes structurelles liées aux dialogues, que les interlocuteurs sont socialement poussés à observer du fait de conventions sociales attachées à l'usage du langage. Les approches sociales, quant à elles, s'articulent autour de l'utilisation de primitives sociales, telles que les engagements sociaux, pour capturer la sémantique des conversations. Comme la tendance actuelle est l'utilisation cumulée de ces deux types d'approches, nous les présentons ensembles en les regroupant sous l'appellation d'approches conventionnelles et sociales.

Pour chacune de ces familles, nous insisterons sur les fondements théoriques, philosophiques, sur lesquels elles reposent (sous-sections 2.2.1 pour les approches intentionnelles, 2.3.1 et 2.3.3 pour les approches conventionnelles et sociales, respectivement). Pour chacune, nous présenterons les contributions saillantes pour ce qui est de l'intelligence artificielle en général et pour les systèmes multi-agents en particulier (sous-sections 2.2.2 et 2.2.3 pour les approches intentionnelles, sous-sections 2.3.4, 2.3.5 et 2.3.6 pour les approches conventionnelles, les approches sociales et les approches conventionnelles et sociales, respectivement). Finalement, nous discuterons les avantages (sous-sections 2.2.5 pour les approches intentionnelles et 2.3.7 pour les approches conventionnelles et sociales) et inconvénients (sous-sections 2.2.4 et 2.2.6 pour les approches intentionnelles et 2.3.8 pour les approches conventionnelles et sociales) qui sont attachés à chacune de ces familles avant de conclure cet état de l'art (section 2.4).

Ce chapitre, comme le précédent, vise à fournir les pré-requis nécessaires à : (1) la compréhension de notre problématique (présentée au chapitre 3) qui viendra conclure cette partie et (2) une meilleure compréhension de nos apports quand à cette problématique tels qu'ils sont présentés dans la seconde partie de ce document.

2.2 Approches intentionnelles

2.2.1 Fondements philosophiques

L'intention individuelle

L'intention individuelle[1] est une notion très étudiée en philosophie de l'esprit. Les travaux les plus influents en intelligence artificielle sont ceux de Searle [1983] et Bratman [1987]. On distingue classiquement : l'intention dirigée vers le futur (avoir l'intention de faire quelque chose) et l'action intentionnelle (faire quelque chose intentionnellement). Ces deux notions sont liées puisque l'intention dirigée vers le futur mène généralement à la réalisation d'actions intentionnelles. L'intention qui nous intéresse ici est l'intention dirigée vers le futur.

Il faut se garder de confondre désir et intention. Les désirs représentent des états souhaités du monde, ils peuvent être irréalisables ou bien contradictoires. Le processus qui sélectionne, sur la base des croyances, les désirs qui pourront être poursuivis est *la délibération*. La dé-

[1] Par le qualificatif « individuelle », on souhaite exclure de notre champ d'investigation les travaux sur les intentions partagées et les intentions collectives. On pense à la notion de We-intention telle que développée par Searle [1990] ou par Tuomela et Miller [1988].

libération concerne la production des intentions ou buts (les liens entre les notions de but et d'intention soulèvent de nombreux débats) et a été beaucoup étudiée en intelligence artificielle. Les travaux de références à ce niveau sont ceux de Rao et Georgeff [1991] qui ont fourni le cadre intentionnel classique nommé BDI [Belief, Desire, Intention] et le modèle de Cohen et Levesque [1990a].

Dans la délibération de sens commun (également appelée raisonnement pratique), les buts peuvent être interprétés comme des désirs « choisis » suite à une délibération. En outre, les buts doivent être consistants. Finalement, les intentions peuvent être comprises comme des engagements individuels sur des buts, pour lesquels des ressources ont été allouées.

Une des caractéristiques principales de l'intention individuelle est donc qu'elle implique un type spécial d'engagement à l'action [Bratman, 1990; von Wright, 1980]. C'est-à-dire que tant que l'agent a l'intention d'atteindre un certain état, il est engagé individuellement à réaliser les actions qu'il pense appropriées pour atteindre cet état. Classiquement, on distingue deux types d'intention, *l'intention de* et *l'intention que*, selon que l'argument est une (ou plusieurs) action ou une proposition [Bell, 1995; Grosz et Kraus, 1996]. L'*intention que* s'applique à une proposition (éventuellement complexe) tandis que l'*intention de* est une intention de réaliser une action ou l'intention qu'une action soit réalisée (éventuellement par un autre agent ou un groupe d'agents).

Cette notion d'intention individuelle est centrale dans la définition de la signification non-naturelle qui est à la source des approches dites « intentionnnelles » de la communication dialogique.

Signification non-naturelle

En linguistique, la notion de signification a longtemps été considérée comme le seul fait des énoncés[2]. Dans la tradition « codique » de la communication, le locuteur « code » un sens dans un énoncé qu'il transmet à un auditeur qui le « décode ». Or, le sens littéral de l'énoncé ne représente généralement pas la totalité de sa signification. La thèse de Grice [1957], père de la vision inférentielle de la communication, est qu'il existe un rapport entre le « vouloir dire » du locuteur et ses intentions. C'est ce qu'il appelle la *signification non-naturelle* en opposition à la relation naturelle qui existe entre signe et signification. La formulation de Levinson [1983] est la suivante. Le locuteur A a voulu dire z en exprimant e, si et seulement si :

[2] L'annexe A présente en détail les différents niveaux de sens tels que considérés en linguistique.

1. A a l'intention que e provoque l'effet z chez B ;
2. A a l'intention que l'intention dont il est question en 1 soit réalisée par B en reconnaissant celle-ci (cette seconde intention est aussi appelée *intention communicative*).

L'intention est donc placée doublement au cœur du processus communicatif. On considère, d'une part, l'intention qu'a le locuteur en produisant l'énoncé et d'autre part, la reconnaissance de cette intention par l'interlocuteur. C'est cette reconnaissance d'intention qui dans un contexte coopératif va permettre à l'interlocuteur de réagir correctement. Lorsque, par exemple, je demande si p, (1) j'ai l'intention de savoir si p et (2) je souhaite que ceci soit réalisé par la reconnaissance de cette intention. C'est-à-dire que je souhaite qu'ayant reconnu mon intention première[3], l'auditeur réagisse coopérativement en me donnant, si possible et conformément à ses croyances, la réponse à ma question.

Coopération

Grice [1975] a le premier défini la notion de coopération dans le cadre du dialogue. Des interlocuteurs rationnels et coopératifs, désireux de maximiser l'échange d'information lors d'un dialogue devront se conformer au *principe de coopération* :

« Que votre contribution conversationnelle corresponde à ce qui est exigé de vous, au stade atteint par celle-ci, par le but ou la direction acceptée de l'échange dans lequel vous êtes engagé. », [Grice, 1975, p.61].

Ce principe peut se décliner sous la forme de quatre maximes de coopération :

– *maxime de qualité* : ne pas dire ce que vous croyez faux ou ce pourquoi vous manquez de preuve ;
– *maxime de quantité* : faites que votre contribution soit aussi informative que le but de l'échange le requiert, mais pas plus ;
– *maxime de pertinence* : soyez pertinent ;
– *maxime de manière* : soyez bref et ordonné, évitez les ambiguïtés.

[3] Dans la littérature, les situations dans lesquelles l'intention (le but) du locuteur n'a pas pour vocation d'être reconnue sont appelées « non manifestes » (*covert*). C'est le cas, par exemple, lorsque le locuteur essaie d'impressionner l'interlocuteur.

Ces maximes ne doivent pas être comprises comme strictement normatives. En effet, c'est en se basant sur le fait qu'elles sont régulièrement transgressées que peut être amorcé un processus inférentiel (appelé implicature) qui va permettre de retrouver le véritable sens des énoncés. Lorsqu'une (ou plusieurs) maxime est violée, les interlocuteurs réagissent coopérativement (le principe de coopération s'applique toujours) et essaient de comprendre/interpréter la violation : ce sont *les implicatures*. Les interlocuteurs cherchent à savoir ce qui est impliqué par cette violation apparente et manifeste.

Lorsque, par exemple, on dit : « Il pleut des cordes. », on viole la maxime de qualité, mais les autres comprennent bien que l'on parle métaphoriquement. C'est un exemple d'implicature conversationnelle. Les actes indirects sont un autre type d'implicature. Si on me demande : « Combien ça t'a coûté ? », et que je réponds : « Assez cher. ». Un interlocuteur comprendra, en supposant que je reste coopératif, que ce viol de la maxime de quantité est à interpréter comme signifiant : « Ce ne sont pas tes oignons ». Souvent, les violations sont indiquées, anticipées explicitement. Par exemple, on dit : « J'exagère peut-être un peu, mais… » pour prévenir d'un viol de la maxime de quantité, ou encore : « Tu vois ce que je veux dire… » pour signifier que l'on a violé la maxime de manière.

Lorsque le locuteur ne suit pas le principe de coopération, mais qu'il fait en sorte que les autres croient qu'il le suit, les violations peuvent servir à tromper les autres. Pour mentir, on viole discrètement la maxime de qualité, pour obscurcir ou embrouiller, on viole la maxime de quantité et pour distraire, on viole la maxime de pertinence.

Une des approches néo-gricéenne célèbre est la théorie de la pertinence qui ne conserve que la maxime de pertinence pour en faire un principe : *le principe de présomption optimale de pertinence*. Dans cette théorie, due à Sperber et Wilson [1986], la pertinence d'un énoncé est évaluée en fonction du rapport entre les effets qu'il produit et les efforts que la production de ces effets demande. Les éléments qui permettent d'évaluer cette pertinence sont malheureusement difficiles à déterminer.

Ajoutons, pour clore cette introduction aux travaux sur le concept de coopération, qu'il ne faut pas confondre la coopération dialogique et la coopération comportementale. La première est nécessaire pour assurer la cohérence conversationnelle du dialogue alors que la seconde est une hypothèse sur le comportement extra-linguistique de l'agent [Airenti et al., 1993].

2.2.2 Modélisation de la structure intentionnelle du dialogue

L'idée qui sous-tend toutes les approches intentionnelles est la suivante : la structure du dialogue n'est qu'un épiphénomène qui résulte des intentions (et éventuellement de la coopérativité) des interlocuteurs. Cette idée a été généralisée par Grosz et Sidner [1986] pour qui les trois éléments constitutifs de la structure du dialogue sont :

- *la structure linguistique* : ce sont les énoncés d'un dialogue agrégés en segments de dialogue (DS, Dialogue Segment) ; plus éventuellement des relations d'emboîtement entre eux ;
- *la structure intentionnelle* : il est possible de distinguer pour chaque segment de dialogue un but propre (DSP, Discourse Segment Purpose), sous-but du but global du dialogue. Les DSPs peuvent être liés entre eux par deux types de liens : (1) les liens de domination et (2) les liens de satisfaction ;
- *l'état attentionnel* : c'est ce sur quoi se focalise dynamiquement l'attention à un moment du dialogue. Cet état est généralement modélisé par une pile.

Selon Grosz et Sidner, la structure linguistique traduit la structure intentionnelle. En particulier, les relations d'emboîtement de la structure linguistique sont le reflet de cette structure intentionnelle. Ces auteures précisent toutefois : « la structure intentionnelle n'est pas isomorphe à la structure de la tâche sous-jacente ».

De nombreuses recherches ont été effectuées dans l'espoir de formaliser cette structure intentionnelle. Le domaine de l'IA concerné par la coordination d'actions en vue d'atteindre un but est *la planification*. C'est donc une approche par la planification qui s'est naturellement imposée comme première approche pour formaliser la structure intentionnelle.

Approches par la planification classique

Cohen et Perrault [1979] et Allen et Perrault [1980] ont développé les premiers systèmes utilisant explicitement la planification pour le dialogue, inspirés des travaux de Bruce [1975] qui fut le premier à utiliser les actes de langage pour des travaux d'IA sur les actions et les plans. Dans ces systèmes, les intervenants sont dotés d'états mentaux ($Want$ pour l'intention, Bel pour la croyance, $Know$ pour la connaissance, c'est-à-dire la croyance justifiée et MB pour la croyance mutuelle). Les actes de langage sont représentés comme des actions

Entête	$Inform(A, B, p)$
Préconditions	$Know(A, p)$
Corps	$MB(A, B, Want(A, Know(B, p)$
Effets	$Know(B, Know(A, p))$ et $Know(B, p)$

TAB. 2.1 – Caractéristiques de l'action $inform$.

X	Y
Action	Effet
Précondition	Action
Corps	Action

TAB. 2.2 – Éléments pour l'inférence de plans.

quelconques (c'est-à-dire en termes de préconditions, de corps et d'effets) qui affectent les croyances des interlocuteurs, comme l'illustre le tableau 2.1 pour l'opérateur $inform$.

Le système de planification utilisé est une version modifiée du STRIPS de Fikes et Nilsson [1971]. L'hypothèse centrale est que les actes de langage sont planifiés au même titre que les autres actions. Les plans sont reconstruits par l'interlocuteur à partir de la connaissance des opérateurs et de règles de construction de plan de la forme :

$Want(A, X) \Rightarrow Want(A, Y)$ où X et Y sont donnés par le tableau 2.2.

On serait donc tenté de dire que la reconnaissance d'intention implique simplement la mise en oeuvre de règles permettant à l'interlocuteur, sur la base de ces règles de construction, d'inférer le plan du locuteur, comme avec la règle suivante :

$Bel(B, Want(A, X)) \Rightarrow Bel(B, Want(A, Y))$ où X et Y sont de nouveau donnés par le tableau 2.2.

C'est vrai, mais ce n'est pas si simple, car il convient de faire une distinction entre deux types de reconnaissance de plans :

– *la reconnaissance à l'insu* : en inversant les règles d'élaboration de plan, l'auditeur peut retrouver le plan du locuteur. Cette reconnaissance est indépendante de l'intention communicative du locuteur. C'est un aspect très important de la communication

puisque c'est ce qui permet de devancer le locuteur en donnant des réponses pleinement coopératives.

– *la reconnaissance d'intention communicative* : la reconnaissance à l'insu n'est pas suffisante pour reconnaître le plan que le locuteur souhaite transmettre. Celle-ci nécessite un niveau d'imbrication supplémentaire :
$Bel(B, Want(A, Bel(B, Want(A, X))))$. Cette reconnaissance est nécessaire pour prendre en compte les actes de langage indirects.

Dans ce cadre, le principe de coopération prend la forme de l'*adoption de but* : après avoir inféré le but du locuteur et son plan pour le réaliser, un auditeur coopératif donnera l'information manquante du plan de manière à ce que le locuteur puisse réaliser son but.

Cette approche d'une grande importance historique souffre d'au moins deux problèmes :

– on se cantonne à l'analyse d'un seul énoncé (une paire d'adjacence si on considère la réponse donnée). Or, la reconnaissance de plan est souvent le résultat de véritables échanges (reconnaissance incrémentale [Carberry, 1990])

– avec cette approche, la structure du dialogue est calquée sur la structure de la tâche sous-jacente. De nombreux dialogues ne respectent pas cet isomorphisme. C'est notamment le cas lorsqu'un sous-dialogue de clarification est entamé (il n'est généralement pas prévu dans la tâche, ...).

Cette seconde difficulté a donné lieu à la différenciation explicite par Litman et Allen [1990] de deux types de plans :

– les *plans de domaine* modélisent les tâches extra-linguistiques sous-jacentes au dialogue ;

– les *plans de discours* sont des méta-plans qui permettent de manipuler la structure d'autres plans.

Il y a trois types de relations entre les plans du discours et ceux du domaine : (1) relation de continuation (permet de commencer l'exécution du plan de domaine ou d'en poursuivre le déroulement), (2) relation de clarification (identifier un paramètre, proposer une correction du plan courant, ...) et (3) relation de changement de sujet (permet d'introduire un nouveau

plan de domaine). Sur la base de ces trois relations, cinq méta-plans ont été proposés par Litman et Allen [1990] : `suivre_plan`, `identifier_paramètre`, `corriger_plan`, `introduire_plan`, `modifier_plan`.

La structure globale du dialogue est modélisée par une pile de plans du domaine et de méta-plans, elle-même composée d'autant de sous-piles que le plan de domaine a de pas ou d'étapes. Carberry [1990] précise que ce modèle, bien que restrictif, semble assez conforme à la majorité des comportements dialogaux observés.

Pourtant, il peut arriver que les participants au dialogue possèdent plusieurs solutions alternatives pour un même problème, et qu'en conséquence, ils souhaitent évaluer les différents plans à leur disposition. Pour ce faire, le modèle de Ramshaw [1989], repris depuis par Lambert et Carberry [1991], propose donc un troisième niveau (intermédiaire) de résolution de problème (ou élaboration de plan) qui permet aux interlocuteurs de considérer plusieurs possibilités de plans leur permettant de réaliser les buts du niveau domaine.

Approches des plans par les attitudes mentales

Une autre approche de la planification préconise la traduction des plans en termes d'états mentaux plutôt qu'en termes de structures de données. En effet, les plans vus comme de simples structures de données posent des problèmes :

- il y a redondance des liens de cause à effet, car la même action représentée différemment évoquera l'une ou l'autre ;
- la notion de pré-condition est ambiguë : est-ce que l'action est déclenchée lorsque les pré-conditions sont remplies ou est-ce que cela rend juste l'action possible ?

Ainsi, pour Pollack [1990], il existe deux types de relations entre les actions : « rendre possible » et « générer ». Il existe une relation « rendre possible » entre deux actions si la réalisation de la première permet la réalisation de la seconde. La relation « générer » est plus forte puisqu'elle indique qu'une action sera réalisée par la réalisation d'une autre.

En outre, Pollack distingue les *recettes* des plans. Formellement, une recette est une suite d'actions pour réaliser une action plus complexe. Comme certaines des actions de la recette peuvent être des actions complexes, une recette complète se représente sous la forme d'une arborescence appelée graphe de recettes (Rgraph). Dans ce cadre, un agent A a un plan P pour réaliser une action b en effectuant une recette R si (cette définition simplifiée ne tient pas compte des relations « rendre possible ») :

1. A croit qu'il peut réaliser chaque action de R ;
2. A veut réaliser chaque action de R ;
3. A croit que réaliser R provoquera la réalisation de b ;
4. A veut exécuter R comme moyen de réaliser b ;
5. A croit que chaque action de R joue un rôle dans le plan P ;
6. A veut que chaque action de R joue un rôle dans le plan P.

Dans le système de Pollack, des règles d'inférence de plan permettent à l'interlocuteur de se représenter le plan du locuteur en termes de ses attitudes mentales. Ainsi, ce modèle permet de traiter les dialogues au cours desquels le plan du locuteur est reconnu comme étant invalide. C'est le cas dans le dialogue suivant :

1.A : Je dois aller à Paris. Puis-je avoir un billet pour le train 77 ?
2.B : Le train 77 ne va pas à Paris. Vous feriez mieux de prendre le train 88.

Où B a pu attribuer les états mentaux suivant à A et ainsi inférer que la troisième des croyances attribuées à A est erronée avant de générer sa réponse :

1. $Bel(B, Bel(A, Exec(aller\_Paris)))$
2. $Bel(B, Bel(A, Exec(prendre\_train\_77)))$
3. $Bel(B, Bel(A, Generation(prendre\_train\_77, aller\_Paris)))$
4. $Bel(B, Int(A, (aller\_Paris))$
5. $Bel(B, Bel(A, Int(prendre\_train\_77)))$
6. $Bel(B, Bel(A, Int(en(prendre\_train\_77, aller\_Paris)))$

Plus récemment, les formalismes de plans ont été adaptés de manière à traduire la nature intrinsèquement collective et collaborative de certaines activités. Grosz et Sidner [1990] ont ainsi proposé la notion de *plan partagé* en s'appuyant sur les travaux de Pollack. Un plan partagé est une collection d'attitudes mentales faisant intervenir des désirs et des croyances des différents protagonistes (les travaux de ces auteurs se sont limités à deux interlocuteurs). Suite à de nombreuses critiques, cette notion a été reprise et améliorée par Grosz et Kraus [1996] pour obtenir la structure récursive de plan partagé suivante. Un groupe d'agents G a le plan partagé de réaliser a en utilisant la recette R_a si et seulement si :

1. Les membres du groupe G ont la croyance mutuelle des actions et des contraintes qui composent la recette R_a ;

2. Pour toute action individuelle, il existe un agent de G tel que tous les agents de G croient mutuellement que :
   (a) cet agent a l'intention de réaliser cette action ;
   (b) cet agent est capable de réaliser cette action en utilisant une certaine recette ;
   (c) cet agent a un plan individuel pour réaliser cette action ;
   (d) le groupe G est engagé sur la réussite de cet agent pour l'action considérée.

3. Pour les actions non-individuelles, il existe un sous-groupe de G tel que les membres de G croient mutuellement que :
   (a) ce sous-groupe a un plan partagé pour réaliser cette action en utilisant une certaine recette ;
   (b) ce sous-groupe est capable de réaliser cette action en utilisant cette recette ;
   (c) le groupe G est engagé sur la réussite de ce sous-groupe concernant cette recette.

Cette définition requiert donc que les protagonistes s'accordent sur une recette, une distribution des tâches et l'engagement à la réussite des autres. Notons que cette approche ne nécessite pas d'intention collective, car la structure du plan capture la dimension collective. Aussi, si cette représentation permet de justifier les comportements coopératifs au cours de l'exécution du plan, il reste toutefois à déterminer :

1. comment les individus identifient-ils le besoin de collaborer à propos d'une action ?

2. comment les groupes se forment-ils pour réaliser l'action en question ?

3. comment le plan partagé est-il élaboré ?

4. comment le plan partagé est-il exécuté ?

Ces quatre étapes sont ce que Wooldridge et Jennings [1994] appellent le processus de résolution coopérative de problème. Comme le dialogue est une activité dynamique, les agents ne peuvent pas planifier de manière définitive et complète. C'est pourquoi la notion de plan partiel à été introduite.

L'application de la théorie des plans partagés au dialogue à été étudiée par Lochbaum [1994]. Selon celle-ci, si les interlocuteurs sont engagés dans un dialogue, ils ont une bonne raison pour cela : c'est soit qu'ils ne peuvent pas réaliser un certain plan seul, soit que leur plan est incomplet en l'état, ... Autrement dit, les interlocuteurs dialoguent pour compléter

des plans partiels. Un plan partiel est un plan pour lequel, les agents, soit (1) n'ont qu'une recette ou un graphe de recette partiel ; soit (2) une ou plusieurs actions n'ont été attribuées à aucun agent. Pour Lochbaum, il existe deux relations entre plans (individuels ou partagés, partiels ou pas) :

- *contribution* : un plan contribue à l'établissement de l'une des croyances ou intentions requises pour le second plan. En d'autres termes, l'exécution du premier plan contribue à celle du second.
- *pré-satisfaction* : un plan doit être complété avant l'autre. En d'autres termes, l'exécution du premier plan est une condition pour l'exécution du second.

Il est ainsi possible d'analyser les sous-dialogues comme l'exécution de plans contributoires à des sous-tâches. En outre, les sous-dialogues de correction et de clarification (qui avaient menés Litman et Allen à distinguer les plans du discours des plans du domaine) visent à établir des pré-conditions de croyances liées à un plan, essentiellement pour identifier des paramètres ou des recettes à utiliser.

Selon Lochbaum, les plans partagés modélisent donc exactement les buts et sous-buts du dialogue de Grosz et Sidner (DSP, introduits section 2.2.2) en un seul et même formalisme qui suffit pour prendre en compte de multiples types de dialogues.

Approches de l'interaction rationnelle basées sur la logique

Ces approches, connues sous le nom de *théories de l'interaction rationnelle*, sont une autre application des théories philosophiques et informatiques présentées ci-dessus. Elles ont, sur les approches par planification, l'avantage d'introduire des sémantiques clairement définies. Elles se basent généralement sur des logiques modales épistémiques[4] pour les attitudes mentales et sur la logique dynamique pour la représentation des actions. Elles trouvent leur principale application dans les sémantiques des ACLs (voir section 2.2.3).

Cohen et Levesque [1990c] ont bâti une théorie de l'interaction rationnelle sur leur théorie de l'action [Cohen et Levesque, 1990a]. Dans leur théorie de l'action, l'intention n'est pas

[4] Plus précisément, la croyance $BEL(x,p)$ est généralement définie grace au système $S5$ faible, le but individuel $GOAL(x,p)$ avec le système K et la connaissance $KNOW(x,p)$ est définie comme la croyance vraie. Le temp est exprimé à l'aide d'une logique temporelle linéaire et les opérateurs standard $HAPPEN$, $DONE$ et $UNTIL$ sont introduits.

une primitive, mais est exprimée en termes de but persistant. Un agent A a le but persistant G relatif à la motivation M (noté $PGoal(A, G, M)$[5]) , si et seulement si :

1. A croit que G n'est pas vrai actuellement ;
2. A veut que G soit réalisé ;
3. l'item précédent (2) restera vrai tant que : (i) G n'est pas vrai, (ii) G est jugé atteignable et (iii) la motivation M de l'atteindre est présente.

L'introduction de la motivation M permet d'éviter que tous les buts ne soient irrévocables, c'est-à-dire poursuivis fanatiquement. Le but est un engagement individuel qui ne peut être annulé sans raison. Dans ce cadre, un agent a l'intention de réaliser un but G s'il a le but persistant d'avoir réalisé G, et de l'avoir réalisé intentionnellement (c'est-à-dire, non accidentellement).

L'idée principale de leur théorie de l'interaction rationnelle est que les propriétés des actes illocutoires peuvent être dérivées des seuls états mentaux des interlocuteurs. Autrement dit, il n'est plus nécessaire de supposer que le locuteur a l'intention que l'auditeur reconnaisse son intention de réaliser un acte illocutoire particulier, puisque les effets liés à l'acte reconnu seront dérivés directement de celui-ci. La reconnaissance de l'acte ne sert qu'à inférer l'effet fondamental à partir duquel les autres conclusions seront déduites. Il ne s'agit en aucun cas de la reconnaissance de la force illocutoire au sens de Searle et Vanderveken. Le problème est alors de déterminer quels sont les effets fondamentaux des actes. Techniquement, pour Cohen et Levesque, ces effets sont contextualisés et décrits à l'aide d'axiomes alors que pour Sadek [1991a], ils sont divisés en effets perlocutoires attendus et effets fondamentaux (qui traduisent la préservation des pré-conditions de faisabilité). C'est cette dernière version de la théorie qui a été appliquée aux sémantiques des ACLs telles que décrites dans la section suivante.

Par ailleurs, dans cette approche, sincérité et bénévolat (la disposition d'un agent à adopter les buts d'autrui si ceux-ci ne contredisent pas les siens) sont deux caractéristiques sur lesquelles repose le comportement coopératif des agents. Ces notions dépassent le cadre du comportement rationnel puisque rien a priori ne contraint un agent rationnel à être un tant soit peu coopératif. Elles sont donc capturées par des hypothèses spécifiques, notamment l'hypothèse de sincérité.

[5] Défini par : $PGoal(x, p, q) \triangleq (BEL(x, \neg p)) \wedge (GOAL(x, \Diamond p)) \wedge (UNTIL[(BEL(x, p)) \vee (BEL(x, \Box\neg p)) \vee (BEL(x, \neg q))](GOAL(x, \Diamond p)))$.

Pour adapter leur théorie au cadre des actions collectives, Cohen et Levesque [1991] ont tenté de définir ce que peut être une intention conjointe dans le cadre de l'interaction rationnelle. La solution retenue est intermédiaire entre la version minimale où chaque agent a la croyance mutuelle que chacun possède l'intention que l'action collective soit réalisée et la croyance mutuelle que chacun ait l'intention de tenir son rôle tant que les autres le font et la version maximale qui exige en plus que cette croyance mutuelle persiste jusqu'à ce qu'il soit mutuellement acquis que l'activité soit réalisée, irréalisable ou plus motivée. La première solution n'offre aucune garantie que la croyance mutuelle persiste alors que la seconde ne permet pas aux agents de porter un jugement personnel (et privé) sur le statut du but à atteindre. La notion de but affaibli (weak achievement goal) a été introduite pour contraindre les agents à avertir les autres d'un éventuel changement de leur attitude vis-à-vis de l'objectif commun. Un agent A a le but affaibli de réaliser p relativement à la motivation M et à un groupe d'agents G[6] si et seulement si :

- soit p est un but « normal » pour A ;
- soit si p est déjà atteint (ou inatteignable, ou n'est plus motivé), alors l'agent a le but que ce statut soit mutuellement connu de tous les membres du groupe.

La notion de but affaibli mutuel (notée $WMG(x, y, p, q)$) tient lorsque les agents ont la croyance mutuelle d'avoir les mêmes buts affaiblis. La notion de but persistant conjoint est alors introduite et définie de la manière suivante. Un groupe d'agents T a le but persistant conjoint G, si et seulement si :

1. il est mutuellement cru par les membres de T que G n'est pas vrai actuellement ;
2. il est mutuellement cru par les membres de T que chacun a le but que G soit réalisé[7] ;
3. il est vrai et mutuellement cru que tant qu'il n'est pas mutuellement admis que G est déjà atteint, inatteignable ou plus motivé, il sera toujours mutuellement cru que chaque agent de T possède G comme but affaibli.

[6] noté $WAG(x, y, p, q)$ et défini comme suit : $WAG(x, y, p, q) \triangleq [\neg BEL(x, p) \wedge GOAL(x, \Diamond p)] \vee [BEL(x, p) \wedge GOAL(x, \Diamond MB(x, y, p))] \vee [BEL(x, \Box\neg p) \wedge GOAL(x, \Diamond MB(x, y, \Box\neg p))] \vee [BEL(x, \neg q) \wedge GOAL(x, \Diamond MB(x, y, \neg q))]$, où MB tient pour la croyance mutuelle et est défini par : $MB(x, y, p) \triangleq BMB(x, y, p) \wedge BMB(y, x, p)$ avec BMB, la croyance mutuelle unilatérale définie par : $BMB(x, y, p) \triangleq BEL(x, p \wedge BMB(y, x, p))$.

[7] C'est-à-dire que les agents ont le but mutuel que G, noté $MG(x, y, p)$, défini par : $MG(x, y, p) \triangleq MB(x, y, GOAL(x, \Diamond p) \wedge GOAL(y, \Diamond p))$

Comme dans le cas de l'intention individuelle, l'intention conjointe peut alors être définie comme le but persistant conjoint d'avoir réalisé une action, avec la croyance mutuelle que cette action a été réalisée intentionnellement (consciemment et volontairement). Les actes de langage sont alors définis comme des tentatives. On retrouve ici, comme chez Lochbaum, l'idée que la communication est expliquée par son rôle de contrôle dans une activité collective. La communication provient rationnellement de la nécessité de maintenir les croyances associées à l'intention conjointe.

2.2.3 Sémantiques mentalistes des ACLs

Les sémantiques mentalistes des ACLs (on les nomme ainsi du fait de leur référence systématique aux états mentaux) s'inspirent directement des approches intentionnelles logiques discutées ci-dessus. Comme l'indiquent Chaib-draa et Vanderveken [1998], fournir une sémantique formelle aux actes de langage va permettre d'analyser rigoureusement l'utilisation des communications aussi bien dans les SMAs que dans les sociétés humaines. Avant de parler de la sémantique des ACLs, rappelons que l'on envisage généralement la sémantique des langages formels à l'aide de pré-conditions indiquant à partir de quel état l'action peut être entreprise et de post-conditions indiquant ses effets.

Sémantique de FIPA-ACL

Comme nous l'avons introduit section 1.2.2, les deux ACLs principaux KQML et FIPA-ACL sont basés sur la théorie des actes de langage : les messages sont considérés comme des actions avec leurs conséquences sur l'environnement (en l'occurrence les états mentaux des agents). Des sémantiques mentalistes ont été définies pour ces actes.

La sémantique de FIPA-ACL [Finin et al., 1999] est basée sur celle d'ARCOL [Sadek, 1991b]. Elle définit des pré-conditions de faisabilité des messages et des post-conditions décrivant les effets rationnels attendus. Le langage sémantique de FIPA-ACL (SL) est une logique multimodale avec des opérateurs pour les croyances (B), les désirs (D), les croyances incertaines (U), les intentions (ou but persistant PG). Basée sur les travaux de Cohen et Levesque (voir section 2.2.2) sa forme actuelle est due à Sadek [1991a]. Un acte est sélectionné lorsque les conditions de faisabilité sont remplies et que l'effet rationnel correspond à une intention du locuteur.

Cette sémantique est donc basée sur une spécification logique des notions mentalistes telles que les croyances et les intentions. On a vu que les communications y sont envisagées

	$\texttt{tell}_{A,B}(X)$	**proactive-tell**
$Pre(A)$	$Bel_{\text{A}}(X) \wedge$ $Know_{\text{A}}(Want_{\text{B}}(Know_{\text{B}}(S)))$	$Bel_{\text{A}}(X)$
$Pre(B)$	$Int_{\text{B}}(Know_{\text{B}}(S))$ avec $S = Bel_{\text{B}}(X)$ ou $S = \neg Bel_{\text{B}}(X)$	
$Post(A)$	$Know_{\text{A}}(Know_{\text{B}}(Bel_{\text{A}}(X)))$	$Know_{\text{A}}(Know_{\text{B}}(Bel_{\text{A}}(X)))$
$Post(B)$	$Know_{\text{B}}(Bel_{\text{A}}(X))$	$Know_{\text{B}}(Bel_{\text{A}}(X))$
Complétude	$Know_{\text{B}}(Bel_{\text{A}}(X))$	$Know_{\text{B}}(Bel_{\text{A}}(X))$

TAB. 2.3 – Sémantique des performatifs KQML `tell` et `proactive-tell`.

comme un type d'action. Cela signifie, en pratique, que la logique multimodale sur laquelle repose la théorie sémantique de FIPA-ACL doit être combinée à une théorie de l'action. Le formalisme résultant est complexe. Aussi, pour tirer profit de la sémantique de FIPA-ACL les agents doivent implémenter ces théories. Cela constitue une contrainte majeure qui est un problème puisqu'on ne connait pas de telles architectures. En fait, le fossé entre ces logiques théoriques puissantes et leurs implémentations est encore immense.

Sémantiques de KQML

Contrairement à la sémantique de FIPA-ACL, KQML ne présupposait pas originellement que les agents implantent des architectures logiques complexes. La sémantique informelle attachée aux performatifs impliquait simplement que l'agent devait pouvoir manipuler une base de connaissances virtuelle (ajouter/extraire des assertions). Ces faibles contraintes ont laissé une grande liberté aux concepteurs et c'est pourquoi différentes sémantiques virent le jour [Cohen et Levesque, 1990b]. Labrou et Finin [1997a,b], notamment, ont fourni une sémantique pour chaque performatif en termes de pré-/post-conditions pour le locuteur et l'interlocuteur avec une condition de complétude en plus.

Les pré-conditions indiquent les états mentaux qu'un agent doit nécessairement posséder pour pouvoir utiliser un performatif et pour que le récepteur puisse l'accepter. Les post-conditions indiquent les états mentaux de l'émetteur après un énoncé réussi d'un performatif et ceux du récepteur après la réception du message. Finalement, la condition de complétude indique les états mentaux qui correspondent à la satisfaction de l'intention qui motive l'échange.

Par exemple (voir tableau 2.3), la pré-condition de `tell` pour le locuteur (A) impose que A croit ce qu'il dit (X) et qu'il sache que le récepteur (B) veut/désire savoir quoi croire

sur X. La pré-condition de `tell` pour B est qu'il doit vouloir savoir quoi croire sur X. La post-condition d'un message `tell` pour B est qu'il peut conclure que A croit X. La condition de complétude permet de s'assurer que le performatif a réussi dans le contexte de communication de A et B. En l'occurrence, il s'agit de la post-condition de B. Certains auteurs ont néanmoins préféré redéfinir une version moins contraignante de ce performatif appelée `proactive-tell` dont la spécification sémantique est indiquée sur la droite du tableau 2.3.

La sémantique proposée par Labrou [1996] est basée sur une logique multi-modale sophistiquée qui contraint les agents à être conçus selon une architecture logique compatible avec ce formalisme, ce qui nous ramène au cas FIPA-ACL. On pourrait appliquer la sémantique présentée par Chaib-draa et Vanderveken [1998] à KQML comme proposé dans leur article, mais il en serait de même. Par contre, cette dernière prend en compte le degré d'intensité des forces illocutoires et certaines facettes des conditions préparatoires qui ont été laissées de côté par Labrou et Finin.

2.2.4 Limitations des sémantiques mentalistes des ACLs

L'idée d'un cadre de communication standard est donc réifiée par les ACLs. Inspirés de la représentation formelle de l'énoncé en langage naturel proposée par la théorie des actes de langage et complétés d'une sémantique mentaliste telle que celles décrites précédemment, les ACLs forment un cadre interactionnel complet. Ce sont les sémantiques des ACLs qui dans l'esprit des approches intentionnelles sont censées assurer le passage de l'énoncé au dialogue. En effet, le raisonnement des agents sur la sémantique mentaliste des actes de langage, sous les hypothèses de coopération et de sincérité, est censé assurer l'émergence de conversations.

Pour autant, comme le soulignent Chaib-draa et Dignum [2002], les problèmes soulevés par les ACLs tels qu'ils existent actuellement sont nombreux. Cela va des problèmes d'ontologie aux problèmes de complétude des ACLs. Avant de présenter (sections 2.2.5 et 2.2.6) les critiques générales des approches intentionnelles logiques qui s'appliquent ici, concentrons-nous sur les problèmes concernant spécifiquement la sémantique et la pragmatique des ACLs actuels.

Problèmes concernant la sémantique des ACLs

Problème de la minimalité sémantique : les systèmes de pré/post-conditions habituellement utilisés dans les ACLs permettent de rendre compte du sens minimal des messages.

Malheureusement, il est des situations où on a besoin d'un sens plus précis, spécifique au contexte. C'est un problème général de la théorie sémantique des ACLs. D'un côté, on veut que la sémantique soit suffisamment générique pour rendre compte de toutes les situations d'utilisation des ACLs. De l'autre, les systèmes de pré/post-conditions obtenus sont trop généraux et abstraits pour être adéquats à toutes les situations.

Problème de conformité sémantique : les ACLs sont si génériques et explicites que leur pouvoir d'expression est très grand, mais les sémantiques bien définies qu'ils acceptent sont basées sur des logiques tellement puissantes que le calcul de la signification d'un message arbitraire par un agent demande tellement de déductions qu'il est typiquement formellement intractable [Dignum et Greaves, 2000]. Pourtant, l'usage d'un ACL avec une sémantique complète, facilement extensible, est un énorme atout pour des SMAs hétérogènes et ouverts. Malheureusement, il est très difficile de vérifier si les agents sont dans des états mentaux qui vérifient les pré/post-conditions car en fait, même si le problème est formellement bien défini, le niveau calculatoire ne suit pas.

Problème de l'alignement des sémantiques : l'alignement de la sémantique d'un ACL sur le modèle cognitif et comportemental de l'agent pose problème quand ce dernier permet l'expression d'actes de communication qui ne sont pas sémantiquement définis dans l'ACL utilisé. Ce type de problème peut advenir lorsque les contraintes (pourtant nécessaires) de la sémantique des ACLs sont trop fortes pour l'agent. Cela arrive par exemple avec les conditions de sincérité. La plupart des sémantiques des ACLs (en particulier celles de FIPA-ACL et de KQML) ne permettent pas aux agents d'affirmer quelque chose qu'eux-mêmes ne croient pas. C'est une hypothèse simplificatrice pour les ACLs qui découle de l'analogie avec l'humain, qui ne communique généralement pas en supposant que son interlocuteur ment. Les théories comportementales sophistiquées permettent à l'agent d'agir avec l'intention de tromper (manipulation, mensonge, ...) si cela l'aide à atteindre ses buts[8]. La condition de sincérité le met donc dans l'impossibilité d'exprimer ce qu'il désire exprimer. Notons que d'autres hypothèses simplificatrices (comme la joignabilité sûre qui stipule que les agents reçoivent correctement tous les messages qui leur sont envoyés, ...) pour les ACLs peuvent poser des problèmes avec les modèles comportementaux des agents.

Problèmes liés à la distribution et à l'autonomie des agents : les sémantiques des ACLs doivent tenir compte du fait que les agents sont distribués et autonomes (idéalement). Pour un programme ordinaire, les post-conditions peuvent être précisément calculées, car son contexte est accessible et les différentes actions/instructions ne sont pas autonomes les unes des autres (programmation séquentielle). Cette distinction pose pour les ACLs une barrière entre l'effet espéré d'un acte de langage et son effet réel. Quand un agent A transmet l'in-

[8] Cas fréquents en commerce électronique et plus généralement en économie lors de stratégies de maximisation de gain.

formation X à B, il a l'intention que B va au moins croire que A croit X. Mais comme les agents sont autonomes, un agent ne peut jamais changer directement les croyances d'un autre et l'effet d'un acte de langage n'est jamais garanti. Autrement dit, les agents n'ont pas le contrôle des effets perlocutoires de leurs actes de langage et de leurs communications en général.

S'il n'y a pas de « consensus » autour de la forme syntaxique que les ACLs peuvent prendre, il n'y en a pas non plus pour ce qui est de leurs sémantiques [Dignum et Greaves, 2000]. Il n'y a pas de modèles clairs et calculables de la sémantique des actes de langage et moins encore de la sémantique des conversations dans les SMAs. Les communications agents souffrent donc de l'absence d'une sémantique concise, tractable et universellement acceptée. Comme l'indiquent Kone et al. [2000], cela confine la communication agent à des environnements restreints et cela rend difficile (voir impossible dans certains cas) la communication entre des agents hétérogènes, c'est-à-dire développés par des concepteurs différents.

Problèmes concernant la pragmatique

Problème de prise en compte du contexte social : le contexte social contraint les actions (y compris les actes de langage) que les agents peuvent entreprendre par le biais d'obligations, de normes et d'engagements de toutes sortes. Ce type de contraintes n'est pas pris en compte par les théories sémantiques actuelles des ACLs.

Problème d'expression de la pragmatique : le comportement communicationnel d'un agent est le résultat implicite ou explicite de sa planification. Par exemple, une question est posée en attente d'une réponse et cette réponse est censée participer à la progression de l'agent vers ses buts ou l'achèvement d'une tâche. De même, un service est demandé dans le but que l'autre agisse en conséquence. Idéalement, cela devrait faire partie des pré-conditions de ce type d'actes de langage. Ces pré-conditions de type pragmatique sont pourtant très difficiles à exprimer avec les théories sémantiques actuelles des ACLs.

2.2.5 Avantages et applications des approches intentionnelles

Le principal avantage des approches intentionnelles est leur complétude (informelle). En effet, les approches intentionnelles couvrent trois des composantes majeures du traitement de la communication : la syntaxe est couverte par les ACLs qui découlent de la théorie des actes de langage ; la sémantique des ACLs est bien définie en termes des attitudes mentales des agents et la pragmatique est gérée par un système de planification collective ou par une

théorie de l'interaction rationnelle (complétée par une forte hypothèse de coopération). Cependant, les modèles issus des approches intentionnelles sont complexes à implanter. Bien qu'envisagées spécifiquement pour les agents artificiels via les sémantiques mentalistes développées pour les ACLs, ce sont les approches intentionnelles basées sur les plans qui ont été appliquées aux IHM [Interfaces Homme-Machine], dans un cadre mono-agent.

Par exemple, le modèle développé par Balkanski et Hurault-Plantet [2000] utilise la reconnaissance et l'élaboration de plans pour comprendre et accomplir des actes communicatifs dans le contexte orienté-tâche bien délimité d'un répondeur téléphonique qui redirige les usagers vers le poste téléphonique souhaité. Il existe des systèmes comparables également basés sur les plans comme le système TRAINS [Allen et al., 1995] dû à Allen et son équipe ou encore le raisonneur de plans de domaine de Ferguson [1995]. D'autres sont plus axés sur le raisonnement et utilisent les plans via les approches logiques comme le Circuit Fixit System de Smith et al. [1995] et le système Artimis développé par l'équipe de France Télécom [Sadek et al., 1997]. Toutes ces applications se heurtent aux limites imposées par la planification (voir section suivante). Les approches logiques qui dépassent ces limitations, sont quant à elles trop lourdes et complexes (en termes de complexité informatique) pour les agents et pour leurs concepteurs de sorte qu'on ne leur connaît pas d'applications réelles dans le cas multi-agents.

Aussi, si on compte de nombreux systèmes utilisant les ACLs (en particulier KQML[9] et FIPA-ACL[10]), le raisonnement sur la sémantique n'y est pas implanté et celle-ci sert juste à donner un sens précis (pour le développeur) aux différents actes de langage disponibles. En ce sens cela aide tout de même les équipes de conception dans la réalisation de SMAs qui utilisent les aspects syntaxiques de ces langages, même si le traitement sémantique et le traitement des aspects de contenu est réalisé de manière ad hoc.

2.2.6 Limites des approches intentionnelles

L'approche intentionnelle logique propose une théorie complète de la communication. Cependant, sa version théorique générale est trop complexe pour être implantée telle quelle. Les modèles mentaux impliqués sont généralement exprimés dans des logiques multi-modales

[9] Le site de l'Université du Mariland (UMBC) consacré à KQML et maintenu par Finin, l'un des investigateurs de KQML, propose une liste d'exemples de systèmes utilisant les aspects syntaxiques de KQML : `http ://www.cs.umbc.edu/kqml/`.

[10] Dès lors que de nombreuses plateformes de développement agent se conforment aux standard FIPA, de nombreux développeurs de SMAs utilisent le standard FIPA-ACL pour les communications entre agents. On pourra consulter une liste de ces plateformes sur le site de la FIPA : `http ://www.fipa.org/resources/`.

dont l'implémentation est encore un sujet de recherche. Les simplifications consenties pour arriver à une implantation font perdre une partie de la puissance de ces modèles, notamment leur sémantique. C'est ainsi que, malgré la chronologie, les approches intentionnelles basées sur la planification peuvent être vues comme une simplification des approches logiques dans le cadre computationnel plus facilement implémentable de la planification. Ce faisant, les domaines d'application ont été limités aux domaines orientés tâche, laissant de côté les autres types de dialogues. Les approches intentionnelles basées sur les plans ont été appliquées et donc implémentées avec succès. Cependant, comme le souligne Cohen [1996], il reste des limites théoriques inhérentes aux modèles basés sur les plans, en particulier :

- *la dépendance au domaine* : les modèles basés sur les plans reposent sur la bonne définition des recettes qui doivent prévoir toutes les possibilités, c'est-à-dire couvrir tout le domaine. Cela limite d'autant la portée de ce type de modèle.
- *le manque de bases théoriques* : même si les modèles basés sur la planification ont permis de nombreuses avancées dans la compréhension des conversations, il leur manque des fondements théoriques solides. En ce sens, ce sont des modèles à court terme qui sont efficaces du fait de leur nature procédurale, mais ils ne constituent pas une théorie du dialogue acceptable pour les sciences cognitives.

À cela, viennent s'ajouter des limitations communes à toutes les approches intentionnelles :

- *la complexité des inférences* : les algorithmes de reconnaissance de plan sur lesquels reposent les approches intentionnelles par planification sont combinatoirement intractables dans le pire des cas, et indécidables dans certains cas [Bylander, 1991]. Pour ce qui est des approches logiques, Maudet et Chaib-draa [2002] indiquent que la sémantique des actes communicationnels est tellement riche qu'il est trop complexe de déterminer les réponses possibles en inférant les états mentaux des autres agents.
- *la sémantique des messages* : comme le rappellent Maudet et Chaib-draa [2002], dans les approches intentionnelles logiques, la sémantique des messages est formulée en termes d'états mentaux, aspects privés aux agents. Cela pose le *problème de la vérification sémantique* : pour que la sémantique des messages soit vérifiable, il faudrait avoir accès aux états mentaux privés des agents, ce qui n'est généralement pas possible[11].

[11] La vérification sémantique dont il s'agit ici ne doit pas être confondue avec la vérification formelle que les agents implémentent correctement une sémantique mathématique particulière. La vérification dont il est question ici est plutôt la vérification que peut entreprendre un agent sur un autre pour s'assurer qu'il agit de manière cohérente vis-à-vis des dialogues tenus.

Le second problème majeur posé par cette formulation est *le problème de l'hypothèse de sincérité* nécessaire à la définition d'une telle sémantique. Cette hypothèse est jugée trop contraignante par la communauté SMA [Dignum et Greaves, 2000]. Elle interdit notamment d'envisager correctement certains types de dialogues dans des domaines où une telle hypothèse ne saurait tenir, comme c'est le cas par exemple pour les dialogues de négociation dans le commerce électronique.

– *l'hypothèse de coopération* : les approches intentionnelles reposent toutes sur une hypothèse de coopération. Cette hypothèse a d'abord été formulée comme adoption de but (voir section 2.2.2) puis de manière moins contraignante comme une participation attendue à l'action conjointe (dans les théories de l'interaction rationnelle, voir section 2.2.2). L'idée selon laquelle la communication émerge rationnellement de l'action collective sous-jacente est fondamentale dans les approches intentionnelles. Si cette idée est élégante, elle reste néanmoins liée à la considération d'une activité collaborative. En conséquence, il n'est pas possible d'expliquer par ces approches, ce qui motive la communication lors de situations non-collaboratives [Traum, 1994].

Finalement, le rôle central de l'intention dans toutes ces approches peut aussi être questionné du point de vue de son universalité et ces approches sont parfois qualifiées d'ethnocentriques [Nuyts, 1994]. En effet, l'analyse de conversation montre bien qu'une grande partie des dialogues comprennent des phases ritualisées (ouverture, fermeture, remerciements, ...) qui sont sans rapport avec la reconnaissance d'intention et qui ne sont pas prises en compte par ces approches. On peut donc se demander si ces approches sont pleinement satisfaisantes pour la modélisation du dialogue et plus généralement : les attitudes mentales considérées sont-elles pertinentes ? Les notions classiques de croyances, désirs et intentions ne traduisent nullement l'idée d'engagement social ou d'obligation pourtant cruciale pour considérer l'action collective [Traum, 1994]. Les approches intentionnelles n'indiquent pas comment caractériser cette intuition de « liant collectif » qui explique que les interlocuteurs peuvent généralement compter sur certaines actions les uns des autres.

2.3 Approches conventionnelles et sociales

Les approches conventionnelles et sociales sont apparues en réaction aux approches mentalistes détaillées dans la section précédente. Développées à partir de fondements théoriques différents et complémentaires (section 2.3.1), les approches conventionnelles ont donné lieu à trois courants de recherches actifs : les protocoles de communication (section 2.3.2) ou approches strictement conventionnelles, les approches strictement sociales (section 2.3.4) et les approches hybrides conventionnelles et sociales comme les jeux de dialogues (section 2.3.6).

Cette section présente ces trois types d'approches à partir de leurs fondations théoriques (sections 2.3.1 et 2.3.3).

2.3.1 Fondements philosophiques des approches conventionnelles

Conventions, normes sociales et règles

Une *convention* est une régularité qui existe au sein d'une communauté (sans pour autant avoir nécessairement fait l'objet d'accords explicites). Les *normes sociales* sont généralement définies comme des conventions auxquelles sont associées des obligations sociales et des sanctions et dont le maintien ne peut pas être réduit à des considérations d'ordre rationnel. Nous ne distinguerons pas les deux notions.

S'il est habituel d'exprimer normes et conventions sous forme de règles, la formalisation de ces notions est un domaine de recherche en lui-même. Pour ce qui est des conventions dans le dialogue, Allwood [1994] constate trois types de régularités :

- celles qui dépendent des relations au sein d'un énoncé entre ses différentes parties ;
- celles qui dépendent des relations entre énoncés ;
- celles qui dépendent des relations entre les facteurs globaux et les énoncés.

C'est sur le second type de régularité dans le cours d'un dialogue que les approches conventionnelles se concentrent. L'idée de ces approches est donc que la *cohérence conversationnelle* [Craig, 1983] peut être garantie par les seules conventions sans hypothèses de coopération ni d'intentionnalité. Le problème n'est pas de savoir comment ces régularités sont établies, on suppose les interlocuteurs prêts à s'y conformer. Il s'agit alors de limiter les formes possibles d'expression, de manière à ce que l'intention véhiculée soit non ambiguë et qu'il ne soit plus nécessaire d'effectuer une analyse pragmatique complexe en vue de la reconnaître. Parmi ces modèles, on trouve les protocoles de communications (voir section 2.3.2) et plus récemment les approches par jeux de dialogues que nous détaillerons après avoir introduit un certain nombre d'autres notions fondatrices issues de la dialectique formelle (section 2.3.3 et suivantes).

Engagements communs sur des projets communs

Pour Searle [1992b], toute conversation est l'expression d'une intention conjointe (*We-intention*). Ce type d'intention transcende la conjonction des états intentionnels individuels. Cette intentionnalité partagée est un facteur d'explication important pour tous les comportements sociaux. Searle considère cette intention collective comme un élément de base qui ne se réduit pas à la somme des intentions individuelles (*I-intention*) et de leurs connaissances mutuelles. Néanmoins, des intentions individuelles sont incluses dans la *We-intention*. Dans ce cadre, les conversations sont des formes de *We-intention*. Par exemple, la *We-intention* « nous parlons du prix de X » inclut la *I-intention* (pourtant très différente) « j'offre 5$ pour X » et la *I-intention* « je refuse ton offre ».

Clark [1996], lui, considère l'usage social du langage comme un *type d'activité commune.* Il utilise le concept de *type d'activité* de Levinson [1979], qui est une notion plus générale que la notion de type de discours, car elle inclut des événements sociaux autres que le discours ou bien dans lesquels le discours n'a qu'une place incidente. Clark introduit la notion de projet conjoint. Un *projet conjoint* est une action commune proposée/suggérée par un des participants et acceptée/réalisée par tous. Un *projet conjoint* pourrait être un plan, une recette ou une procédure pour accomplir une activité ensemble [Chaib-draa et Vongkasem, 2000]. Pour avoir des chances d'aboutir, un tel projet nécessite l'engagement de tous les participants. Ainsi, on peut dire que les *We-intentions* se manifestent par des *engagements conjoints* sur des *projets conjoints*.

Lors d'une conversation libre, l'*intention conjointe* doit émerger des interlocuteurs et de la conversation elle-même. Cette émergence est partie intégrante de la conversation et est l'objet d'évaluations et de négociations continues entre les interlocuteurs. Dans le cas des conversations orientées/contraintes, cette intention conjointe pré-existe souvent. En effet, les conversations contraintes ont un propos, un sujet ou un but auquel se raccrocher (comme l'indique le tableau 1.1 de la section 1.3.1). On peut voir les protocoles de communication utilisés avec les ACLs comme l'expression d'autant de projets conjoints. Chaque participant doit s'engager dans le projet conjoint, après quoi il s'agit d'une intention conjointe.

Quand les agents prennent part à une *activité commune*, à l'exécution d'un plan commun ou à un *projet conjoint*, ils accomplissent des actions conjointes. Beaucoup de ces actions conjointes (ou de leurs parties) sont des actes de communication nécessaires au bon accomplissement de l'action (se faire comprendre, tester la compréhension des autres, divulguer une information, ...). Cela nous amène à considérer les actes de langage sous une nouvelle perspective. En effet, on ne peut plus étudier les actes de langage sous la seule perspective du locuteur. Il nous faut étudier *pour le locuteur comme pour l'interlocuteur* : le rôle de l'acte dans la conversation, les liens de l'acte aux autres actes de la conversation. Il convient

d'étudier aussi, la façon dont le locuteur génère l'acte dans le contexte et la manière dont l'interlocuteur peut comprendre ce que veut lui dire le locuteur et qui peut relever de ces liens plus que de l'acte lui-même.

2.3.2 Protocoles et politiques de conversation

L'idée de protocole est une réification de l'idée de projet conjoint conventionnel. Dans le champ de la communication dans les systèmes artificiels, les protocoles de communication pallient à un problème majeur des ACLs, hérité de la théorie des actes de langage. En effet, la théorie des actes de langage est une théorie de l'énoncé « isolé » ; or, la communication langagière donne lieu à des discours ou à des conversations qui sont des suites d'énoncés inter-dépendants (comme nous l'avons vu section 1.4.1). Chaque agent doit posséder une procédure de décision qui lui permet de choisir puis de générer des actes de langage en fonction de ses propres intentions. Il ne s'agit pas simplement de trouver l'acte de l'ACL dont la sémantique s'unifie avec les intentions de l'agent. Pour être pertinent, l'agent doit au minimum prendre en compte le contexte de son acte de langage (les événements en cours et passés, y compris les actes de langage précédents). Cela pose le problème du lien entre la définition de la sémantique d'une primitive d'ACL et la conversation à laquelle elle participe. D'un côté, il semble clair que le sens général d'une conversation (les engagements/promesses/informations qui y circulent) ne peut se passer du sens des performatifs qui la constituent. De l'autre, certains chercheurs pensent que la sémantique de la conversation elle-même doit être vue comme une primitive. Dans cette voie, les éléments de base de la sémantique des conversations doivent être sociaux plutôt qu'individuels pour être compatibles avec les théories et concepts présentés ci-dessus (voir section 2.3.1). Les activités communes, actions communes et intentions communes sont alors les éléments de base de ces sémantiques, au même titre que les croyances, désirs et intentions individuelles.

Actuellement, dans les SMAs concrètement développés, la prise en compte du contexte se fait de manière simplifiée par l'utilisation de conversations pré-planifiées, stéréotypées. Ces séquences d'échanges prédéfinies permettent de réduire considérablement l'espace de recherche pour la poursuite de la conversation tout en restant consistant avec la sémantique. Du fait de cet avantage computationnel, quasiment tous les SMAs utilisant un ACL sont dotés d'une couche « conversation », qu'elle soit standard ou ad hoc[12]. La spécification de

[12] Cette couche conversationnelle peut être considérée comme standard si les protocoles utilisés le sont. Les protocoles définis par la FIPA et attaché à FIPA-ACL (voir ci-dessous), le contract-net [Smith, 1977], le protocole pour la demande d'action de Winograd et Flores [1986] en sont des exemples. Cependant, du fait des besoins spécifiques des applications, les développeurs définissent eux-mêmes les protocoles ou enchaînements de messages que leurs agents utiliseront de manière systématique. Cette couche est donc généralement définie de manière ad hoc.

ces conversations se fait à l'aide de protocoles et de stratégies ou politiques de conversation (traduit de l'anglais : conversation policies).

Les protocoles de communication

Un *protocole* spécifie pour une tâche précise les actions ou réactions (communicatives dans notre cas) autorisées, souvent en nombre limité, en fonction de l'état de la conversation. Les protocoles spécifient des séquences d'actes communicatifs sans rien préciser du contenu des actes. S'ils ont donc l'avantage de simplifier le calcul des réponses possibles à un message donné, ils ont par contre le désavantage de contraindre la forme de la conversation de manière draconienne et ce malgré les efforts pour en favoriser la modularité et la réutilisabilité (voir [Vitteau et Huget, 2004] à ce propos).

Règles de conversation et protocoles de KQML et FIPA-ACL

KQML est muni d'une seule règle de conversation. Elle est simple, même si de nombreuses variantes sont permises. La conversation commence lorsque qu'un agent envoie un message KQML à un autre et se termine lorsque ce dernier répond. Les variantes sont obtenues par l'utilisation de performatifs de régulation de conversation. Comme leur nom l'indique, ces performatifs permettent aux agents d'intervenir dans le cours normal d'une conversation. Elles permettent notamment d'enrichir une conversation en prolongeant la règle par défaut (`standby`, `next`, `rest` ou `discard`) ou au contraire de mettre prématurément fin à une conversation (`eos`, `error` ou `sorry`). La thèse de Labrou [1996], présente un certain nombre de mini-conversations ainsi que leurs sémantiques et les contraintes qui lient les différents performatifs les composant. Ces mini-conversations sont conçues pour servir d'éléments de base pour des échanges plus importants.

FIPA-ACL, contrairement à KQML, fournit un certain nombre de protocoles de communication impliquant chacun plusieurs actes de communication (notés CA pour Communicative Act). Les protocoles retenus sont parmi les plus populaires : réseau de contrats (contract net [Smith, 1980]), demande d'action [Winograd et Flores, 1986], différents types d'enchères, ...

Problèmes concernant les protocoles

Malgré leur apparente simplicité, l'utilisation des protocoles soulève un certain nombre de questions. Même si les conversations peuvent être structurées par enchaînement de protocoles, ce type d'approche semble trop rigide à la majorité des chercheurs. Une conséquence de cette rigidité est le manque de généricité de ce type d'approche. Aussi, comme l'ont constaté Johnson et al. [2003], on assiste à une prolifération de protocoles dont nombre d'entre eux rendent compte des mêmes types de dialogue.

En outre, il existe plusieurs candidats non-équivalents pour la spécification de protocoles : les réseaux de transitions à états finis qui sont utilisés pour les protocoles de bas niveau de l'informatique distribué (TCP/IP, RPC, HTTP, ...), les réseaux de Pétri colorés [Cost et al., 1999], les arbres de sous-buts, les graphes de Dooley [Parunak, 1996] ainsi que d'autres approches fondées sur la logique [Endriss et al., 2004]. Ces formalismes offrent des degrés de souplesse des conversations envisageables très variables.

Finalement, les protocoles sont extrêmement contraignants et les messages non attendus dans le protocole ne seront pas examinés. Reste aussi à savoir comment des agents se mettent d'accord sur l'utilisation d'un protocole. Puis, comment les implémenter dans les systèmes multi-agents ? Doivent-ils faire partie de l'axiomatique de la communication ? Comment apprend-on de nouveaux protocoles ? Comment intégrer les protocoles au sein du fonctionnement de l'agent ?

Politiques de conversation

On regroupe sous le nom de *politiques de conversation* ou stratégies de conversation, les tentatives d'assouplir l'utilisation des protocoles en spécifiant à un plus haut niveau les interactions à utiliser en fonction du contexte. Ce domaine est relativement jeune et aucun consensus sur la définition de ces politiques de conversation n'a été atteint. Ces instructions de haut niveau permettent outre le choix de protocoles adaptés à la situation, la gestion des exceptions et problèmes lors de leur déroulement. Plusieurs modèles ont été proposés, qu'ils soient basés sur l'assemblage de micro-protocoles [Huget, 2001] ou d'échange question-réponse [Elio et al., 2000], sur un système de décomposition du sujet et du type de la conversation en actes de langage [Lin et al., 1999] ou sur la gestion des situations inattendues et des cas d'échec de protocoles [Philips et Link, 1999].

Nous n'approfondirons pas ici ces idées, car l'importante littérature concernant les stratégies de conversation et les protocoles nous fait penser que ces approches souffrent d'une

absence de consensus. En fait, les fondations théoriques de ces techniques que nous avons décrites aux sections 1.3.2 et 2.3.1 ne nous semble pas être suffisantes dans la mesure où en ne considérant que les aspects conventionnels du dialogue : sa dynamique, son émergence tout ce qui fait sa généricité et sa flexibilité se perdent. Aussi, si les fondations des approches conventionnelles sont importantes à prendre en considération, elles ne sont que partielles et doivent être complétées d'autres aspects. En effet, l'utilisation de protocoles revient à éluder la problématique de la modélisation des dialogues agents dans ce qu'elle a à voir avec le reste des sciences cognitives. Ce n'est donc pas un hasard, si la plupart de ces approches sont issues et se réclament de l'ingénierie. On parle d'ailleurs à juste titre d'ingénierie de protocoles, que ce soit pour les SMAs [Huget et Koning, 2003] ou plus généralement en informatique distribué [Holzmann, 1991]. Ces techniques sont de ce fait peu liées aux travaux sur les fondements que nous décrivons dans cet état de l'art.

C'est pourquoi de nombreuses recherches récentes considèrent une approche alternative aux protocoles et aux stratégies de conversation. Il s'agit d'une autre famille de modèles, qui sans être moins techniques sont plus théoriques et plus génériques. Nous les nommerons *approches conventionnelles et sociales*. En effet, si les protocoles sont limitatifs, les agents doivent néanmoins suivre des conventions sociales qui rendent possible la conversation. Tout en acceptant les éléments de fondations des approches strictement conventionnelles que nous venons de décrire, les approches conventionnelles et sociales s'attachent à prendre en compte le caractère social de la communication. Aussi, ces nouvelles approches sont généralement inspirées de la dialectique formelle (dont les concepts seront présentés en section 2.3.5) qui a introduit les notions d'engagements sociaux et de tableaux de conversation qui permettent de tenir compte de l'arrière-plan conversationnel et du fond commun envisagé comme liant social. Les fondements de ces approches conventionnelles et sociales sont présentés dans les sections suivantes.

2.3.3 Fondements des approches conventionnelles et sociales

Arrière-plan, fond commun et établissement

La notion d'*arrière-plan*[13] est la clé de voûte de la philosophie Searlienne [Searle, 1992a]. L'idée en est simple : les mots et les phrases ne suffisent pas en eux-mêmes à générer une interprétation. Le même sens linguistique admettra des interprétations différentes selon les

[13] Les termes arrière-plan, fond commun (parfois appelé terrain commun) et établissement sont les traductions « fragiles » de : background, common ground et grounding dont il est difficile de rendre exactement compte en français.

usages[14]. Par exemple, le verbe ouvrir s'interprète différemment dans « ouvrir les yeux », « ouvrir un restaurant », « ouvrir son cœur », « ouvrir une porte », « ouvrir le débat ». Pour Searle, tout ce qui est sens s'appuie sur un ensemble d'aptitudes, de dispositions et de capacités dites d'arrière-plan. L'arrière-plan ne fait pas partie du sens et pourtant le sens n'existe en tant que tel que par rapport à lui[15]. En première approximation, on réduit l'arrière-plan à ce qui est dit et communément accepté dans le cadre d'une conversation, c'est ce que l'on appelle le fond commun. À cet égard, la notion de croyance commune introduite par Clark et Schaeffer [1987] avec la notion d'*établissement* semble acceptable. Une information est d'abord présentée, comprise (retour de compréhension) puis acceptée : on dira alors qu'elle est socialement établie. Elle devient alors commune. Cette façon de voir évite le problème de la récursivité infinie des croyances mutuelles : je sais qu'il sait que je sais qu'il sait que je…

Ce qui rend néanmoins difficile la formalisation de ces notions dans le cas de la communication humaine, c'est la diversité des types de présentation (directe, co-construite, …) et d'acceptation (implicite par continuation ou explicite comme « OK » ou encore par une sous-conversation, …). Ainsi, il est souvent difficile de dire si un énoncé fait partie de la présentation ou de l'acceptation. Le caractère formel des communications agents lève cette difficulté.

Dans les approches conventionnelles et par opposition aux approches intentionnelles, l'emphase est mise sur l'aspect public de la conversation. Introduit pour rendre compte de la construction collective et publique des conversations, le *tableau de conversation* est un enregistrement (éventuellement structuré) de l'état de la conversation. Il fait partie du fond commun encore appelé arrière-plan conversationnel. Les débats pour savoir ce que doit contenir le tableau de conversation ne sont pas clos et on ne sait s'il doit se limiter à l'histoire de la conversation ou bien être plus général et capable de s'adapter de manière à rendre corrects les énoncés. L'implantation de celui-ci est aussi problématique. La question de savoir si ce tableau doit être centralisé ou bien si chaque agent doit en posséder sa propre version reste en suspend. Dans le second cas, on parlera d'*agenda*.

Les engagements sociaux

Le comportement des agents cognitifs conventionnels est basé sur l'état interne privé, c'est-à-dire les attitudes mentales de l'agent. Or, les agents sont habituellement définis en termes d'autonomie, d'inter-opérabilité et de leur capacité à atteindre ensemble des buts

[14] La notion d'arrière-plan inclut celle de contexte mais ne s'y réduit pas.

[15] Cette idée est déjà en germe chez les philosophes de l'école Gestalt. Notons en outre que Searle pose cet arrière-plan langagier comme infini et non représentable, ce qui l'amène à conclure l'existence d'une complexité irréductible de la conscience.

communs. Le point de vue mentaliste, centré sur l'agent-individu, n'aide pas à définir une sémantique pour tout ce qui est commun et partagé dans un groupe d'agents. Dans ce cadre, la communication agent souffre d'un manque de sémantique formelle concise et universellement acceptée (voir section 2.2.4), de sorte que la communication agent est restreinte à des domaines précis dans des environnements qui ne sont pas ouverts. Un certain nombre de solutions à ce problème ont été récemment proposées dans lesquelles le caractère social des agents est mis de l'avant [Singh, 1998; Colombetti, 1998; Moulin, 1997]. La notion d'*engagement social*, présente notamment en dialectique formelle [Walton et Krabbe, 1995] a alors rapidement émergée comme permettant de capturer un niveau public du dialogue.

La notion d'engagement social ne doit pas être confondue avec la notion d'engagement individuel évoquée ci-dessus (et utilisée, par exemple, par Cohen et Levesque [1990a] dans leur théorie de l'action rationnelle pour traduire la persistance liée à l'intention). Les engagements jouent un rôle central dans le dialogue, ce sont des entités complexes qui alimentent le tableau de conversation. Conceptuellement, les engagements sociaux capturent les obligations que contractent les agents les uns envers les autres [Castelfranchi, 1995]. En effet, les engagements sociaux sont orientés et indiquent les responsabilités d'un agent envers un autre[16] concernant une action à accomplir ou une proposition à maintenir [Singh et al., 1999]. Dans la littérature, on leur attache les caractéristiques suivantes (reprises, adaptées et complétées de [Maudet, 2001]) :

– *Les engagements sont sociaux* : ce sont des engagements vis-à-vis d'autres membres d'une communauté. La particularité principale des engagements sociaux est qu'ils doivent être socialement établis. Ce sont des cognitions partagées, communes. En respectant la propriété d'autonomie des agents, la seule manière pour un agent de déterminer qu'un autre agent partage une cognition (en l'occurrence la connaissance d'un engagement social) est l'observation des actes de communications qui doivent permettre d'établir de nouveaux engagements. Finalement, on nomme créditeur l'agent engagé et débiteurs celui ou ceux envers qui l'engagement est pris.

– *Les engagements sont publics* : les engagements sont publics et accessibles à tous[17]. Pour ce faire, ils sont stockés dans une structure de données commune ou accessible à tous. Dans les systèmes rencontrés jusqu'alors, ces structures sont de deux types, l'une est centralisée et l'autre répartie : tableau de conversation (commitment store) ou agendas personnels publics.

[16] Aucun des formalismes rencontrés ne considère l'engagement conjoint de plusieurs agents envers plusieurs autres. Il s'agit d'une première simplification de modélisation.

[17] Là encore, il s'agit d'une simplification de modélisation.

– *Les engagements sont propositionnels ou en action* : les engagements propositionnels concernent les éléments propositionnels (par exemple : l'énoncé « je crois que p » engage le locuteur envers ses interlocuteurs sur son contenu propositionnel p) alors que les engagements en action concernent les actions futures (par exemple : l'énoncé « demain, je ferais α » engage le locuteur envers ses interlocuteurs à réaliser l'action α).

– *Les engagements peuvent être directs ou conditionnels* : un engagement conditionnel est un engagement qui tient si certaines conditions sont remplies. Les engagements conditionnels sont à rapprocher de l'idée de règle, comme par exemple dans : si tu fais α, je ferais β. Cependant, les engagements conditionnels sont avant tout des engagements et il n'est pas nécessaire d'introduire une nouvelle primitive pour les engagements conditionnels, un simple connecteur suffit.

– *Les engagements sont dialogiques ou extra-dialogiques* : parmi les engagements pris au cours du dialogue, on distingue ceux qui concernent le dialogue en cours qui sont dit « dialogiques » de ceux qui se rapportent au contexte du dialogue (la tâche sous-jacente ou le sujet de dialogue). Par exemple, dans les cadres interactionnels conventionnels, lorsqu'une question est posée, l'interlocuteur est engagé à y répondre. Cet engagement dialogique, qui vise à assurer la cohérence structurale du dialogue et qui disparaîtra lorsque le dialogue sera terminé, ne doit pas être confondu avec l'engagement extra-dialogique qui peut découler de la réponse donnée et qui sera généralement persistant au-delà du dialogue.

– *Les engagements sont posés explicitement ou implicitement* : les engagements posés explicitement résultent d'actes (de langage ici) alors que les engagements posés implicitement correspondent à des habitudes qui sont incarnées par les règles du cadre interactionnel que les agents sont tenus de respecter ainsi que les conventions et normes en cours dans le système considéré. Les engagements dialogiques sont un exemple d'engagements posés implicitement. La possibilité d'un méta-dialogue, c'est-à-dire d'un dialogue sur le dialogue (notamment sur sa structure), indique que ces engagements posés implicitement sont traités de manière explicite par les agents. On ne peut donc pas parler d'engagements implicites mais juste d'engagements posés implicitement.

– *Les engagements sont datés* : un engagement est pris à un moment donné et dans un contexte donné. En outre, dans les engagements en action, un des paramètres de l'action doit indiquer à quel moment l'action doit être effectuée (cette indication est donnée sous la forme d'un instant ou d'un intervalle selon le type de gestion temporelle utilisé dans le système considéré).

– *Les engagements sont ordonnés* : certains engagements en action doivent être satisfaits avant d'autres. C'est un moyen classique pour structurer le tableau de conversation (ou les agendas) qui les contient que de les ordonner.

2.3.4 Approches des communications agents basées sur les engagements sociaux

Ces approches ont en commun de reposer sur la notion d'engagement social pour capturer la dimension publique de la communication. Nous présenterons brièvement le modèle d'agent social de Singh (section 2.3.4) et l'approche par la logique modale de Colombetti (section 2.3.4).

Le modèle d'agent social de Singh

Singh [1998] est le premier à avoir identifié le besoin d'un modèle sémantique formel des ACLs en termes de « notions sociales ». Il a proposé sa propre sémantique sociale pour ACL [Singh, 2000] comme une partie de sa théorie des agents sociaux. Plus précisément, Singh et ses collaborateurs ont introduit une sémantique basée sur l'engagement social intégrée à une logique modale temporelle avec temps ramifié (CTL [Computational Tree Logic]). En s'inspirant d'Habermas [1984], il a défini trois niveaux de sémantique, qui correspondent à trois assertions valides pour chaque acte de langage :

- *assertion objective* : la communication est vraie, c'est-à-dire que le locuteur s'engage sur son acte de langage. Par exemple, si le locuteur informe le groupe que p, alors il s'engage envers le groupe sur cette croyance.
- *assertion subjective* : la communication est sincère. Par exemple, si le locuteur informe le groupe que p, alors il s'engage envers le groupe sur sa sincérité (supposée).
- *assertion pratique* : la communication est justifiée. Par exemple, si le locuteur informe le groupe que p, alors il doit avoir des raisons de penser que p est vrai.

Singh encapsule donc l'approche mentaliste dans le niveau social. Il y a bien une différence entre le fait que le locuteur soit sincère (hypothèse de sincérité dans les approches intentionnelles) et dire qu'il est socialement engagé comme étant sincère. En outre, Singh et ses collaborateurs ont proposé de nombreux raffinements et applications de cette théorie générale. Les aspects temporels ont étés approfondis [Mallya et al., 2004] grâce à une extension de CTL (qui est initialement une théorie d'instants) avec des intervalles à la Allen [1983].

Par contre, cette approche strictement sociale ne tient pas compte des aspects conventionnels des communications. En effet, si elle permet de connaître *a posteriori* le résultat

d'une composition de communication, elle n'indique pas quelles sont les compositions pertinentes comme le fait un protocole. C'est pourquoi Chopra et Singh [2004] reconsidèrent à nouveau les protocoles en proposant les machines à engagement non-monotones (Non-monotonic Commitments Machines) reposant sur la logique causale non monotone (NCL). Ce cadre logique, en particulier la présence d'une sémantique formelle basée sur la notion d'engagement social, permet de définir des protocoles plus flexibles qu'avec les machines à état fini (FSM) généralement utilisées. Cette flexibilité accrue est conséquente de (1) la présence d'une sémantique formelle qui permet aux agents de raisonner sur le protocole utilisé ainsi que de (2) la non-monotonicité du système proposé qui permet de reconsidérer (lorsque le raisonnement le suggère) des états de faits passés.

Approche par la logique modale de Colombetti

Colombetti explore l'idée d'états mentaux sociaux à l'aide de la logique modale. Il distingue différents mécanismes par lesquels les agents pourraient acquérir des croyances communes : déduction, information montrée, observation mutuelle, communication intentionnelle, ... Sur la base de ces mécanismes, Colombetti [2000] a proposé le langage ALBA-TROSS [Agent Language Based on the TReatment of Social Semantics] qui – dans la même veine que les travaux de Singh – donne une sémantique sociale (exprimée dans la logique temporelle CTL*) aux actes illocutoires courants. Une des particularités de ce langage est la notion de pré-engagement, qui est un type d'engagement conditionnel. Par exemple, une requête pré-engage l'interlocuteur à qui elle s'adresse, signifiant que si celui-ci accepte, il sera engagé à agir en conséquence.

Plus récemment, une spécification opérationnelle des primitives de communication dans une terminologie adaptée à la programmation orientée objet [Fornara et Colombetti, 2002] ainsi qu'un cadre logique révisé [Verdicchio et Colombetti, 2004] ont été proposés. Cependant, et à l'instar de celle de Singh, cette approche purement sociale ne tient pas compte des aspects conventionnels des communications. Comme aucune architecture d'agent susceptible de manipuler les primitives définies n'a été produite, cette approche a plutôt été appliquée à la conception de protocoles [Fornara et Colombetti, 2003].

En conclusion, les approches strictement sociales, que ce soit celle Singh ou Colombetti, ont semblé à certains insuffisantes pour rendre compte de manière pratique de l'émergence des conversations entre agents de sorte que d'autres approches hybrides, mêlant les aspects conventionnels et sociaux ont été proposées[18]. La section suivante (section 2.3.5) présente les fondements théoriques pour cette autre avenue ainsi que les différentes propositions faites

[18] Les derniers développements de l'approche de Colombetti et de sa collègue Fornara [Fornara et al., 2005] visent à établir un lien entre la communication agent et la réalité institutionnelle dans laquelle ces communica-

dans ce sens. Les approches basées sur l'idée de jeux de dialogue (section 2.3.6) seront ensuite présentées.

2.3.5 Les systèmes dialectiques

De son étude des arguments formellement fallacieux, Hamblin [1970] déduit que certains arguments sont inadéquats sans être formellement non valides, c'est pourquoi la logique formelle n'est pas adaptée pour rendre compte de l'argumentation. C'est la naissance de la logique informelle (ou logique dialectique). Un système dialectique est un système normatif de régulation du dialogue[19] considéré comme un jeu entre les participants. Un tel système est constitué de :

- *un ensemble de coups* : idéalement, les différents coups ou locutions définies couvrent les différents types d'actes de langage. On trouve par exemple des coups pour l'assertion, la question, le défi, ...
- *une liste d'engagements par participant* : Hamblin suppose l'existence d'une liste d'engagements qui permet notamment de gérer la cohérence du dialogue. Pour autant, il faut qu'un joueur soit capable d'anticiper/détecter les inconsistances dans les listes d'engagements. Afin de ne pas doter les agents de capacités irréalistes (comme l'omniscience logique) Hamblin isole un ensemble de schémas d'axiomes inconsistants que le joueur essaye d'instancier dans le tableau d'engagement pour détecter les différents types d'inconsistance correspondants.
- *un ensemble de règles de mise à jour des engagements* : ces règles spécifient l'effet de chacun des coups sur les listes d'engagements ;
- *un ensemble de règles de dialogue* : les règles du dialogue régissent la structuration globale du dialogue et interdisent de jouer certains coups. Ces règles indiquent quels sont les coups licite dans une situation donnée.

Notons que les systèmes dialectiques ne traitent pas de la production des arguments, ils ne font pas le lien entre les états mentaux de l'agent et son activité dialogique. Ils se contentent de représenter les énoncés et d'en vérifier l'acceptabilité. La sous-section suivante présente brièvement un exemple de système dialectique.

tions prennent place. Une institution y étant définie de manière générique comme un ensemble de conventions, ces derniers développements reviennent à prendre en compte la dimension conventionnel.

[19] Initialement, Hamblin ne traitait que les dialogues de persuasion.

Un exemple de système dialectique

Le célèbre système DC dû à MacKenzie [1979] modélise les interactions de type argumentatives entre deux opposants qui soutiennent des thèses contradictoires. Le système DC permet, entre autres, de prévenir les esquives de questions comme dans le petit dialogue suivant :

1.A : Je l'ai peint en rouge.

2.B : Pourquoi ?

3.A : Parce que je ne l'ai pas peint en bleu.

4.B : Pourquoi ?

5.A : Parce que je l'ai peint en rouge.

Dans DC, cinq types de coups sont possibles : affirmation, retrait, question, défi et résolution. Ces coups portent sur des contenus propositionnels exprimés en logique propositionnelle. Ainsi, chaque participant X possède une structure d'engagement $CS(X)$ qui vérifie les règles de mise à jour suivantes :

1. les CS sont vides au commencement du dialogue ;
2. les questions et les demandes de résolution ne modifient par les CS ;
3. une assertion ajoute la proposition dans les CS du locuteur et de l'interlocuteur ;
4. un retrait retire la proposition du CS du locuteur ;
5. un défi marque la proposition défiée (à l'aide du symbole ?) dans le CS du locuteur ;
6. une assertion sur q, lorsqu'elle est une défense de p, ajoute q et $q \rightarrow p$ aux CS des deux interlocuteurs ;

Dans ce cadre, les règles de dialogue sont les suivantes :

1. chaque joueur joue à son tour une des cinq locutions autorisées par le système ;
2. après une question à propos de p, l'interlocuteur peut : (a) affirmer p, (b) affirmer $\neg p$ ou (c) défier p ;

Tour	Joueur	Coups	CS(A)	CS(B)
1	A	affirmation(r)	r	r
2	B	défi(r)	r	$?r$
3	A	affirmation($\neg b$)	$r, \neg b, \neg b \rightarrow r$	$?r, \neg b, \neg b \rightarrow r$
4	B	défi($\neg b$)	$r, \neg b, \neg b \rightarrow r$	$?r, ?\neg b, \neg b \rightarrow r$
5	A	affirmation(r)	esquive de question	$?r, ?\neg b, \neg b \rightarrow r$

TAB. 2.4 – Évolution du tableau des engagements dans le dialogue de A et B.

3. il est interdit d'affirmer une proposition déjà présente dans les deux CS ;
4. il est interdit d'affirmer une proposition sur laquelle les deux interlocuteurs sont déjà engagés ;
5. après un défi sur p, l'interlocuteur peut : (a) retirer p, (b) asserter un fait non défié ou (c) demander une résolution sur une implication de conclusion p et pour laquelle toutes les prémisses sont des faits sur lesquels le défiant est engagé ;
6. un interlocuteur ne peut demander de résolution sur p que si : (a) p est une conjonction de faits directement inconsistante ou (b) p est une implication et le locuteur est engagé sur toutes les prémisses tandis que le coup précédent était un retrait ou un défi de la conclusion ;
7. après une demande de résolution sur p, l'interlocuteur peut : (a) retirer un des faits de p, (b) retirer une des prémisses de p ou (c) affirmer p.

On remarque que l'affirmation engage aussi bien le locuteur que l'allocutaire (troisième règle de mise à jour). Cela permet de gérer l'acceptation de manière implicite. Les interlocuteurs doivent alors respecter le principe de coopération suivant : « Exprimer un désaccord de croyance dès que possible ».

Le tableau 2.4, représente l'évolution du tableau d'engagement pour l'exemple de dialogue précédent, où r et b tiennent pour les propositions « peint en rouge » et « peint en bleu », respectivement. L'esquive de question est repérée puisque la règle de dialogue 5.c est enfreinte lors du dernier énoncé de l'agent A.

Mouvement/glissement dialectique

En dialectique, il est admis que chaque type de dialogue doit disposer de son système dialectique [Walton et Krabbe, 1995]. Or, l'analyse de conversation montre bien qu'un dialogue

est rarement d'un seul et même type du début à la fin. Il est courant d'imbriquer des types de dialogues. On appelle ces changements de type de dialogue (et donc de contexte) au cours de celui-ci des *glissements dialectiques*. Pour que ces glissements se réalisent sans malentendus, il faut qu'ils soient acceptés par tous les participants, c'est-à-dire établis.

En conclusion de cette section, ajoutons que la dialectique formelle a permis le développement de la notion de jeux de dialogue [Walton, 1984]. En effet, l'idée de système dialectique étendue à tous les types de dialogues est présente dans de nombreux domaines où elle est nommée *jeux de dialogues*. On prendra pour exemple, les systèmes tutoriels intelligents [Moore, 1993], d'assistant informatique ou encore les systèmes de génération d'explication [Moore, 1990a].

Les sections suivantes présentent les jeux de dialogues tels qu'envisagés spécifiquement pour les SMAs tandis que la thèse de Quignard [2000] offre un état de l'art de ces systèmes pour les interactions homme-machines.

Approche par la dialectique formelle de Amgoud et al.

Amgoud et al. [2000, 2002] ont définis des règles de dialogue et des règles de mise à jour des connaissances pour les différentes locutions de leur système. Dans ces travaux, directement inspirés par ceux de MacKenzie sur les systèmes dialectiques, les règles de dialogue indiquent les séquences de locutions autorisées (comme pour les protocoles) tandis que les règles de mise à jour capturent les effets de ceux-ci sur l'état du dialogue (représenté par les tableaux d'engagement des participants). Les locutions définies étendent celles de MacKenzie de sorte à couvrir tous les types de dialogues proposés par Walton et Krabble (présentés en section 1.3.3).

Ce modèle a été raffiné, mais reste étroitement lié aux dialogues d'argumentation. D'autres propositions basées sur ces systèmes d'argumentation ont été proposées, comme celles de Parsons et McBurney [2003] ou Bentahar [2002]; Bentahar et al. [2004]. Cependant, l'étude de l'argumentation et de l'explication (dont un état de l'art est donné dans [Moulin et al., 2002]) est une des perspectives de cette thèse que nous n'explorerons pas ici.

Approche de Flores et Kremer

Flores et Kremer [2001] ont également proposé une approche qui vise à définir les protocoles et la sémantique des actes de langage dans un modèle unifié basé sur la notion d'en-

Énoncé	But illocutoire	Opération	Créditeur	Débiteur
Demande	Propose	Ajout	locuteur	interlocuteur
Offre	Propose	Ajout	interlocuteur	locuteur
Retire	Propose	Retrait	locuteur	interlocuteur
Annule	Propose	Retrait	interlocuteur	locuteur

TAB. 2.5 – Sémantique des actes de dialogue dans le système de Flores et Kremer.

gagement social. Comme dans les travaux d'Hamblin, les agents maintiennent un tableau des engagements partagés qui peuvent être ajoutés ou retranchés. Un unique protocole (appelé « Protocol for Proposal ») définit comment ces engagements sociaux peuvent être négociés [Flores, 2002]. Dans le système de Flores et Kremer, tous les actes de dialogue (proposer, accepter, rejeter, contrer qui s'appliquent aux actions sur les engagements et informer pour transmettre un contenu propositionnel) sont conjointement produits et résolus via ce protocole. Ainsi, les actes de dialogue sont formulés en termes d'opérations génériques sur les tableaux d'engagement, comme indiqué par le tableau 2.5. Par exemple, une demande est une proposition d'adopter un engagement social fourni par le locuteur, alors qu'une offre est une proposition par le locuteur d'adopter un des engagements sociaux de l'interlocuteur. Cette formulation, qui peut être mise en question, à l'avantage d'être compacte et mieux adaptée aux SMAs que les systèmes dialectiques décrits ci-dessus. Flores et Kremer [2004] ont également utilisé cette approche pour la définition de protocoles.

2.3.6 Approches des protocoles basées sur les jeux de dialogue

Les jeux de dialogue formels sont des jeux dans lesquels les coups sont des locutions régies par des règles. L'idée de jeux de dialogue remonte à Aristote, elle fut successivement exploitée par Wittgenstein (dans un sens différent de celui donné ici) puis par Hamblin [1970], père de la dialectique formelle. L'approche par les jeux de dialogue passe par la définition de systèmes dialectiques dont les désiratas, lorsqu'ils sont appliqués aux SMAs sont (adapté de [Wooldridge et al., 2002]) :

- *L'inclusivité* : le système dialectique ne doit pas empêcher un agent de participer au dialogue s'il est qualifié et qu'il en a la volonté ;
- *La transparence* : les participants d'un dialogue doivent avoir une connaissance *a priori* des règles et de la structuration du système. En particulier, toutes les références du système dialectique vers la réalité extérieure doivent être explicites (principalement les engagements sociaux extra-dialogiques) ;

– *La justice* : le système doit traiter tous les participants de manière équivalente ou en cas d'asymétries (dues à la structure organisationnelle, aux rôles), elles devront être explicites ;

– *La clarté de la théorie dialectique sous-jacente* : le système dialectique devra être basé sur une théorie des conventions dialogiques partagée, de sorte que les obligations dialogiques soient connues et acceptées. Ainsi, les agents peuvent raisonnablement anticiper le comportement des autres (dans une fourchette définie par le jeu). Par exemple, un agent qui pose une question doit savoir s'il va recevoir une réponse ou non, ...

Historiquement, l'idée de jeux de dialogue s'est concrétisée sous différentes formes : scripts partagés de Levin et Moore [1978], recettes partagées de Mann [1988], réseaux de transition de Lewin [2000] et méta-règles de conversation de Airenti et al. [1993]. On présente ici les approches proposées dans le cadre de la communication agent. À l'instar de Maudet et Chaib-draa [2002], on indique pour chacune de ces approches : quel type de *structure* est utilisée pour représenter les jeux, quelle méthode est utilisée pour en assurer l'*établissement* et quels types de *composition* des jeux sont considérés.

Approche de Reed

Reed [1998] a proposé la notion de cadre de dialogue comme structure d'échange abstraite. Ses travaux sont directement basés sur ceux des chercheurs en dialectique formelle Walton et Krabbe [1995].

Structure – Un cadre de dialogue est formellement défini comme un triplet :

$$F = << t, \delta >\in D, \tau \in \Delta, \{u^0_{x\to y}, u^1_{y\to x}, \ldots, u^n_{x\to y}\} >, \text{ où :}$$
$$D = \{ < persuade, B >, < negociate, C >, < inquire, B >, < infoseek, B >$$
$$, < deliber, P > \}$$

C'est-à-dire que D est l'ensemble des couples $< t, \Delta >$ où t est un type de dialogue et Δ indique le type de notion concerné par le type de dialogue considéré (B pour les croyances, C pour les contrats et P pour les plans). τ est le sujet du dialogue et les u^n sont les types d'énoncés qui peuvent être produits au $n$ième tour du dialogue par les interlocuteurs x et y. Les u^n définissent donc un protocole (au sens de l'ensemble des séquences de messages possibles) pour chaque cadre de dialogues. Ce protocole peut être vide. Cette approche permet de couvrir la classification des types de dialogue de Walton et Krabbe (voir section 1.3.3) [Reed et Long, 1997].

Établissement – Les structures proposées par Reed sont manipulées par des méta-actes de communication : $propose$ et $accept$. Ces énoncés ont pour but de permettre aux agents de manipuler les cadres de dialogues tout en assurant l'établissement de ces manipulations. Par exemple, l'énoncé suivant est bien formé et indique que x propose à y de négocier le prix d'un produit ($item25$), avec 50 comme proposition initiale :

$$u^0_{x \to y} :< propose(negociate, < buy(y, intem25), < price, 50 > >), \{\} >$$

Dans l'approche de Reed, développée pour introduire l'argumentation dans les SMAs, chaque énoncé inclus un support, c'est à dire une liste de propositions qui permettent de justifier l'énoncé. Ce support est éventuellement vide, comme c'est le cas ici. Les actes $concede$ et $accept$ ferment automatiquement le cadre.

Composition – Reed considère deux types de compositions : séquentielle et imbriquée, toutes deux capturées d'emblée par la structure définie ci-dessus. En effet, comme les propositions sont des coups comme les autres, elles peuvent être faites en cours de dialogue. Quand un nouveau cadre de dialogue est proposé au tour i et accepté au tour $i + 1$, si aucun cadre n'est ouvert, c'est un séquencement sinon c'est une imbrication. Dans ce dernier cas, le cadre courant est simplement suspendu jusqu'à ce que le nouveau cadre soit terminé. C'est au concepteur de s'assurer que ses cadres termineront tous et de réguler les possibilités d'imbrication (qui doivent rester en nombre fini).

Approche de Dastani et al.

Dastani et al. [2000] ont proposé une méthodologie pour la construction de protocoles de négociation flexibles. Bien qu'ils négocient, les agents partagent le but commun de coordonner leurs actions. Une représentation partielle des actions coordonnées, sous la forme de recette, est donnée et les jeux de dialogue en sont un type particulier. Les auteurs mettent également de l'avant la notion de *cohérence*. Un dialogue est cohérent dans son contexte : (1) s'il correspond à un plan qui peut permettre d'atteindre le but apparent d'un agent et, (2) s'il suit les règles d'interaction courantes. Dépendamment des attitudes des agents et de celui qui a l'initiative, (1) ou (2) prend le dessus. Même si le travail effectué concerne la négociation, le cadre proposé est sensé être suffisamment générique pour permettre d'autres types de dialogues orientés tâche.

Les énoncés sont formés d'actes de dialogue qui sont composés d'un contenu sémantique et d'une fonction communicative. Un enregistrement conversationnel garde trace des actes de dialogue et des engagements associés qui circulent (représentation du contexte). La fonction communicationnelle est liée à la tâche sous-jacente et/ou à la régulation de l'interaction.

La cohérence tient donc autant à la cohérence au niveau de la tâche qu'à celle du niveau interactionnel. Dans leur système, la cohérence orientée tâche est assurée par l'inférence de plans alors que celle du niveau interactionnel repose sur des recettes pré-planifiées pour l'action communicative commune telles que présentées par Hulstiln [2000].

Structure – Un acte de dialogue est soit initiatif, soit réactif et la structure de base des jeux développés est une unité initiation/réaction. Cependant, les échanges sont régulés par des conditions de cohérence sur le contenu sémantique.

Composition – Les jeux peuvent être composés de manière statique (au moment de leur conception) ou dynamique par séquencement ou chaînage.

Établissement – Bien que l'utilité d'une phase de négociation du jeu courant soit mentionnée, aucune indication n'est fournie sur ce point.

Approche de McBurney et Parsons

McBurney et Parson [2001, 2002] ont proposé une autre approche utilisant explicitement des structures de jeux ayant pour ambition de représenter les types de dialogues proposés par Walton et Krabbe (voir section 1.3.3) ainsi que certains méta-dialogues. Pour ce faire, ils proposent un modèle en trois couches : (1) une couche de topiques[20] qui définit quels sont les sujets possibles du dialogue, (2) une couche de dialogue et (3) une couche de contrôle.

Structure – Les jeux de dialogue sont définis dans la couche de dialogue et consistent en un système dialectique traditionnel composé de : (1) règles d'ouverture, (2) règles de locution, (3) règles de dialogue, (4) règles de mise à jour et (5) règles de terminaison.

Établissement – C'est la couche de contrôle qui assure l'acceptation via un méta-dialogue de contrôle auquel on suppose que les agents sont prêts à participer. Ce méta-dialogue de contrôle permet aux agents de décider conjointement des jeux de dialogue à jouer. Pour ce faire, des coups de méta-niveau sont définis ($begin(G(p))$,$end(G(p))$) et un coup spécial ($propose.return.control$) permet aux agents de remonter au niveau contrôle alors qu'ils jouent un jeu de dialogue, assurant ainsi la liaison entre les deux couches.

Composition – C'est aussi la couche de contrôle qui permet la composition des jeux. Différents types de composition sont permis :

[20] En linguistique, la notion de topique indique ce dont on dit quelque chose.

- *itération* G^n : répétition de n dialogues du type G, chacun des jeux débutant après la fermeture du précédent ;
- *séquence* $G; H$: H débute après la fermeture de G ;
- *emboîtement* $G[H : I]$: H débute pendant G après la séquence de coups I ;
- *parallélisation* $G \cup H$: G et H débutent simultanément ;
- *test* $< p >$: $< p >$ est un dialogue de contrôle pour tester le statut de vérité de p. Le dialogue courant termine si p s'avère faux.

On note que la parallélisation est un mode de composition original, qui n'est pas présent dans la littérature classique, mais peut s'avérer utile pour la communication agent. En outre, et contrairement à la version de Reed, l'emboîtement ne suspend pas le jeu emboîtant.

Approche de Maudet

Le modèle des jeux de dialogue de Maudet [2001], avec qui nous avons eu l'occasion de travailler lors de son séjour post-doctoral au laboratoire DAMAS [Dialogue, Apprentissage et systèmes Multi-AgentS][21], entrerait dans cette catégorie des approches reposant sur les jeux de dialogue. Initialement, développé pour le langage naturel dans le laboratoire GRAAL (Groupe Raisonnement, Action et Actes de Langage à Toulouse), c'est ce modèle que nous avons repris, raffiné et implémenté collectivement avec Maudet, Chaib-draa, Andrillon, Bourget, Labrie, Bergeron et Flores. Le chapitre 5 présente l'approche résultante, adaptée aux systèmes multi-agents, à laquelle nous avons contribuée. Les sections suivantes présentent les avantages (section 2.3.7) et limites (section 2.3.8) des approches conventionnelles et sociales.

2.3.7 Avantages des approches conventionnelles et sociales

Avantages des approches sociales

Parmi les avantages des approches sociales sur lesquels il est bon d'insister, car ils sont pour une bonne part de l'intérêt porté à ces approches par la communauté SMA, il faut relever

[21] http ://www.damas.ift.ulaval.ca/

que les approches sociales résolvent le problème de l'hypothèse de sincérité qui était attaché aux approches intentionnelles et simplifient la vérification sémantique. C'est une double conséquence inhérente à l'utilisation d'engagements sociaux. Le problème de l'hypothèse de sincérité est résolu puisque les engagements ne sont pas nécessairement sincères. Par contre, un engagement doit idéalement être respecté et dans le cas contraire, son créditeur s'expose à des sanctions sociales ou matérielles. En cela, les engagements sont des attitudes sociales indépendantes mais pas indifférentes des spécificités internes aux agents. Le traitement des aspects sociaux du dialogue en est simplifié.

Quant au problème de la vérification, il est théoriquement résolu grâce au caractère public des engagements qui les rend accessibles, en particulier pour vérification. En outre, les approches sociales parviennent, grâce à la notion d'engagement, à éviter toute spécification mentaliste dans la sémantique des langages de communication utilisés. Ainsi, les agents n'ont plus nécessairement à implémenter les attitudes mentales qui sont attachées aux sémantiques mentalistes. Cela permet d'envisager de faire communiquer des agents hétérogènes, d'architectures internes variées. Nous reviendrons sur ces avantages aux chapitres 4 et 5, lorsque nous présenterons notre approche des communications agents pour les SMAs ouverts.

Aspects positifs spécifiques aux jeux de dialogue

Même s'il reste du travail de formalisation, d'implémentation et de validation, les approches par jeux de dialogue semblent être une avenue prometteuse pour pallier la rigidité des protocoles traditionnels. En effet, ces approches adoptent un formalisme plus flexible que les automates à états finis : l'utilisation des engagements garde une trace plus riche du dialogue que le simple dernier coup considéré dans les protocoles traditionnels. Ces derniers contraignent les agents à se conformer aux transitions attendues alors que les engagements motivent les agents à se conformer à un comportement attendu. Les approches par engagements sont donc capables de considérer les messages inattendus ou exceptionnels mieux que les protocoles classiques. En outre, et même si de nombreuses clarifications seraient nécessaires de ce côté, les jeux de dialogue peuvent être composés. Sous l'impulsion de Reed, des méta-actes (jeux de contextualisation, couche de contrôle, ...) ont été définis pour négocier et établir le jeu de dialogue courant. En outre, l'approche par les jeux de dialogue est déclarative (les règles sont explicitées) : cela accroît leur clarté et simplifie leur utilisation informatique.

Comme le signale Maudet [2001], les jeux sont utiles au niveau du dialogue aussi bien dans les phases d'interprétation que dans les phases de production. De plus, ils sont une contribution intéressante pour ce qui est de la structuration du dialogue. La structure locale est donnée par la structure intra-jeu (ou par les obligations langagières correspondantes). La structuration globale est donnée par la structure inter-jeux. Mais, le niveau jeu est-il réelle-

ment nécessaire ? Qu'est-ce que la notion de jeu apporte de plus qu'un modèle intentionnel augmenté d'obligations ? En guise de réponse à ces questions, Maudet indique que :

- Les jeux sont empiriquement fondés, comme le montrent les expériences menées sur les annotations de dialogues [Kowtko et al., 1991] ;
- Les jeux sont des structures prédictives : Poesio et Mikheev [1998] ont montré que sur le corpus de MAPTASK les prédictions sont de 50% avec les jeux contre 38% si l'on considère uniquement le coup (l'énoncé) précédent ;
- Les jeux raffinent et concrétisent la notion de coopération dialogique. Ils réalisent concrètement les notions d'établissement (grounding) et de projet conjoint discutées en sections 1.3.2 et 2.3.1.

2.3.8 Limites des approches conventionnelles et sociales

Limitations des approches sociales

On peut s'étonner du peu de concepts communs aux approches intentionnelles et aux approches sociales. En fait, en introduisant une couche publique via les engagements sociaux, les normes et les conventions (qui se déclinent en obligations, permissions, interdictions), les approches conventionnelles et sociales séparent bien le niveau public du niveau privé, mais au prix d'une perte de complétude. En effet, l'introduction d'un niveau public ne dispense pas de la définition d'un niveau privé et il reste alors à articuler ces deux niveaux. Il s'agit donc de définir un système de gestion du niveau public cohérent sur lequel les agents savent raisonner.

En outre, de nombreuses notions connexes aux engagements méritent d'être approfondies, il s'agit de :

- *clarifier la notion d'engagement social* : si différents modèles d'engagements ont été proposés, certains points du cycle de vie des engagements restent à clarifier. Par exemple, la vérification de la satisfaction des engagements n'a pas été clairement définie ;
- *définir des mécanismes de contrôle social adaptés à ces approches* : toutes les approches sociales détaillées dans ce chapitre reposent sur l'hypothèse que les engagements seront généralement respectés. Pourtant, aucun mécanisme susceptible de valider cette hypothèse n'a été exhibé dans les travaux sur la communication agent ;

Limitations des approches par les jeux de dialogue

L'approche par les jeux de dialogue nous semble, de par sa simplicité, la plus prometteuse des approches conventionnelles. Cependant, ses développements multi-agents sont récents et donc encore incomplets. Parmi les éléments qui restent à définir pour en faire un cadre interactionnel complet, on note les besoins suivants :

- régler les problèmes d'implémentation : implantation distribuée du gestionnaire de dialogue et des agendas (dans un cadre multi-agents) ;
- considérer le cycle de vie complet des engagements : parmi les approches énumérées dans cet état de l'art, aucune ne tient compte des éventuels dialogues de décharge des engagements satisfaits ou violés ;
- étendre les cadres existants (dilogiques) aux conversations multi-parties (plus de deux interlocuteurs) : ce besoin n'a été abordé dans la littérature SMA que très récemment [Dignum et Vreeswick, 2003; Traum, 2004; Huget et Demazeau, 2005] ;
- valider les approches sur des exemples réels, d'envergure : à notre connaissance aucune application réelle issue des propositions de la communauté SMAs concernant les jeux de dialogue n'a été réalisée ;

Il reste, en outre, un certain nombre de points non expliqués par les approches sociales en général. En effet, ces approches se concentrent sur les aspects sociaux - publics - de la communication sans définir comment ces derniers seront pris en charge par les agents cognitifs tels qu'ils sont définis actuellement. C'est-à-dire que rien n'est dit sur la manière dont les agents devraient utiliser ces langages et structures afin de gérer leurs engagements d'une manière utile à leurs objectifs individuels et collectifs. Maudet [2001] indique simplement que les agents devront être normatifs et délibératifs. Normatifs pour suivre les règles de dialogue, se conformer aux jeux de dialogue et agir en fonction de leurs engagements et délibératifs pour prendre en compte leurs propres besoins/intentions. Reste donc à savoir comment ces différents niveaux - privé et public - peuvent être combinés.

2.4 Conclusion et discussion

Pour ce qui est des communications, la communauté SMA se concentre, et ce depuis ses débuts, sur l'élaboration d'un hypothétique cadre interactionnel standard. Les principaux

langages de communication agent actuels (section 1.2.2), KQML et FIPA-ACL, sont basés sur la théorie des actes de langage (section 1.2.1) augmentée d'une sémantique mentaliste (section 2.2.3). Le dialogue est censé émerger de l'enchaînement des productions d'actes issus des intentions de chaque agent via la reconnaissance et le raisonnement sur les intentions des autres (section 2.2). Cette approche, dite « mentaliste », a été critiquée (section 2.2.6) et une redéfinition de la sémantique en termes plus sociaux ainsi que la construction d'une surcouche conversationnelle ont été rendues nécessaires à différents égards (section 2.3). Les protocoles (section 2.3.2) se sont fait reprocher leur manque de souplesse et les politiques de conversation puis plus récemment les jeux de dialogue ont été proposés pour pallier les défauts de ceux-ci (section 2.3.6).

Dans cet état de l'art, nous avons délimité, grâce à une revue de la littérature fournie, un certain nombre de problématiques concernant les conversations entre agents. Ces problématiques se déploient selon quatre dimensions pertinentes dans l'étude de la communication inter-agent qui correspondent à quatre facettes générales de l'étude du langage : syntaxe, structure, sémantique et pragmatique. Pour ce qui est de la syntaxe des énoncés, les ACLs fournissent un bon outil puisqu'ils ont la puissance expressive suffisante pour rendre compte de tous les énoncés. Ce résultat découle du fait qu'ils sont l'application de la théorie des actes de langage qui couvre théoriquement l'ensemble des énoncés en langage naturel[22].

Au niveau structurel, une alternative semble émerger entre les approches strictement cognitives comme les approches intentionnelles ou les approches strictement sociales qui ne spécifient rien de la structure (vue comme émergeant des enchaînements des énoncés des agents) et les approches par protocoles qui réduisent l'espace de recherche des continuations possibles au strict minimum, mais font perdre souplesse et adaptativité aux conversations. Cette alternative est proposée par les jeux de dialogue. Pour autant, ceux-ci reste à parfaire avant qu'une implantation réaliste soit possible (voir section 2.3.8).

Pour ce qui est de la sémantique des unités conversationnelles, la communauté scientifique spécifiquement multi-agents s'est déplacée d'une sémantique mentaliste vers une sémantique sociale exprimée en termes d'engagements permettant de résoudre le problème de la vérifiabilité, de lever l'hypothèse de sincérité et de faciliter le traitement des aspects sociaux de la communication. Il reste cependant de nombreux débats pour savoir à quel niveau on place la sémantique (au niveau de l'énoncé ou au niveau de la conversation) et déterminer quelle forme peut prendre cette sémantique. Soulignons au passage l'ambiguïté qui réside entre sémantique linguistique et sémantique mathématique dans l'ensemble des recherches sur la communication agent (voir à ce propos l'annexe A).

[22] En réalité, ce résultat n'est que potentiel, car les ACLs développés pour l'heure ne couvrent pas tous les types d'actes de langage définis dans la théorie. En particulier, Chaignaud et El Fallah-Seghrouchni [2001] ont montré que KQML et FIPA-ACL sont insuffisants pour les interactions homme-machines.

Concernant la pragmatique, ces déplacements n'invalident sans doute pas complètement la pragmatique Gricéenne et certaines des considérations sur la coopération restent valables, mais l'introduction de la couche publique des engagements mérite que celle-ci soit réexaminée. Dans les approches sociales et conventionnelles, les concepts de base fournissant la pragmatique des approches intentionnelles sont complétés d'une couche sociale constituée d'engagements et d'une couche conventionnelle (constituée de jeux de dialogues, par exemple). L'introduction de ce niveau social et public nécessite de repenser une pragmatique étendue à ce cadre plus général dans lequel l'hypothèse de coopération attachée aux approches mentalistes pourra être assouplie.

Effectivement, selon ces approches sociales, l'agent ne doit plus raisonner directement sur les intentions (communicationnelles ou pas) des autres, mais sur les engagements pris et à prendre. Ces engagements sont autant les engagements issus des conventions liées au système, ceux attachés aux rôles des agents, que ceux issus des conversations avec d'autres agents. Ainsi, trop occupée à définir un cadre interactionnel standard, la communauté SMA a quelque peu délaissé les aspects cognitifs liés à la pragmatique des communications, laissant ainsi au concepteur la majeure partie du travail quand vient le moment d'indiquer comment les agents vont utiliser le cadre interactionnel sélectionné. En conséquence, un objectif principal pour cette communauté, sera de fournir une pragmatique, au sens d'une théorie des aspects cognitifs de l'usage de le communication, qui soit adaptée aux cadres interactionnels pour agents que proposent les approches sociales et conventionnelles, en particulier les jeux de dialogue. C'est à cette problématique que nous souhaitons contribuer. Le chapitre suivant introduit et motive de manière plus complète cette problématique et définit quels ont été nos objectifs.

Chapitre 3

Problématique, motivations et objectifs

3.1 Introduction

Puisque les approches intentionnelles ne tiennent pas compte des aspects conventionnels du dialogue et que de fortes hypothèses de coopérativité et de sincérité leurs sont nécessaires, les chercheurs en intelligence artificielle et en système multiagents (desquels sont issues ces critiques, détaillées au chapitre précédent) s'accordent à penser que le niveau des engagements publics est une nécessité[1]. Dans cette veine, on souhaite donc contribuer à compléter les approches conventionnelles et sociales. Avec les approches conventionnelles, il est acquis que le cadre interactionnel est en définitive un ensemble de contraintes visant à simplifier la structuration des conversations grâce à une couche conventionnelle qui contraint les enchaînements conversationnels. Il est également acquis que les modèles cognitifs des agents fournissent et produisent un ensemble de contraintes qui devront être satisfaites au mieux pour assurer la satisfaction de l'agent (ou du groupe d'agents). Reste donc à établir les liens entre le modèle cognitif et le cadre interactionnel, c'est-à-dire à définir comment l'agent va utiliser le cadre interactionnel (en en respectant les contraintes) de façon à satisfaire ses contraintes cognitives (c'est-à-dire à maximiser sa satisfaction ou encore celle du groupe).

En effet, comme nous l'avons vu section 2.3.8, les cadres interactionnels proposés par les approches conventionnelles et sociales ne fournissent pas les éléments nécessaires à leur utilisation automatique par des agents cognitifs. De la même manière, ils ne fournissent aucune garantie quant à l'utilité des conversations tenues et tel n'est pas leur objectif. Pourtant, ce

[1] À ce propos, il est intéressant de noter que les derniers développements de la théorie de l'interaction rationnelle de Cohen et Levesque incluent la définition de la notion d'engagement social unilatéral et d'engagement conjoint et que les aspects conventionnels sont pris en compte via la définition de protocoles reposant sur cette notion d'engagement conjoint [Kumar et al., 2002].

n'est généralement pas l'habileté des agents à structurer leurs dialogues qui nous intéresse mais leur habileté à communiquer de manière utile à leurs objectifs individuels et collectifs. On pourrait donner une première formulation de notre problématique sous la forme des deux questions suivantes : de quelle manière pourrait-on automatiser la communication agent (au sens de l'utilisation automatique du cadre interactionnel) ? Comment un agent pourrait-il procéder pour déterminer si une conversation lui a été profitable ou pas et agir en conséquence ?

Avant de détailler davantage cette problématique, nous souhaitons introduire une nouvelle distinction dans les communications agents (section 3.2). Nous formulerons ensuite de manière plus précise notre problématique (section 3.3), nos objectifs (section 3.4) et notre méthodologie (section 3.5).

3.2 Cohérence structurale et cohérence cognitive

Notre problématique peut être reformulée autour de la notion de cohérence. Dans les théories de la communication, on distingue les théories cognitives des théories interactionnelles [Littlejohn, 2002]. Les théories interactionnelles traitent de la forme de la communication : comment modéliser un énoncé, un discours, une conversation ? Quelles sont les régularités structurelles des conversations ou quelles sont les contraintes conventionnelles qui pèsent sur la forme du dialogue et sur les enchaînements permis (enchaînements d'énoncés, d'actes de langage) ? Il existe de nombreuses théories interactionnelles : analyse de conversations, théories des actes de langage, ...

Les théories cognitives s'intéressent, quant à elles, à la production des messages (quoi dire, quand le dire et à qui le dire) ainsi qu'à la réception et au traitement cognitif des messages (quoi comprendre, comment le comprendre et comment réagir). Elles adressent l'aspect fonctionnel de la communication aux niveaux interne et externe. Quels sont les éléments qui poussent un agent à former tel énoncé plutôt que tel autre ? Comment un agent réagit-il à un énoncé au niveau interne ainsi qu'en terme de mise à jour du modèle d'autrui et de ses propres croyances ? Au niveau externe et public (c'est-à-dire, vis-à-vis de son environnement), quels sont les engagements que l'agent veut obtenir ? Pourquoi ? Quelle est l'utilité de la conversation, quelle est son importance ? L'agent et plus généralement le groupe d'agents conversant est-il satisfait par la conversation ? Il y a bien une différence entre la satisfaction des conventions qui pèsent sur le dialogue (par exemple, satisfaire un jeu de dialogue) et la satisfaction des agents.

Si dans les deux approches la cohérence est une notion centrale, il faut se garder de confondre la cohérence structurale du dialogue (souvent appelée cohérence conversation-

nelle [Craig, 1983]) - est-il permis de poursuivre le dialogue de cette façon ? - de sa cohérence de fond encore appelée cohérence cognitive. Le contenu du message est-il approprié à la vue des messages précédents et des états mentaux de l'agent ? Est-ce que le contenu du message est cohérent avec l'état interne de l'agent ? Est-ce que les agents ont tenu des propos pertinents quant à leurs objectifs ? Est-ce que la conversation leur est profitable ? Évidemment, ces deux dimensions de la cohérence sont souvent liées. Et, travailler sur une théorie cognitive comme le prescrit notre problématique ne signifie pas nier le besoin d'une théorie interactionnelle. En effet, lorsque l'on a déterminé quoi dire, quand le dire et à qui le dire, reste à savoir comment le dire. Par contre, cela permet de dépasser ce niveau et les idées avancées devront être valables pour tout cadre de communication conventionnel suffisamment riche au niveau interactionnel. Il y a là un *objectif de généricité*.

Or on retient de notre état de l'art que dans les recherches récentes, l'emphase a été mise sur la cohérence structurelle puisque ce sont essentiellement les aspects syntaxiques, structurels et sémantiques des communications qui ont été étudiés. À ce niveau, il est important de noter qu'une sémantique du cadre interactionnel, quelle qu'en soit la forme, ne garantit pas l'utilité des conversations tenues dans ce cadre. Pour nous, comme pour d'autres (philosophes du langage [Wittgenstein, 1953], épistémologues [Barreau, 1995] ou encore chercheurs en sciences cognitives [Vignaux, 1991]), la possibilité d'aboutir à un modèle complet qui soit purement structurel/interactionnel est hypothétique, et ce du fait même de l'existence et de la prédominance de la cohérence cognitive du dialogue, c'est-à-dire des aspects cognitifs, psychologiques et des conditions de satisfaction qui s'y rapportent. Dans de nombreux cas, c'est cet aspect sémantique, fonctionnel et pragmatique qui domine. L'important est ce qui est dit, pas comment cela est dit ni même dans quel ordre. Examinons un exercice classiquement réalisé en cours de linguistique pour s'en convaincre : à la suite de l'écoute d'une intervention radio d'une dizaine de minutes, force est de constater qu'aucun des élèves n'est capable de se souvenir ou de reproduire les cinq premiers énoncés de la conversation, ni de savoir s'ils constituaient un enchaînement structurellement cohérent alors que tous savent ce qui s'est dit à ce moment. La cohérence et la pertinence du dialogue sont jugées au niveau des idées qui transitent dans le dialogue, de leur impact cognitif[2], pas au niveau de sa forme qui n'en est que le médium. Une étude approfondie serait nécessaire pour étudier les rapports entre ces deux types de cohérence. Prenons deux exemples pour bien les différencier :

Exemple 3.1

1. *A : Est-ce que je peux te poser une question ?*
2. *B : Oui, vas-y.*

[2] Les approches intentionnelles rendent en partie compte de cette dimension cognitive dans un cadre coopératif, mais sans offrir la souplesse des approches conventionnelles en terme de structuration des dialogues.

3.A *: Est-ce que tu as l'heure ?*

4.B *: Non, j'aimerais bien la connaître...*

L'exemple 3.1 précédent montre une conversation dans laquelle tous les aspects interactionnels conventionnels sont remplis. Le dialogue est bien formé et les engagements pris par les agents sont respectés (notamment l'engagement de répondre à une question pris par B en 2.B). Pourtant, cette conversation a en définitive une assez faible utilité si elle s'arrête à ce moment-là. En effet, bien que les intervenants finissent par partager leur problème (ils désirent connaître l'heure, mais ne disposent pas de l'information), ce problème n'est pas résolu. De manière un peu simpliste, on pourrait dire que les agents communiquent pour résoudre ou éviter des problèmes (pas de problème, pas de communication, l'inverse étant faux). Dans ce cadre, les communications ne sont que des tentatives pour résoudre ou éviter ces problèmes et comme telles, elles peuvent échouer. C'est ce qui se passe dans cet exemple. Cette vision des choses permet de bien comprendre pourquoi les cadres interactionnels actuels (protocoles, jeux de dialogue, ...) ne garantissent pas l'utilité des conversations et qu'un mécanisme est requis pour la mesurer et permettre aux agents d'agir en conséquence.

Exemple 3.2

1.A *: Est-ce que tu peux me prêter ta règle ?*

2.B *: Si on ne part pas tout de suite, nous serons en retard.*

À l'inverse, dans l'exemple 3.2, la cohérence structurale n'est pas respectée, le dialogue est mal formé du point de vue de la dialectique. Les règles conventionnelles de gestion du sujet dans les paires d'adjacences ne sont pas respectées, mais l'échange est utile, car A et B vont peut-être pouvoir éviter un problème (on suppose que A et B tiennent à être ponctuels). En effet, la cohérence de fond est indéniable dès lors que B fait allusion à une urgence et se permet donc de violer les règles structurelles les plus élémentaires pour expliciter un problème commun qu'il juge plus important que le problème pour lequel A réclame son aide. C'est un type de « court-circuit » utilitariste et opportuniste très courant dans l'utilisation du langage naturel.

Ainsi, il est probable qu'une théorie purement structurale des conversations énoncerait des règles qui, si elles sont souvent respectées, sont également très largement enfreintes pour des raisons de cohérence cognitive. Comme le montre l'exemple 3.2 : pour augmenter l'utilité du dialogue, un agent peut être amené à ne pas respecter les règles interactionnelles standards. Ces infractions sont comparables aux violations des maximes de Grice [1957] (ici, les

maximes de qualité et de manière sont violées). En effet, comme elles, les théories interactionnelles définissent généralement un cadre normatif qui peut (et parfois doit) être enfreint. Cependant, pour des agents rationnels, ces infractions (si elles sont permises) tout comme les dialogues bien formés doivent supporter une caractéristique plus générale d'utilité. Notre problématique inclut donc naturellement la question de la définition d'une métrique pour juger de l'*utilité des dialogues*.

3.3 Problématique

Si, comme nous l'avons décrit dans les deux premiers chapitres de cette thèse, de nombreux travaux se sont préoccupés de définir des langages de communication agents, peu se sont concentrés sur les aspects cognitifs de la communication agent, c'est-à-dire sur les procédés par lesquels les agents vont utiliser dynamiquement et automatiquement les langages de communication proposés. Pourtant, si on suppose la souplesse et la cohérence structurale des dialogues engagés garanties par les contraintes fournies par le cadre interactionnel conventionnel, un certain nombre de problèmes subsistent :

- *les problèmes liés à la gestion de la dynamique du dialogue* : Quand et quoi communiquer, à qui et pourquoi ? Comment choisir un interlocuteur ? Quel type de dialogue tenir ? À quel sujet ? Dans quel espoir ? Au sein d'un dialogue, comment sélectionner parmi les continuations possibles ? Comment choisir les degrés d'intensité des actes de langage utilisés ? Comment et pourquoi composer différents types de dialogues ?
- *les problèmes liés à l'indépendance des modèles cognitifs des agents et des cadres interactionnels pour la communication entre agents* : Comment garantir l'alignement de la satisfaction des aspects conventionnels des conversations, dictés par le cadre interactionnel, avec les aspects cognitifs des agents dialoguant ?
- *le problème de la définition et de la mesure de l'utilité du dialogue* : Quelle est l'utilité du dialogue pour un agent ? Quelle est l'utilité de la conversation pour le groupe d'agents impliqués ? Les agents sont-ils satisfaits d'une conversation ?
- *les problèmes liés à la gestion des conséquences sociales et cognitives des communications entre agents* : Quels sont les impacts des communications sur les agents au niveau cognitif de leurs états mentaux comme au niveau social de leurs accointances ?

C'est à ces questions que l'on souhaite contribuer. Il y a là un certain nombre d'enjeux dont le principal pour l'informatique concerne la conception des SMAs : pour l'heure, dans

le cas général, le comportement communicationnel des agents n'est pas automatisé. C'est au concepteur de définir quels dialogues seront tenus dans telle ou telle circonstance. Bien souvent, seul le concepteur connaît les raisons de ces dialogues et lui seul est garant de leur utilité. C'est également au concepteur de déterminer quelles seront les poursuites du dialogue, si poursuite il y a. Certains concepteurs ont évidemment imaginé des algorithmes *ad hoc* pour automatiser certains enchaînements conversationnels ou encore ont proposé des approches pour des types de dialogues précis (on pense aux nombreux travaux sur la négociation ou l'argumentation), mais il n'y a pas de théorie générale de la communication dialogique entre agents (autre que les approches mentalistes intentionnelles, dont on a vu les limitations sections 2.2.4 et 2.2.6). Ainsi, du point de vue informatique, notre problématique est de chercher à fournir des éléments théoriques et pratiques pour permettre aux agents de :

- manier automatiquement le cadre interactionnel conventionnel pour répondre à leurs besoins ;
- juger par eux-mêmes de l'utilité de leurs actions et en particulier des conversations tenues ;
- décider de manière automatique des suites à donner à une conversation.

Nous pensons que ces éléments pourraient être utiles à la conception des SMAs et donc susceptibles d'alléger le travail du concepteur. La (ou les) réponse aux questions évoquées dans le paragraphe précédent sera d'autant plus satisfaisante qu'elle sera générique, c'est-à-dire que ce soit bien un mécanisme général qui soit exhibé et non une simple réponse *ad hoc* supplémentaire. Pour parvenir à cette généricité, nous devrons prendre garde de *fonder notre approche sur des bases théoriques solides et bien acceptées dans les sciences cognitives*. On saisira au passage la chance de contribuer, en retour de cet apport à l'informatique, aux sciences cognitives en général et à la pragmatique en particulier.

En effet, la pragmatique des cadres interactionnels des SMAs est un aspect souvent négligé. Soulignons que ce n'est pas complètement un hasard puisque ce domaine est réputé difficile, et ce, dans tous les domaines d'étude de la communication (linguistique, philosophie du langage, théories de la communication, intelligence artificielle, . . .). « Le besoin d'une véritable théorie du langage comme moyen d'expression se fait sentir » [Lienard, 1991]. C'est en fait, une théorie de l'utilisation de ce système d'action qu'est le langage qui est à produire.

3.4 Objectifs

Niveau d'analyse du langage appelé de ses voeux par le philosophe et sémioticien Morris[3] en 1938 [Morris, 1938, 1946], *la pragmatique* rassemble tous les éléments de sens du langage. Elle prend en compte la théorie de la référence qui caractérise la sémantique, mais aussi une théorie de l'usage et du contexte que la langue anime pour faire sens.

De nos jours, la pragmatique en tant que domaine de recherche, parfois qualifiée de poubelle de la linguistique [Leech, 1983], est un domaine difficile à cerner tant ses préoccupations sont variées. Tandis que certains étudient la deixis, c'est-à-dire la signification des énoncés en contexte, d'autres se consacrent à l'étude des aspects performatifs du langage, illustré par la théorie des actes de langage ou encore aux aspects normatifs ou cognitifs de l'usage du langage. Le principe de coopération de Grice [1957], le principe de politesse de Leech [1983] ou la théorie de la pertinence de Sperber et Wilson [1986] donnent un bon aperçu de cette dernière catégorie. L'étude de l'utilisation de la communication en contexte a donné naissance à nombre de domaines connexes comme l'analyse conversationnelle et la pragmatique inter-langage. La capacité à comprendre et à produire des actions communicatives est nommée *compétence pragmatique* [Kasper, 1997]. Dans le cas des SMAs, *notre objectif principal est précisément de fournir une théorie des aspects cognitifs de l'usage du langage qui soit adaptée aux cadres interactionnels agent que proposent les approches conventionnelles et sociales, en particulier aux jeux de dialogue.*

Il s'agit donc de fournir des outils de méta-niveau pour guider l'agent dans son comportement communicationnel. On devra donc définir des outils de gestion de la dynamique du dialogue pour guider l'agent dans son comportement communicationnel, autant au niveau de l'utilisation du cadre interactionnel (quel type de dialogue choisir ? quelle structuration choisir ?...) qu'au niveau cognitif (avec qui communiquer, quand et quoi ? pourquoi ?). Ce faisant, on souhaite contribuer à établir un lien entre les aspects privés et les aspects publics dans les communications agents. En effet, on a vu à la section 2.2.6 que pour l'heure, les approches conventionnelles échouent à :

- indiquer comment et pourquoi le dialogue est initié. De la même manière, la terminaison du dialogue pose problème puisqu'aucun critère satisfaisant n'a été proposé ;
- établir les liens entre les aspects privés (états mentaux) et la couche publique (les engagements). Il s'agit essentiellement de définir un système de gestion des engagements et une théorie de l'utilisation du cadre interactionnel par les agents ;

[3] C'est Morris qui est à l'origine de la division de la sémiotique dans ses trois branches actuelles : syntaxe, sémantique et pragmatique [Littlejohn, 2002].

– indiquer comment gérer la structuration dynamique du dialogue et ce aussi bien au niveau (1) local (comment un agent choisit dynamiquement parmi les différentes structurations possibles dans une conversation ?) que (2) global (comment un agent choisit son prochain interlocuteur et le type de dialogue dans lequel il souhaite s'engager et pourquoi ?).

De manière plus synthétique, on peut reprendre les différents aspects informatiques de notre problématique en indiquant les objectifs qui en découlent :

– *l'utilité dans les communications agent* : les objectifs sont de la définir, de fournir des moyens de la calculer et de montrer en quoi cette notion peut être utile pour la structuration des communications aussi bien au niveau du choix des locutions/coups dans un dialogue, qu'au niveau du choix d'un interlocuteur et d'un type de dialogue (capturés par des jeux de dialogue ou des protocoles) ;

– *l'automatisation des communications agents* : il s'agit pour un agent de maximiser la satisfaction de ses contraintes cognitives, tout en respectant les contraintes normatives imposées par les conventions ayant cours dans le système multi-agent, ainsi que par le cadre interactionnel utilisé. Un autre objectif est donc de fournir des éléments d'automatisation des communications agents via l'utilisation des outils de satisfaction de contraintes. Cet aspect devra découler de l'étude entreprise sur l'utilité dans les communications et être en accord avec cette dernière.

Ces objectifs découlent de notre problématique, elle-même dérivée de l'état de l'art présenté dans la partie précédente. La section suivante indique quelques éléments méthodologiques qui nous ont guidés dans l'élaboration de nos contributions.

3.5 Remarques méthodologiques

Les systèmes multiagents sont l'approche paradigmatique de l'intelligence artificielle distribuée moderne. Les techniques multiagents permettent la conception et le développement d'applications complexes. La notion d'agent, centrale dans ces systèmes, est issue de la philosophie, des sciences sociales et cristallise, en définitive, les apports mutuels et réciproques de l'intelligence artificielle et des sciences cognitives. Si une diversité de types d'agents peuvent être conçus, nous nous cantonnerons aux agents cognitifs issus du cognitivisme classique tels que nous les avons décrits en section 1.1.2. L'agent cognitif, entité autonome, proactive

et dotée de capacités de représentation des connaissances, de raisonnement et de communication est fondamentalement anthropomorphique. L'intérêt commun des sciences cognitives et de l'ingénierie pour les systèmes multi-agents ou plusieurs agents coopèrent (ou compétitionnent) autour d'une tâche commune est l'étude de la notion d'émergence et l'accroissement du savoir théorique et pratique autour des systèmes complexes ainsi formés. La caractéristique principale des systèmes multiagents est la capacité des agents à communiquer[4] les uns avec les autres de manière utile relativement à leurs objectifs individuels et collectifs. Cette capacité nécessite de définir pour les agents les trois dimensions généralement associées au langage et à son usage : syntaxe, sémantique et pragmatique. Si pour les deux premières dimensions, de nombreux apports et débats sont parvenus à fournir des outils convenables[5], la dimension pragmatique (au sens d'une théorie de l'utilisation dynamique du langage en contexte) est moins avancée.

De par notre formation initiale et notre intérêt dans la multidisciplinarité scientifique qui nous semble faire la force des sciences cognitives, nous souhaitons contribuer en fournissant une théorie de la pragmatique des communications agents qui soit fondée sur les sciences cognitives. En particulier, nous souhaitons introduire dans le champ des SMAs, un certain nombre de notions et concepts issus de la psychologie sociale et jusqu'ici ignorés en dépit de leurs intérêts théoriques et pratiques indéniables. Cette absence se comprend pour des raisons essentiellement historiques (pas nécessairement rationnelles) : l'intelligence artificielle a favorisé les apports de la philosophie de l'esprit, en particulier de son versant analytique et logique qui produit des théories logiques formelles (bien que celles-ci ne soient généralement pas directement applicables pour des raisons de complexité informatique). Aussi, il est habituel de fonder les travaux concernant la communication entre agents cognitifs sur les apports de la philosophie analytique du langage. L'effort de transfert de connaissances s'en trouve réduit. Si nous ne renonçons pas à ces apports, nous considérons que la psychologie cognitive et la psychologie sociale doivent aussi être prises en compte dans l'élaboration des théories d'agent et dans la modélisation des systèmes multi-agents.

Cela est d'autant plus vrai que dans le cas de notre problématique, la philosophie du langage a explicitement délaissé les aspects cognitifs de la pragmatique du langage pour se concentrer sur ses aspects conventionnels (l'approche Gricéenne est un bon exemple de cette orientation). En effet, comme l'indique Marconi [1997] dans son ouvrage consacré à la philosophie du langage au XXième siècle, les philosophes du langage ont explicitement exclu les aspects cognitifs de la pragmatique de leur champ d'investigation, laissant à la psychologie sociale et cognitive le soin de les étudier. Nous irons donc chercher dans ces autres domaines des sciences cognitives les résultats scientifiques nécessaires à notre modélisation.

[4] et plus généralement à interagir.

[5] Un état de l'art de ces contributions a été présenté dans les deux premiers chapitres de cette thèse (on pourra également consulter [Pasquier, 2001b] ou [Pasquier, 2001a]) et les chapitres 4 et 5 présentent notre contribution à cet égard.

D'un point de vue méthodologique, il nous semble sain de venir compléter, du point de vue des fondements des approches de la communication agent, les analyses (hypothétiques) des philosophes avec les résultats empiriques des psychologues.

La problématique que nous avons soulevée est très générale et cela fait parti de nos objectifs que de la traiter de manière générique. Si l'intelligence artificielle sert conjointement les sciences cognitives et l'ingénierie du logiciel, nous insisterons lors de la présentation de nos contributions sur les aspects théoriques. Nous mettrons de l'avant ce qui nous semble fournir des fondations solides, résultant en un modèle crédible et réaliste pour les sciences cognitives. Nous pensons que ce faisant, notre contribution aux SMAs n'en sera que plus pérenne. On prend ainsi la chance de contribuer en retour aux sciences cognitives[6].

La seconde partie de cette thèse présente nos apports à l'égard de la problématique que nous avons isolée et discutée dans ce chapitre. Comme l'étude des aspects cognitifs de la pragmatique (chapitre 6 pour la théorie et chapitre 7 pour sa validation informatique) ne dispense pas de la définition d'un cadre intéractionnel, nous commencerons par présenter notre cadre interactionnel conventionnel et social (chapitre 5) ainsi que le modèle de l'engagement social sur lequel il repose (chapitre 4).

[6] On se référera au travail de Sun [2001] pour une discussion de la fécondité et de la nécessité des liens entre SMAs et sciences cognitives.

Deuxième partie

Contributions

Chapitre 4

Modéliser l'engagement social et son respect

4.1 Introduction

Avant de développer notre approche des aspects cognitifs de la pragmatique des communications agents (chapitres 6 et 7), nous devons définir un langage de communication agent dont nous discuterons l'usage. Ce langage de communication, basé sur les jeux de dialogue, est présenté au chapitre suivant tandis que nous introduisons et discutons dans ce chapitre le modèle de l'engagement social sur lequel il repose[1]. Ce modèle a été développé en collaboration avec Roberto Flores lors de son séjour post-doctoral au sein du laboratoire DAMAS [Dialogue, Apprentissage et systèmes Multi-AgentS][2].

Dans les années passées et comme indiqué dans les sections 2.3.3 et 2.3.7, la communauté de recherche sur les systèmes multi-agents s'est concentrée, à produire un cadre interactionnel susceptible de résoudre les problèmes attachés aux approches intentionnelles, strictement mentalistes. Cela a mené à l'apparition des modèles de communication entre agents basés sur les engagements sociaux (tels qu'introduits en section 2.3).

Dans ce chapitre, nous analysons les avantages des approches conventionnelles et sociales pour la communication dans les systèmes multi-agents ouverts et hétérogènes. À cet effet, nous indiquons les conditions sous lesquelles les modèles basés sur les engagements sociaux peuvent permettre la communication entre agents hétérogènes dans les systèmes multi-agents

[1] Ce chapitre reprend et approfondi des éléments publiés dans le cadre du cinquième workshop international : Engineering Societies in the Agent World [Pasquier et al., 2004b].

[2] http ://www.damas.ift.ulaval.ca/

ouverts (section 4.2). Nous introduisons ensuite la problématique du respect des engagements (section 4.3) qui nous semble conditionner les approches basées sur les engagements sociaux. Cette problématique est analysée et une ontologie des outils de contrôle social (section 4.4.1) et des philosophies de punitions (section 4.4.2) est introduite. On présente ensuite notre modèle générique de l'engagement social (section 4.5.1), en indiquant quand et pourquoi les éléments de contrôle social peuvent y être attachés. Le méta-problème du respect des sanctions est alors introduit et discuté (section 4.5.2).

4.2 Avantages des approches sociales pour la communication dans les systèmes multi-agents ouverts

Un système multi-agents est dit *ouvert* si les propriétés suivantes sont réunies [Hewitt, 1991] :

1. Le comportement des agents et leurs interactions ne peuvent être prévus à l'avance ;
2. L'architecture interne des agents n'est pas connue publiquement ;
3. Les agents n'ont pas nécessairement de buts, d'intentions ou de désirs communs.

La première propriété implique que l'exécution d'un système multi-agents ouvert est *non-déterministe*, cela signifie que les interactions dans les sociétés d'agents ouvertes ne peuvent être anticipées. La seconde propriété implique qu'un système multi-agent ouvert peut regrouper des agents *hétérogènes*, c'est-à-dire d'architectures internes différentes. La troisième propriété implique que les membres d'un tel système ouvert peuvent échouer à ou refuser de se conformer aux aspects normatifs du système multi-agent afin d'atteindre leurs buts. Une dernière particularité des systèmes multi-agents ouverts, présente dans la littérature, est que les agents peuvent y entrer et en sortir à n'importe quel moment. Généralement, un protocole d'assignation de rôle permet d'entrer dans le système.

L'autonomie, la pro-activité et la rationalité des agents délibératifs traditionnels (introduits en section 1.1.2) reposent sur la manipulation et l'utilisation appropriée d'états mentaux. Dans les systèmes utilisant ce type d'agents, l'ordre social est généralement maintenu par de fortes hypothèses de sincérité, de coopération et/ou de collaboration. Aussi, les langages de communications agents dotés de sémantiques mentalistes, qui ont été proposés dans ce cadre, ont été critiqués puisque que l'hypothèse de sincérité et le problème de vérifiabilité qui leurs sont attachés interdisent de les utiliser dans des systèmes ouverts. En effet, la

seconde propriété des systèmes ouverts énoncée ci-dessus rend alors la vérifiabilité de ces systèmes impossible et l'hypothèse de sincérité intenable. Pire, la formulation des sémantiques mentalistes en termes d'états mentaux privés, internes aux agents, contraint l'architecture interne des agents (qui doit alors implanter des états mentaux tels que ceux utilisés dans la sémantique ou être au moins capable de les manipuler) ce qui enfreint la seconde propriété des systèmes multi-agents ouverts, et nuit à l'hétérogénéité.

Nous avons vu (section 2.3.7) que l'introduction de la notion d'engagement social dans les modèles de dialogue entre agents résout partiellement ces problèmes : l'hypothèse de sincérité n'est plus nécessaire et la vérifiabilité est facilitée. La notion d'engagement social rend compte des aspects publics de la conversation sans référer explicitement aux états mentaux des agents ni à leur architecture interne, ce qui autorise le dialogue d'agents hétérogènes. Pour autant, si les engagements sociaux, capturent les responsabilités contractées par les agents les uns envers les autres, le système résultant ne fonctionnera que dans la mesure ou ces engagements sociaux sont généralement respectés. La section suivante introduit le respect des engagements sociaux comme problématique à part entière.

4.3 Le problème du respect des engagements flexibles

Nous considérons que les engagements sociaux sont des attitudes sociales, publiques qui doivent être discriminées des autres attitudes propositionnelles normatives. Plus précisément, la notion d'engagement social doit offrir plus de *flexibilité* que les obligations, interdictions ou nécessités tout en étant plus contraignante que de simples permissions ou possibilités. Ainsi :

1. *un engagement peut ne pas être respecté* : la possibilité de violation d'un engagement n'est pas prise en compte par les formalisations considérant l'engagement social comme une nécessité. En effet, de par leur nature et leur objet, les logiques alétiques ne permettent pas de considérer explicitement la violation. On pense à l'axiome T : $\Box(p) \rightarrow p$, qui indique que les nécessités sont nécessairement respectées. Si les logiques déontiques écartent l'axiome T (lui préférant l'axiome D), elles sont tout de même trop contraignantes pour modéliser l'engagement social tel que nous l'envisageons. En effet, par nature, les obligations ne doivent pas être violées. Ceci est exprimé au sein des logiques déontiques par la définition même des notions duales d'obligation et de permission qui indique clairement qu'il n'est pas permis de ne pas faire ce qui est obligatoire : $O(p) \equiv \neg P(\neg p)$;

2. *un engagement peut-être annulé unilatéralement* : le désengagement unilatérale doit être permis et considéré dans la modélisation de la notion d'engagement social. Cette possibilité est male prise en compte par les approches qui réduisent la notion d'engagement à celle d'obligation dirigée ;

3. *un engagement préalablement accepté peut-être modifié par le dialogue* : la possibilité de modifier par le dialogue les engagements préalablement acceptés est également difficilement prise en compte par les approches déontiques des engagements. De manière générale, la non-monotonie est encore un sujet de recherche en logique et aucune approche ne c'est imposé comme solution satisfaisante.

Ces trois possibilités dont nous souhaitons rendre compte dans notre modèle invalident la plupart des modèles de l'engagement rencontrés, comme ceux de Castelfranchi [1995] ou de Royakkers et Dignum [2000], qui réduisent la notion d'engagement à celle d'obligation orientée (par exemple à l'aide de l'axiome suivant : $C(A, B, p) \rightarrow O_A^B(p)$ qui indique que l'engagement de A envers B sur le contenu p implique une obligation de A envers B à ce que p tienne.).

La flexibilité des engagements sociaux, conséquence directe des ces trois possibilités de manipulation à des conséquences théoriques et pratiques importantes pour les systèmes multi-agents reposant sur cette notion. En effet, les engagements sociaux tiennent pour les cognitions sociales primitives susceptibles de capturer l'interprétation commune résultant des conversations entre agents, c'est-à-dire leur sémantique. En fait, les engagements sociaux sont utilisés, entre autres, pour capturer les conséquences des dialogues entre agents. Dès lors, c'est cette *flexibilité sémantique*[3] des engagements qui permet aux agents de revenir sur les dialogues passés et leurs conséquences pour mieux prendre en compte la dynamique du système et de son environnement. Ces possibilités d'annulation, de modification ou de violation sont donc des caractéristiques essentielles d'un modèle de l'engagement social. Dans l'exemple de dialogue suivant, l'agent A s'engage envers l'agent B à envoyer un courriel à $20h$, puis modifie cet engagement pour $21h$:

1.A : Je t'envoie les résultats de l'analyse par courriel à 20h.

2.B : Ok.

. . . dix minutes plus tard, . . .

3.A : Heu, pour les résultats de l'analyse.

[3] On prendra soin de ne pas confondre cet objectif de flexibilité sémantique des engagements avec l'objectif de flexibilité structurale du cadre dialogique dont on reproche le manque aux protocoles (section 2.3.2), par exemple.

4.B : Oui ?

5.A : Finalement, j'enverrai le mail à 21h, est-ce que cela te vas ?

6.B : Oui, parfait.

Ce type de flexibilité sémantique, envisagée comme un moyen de revenir sur les dialogues passés, si elle peut être couramment observée dans le langage naturel nous semble également souhaitable pour les systèmes multi-agents. Elle marque la non-monotonie des engagements sociaux, absente des cadres déontiques auxquels ils sont trop souvent assimilés.

Une conséquence de cet objectif de flexibilité pour les engagements est que leur respect ne peut être imposé, comme c'est le cas avec les obligations dans les approches déontiques classiques. En effet, le respect des obligations est généralement assuré par réglementarisme (au sens du terme anglais *regimentation*), c'est-à-dire sans même qu'il soit possible de les violer. Une question reste alors en suspens : que se passe-t-il si les agents ne respectent pas leurs engagements ? Les approches des systèmes multi-agents basées sur les engagements sont valides et utiles sous l'hypothèse que les engagements sont généralement respectés. Cette hypothèse mérite d'être étudiée plus avant dans le cas des engagements flexibles. En particulier, il convient de préciser les éléments par lesquels cette hypothèse pourra être rendue vraie, garantie dans les systèmes multi-agents ouverts et hétérogènes. C'est le problème du *respect des engagements sociaux*. La prochaine section introduit et discute une ontologie des mécanismes de contrôle social qui peuvent être utilisés pour garantir conjointement flexibilité et parité quand au respect des engagements. Nous introduirons ensuite notre modèle de l'engagement social (section 4.5.1) défini précisément pour permettre l'introduction dans les systèmes multi-agents ouverts des outils de contrôle social présentés ci-bas.

4.4 Ontologie des sanctions et des mécanismes de contrôle social

Introduit en sociologie à la fin du 19ième siècle, le concept de *contrôle social* dénotait déjà la capacité d'un groupe ou d'une société à ce réguler elle-même en assurant la cohérence et l'unité de la vie sociale [Martindale, 1978]. En ce sens, le contrôle social s'interroge sur la coordination d'actions et soulève la question de l'origine et de la possibilité de l'ordre social. Le contrôle social est généralement envisagé comme un processus englobant, représentant pratiquement tous les phénomènes susceptibles d'assurer la conformité sociale. Considéré comme l'ensemble des mécanismes sociaux contraignant le comportement des membres du groupe étudié, il peut-être vu comme le liant qui tient la société unifiée [Hechter et Opp,

2001]. Plus formellement, seront considérés de contrôle social, tous les mécanismes mis en jeu pour réagir, prévenir, réduire et détecter les violations afin d'assurer le respect des normes sociales.

Les théories modernes du contrôle social se concentrent sur les stratégies et techniques qui visent à réguler le comportement des agents pour assurer la conformité et la soumission aux règles de la société aux niveaux micro et macro. Dans le reste de cette section, nous allons détailler ce que nous pensons être les composantes du contrôle social qui peuvent être utiles pour assurer la conformité dans les systèmes multi-agents basés sur les engagements sociaux. Cette conformité passe par le respect de la norme sociale indiquant que les engagements doivent être généralement respectés. Les principaux outils que nous avons isolés pour assurer le respect des engagements sociaux sont :

- *les sanctions* (section 4.4.1), qui sont considérées ici autant pour leur sens général d'incitatif que pour leurs effets compensatoires. La section suivante présente une ontologie des sanctions selon leur différentes dimensions ;
- *les philosophies de la punition* (section 4.4.2) qui sont les techniques qu'une société utilise pour faire respecter ses normes. On présentera les différentes philosophies de punition et les stratégies de punition résultantes qui permettent de déterminer quels type et magnitude de sanction appliquer ainsi que la manière dont de telles sanctions sont établies.

4.4.1 Les sanctions

Par simplification, nous ne considérerons que les *sanctions individuelles*, laissant les *sanctions collectives* [Levinson, 2003], qui peuvent être associées aux équipes, rôles ou groupes d'agents, de coté. Les sous-sections suivantes introduisent les trois dimensions principales des sanctions que nous avons isolées : la direction, le type et le style.

La direction des sanctions

Deux directions sont envisageables pour les sanctions :

- *les sanctions positives* : les sanctions positives sont des récompenses qui encouragent la poursuite du comportement qui les suscite. Par exemple, il est courant dans les systèmes

ouverts qu'un agent accepte de s'engager sur une tâche seulement si la récompense associée est suffisante ;

– *les sanctions négatives* : au contraire, les sanctions négatives sont utilisées pour prévenir les comportements susceptibles de violer la norme.

Les sanctions positives sont donc des récompenses incitatrices tandis que les sanctions négatives, outre leur caractère compensatoire, visent à prévenir les violations. Par simplification, dans le reste du texte, nous nommerons sanctions les sanctions négatives et récompenses les sanctions positives.

Types de sanction

Le premier type de sanction regroupe les sanctions dites *automatiques*, qui surviennent lorsque les actions violatrices portent en elle-même une pénalité du fait qu'elles ne sont pas coordonnées avec le reste de la société. Par exemple, un automobiliste qui conduit du mauvais côté de la route a une probabilité plus forte que la normale d'avoir un accident. Nous ne considérerons pas ce type de sanctions structurelles et résiduelles, qui surviennent de manière involontaire (en ce sens que personne ne décide de les appliquer).

Dans la vaste littérature abordant ce sujet, avec des perspectives aussi différentes que celles de l'économie, de la criminologie, de la sociologie, de la psychologie sociale ou encore de l'intelligence artificielle et des systèmes multi-agents, on rencontre généralement trois types de sanctions non-automatiques : (1) les sanctions matérielles, (2) les sanctions sociales et (3) les sanctions psychologiques.

Les *sanctions matérielles* rassemblent les sanctions physiques comme la violence ou les corvées et autres actions réparatrices ainsi que les sanctions impliquant des donations de biens matériels comme les sanctions financières. Bien que cela soit discutable, les interdictions ou obligations de toutes nature peuvent être considérées comme sanctions matérielles. Le bannissement est un exemple d'interdiction tenant pour une sanction matérielle qui peut s'averrer utile dans les SMAs ouverts de la même manière qu'il l'est dans la plupart des communautés virtuelles.

Il existe également différents types de *sanctions sociales*. La confiance, la crédibilité et la réputation sont les trois principales valeurs qui peuvent être affectées par ce type de sanction. Posner et Rasmusen [1999] indiquent que les sanctions sociales sont généralement la conséquence d'une divulgation d'informations associées à la violation elle-même que le violateur aurait voulu tenir secrètes autrement. Par exemple, un agent qui viole un engagement sans

raison apparente indique (généralement non-intentionnellement) aux autres agents qu'il n'a que peu faire de la réalisation de cet engagement. Pour autant, cette violation sera prise en compte négativement lors de l'évaluation de sa réputation par les agents concernés.

Les *sanctions psychologiques*, si elles sont plus utilisées pour les agents impliqués dans des communautés mixtes ou dans les systèmes dédiés à la simulation sociale pour des raisons de réalisme peuvent s'avérer importantes. On distingue alors :

- *la culpabilité* : le violateur se sent mal concernant sa violation du fait de sa connaissance de la norme sociale enfreinte, et ce indépendamment des conséquences externes que sa violation pourrait avoir. Dans les sociétés humaines, par exemple, la plupart des individus se sentent coupables d'avoir volé, et ce, même s'ils sont certains de ne pas être découverts ;
- *la honte* : le violateur à l'impression que son action l'a abaissé, soit à ces propres yeux ou bien à ceux des autres. Dans sa forme la plus commune, la honte survient lorsque la violation est découverte par les autres et que leur jugement négatif s'exprime. Cette condition n'est cependant pas nécessaire puisque que le violateur peut aussi bien se sentir honteux sans que la violation ai été découverte.

L'horizon temporel des sanctions est également un facteur d'importance que nous associons au type. Les sanctions peuvent avoir un effet immédiat, ponctuel, après la violation (ou sa découverte) ou bien s'appliquer sur une durée plus longue. En particulier, il convient de distinguer les sanctions susceptibles de persister dans le temps comme les sanctions psychologiques et certaines sanctions sociales des sanctions ponctuelles comme les sanctions matérielles d'application immédiate. Il n'y a cependant pas lieu de généraliser et les sanctions matérielles peuvent être permanentes (comme certaines interdictions) ou s'étaler dans le temps (comme des versements réguliers) et symétriquement les sanctions sociales ou psychologiques peuvent être ponctuelles. Dans tous les cas, des phénomènes complexes et subtils comme le pardon nécessitent que ces aspects temporels soient pris en compte.

Styles de sanction

Pour les besoins formels spécifiques aux systèmes multi-agents, il convient de distinguer les sanctions explicites des sanctions implicites. Les *sanctions implicites* sont déterminées de manière autonome, individuelle et unilatérale par l'agent qui sanctionne. La principale difficulté associée à ce style de sanctions est qu'elles ne sont pas connues publiquement et que l'agent sanctionné doit le découvrir par lui-même (par exemple, en constatant qu'un autre

agent ou groupe d'agents ne communique plus avec lui). Les sanctions sociales et psychologiques tombent généralement dans cette catégorie.

Les *sanctions explicites*, au contraire des précédentes, sont publiquement connues (au moins parmi les agents conversant). Une autre distinction utile peut être faite entre sanctions décidées *a priori* et sanctions déterminées *a posteriori* selon que le type et la nature de la sanction ont été déterminés explicitement avant la violation ou après. En particulier, les sanctions décidées a posteriori devraient être évitées dans les systèmes multi-agents, puisqu'elles ne permettent pas aux agents de raisonner sur les tenants et les aboutissants du respect de leurs engagements et donc d'agir en conséquence. Dans ce cas, l'agent puni peut donc contester la sanction assignée a posteriori, ce qui entraînera nécessairement des complications.

Dans la suite du texte, nous ne considérerons que les sanctions explicites *définies a priori*. Parmi celles-ci, nous pouvons encore distinguer entre les sanctions explicites *statiques*, décidées a priori selon un système de sanction commun et implanté à la conception du système et les sanctions explicites a priori négociées *dynamiquement* par les agents au cours de leurs dialogues.

4.4.2 Les différentes philosophies de punition

Un mécanisme de contrôle social pour le respect des engagements doit être conçu selon une philosophie de la punition. Par punition, on désigne l'imposition de sanctions afin de satisfaire le désir de rétribution des concepteurs envers les agents fautifs. Les théoriciens du contrôle social distinguent cinq philosophies de la punition desquelles toutes les stratégies de punition peuvent être dérivées [Vold et al., 2002]. Nous présentons brièvement chacune de ses familles en discutant leur pertinence pour les systèmes multi-agents :

- *la dissuasion* : issue de l'école classique de criminologie et supportée par des philosophes tels que Beccaria [1963] ou Bentham [1970], la dissuasion est un principe utilitariste qui postule que le rôle des sanctions est de prévenir les violations futures. Pour que la dissuasion fonctionne, il faut que la punition soit sévère et certaine. Appliqué au respect des engagements dans les systèmes multi-agents, cela signifie que des sanctions explicites importantes doivent être associées aux engagements. Cette position extrême, insistant sur l'effet prohibitif de sanctions sévères, a pour conséquences de réduire les engagements sociaux à de simples obligations au détriment de l'objectif de flexibilité des engagements sociaux énoncé en section 4.3.
- *la réparation* (*au sens de l'anglais retribution*) : cette philosophie considère que la violation doit être réparée par une pénalité dont la sévérité doit équivaloir au tort commis.

Pour des raisons qui seront détaillées dans la prochaine section, cela semble un choix praticable et sous certaines conditions optimal pour les systèmes multi-agents ouverts.

– *l'incapacitation*[4] : l'incapacitation vise à altérer ou restreindre la capacité de l'agent violateur à récidiver. Dans les sociétés humaines : l'exclusion, le licenciement et l'emprisonnement sont les principales méthodes d'incapacitation. Appliquée au respect des engagements sociaux dans les systèmes multi-agents, cela peut signifier la perte de certains privilèges ou l'exclusion des agents fautifs. Bien que cela dépende grandement du système considéré, l'incapacitation temporaire ou définitive résulte en une perte d'activité pendant le temps de la punition qui se traduit elle-même par une sanction matérielle. Si l'on considère, par exemple, un système de commerce électronique, il est clair qu'à une perte de temps est associée une perte d'opportunité qui a des conséquences économiques. Ainsi, puisque cette philosophie se réduit à l'application de sanctions matérielles explicites, nous la considérerons dans la suite du texte comme un cas particulier de réparation ou de dissuasion.

– *la réhabilitation* : le processus de réhabilitation vise à corriger le comportement fautif, notamment en re-conditionnant le sujet au respect de la norme. Cette philosophie semble plus difficile à implanter dans les systèmes multi-agents ouverts, puisqu'elle suppose que le mécanisme de décision qui a mené au comportement fautif est modifiable. Or, dans un système multi-agent ouvert, ce mécanisme de décision n'est pas directement accessible. En fait, cette philosophie requerrait des mécanismes d'apprentissage pro-sociaux que l'on ne peut imposer aux agents dans un système multi-agent ouvert.

– *la restauration* (restoration) : la restauration vise à ré-unifier le fautif et la communauté, généralement au travers de rituels. Cette philosophie, présente essentiellement dans les pays les moins industrialisés ne sera pas traitée ici car elle pré-suppose des dimensions sociales et culturelles qui ne sont pas (encore) considérées pour les systèmes multi-agents.

On note que l'incapacitation, la réhabilitation et la restauration sont caractérisées par la nature de la punition plutôt que par le choix de sa magnitude. On peut, par exemple, songer à une incapacitation dissuasive ou simplement réparatrice. Ainsi, la dissuasion et la réparation qui sont définies en terme de l'intensité des sanctions à appliquer en laissant ouvert le choix du type et style de sanction, nous semble mieux adaptées à une spécification générique des systèmes multi-agents ouverts et hétérogènes. La section suivante présente notre argumentaire pour favoriser une philosophie de la punition réparatrice plutôt que dissuasive.

[4] De l'anglais *incapacitation*.

4.4.3 Sanctions et optimalité

Les travaux de droit et d'économie des dernières décennies ont fourni au moins deux raisons pour que l'intensité des sanctions égale celle des dommages. Dans notre cas, le dommage est la violation d'un engagement social et sa magnitude est au moins équivalente à l'effort nécessaire à sa satisfaction.

Le premier argument concerne le *degré de précaution* des parties, où le terme précaution est à prendre dans toute sa généralité. Si les sanctions sont moindres que les dommages (c'est à dire, dans notre cas, que l'effort nécessaire pour respecter l'engagement), le degré de précaution sera trop faible et les agents risquent de (et même auront intérêt à) ne pas respecter les engagements pris, lorsque cela est à leur avantage. Symétriquement, si la magnitude des sanctions excède les dommages, le degré de précaution sera trop élevé et les agents ne s'engageront plus sur les engagements désirés. En effet, un agent rationnel ne prendra pas, même s'il le souhaite, un engagement dont la sanction attachée en cas de violation, même involontaire, est prohibitive. Ainsi, il a été montré que si les sanctions égalent les dommages, cela donne aux parties un incitatif optimal pour avoir le bon degré de précaution [Polinsky et Shavel, 1998].

La seconde raison pour laquelle il est désirable que les sanctions égalent les dommages implique *le niveau d'activité* des agents, c'est-à-dire la propension de l'agent à participer à des activités risquées. Indépendamment du niveau de précaution de l'agent, son niveau d'activité affecte la magnitude des dommages totaux causés. Plus grand est le nombre d'engagements contractés, plus grand sera le nombre d'accidents (c'est-à-dire de violations involontaires), plus grand seront les dommages causés, et ce, indépendamment du niveau de précaution de l'agent (qui influence le dommage espéré par engagement) [Polinsky et Shavel, 1998].

Signalons que ces résultats tiennent sous l'hypothèse implicite que les parties sont *neutres envers le risque* (risk neutral). Si, en revanche, le violateur a une aversion au risque, l'intensité optimale pour les sanctions devient inférieure à celle des dommages, permettant ainsi de réduire l'intensité du risque pour compenser l'aversion de l'agent. Autrement dit, dans ce cas, les sanctions n'ont pas besoin d'être aussi élevées pour jouer leur rôle d'incitation au respect des engagements. Finalement, une seconde hypothèse est nécessaire pour que ce résultat d'optimalité tienne : la *responsabilité sûre* (strict liability). La responsabilité sûre postule que les agents coupables seront identifiés et punis de manière certaine, c'est-à-dire qu'ils ne peuvent pas échapper aux sanctions lorsque celles-ci s'appliquent.

4.5 Modélisation pour les systèmes multi-agents

Rappelons que les engagements sociaux sont des attitudes sociales qui doivent être socialement établies. Dans les approches conventionnelles et sociales (voir section 2.3.3), la communication agent est le processus social par lequel la couche des engagements sociaux - qui capture la majeure partie des dépendances entre les agents - est manipulée. Le succès d'une unité de dialogue[5] est l'acceptation sociale de l'opération proposée (création, annulation, modifier, décharger) sur la couche des engagements sociaux tandis que la satisfaction d'une unité dialogique est liée aux conditions de satisfaction des éventuels engagements résultant, qui sont les conditions sous lesquelles l'engagement est satisfait.

Trois éléments nous semblent d'une importance cruciale pour pouvoir utiliser un langage de communication basé sur les engagements sociaux dans des systèmes multi-agents ouverts et hétérogènes. Il s'agit de fournir un modèle clair et générique des engagements sociaux qui :

1. supporte la flexibilité : cela signifie, entre autres, que le respect des engagements doit être pris en compte dans leur modélisation ;
2. indique la manière dont les sanctions explicites sont traitées, par exemple en indicant quand elles sont attachées aux engagements et quand elles doivent s'appliquer, le cas échéant ;
3. supporte une variété de stratégies de punitions.

En effet, les concepteurs d'un système multi-agent peuvent vouloir utiliser un système de sanctions explicite et statique qui s'applique pour tout le système multi-agent et qui n'implique que des sanctions monétaires, tout comme ils peuvent décider que ce sont les agents qui devront décider dynamiquement d'actions réparatrices (un autre type de sanction). La section suivante introduit notre modèle de l'engagement social qui sous l'hypothèse de n'utiliser que des sanctions explicites définies a priori, respecte les trois désiratas précédents.

4.5.1 Modèle de l'engagement social et de son respect

Afin d'introduire les sanctions dans les approches de la communication agent basés sur les engagement sociaux, il nous faut développer un modèle de l'engagement social adapté. En particulier, puisque les sanctions explicites définies à priori font partie intégrante de la

[5] Une telle unité de dialogue sera représentée par un jeu de dialogue dans notre approche.

sémantique des engagements, le modèle doit indiquer clairement quand les sanctions doivent être attachées à l'engagement et quand, dans le cycle de vie de l'engagement, elles doivent s'appliquer. La nature des sanctions doit apparaître clairement et être socialement établie et le modèle doit permettre une grande variété de types et de styles de sanctions afin de respecter le désirata numéro 3 (c'est-à-dire, donner la liberté aux designers de déterminer le système de sanctions le plus adapté à leur système). Cette section présente un tel modèle, celui-ci est compatible avec ce qui a été dit des engagements sociaux en section 2.3.3.

Conceptuellement, les engagements sociaux sont des responsabilités orientées envers un agent ou un groupe d'agents. Les engagements sont représentés par un prédicat d'arité 6. Un engagement accepté prend la forme :

$$C(x, y, \alpha, t, s_x, s_y)$$

signifiant que le débiteur x est engagé envers le créditeur y sur le contenu α au temps t, avec les sanctions s_x et s_y. Nous détaillerons quand et comment ces sanctions sont attachées aux engagements ainsi que quand elles s'appliquent plus loin dans cette section. La notation précédente, inspirée des travaux de Singh [2000], permet la composition des actions et des propositions impliquées dans l'engagement grâce à l'opérateurs de choix ($\alpha_1|\alpha_2$ tient pour le choix) et l'opérateur conditionnel ($\alpha_1 \Rightarrow \alpha_2$ indique que α_2 est conditionnel à l'occurrence (ou la vérité) de α_1). Finalement, chacun des agents impliqué dans l'engagement garde trace des engagements pour lesquels il est débiteur ou créditeur dans son *agenda* qui est un type de tableau d'engagement (notion introduite en section 2.3.3) distribué sur lequel nous reviendrons.

La figure 4.1 présente notre modèle d'engagement sous la forme d'une machine à états finis, c'est-à-dire d'un graphe d'états/transitions. On y indique qu'un engagement peut être accepté (noté $C(x, y, \alpha, t, s_x, s_y)$) ou rejeté (noté $\neg C(x, y, \alpha, t, s_x, s_y)$). On se gardera de confondre un engagement rejeté ($\neg C(x, y, \alpha, ...)$) d'un engagement accepté ayant un contenu négatif ($C(x, y, \neg\alpha, ...)$). Si un engagement peut être accepté parce qu'il résulte d'une des conventions du système, qu'il fait parti d'une structure organisationnelle ou qu'il est associé au rôle d'un agent, nous nous limiterons dans la suite aux engagements liés à l'activité dialogique.

La figure 4.1 détaille les différents états dans lesquels un engagement peut se trouver :

- *inactif* : par défaut, un engagement social est rejeté (c'est un non engagement) et inactif. C'est une hypothèse de monde clos[6] que nous supposons pour les engagements sociaux : tout engagement social non explicitement accepté est rejeté et inactif.

- *actif* : un engagement social est actif s'il a été explicitement accepté et que ses conditions de satisfaction (notées CoS sur la figure 4.1) peuvent être remplies.

- *violé* : un engagement violé est un engagement socialement accepté dont les conditions de satisfaction ne peuvent plus être remplies selon les modalités indiquées par l'engagement.

- *satisfait (ou rempli)* : un engagement social est satisfait si ses conditions de satisfaction sont remplies. Un engagement social est donc satisfait si son contenu a été effectué (dans le cas d'un engagement en action) ou s'il s'avère être vrai (dans le cas d'un engagement propositionnel).

- *annulé* : un engagement social est annulé s'il a été socialement établi comme étant rejeté. Cet état se rencontre dans des circonstances variées : (1) quand un engagement actif a été rejeté par le dialogue, (2) quand les sanctions (éventuellement nulles) associées avec un engagement violé ont été acquittées, signifiant que leur effet compensatoire permet d'annuler l'engagement ou (3) quand les récompenses (éventuellement nulles) associées à un engagement rempli ont été acquittées. Un engagement annulé redevient un non-engagement, c'est-à-dire un engagement rejeté.

La figure 4.1 indique également les différentes transitions existantes entre ces états. Ce sont les transitions nécessaires pour que la flexibilité des engagements soit supportée. Sans détailler l'analyse qui nous a menés à ce modèle, nous allons indiquer comment le traitement des sanctions s'intègre à ces transitions afin de satisfaire le désirata numéro 2. Ces transitions sont :

1. *la création* : la création n'implique pas l'application de sanction, mais c'est la transition pendant laquelle les sanctions sont attachées à l'engagement social. La négociation de l'engagement social inclue celle des sanctions attachées. Cette dernière peut être complexe dans le cas d'un système de sanctions explicites, dynamiquement négociées, ou bien triviale dans le cas d'un système de sanctions explicites statique partagé et commun (les agents n'ont alors qu'à vérifier que les sanctions attachées sont bien celles prescrites par le système de sanction commun).

[6] Nous référons le lecteur aux ouvrages de Thaise et al [1988] pour les pré-requis concernant les approches logiques de l'intelligence artificielle.

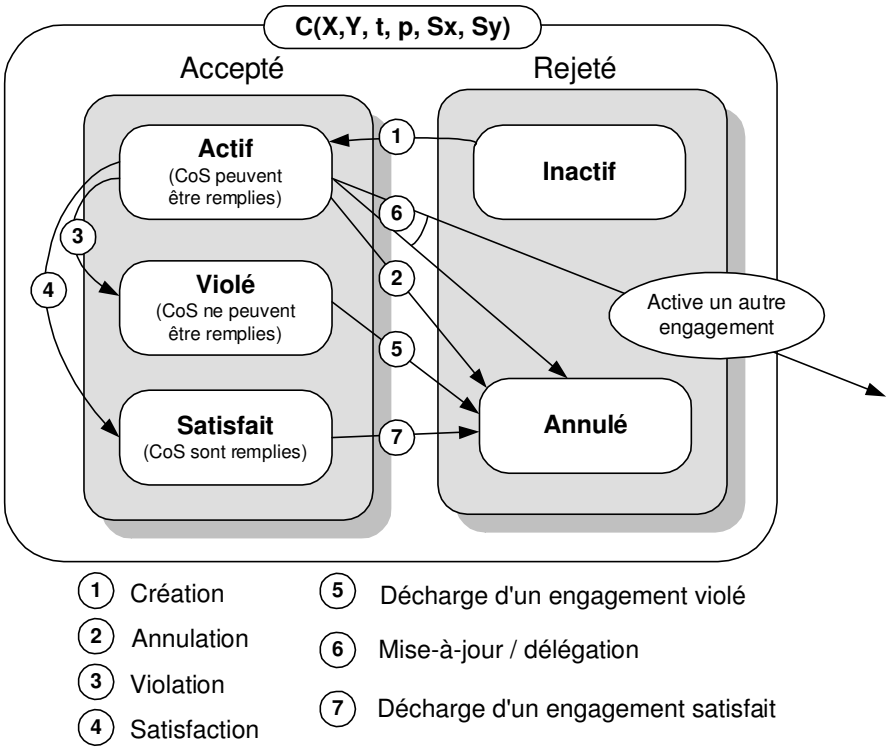

FIG. 4.1 – Un modèle de l'engagement social.

2. *l'annulation* : l'annulation concerne le rejet d'un engagement accepté et actif et implique l'application éventuelle de sanctions. Puisque la motivation de certaines annulations peut être partagée, le cadre interactionnel autorise l'interlocuteur (l'agent non-initiateur de l'annulation) à décider si les sanctions prévues à cet effet s'appliquent ou non. Seules les sanctions associées à l'agent qui initie l'annulation s'appliquent.

3. *la violation* : la violation elle-même n'est pas nécessairement consommée par un processus social (comme le dialogue) ce qui signifie que l'état violé n'est pas toujours socialement établi. Il est même possible d'établir socialement (par le dialogue) un engagement qui soit déjà violé, par exemple en ce référant au passé. Ainsi, la violation elle-même n'implique aucune sanction, mais c'est la décharge (voir ci-bas), socialement établie, de l'engagement violé qui impliquera l'application des sanctions.

4. *la satisfaction* : de manière analogue à la violation, la satisfaction n'est pas nécessairement supportée par un processus social et il est même possible de créer un engagement qui soit déjà satisfait, par exemple en se référant au passé. La satisfaction elle-même n'implique pas de traitement sur les sanctions et c'est lors de la décharge, socialement établie, que les éventuelles récompenses interviendront.

5. *la décharge d'engagement violé* : la décharge d'un engagement violé se fait par un dialogue pendant lequel : (1) il est socialement accepté que l'engagement a été violé (c'est-à-dire que ses conditions de satisfaction ne peuvent plus être remplies et c'est irréversible), (2) des sanctions s'appliquent. Généralement, seules les sanctions associées au débiteur de l'engagement s'appliquent puisqu'il était responsable de la satisfaction de l'engagement. L'application des sanctions est généralement indépendante de l'implication du débiteur dans la violation.

6. *la modification*[7] : la modification d'un engagement actif est une double transition qui consiste à simultanément (1) annuler l'engagement considéré et (2) créer un nouvel engagement (supposément différent de l'original, ce sans quoi il n'y a pas modification). La modification doit être socialement acceptée par le dialogue et des sanctions peuvent s'appliquer. Cependant, la modification se différencie d'une séquence < annulation ; création > précisément par la manière dont les sanctions peuvent être appliquées. C'est d'ailleurs la raison pour laquelle cette transition a été introduite dans le modèle. En fait, il est fréquent que la modification d'un engagement (qui peut être motivée de différentes manières) permette d'éviter tout ou partie des sanctions qui seraient associées à son annulation. C'est, une fois de plus, à l'interlocuteur (qui n'est pas à l'initiative de la modification) qu'il revient de décider si les sanctions s'appliquent et dans quelle mesure.

7. *la décharge d'engagement satisfait* : la décharge d'un engagement satisfait (ou rempli) donne lieu à un dialogue au cours duquel : (1) il est socialement reconnu et accepté que l'engagement a été satisfait (et c'est irréversible) et (2) l'éventuelle récompense (incluse comme sanction positive au sein des sanctions du débiteur) s'applique.

Pour un engagement, on distingue donc (cela n'apparaît pas explicitement sur la figure) les états socialement établis des états non-socialement établis. Les états : inactif, violé et satisfait ne sont pas nécessairement socialement établis tandis que tous les autres doivent être atteints par un processus social d'établissement. Cela signifie que les transitions 1, 2, 5, 6 et 7 (repérée par la figure 4.1) sont consommées par le dialogue (ou un autre processus d'établissement social) à l'inverse des transitions 3 et 4. Dans toutes les modélisations des engagements sociaux rencontrées jusqu'alors, ces dernières transitions sont consommées de manière automatique. Même si spécifier comment implanter ces deux transitions dans un système distribué réel est une question de recherche ouverte, nous suggérerons une solution (celle que nous avons mise en pratique) dans le chapitre suivant (section 5.4.3).

[7] Certains auteurs [Singh, 1999] considèrent la délégation (changement de débiteur) et l'assignation (changement de créditeur) d'engagements comme des opérations à part entière. Nous les considérerons plutôt comme des cas particuliers de la modification. Néanmoins, parce que ces opérations impliquent trois agents plutôt que deux, nous ne les considérerons pas dans le reste de ce texte, laissant l'extention de notre approche au cas multi-parties comme perspective.

Finalement, dans nos engagements, de la forme $C(x, y, \alpha, t, s_x, s_y)$, s_x est un triplet associé à l'engagement à sa création qui indique les sanctions pour le débiteur en cas de violation, d'annulation et de modification respectivement, tandis que s_y est un triplet indiquant les sanctions pour le créditeur en cas de modification ou d'annulation et les récompenses en cas de satisfaction. Par exemple, l'engagement :

$$C(A, B, publish(SQL(req1), 0am), dc, (12\$, 8\$, 4\$), (6\$, 4\$, 4\$))$$

indique que l'agent A est engagé envers l'agent B depuis la date de création de l'engagement (dc) à publier les données indiquées par la requête SQL à minuit. L'agent A devra payer 12$ à B s'il viole cet engagement, 8$ s'il annule et 4$ s'il le modifie tandis que l'agent B devra payer 6$ à A s'il annule l'engagement, 4$ s'il modifie celui-ci et 4$ si l'engagement est respecté. Peu importe que ces sanctions aient été attachées à l'engagement par les agents eux-mêmes ou qu'elles aient été déduites d'un système de sanctions commun et partagé, la sémantique de l'engagement reste celle indiquée ci-dessus. De manière similaire, l'engagement :

$C(A, B, \Delta, dc, (-, -, -), (-, -, -))$, où
$\Delta = lever_B(ObjetA, t) \rightarrow C(B, A, poser{-}sous_A(ObjetX, ObjetY, t{+}1), dc, -, -)$

signifie que A est engagé envers B à ce que si (opérateur conditionnel, $\rightarrow$) A soulève l'$ObjetX$ au temps t alors B est engagé envers A à glisser l'objet $ObjetY$ sous l'objet $ObjetX$ tenu en l'air par A et ce sans qu'aucune sanction ne s'applique en cas de viol, de modification, d'annulation, ou de satisfaction (peut-être que des sanctions sociales, implicites, s'appliqueront).

4.5.2 Le problème du respect du système de contrôle social

Le problème du respect des engagements est - tel que discuté dans les sections précédentes - théoriquement résolu par l'utilisation d'un système de contrôle social et l'utilisation de sanctions attachées aux engagements lors de leur création qui assure la parité souhaitée pour les systèmes multi-agents ouverts. Si cette solution permet de conserver la flexibilité désirée pour les engagements, elle lève un autre problème. Il s'agit du problème du respect du système de contrôle social qui peut être formulé comme suit :

1. s'il sont respectés, les engagements sociaux assurent que les agents agissent conformément à leurs dialogues (le problème de la vérification est résolu) ;

2. si elles sont respectées, les sanctions (au travers des mécanismes de contrôle social discutés précédemment) assurent que les agents tendent à respecter leurs engagements ;

3. si (1) et (2) tiennent, qu'est-ce qui assure la prémisse de (2) qui permettrait d'inférer la conséquence de (1) ? C'est-à-dire, qu'est-ce qui assure le respect des sanctions ?

Ce problème peut être reformulé de la manière suivante : comment l'hypothèse de responsabilité sûre[8] peut-être garantie ? Il est clair que considérer la possibilité d'échapper aux sanctions rend les choses beaucoup plus compliquées. En effet, dans ce cas les sanctions doivent être sur-estimées pour prendre en compte cette possibilité d'impunité, ce qui a pour effet d'augmenter le niveau de précaution des agents. Si le monde réel ne peut sous-estimer cette possibilité[9], nous pensons que le cadre artificiel et formel des systèmes multi-agents permet de résoudre ce problème.

La solution que nous proposons consiste à ajouter des contraintes d'implantation au système multi-agents ouvert considéré. Nous distinguerons parmi les trois types de sanctions introduits en section 4.4.1 : (1) les sanctions matérielles, (2) les sanctions sociales et (3) les sanctions psychologiques.

Pour les sanctions de type (2) ou (3), nous postulons que c'est au mécanisme de décision de l'agent de les prendre en compte ou non. Rien ne peut forcer un agent à prendre en compte des sanctions sociales ou psychologiques dans son comportement futur. Cela dit, comme les sanctions sociales sont d'application externe à l'agent auquel elles s'appliquent, il ne pourra les éviter. Les sanctions sociales prendront effet même s'il ne le réalise pas. Le problème du respect des sanctions n'est donc pas à considérer dans le cas de sanctions psychologiques ou sociales.

Une solution pour le problème de la régression infinie du respect des sanctions matérielles explicites, lorsque celles-ci s'appliquent, est de considérer les sanctions matérielles comme de strictes obligations. Dans ce cadre, si les sanctions ne sont pas respectées, cela sera découvert et considéré comme une erreur système[10]. Cela peut-être implanter en s'assurant que :

1. *les règles de fonctionnement du cadre dialogique utilisé sont des obligations* : dans ce cas, les agents ne peuvent enfreindre les règles de fonctionnement du cadre dialogique.

[8] Rappelons que cette hypothèse, introduite section 4.4.3, postule que les agents fautifs seront découverts et sanctionnés de manière certaine, c'est-à-dire que tout dommage sera sanctionné.

[9] Cette possibilité d'échapper aux sanctions est un des arguments majeurs des défenseurs de la philosophie de punition par dissuasion.

[10] Ajoutons que nous verrons en section 5.4.3 comment cette partie de l'hypothèse de responsabilité sure est rendue vraie dans notre implémentation.

Le système d'application des sanctions faisant partie intégrante du langage de communication agent utilisé, via le modèle d'engagement décrit précédemment, le problème de l'établissement social de l'application des sanctions est ainsi résolu.

2. *le respect des sanctions applicables est une obligation stricte* et toute infraction est considérée comme une erreur système.

Les sanctions matérielles explicites sont plus simplement vérifiables que les autres types de sanctions. Il est possible d'utiliser un système de vérification central. De manière générale, les engagements sociaux publics peuvent être mémorisés sur un support sécurisé et le traitement des aspects légaux des systèmes multi-agents s'en trouve grandement facilité. Les engagements sociaux tiennent alors pour des contrats et l'on peut leur appliquer une variante du droit des contrats [Jaccard, 1996]. Dans ce cadre, l'engagement social est vu comme une relation entre trois et non plus deux agents, le troisième étant appelé témoin dans la littérature et joue alors un rôle crucial en ce qui concerne la vérification du respect des engagements [Conte et Castelfranchi, 1995].

4.5.3 Travaux connexes et discussion

Les aspects normatifs des systèmes multi-agents et le besoin de mécanismes de contrôle social [Castelfranchi, 2000] ont été étudiés dans le cadre de la modélisation des systèmes multi-agents et de la spécification des sociétés computationnelles, des organisations et des institutions [Artikis et al., 2002]. En particulier, puisque les sanctions et les mécanismes de punition influencent le modèle de décision des agents, plusieurs applications les utilisent déjà pour garantir le respect des engagements sociaux. Les contrats à niveaux d'engagement (levelled commitment contract), issus de la micro-économie, se sont montrés utiles pour prendre en compte la possibilité de désengagement unilatéral – présente chez les agents égoïstes (self-interested) – dans l'implantation de systèmes multi-agents [Sandholm et Lesser, 1995]. Il a même été démontré, à l'aide de la théorie des jeux, que cette possibilité de désengagement peut accroître les gains de tous les agents [Sandholm et Lesser, 1996], ce qui constitue un argument en faveur de la flexibilité des engagements introduite en section 4.3. D'autres modélisations des sanctions pour le désengagement d'agents non complètement coopératifs, utilisant la théorie de la décision, ont été relevées [Excelente-Toledo et al., 2001]. Cela montre l'intérêt pratique de ces problématiques. Cependant, aucune des approches rencontrées ne discute le problème du respect des engagements de manière générique et aucune ne lève explicitement ce problème dans le cadre des langages de communication agent basés sur les engagements sociaux. En particulier, aucun modèle des engagements en termes d'états et de transitions, intégrant l'application des sanctions, tel que celui proposé précédemment, n'a été produit.

4.6 Conclusion

Dans ce chapitre, nous avons discuté du problème du respect des engagements et introduit les notions de sanction et de philosophie de punition pour le traiter. Nous avons détaillé les conditions sous lesquelles le méta-problème du respect des sanctions peut-être évité.

En guise de conclusion, rappelons qu'un des intérêts majeurs du modèle de l'engagement social flexible et de son respect présenté dans ce chapitre est de sous-tendre notre approche de la modélisation dialogique basée sur les jeux de dialogue autant que celle de Flores reposant sur le *Protocol for Proposal* (introduite section 2.3.5). Aussi différents langages de communication agents, envisagés comme outils de manipulation des engagements sociaux, peuvent être proposés à partir de ce modèle. Le chapitre suivant détaille nos propositions théoriques et pratiques à cet égard.

Chapitre 5

Langage de communication agent par les jeux de dialogue (DIAGAL)

5.1 Introduction

Ce chapitre[1] présente le langage de communication agents basé sur les jeux de dialogue développé collectivement au laboratoire DAMAS [Dialogue et Apprentissage Multi-Agents][2]. Cet ACL a été nommé DIAGAL (section 5.2), acronyme anglophone pour *DIAlogue Games Agent Language*. Ce langage repose sur le modèle de l'engagement social développé au chapitre précédent. Son implémentation et son utilisation concrète via le simulateur de dialogue DGS, qui tient pour l'acronyme de *Dialogue Game Simulator*, sont présentées (section 5.3). Finalement, les possibilités d'utilisations (section 5.4), les avantages et limites de ce cadre dialogique sont détaillés et analysés en regard des autres approches existantes (section 5.5).

5.2 Le langage DIAGAL

Cette section ainsi que la suivante présentent le langage DIAGAL (DIAlogue Game Agent Language) tel que développé collectivement au laboratoire DAMAS entre 2001 et 2005. Le langage DIAGAL et son implémentation via le simulateur DGS (Dialogue Game Simulator)

[1] Le contenu de ce chapitre est une version largement révisée et complétée de certains éléments présentés dans les publications [Pasquier et Chaib-draa, 2004a] et [Pasquier et al., 2004a].

[2] http ://www.damas.ift.ulaval.ca/

sont issus d'un travail collectif qui déborde largement le cadre de cette thèse. Il inclut notamment des éléments de la thèse de Maudet [2001], de la maîtrise de Labrie [2003], d'une partie de la maîtrise de Bergeron [2005] et des stages de Bourget [2002] et Andrillon [2003] que nous avons encadrés. Ce chapitre présente la version actuelle de DIAGAL, envisagée spécifiquement pour cette thèse, c'est-à-dire pour sous-tendre les avancées théoriques concernant la pragmatique.

Nous allons commencer par décrire le mécanisme par lequel les engagements sociaux sont manipulés. Ce mécanisme est modélisé par la structure des jeux de dialogue.

5.2.1 Structure des jeux

La particularité principale des engagements sociaux est d'être des cognitions sociales qui doivent être socialement établies pour pouvoir être considérées comme telles. Cela signifie que tout changement de la couche sociale des engagements (réifiée dans les agendas) doit être établi socialement (*grounded*). On partage avec la majorité des autres approches sociales et conventionnelles, présentées section 2.3, l'idée que les jeux de dialogues sont les outils de manipulation de cette couche publique. Par contre, à l'inverse des autres modèles et dans la lignée des travaux de Maudet [2001], nous utilisons une approche homogène entièrement basée sur les engagements sociaux. En effet, dans notre approche des jeux de dialogue, les règles de dialogues - qui capturent la dimension conventionnelle du dialogue - sont elles-aussi exprimées comme des engagements sociaux. Ces engagements dits dialogiques, contractés dans le cadre d'un jeux et n'ayant de sens que dans ce cadre prennent l'identificateur du jeu en indice (concrètement, ces engagements seront notés C_g dans la suite du texte).

Dans DIAGAL, les jeux sont des structures bilatérales définies par quatre ensembles d'engagements :

- *des conditions d'entrée*, (E) : exprimées en termes d'engagements extra-dialogiques, les conditions d'entrée indiquent les conditions qui doivent être respectées pour qu'un jeu puisse être joué ;
- *des règles de dialogue*, (R) : exprimées en termes d'engagements dialogiques, les règles du dialogue spécifient ce que les agents, qui jouent le jeu courant, sont dialogiquement engagés à faire. La satisfaction de ces règles de dialogue mène aux conditions de succès ou aux conditions d'échec du jeu ;

– *des conditions de succès*, (S) : les conditions de succès d'un jeu de dialogue indiquent le résultat, c'est-à-dire l'effet en termes d'engagements extra-dialogiques, du jeu de dialogue si la manipulation proposée à été acceptée ;

– *des conditions d'échec*, (F) : les conditions d'échec indiquent l'effet du jeu en termes d'engagements extra-dialogiques si la manipulation de la couche des engagements sociaux proposée est rejetée.

5.2.2 Établissement et composition des jeux

La question spécifique de savoir comment les jeux sont établis lors du dialogue est l'une des plus délicates. L'établissement, rappelons-le, dénote le processus par lequel les croyances mutuelles (réunies dans ce qui forme le terrain commun) sont établies [Clark, 1996]. On le modélise par un processus en deux phases : présentation et acceptation (ou rejet). Dans l'esprit de [Reed, 1998], on met à la disposition des agents une série de méta-actes de communication qui permettent de manier les jeux de dialogues. Un jeu de dialogue peut avoir différents statuts, il peut être : *ouvert*, *fermé* ou simplement *proposé*. Un jeu de contextualisation, introduit par Maudet dans sa thèse [Maudet, 2001] et raffiné par la suite, régule ce méta-niveau de communication et indique comment le statut des jeux de dialogue est discuté en pratique.

Coups	**Opérations**
$prop.in(x, y, g)$	$create(C_g(y, x, acc.in(y, x, g)\|ref.in(y, x, g)))$
$prop.out(x, y, g)$	$create(C_g(y, x, acc.out(y, x, g)\|ref.out(y, x, g)))$
$acc.in(x, y, g)$	créé les engagements dialogiques pour le jeu g
$acc.out(x, y, g)$	supprime les engagements dialogique pour le jeu g
$ref.in(x, y, g)$	sans effet sur la couche publique
$ref.out(x, y, g)$	sans effet sur la couche publique

TAB. 5.1 – Le jeu de contextualisation de DIAGAL.

La figure 5.1 indique les différents coups du jeu de contextualisation que nous utiliserons dans le reste du texte ainsi que leur effets en termes d'engagements dialogiques[3]. Par exemple, lorsqu'une proposition d'entrer dans un jeu de type g ($prop.in(x, y, g)$) est énoncée par l'agent x, l'engagement dialogique $C_g(y, x, acc.in(y, x, g)|ref.in(y, x, g))$ est ajouté à l'agenda des deux agents conversant ce qui signifie que l'agent y est dialogiquement engagé envers l'agent x à accepter ou à refuser de jouer un jeu de type g.

[3] Dans la suite de ce document, les opérateurs $create$ ou $supress$ seront omis pour gagner en lisibilité.

En ce qui concerne les possibilités de composition des jeux, nous ne considérerons que ce que permet le jeu de contextualisation présenté Figure 5.1, à savoir :

- *le séquencement*, noté $g_1; g_2$, signifie que g_2 sera joué immédiatement après la fermeture de g_1 ;
- *l'imbrication*, notée $g_1 < g_2$, signifie que g_1 est ouvert et joué alors que g_2 est ouvert ;

Nous nous conformons ainsi à l'analyse de Walton et Krabbe [1995] et la formalisation subséquente de Reed Reed [1998] ne considéraient que ces deux types de structurations. Cependant, de récents travaux (incluant les nôtres) ont introduit de nouvelles combinaisons [Pasquier et al., 2004a; Chaib-draa et al., 2003; McBurney et Wooldridge, 2002] :

- *le pré-séquencement*, noté $g_2 \rightsquigarrow g_1$, signifie que g_2 est ouvert et joué alors que g_1 est proposé ;
- *le choix*, noté $g_1|g_2$, signifie que les participants peuvent jouer g_1 ou g_2
- l'itération est notée g^n lorsque le nombre de répétitions est fixe et g^* autrement.

Si, comme nous le verrons à la section suivante, celles-ci ne changent pas la puissance expressive du langage en terme de sa capacité à manipuler des engagements extra-dialogiques, cela peut s'avérer utile pour la spécification de certaines conversations. En effet, ces autres structurations peuvent être nécessaires pour spécifier certains protocoles pré-existants. Dans certains cas, il est également nécessaire de rendre explicite l'initiateur et le partenaire de chacun des jeux, la notation peut alors être simplement étendue : $[x, y]g_1$ signifie que l'agent x (x pouvant aussi bien être une variable libre qu'un nom d'agent précis) est l'initiateur du jeu g_1 tandis que y est son partenaire. Ainsi, $[x, y]g_1; [y, x]g_2$ signifie que c'est le partenaire de x dans le jeu g_1 qui devra être l'initiateur du jeu g_2.

La section suivante présente les jeux de DIAGAL dont il est question lors de la contextualisation.

5.2.3 Les jeux de dialogue

Dans cette section, nous allons présenter en détail les jeux de dialogues qui permettent de manipuler les engagements sociaux. Pour améliorer la lisibilité des jeux, les sanctions

attachées (les triplets s_x et s_y) ont été omises. Le temps de création des engagements est représenté en utilisant une simple théorie d'instants dans laquelle $<$ est la relation de précédence. t_j est l'instant auquel le jeu est ouvert. Notez que la structure des jeux offre un élégant mécanisme de gestion des tours de parole puisqu'il est naturellement impliqué que $t_j < t_k < t_l < t_f$ (lorsque ces instants existent).

Le jeu de requête : $Request\ Game$ **(**rg**)**

Ce jeu de dialogue modélise la requête pour l'action standard. Si les conditions d'entrée de celui-ci sont remplies et que les interlocuteurs ont accepté de jouer le jeu, l'initiateur (x) est engagé à adresser une requête à son partenaire (y) lui demandant d'effectuer une action α, ce dernier est alors engagé à accepter ($accept$) ou refuser ($refuse$) celle-ci. L'engagement extra-dialogique correspondant sera alors accepté ou rejeté selon la réponse fournie. Les conditions et règles précises de ce jeu sont les suivantes :

E_{rg}	$\neg C(y, x, \alpha, t_i)$ and $\neg C(y, x, \neg\alpha, t_i)\ \forall\ t_i,\ t_i < t_j$
S_{rg}	$C(y, x, \alpha, t_f)$
F_{rg}	$\neg C(y, x, \alpha, t_f)$
R_{rg}	1) $C_g(x, y, request_{d_1}(x, y, \alpha), t_j)$
	2) $C_g(y, x, request_{d_1}(x, y, \alpha) \Rightarrow$
	$\quad C_g(y, x, accept_{d_2}(y, x, \alpha) \mid refuse_{d_3}(y, x, \alpha), t_k), t_j)$
	3) $C_g(y, x, accept_{d_2}(y, x, \alpha) \Rightarrow C(y, x, \alpha, t_f), t_j)$
	4) $C_g(y, x, refuse_{d_3}(y, x, \alpha) \Rightarrow \neg C(y, x, \alpha, t_f), t_j)$

Rappelons que l'indice g permet de distinguer les engagements dialogiques (notés C_g) des engagement extra-dialogiques (notés C). Les quatres ensembles d'engagements que sont les condition d'entrée (E_{rg}), de succès (S_{rg}), d'échec (F_{rg}) et les règles (R_{rg}) du jeu prennent en indice le nom du jeu (ici rg pour request game). Les paramètres en indice, d_1, d_2, ... tiennent pour les degrés d'intensité des forces illocutoires des actes de langage produits. Ils sont optionnels et nous en indiquerons le sens précis en sections 5.4.1 et 5.4.2.

Le jeu d'offre : $Offer\ Game$ **(**og**)**

Une offre est une promesse (un acte commissif) conditionnelle à l'acceptation du destinataire. Plus précisément, offrir de faire l'action α c'est, si l'interlocuteur l'accepte, s'engager à réaliser cette action. Les conditions et règles du jeu d'offre de DIAGAL sont les suivantes :

E_{og}	$\neg C(x,y,\alpha,t_i)$ and $\neg C(x,y,\neg\alpha,t_i) \; \forall \; t_i, \; t_i < t_j$
S_{og}	$C(x,y,\alpha,t_f)$
F_{og}	$\neg C(x,y,\alpha,t_f)$
R_{og}	1) $C_g(x,y,\mathit{offer}_{d_1}(x,y,\alpha),t_j)$
	2) $C_g(y,x,\mathit{offer}_{d_1}(x,y,\alpha) \Rightarrow$
	$\quad C_g(y,x,\mathit{accept}_{d_2}(y,x,\alpha) \mid \mathit{refuse}_{d_3}(y,x,\alpha),t_k),t_j)$
	3) $C_g(x,y,\mathit{accept}_{d_2}(y,x,\alpha) \Rightarrow C(x,y,\alpha,t_f),t_j)$
	4) $C_g(x,y,\mathit{refuse}_{d_3}(y,x,\alpha) \Rightarrow \neg C(x,y,\alpha,t_f),t_j)$

Le jeu de question fermé : $Ask\ Game$ **(**ag**)**

Une question fermé est une question par laquelle un agent demande à un autre de s'engager sur la véracité d'une assertion propositionnelle. Notons que le fait de s'inscrire en désaccord avec la proposition énoncée ne signifie pas être en accord avec sa négation. Concrètement, les conditions et règles du jeu de question fermé de DIAGAL sont les suivantes :

E_{ag}	$\neg C(y,x,p,t_i)$ and $\neg C(y,x,\neg p,t_i) \; \forall \; t_i, \; t_i < t_j$
S_{ag}	$C(y,x,p,t_f)$
F_{ag}	$\neg C(y,x,p,t_f)$
R_{ag}	1) $C_g(x,y,\mathit{ask}_{d_1}(x,y,p),t_j)$
	2) $C_g(y,x,\mathit{ask}_{d_1}(x,y,p) \Rightarrow$
	$\quad C_g(y,x,\mathit{agree}_{d_2}(y,x,p) \mid \mathit{disagree}_{d_3}(y,x,p),t_k),t_j)$
	3) $C_g(y,x,\mathit{agree}_{d_2}(y,x,p) \Rightarrow C(y,x,p,t_f),t_j)$
	4) $C_g(y,x,\mathit{disagree}_{d_3}(y,x,p) \Rightarrow \neg C(y,x,p,t_f),t_j)$

Le jeu d'assertion : $Inform\ Game$ **(**ig**)**

Le jeu d'assertion de DIAGAL permet à un agent de s'engager sur une assertion propositionnelle quelconque, si son partenaire le lui permet. Les conditions et règles du jeu d'assertion de DIAGAL sont les suivantes :

E_{ig}	$\neg C(x,y,p,t_i)$ and $\neg C(x,y,\neg p,t_i) \; \forall \; t_i,\; t_i < t_j$
S_{ig}	$C(x,y,p,t_f)$
F_{ig}	$\neg C(x,y,p,t_f)$
R_{ig}	1) $C_g(x,y,\mathit{inform}_{d_1}(x,y,p),t_j)$
	2) $C_g(y,x,\mathit{inform}_{d_1}(x,y,p) \Rightarrow$
	$\quad C_g(y,x,\mathit{accept}_{d_2}(y,x,p) \mid \mathit{refuse}_{d_3}(y,x,p),t_k),t_j)$
	3) $C_g(x,y,\mathit{accept}_{d_2}(y,x,p) \Rightarrow C(x,y,p,t_f),t_j)$
	4) $C_g(x,y,\mathit{refuse}_{d_3}(y,x,p) \Rightarrow \neg C(x,y,p,t_f),t_j)$

Le jeu de désengagement en action : $Cancel.ActionC\ Game$ **(**cag**)**

Ce jeu peut être utilisé pour rejeter socialement un engagement en action préalablement établi et accepté, c'est-à-dire pour annuler un engagement en action. Dans ce jeu, le débiteur (x) de l'engagement $C(x,y,\alpha,t_i)$ propose son annulation. Le créditeur (y) doit alors signaler son accord ou son désaccord. Dans le cas d'un désaccord du créditeur, l'engagement pourra être rétracté tout de même pour respecter le droit de désengagement unilatéral des agents mais le débiteur aura à faire face aux sanctions. Ce ne sera pas le cas si le créditeur donne son aval. Ensuite, et suivant l'opinion du créditeur ($agree$ ou $disagree$), le débiteur peut confirmer son désengagement et faire face aux éventuelles sanctions ou bien changer d'avis et rester engagé. Les conditions et règles du jeu de désengagement en action $Cancel.ActionC$ sont les suivantes :

E_{cag}	$\exists\, t_i,\; t_i < t_j : C(x,y,\alpha,t_i)$
S_{cag}	$\neg C(x,y,\alpha,t_i)$
F_{cag}	$C(x,y,\alpha,t_i)$
R_{cag}	1) $C_g(x,y,\mathit{cancel}_{d_1}(x,y,(\alpha,t_i)),t_j)$
	2) $C_g(y,x,\mathit{cancel}_{d_1}(x,y,(\alpha,t_i)) \Rightarrow$
	$\quad C_g(y,x,\mathit{agree}_{d_2}(y,x,\mathit{cancel}_{d_1}(\alpha,t_i)) \mid$
	$\qquad \mathit{disagree}_{d_3}(y,x,\mathit{cancel}_{d_1}(\alpha,t_i)),t_k),t_j)$
	3) $C_g(x,y,\mathit{disagree}_{d_3}(y,x,\mathit{cancel}_{d_1}(\alpha,t_i)) \Rightarrow$
	$\quad C_g(x,y,\mathit{confirm}_{d_4}(x,y,\mathit{cancel}_{d_1}(\alpha,t_i)) \mid$
	$\qquad \mathit{decline}_{d_5}(x,y,\mathit{cancel}_{d_1}(\alpha,t_i)),t_l),t_j)$
	4) $C_g(x,y,\mathit{agree}_{d_2}(y,x,\mathit{cancel}_{d_1}(\alpha,t_i)) \Rightarrow \neg C(x,y,\alpha,t_i),t_j)$
	5) $C_g(x,y,\mathit{confirm}_{d_4}(x,y,\mathit{cancel}_{d_1}(\alpha,t_i)) \Rightarrow \neg C(x,y,\alpha,t_i),t_j)$
	6) $C_g(x,y,\mathit{decline}_{d_5}(x,y,\mathit{cancel}_{d_1}(\alpha,t_i)) \Rightarrow C(x,y,\alpha,t_i),t_j)$

Le jeu d'annulation en action : $Release.ActionC\ Game$ **(**rag**)**

Comme le jeu $Cancel.ActionC$ vu dans la sous-section précédente, le jeu d'annulation en action $Release.ActionC$ permet, lorsqu'il est joué avec succès, d'annuler un engagement en action. Par contre, et de manière symétrique au jeu $Cancel.ActionC$, il offre cette possibilité au créditeur plutôt qu'au débiteur. Les conditions et règles du jeu $Release.ActionC$ sont les suivantes :

E_{rag} | $\exists\, t_i,\ t_i < t_j : C(y, x, \alpha, t_i)$

S_{rag} | $\neg C(y, x, \alpha, t_i)$

F_{rag} | $C(y, x, \alpha, t_i)$

R_{rag} | 1) $C_g(x, y, release_{d_1}(x, y, (\alpha, t_i)), t_j)$

2) $C_g(y, x, release_{d_1}(x, y, (\alpha, t_i)) \Rightarrow$
$C_g(y, x, agree_{d_2}(y, x, release_{d_1}(\alpha, t_i))|$
$disagree_{d_3}(y, x, release_{d_1}(\alpha, t_i)), t_k), t_j)$

3) $C_g(x, y, disagree_{d_3}(y, x, release_{d_1}(\alpha, t_i)) \Rightarrow$
$C_g(x, y, confirm_{d_4}(x, y, release_{d_1}(\alpha, t_i))|$
$decline_{d_5}(x, y, release_{d_1}(\alpha, t_i)), t_l), t_j)$

4) $C_g(x, y, agree_{d_2}(y, x, release_{d_1}(\alpha, t_i)) \Rightarrow \neg C(y, x, \alpha, t_i), t_j)$

5) $C_g(x, y, confirm_{d_4}(x, y, release_{d_1}(\alpha, t_i)) \Rightarrow \neg C(y, x, \alpha, t_i), t_j)$

6) $C_g(x, y, decline_{d_5}(x, y, release_{d_1}(\alpha, t_i)) \Rightarrow C(y, x, \alpha, t_i), t_j)$

Le jeu de désengagement propositionnel : $Cancel.PropC\ Game$ **(**cpg**)**

Le jeu de désengagement propositionnel permet au créditeur d'un engagement propositionnel d'annuler celui-ci. Identique au jeu de désengagement en action présenté précédemment, les conditions et règles en sont les suivantes :

E_{cag}	$\exists\, t_i,\ t_i < t_j : C(x, y, p, t_i)$
S_{cag}	$\neg C(x, y, p, t_i)$
F_{cag}	$C(x, y, p, t_i)$
R_{cag}	1) $C_g(x, y, \mathit{cancel}_{d_1}(x, y, (p, t_i)), t_j)$
	2) $C_g(y, x, \mathit{cancel}_{d_1}(x, y, (p, t_i)) \Rightarrow$
	$\quad C_g(y, x, \mathit{agree}_{d_2}(y, x, \mathit{cancel}_{d_1}(p, t_i))\vert$
	$\qquad \mathit{disagree}_{d_3}(y, x, \mathit{cancel}_{d_1}(p, t_i)), t_k), t_j)$
	3) $C_g(x, y, \mathit{disagree}_{d_3}(y, x, \mathit{cancel}_{d_1}(p, t_i)) \Rightarrow$
	$\quad C_g(x, y, \mathit{confirm}_{d_4}(x, y, \mathit{cancel}_{d_1}(p, t_i))\vert$
	$\qquad \mathit{decline}_{d_5}(x, y, \mathit{cancel}_{d_1}(p, t_i)), t_l), t_j)$
	4) $C_g(x, y, \mathit{agree}_{d_2}(y, x, \mathit{cancel}_{d_1}(p, t_i)) \Rightarrow \neg C(x, y, p, t_i), t_j)$
	5) $C_g(x, y, \mathit{confirm}_{d_4}(x, y, \mathit{cancel}_{d_1}(p, t_i)) \Rightarrow \neg C(x, y, p, t_i), t_j)$
	6) $C_g(x, y, \mathit{decline}_{d_5}(x, y, \mathit{cancel}_{d_1}(p, t_i)) \Rightarrow C(x, y, p, t_i), t_j)$

Le jeu d'annulation propositionnel : $Release.PropC\ Game$ **(**rpg**)**

Ce jeu est le pendant du jeu d'annulation d'engagement en action $release.ActionCGame$ présenté ci dessus. Mise à part le fait qu'elles concernent un contenu propositionnel, les conditions et règles en sont identiques :

E_{rag}	$\exists\, t_i,\ t_i < t_j : C(y, x, \alpha, t_i)$
S_{rag}	$\neg C(y, x, p, t_i)$
F_{rag}	$C(y, x, p, t_i)$
R_{rag}	1) $C_g(x, y, \mathit{release}_{d_1}(x, y, (p, t_i)), t_j)$
	2) $C_g(y, x, \mathit{release}_{d_1}(x, y, (p, t_i)) \Rightarrow$
	$\quad C_g(y, x, \mathit{agree}_{d_2}(y, x, \mathit{release}_{d_1}(p, t_i))\vert$
	$\qquad \mathit{disagree}_{d_3}(y, x, \mathit{release}_{d_1}(p, t_i)), t_k), t_j)$
	3) $C_g(x, y, \mathit{disagree}_{d_3}(y, x, \mathit{release}_{d_1}(p, t_i)) \Rightarrow$
	$\quad C_g(x, y, \mathit{confirm}_{d_4}(x, y, \mathit{release}_{d_1}(p, t_i))\vert$
	$\qquad \mathit{decline}_{d_5}(x, y, \mathit{release}_{d_1}(p, t_i)), t_l), t_j)$
	4) $C_g(x, y, \mathit{agree}_{d_2}(y, x, \mathit{release}_{d_1}(p, t_i)) \Rightarrow \neg C(y, x, p, t_i), t_j)$
	5) $C_g(x, y, \mathit{confirm}_{d_4}(x, y, \mathit{release}_{d_1}(p, t_i)) \Rightarrow \neg C(y, x, p, t_i), t_j)$
	6) $C_g(x, y, \mathit{decline}_{d_5}(x, y, \mathit{release}_{d_1}(p, t_i)) \Rightarrow C(y, x, p, t_i), t_j)$

Le jeu de modification en action : $Update.ActionC\ Game$ **(**uag**)**

Si un agent souhaite modifier un engagement (changer la valeur d'un attribut de l'engagement autre que le créditeur ou le débiteur), il peut tenter d'annuler l'engagement et d'en créer un nouveau en tenant compte de la modification désirée. Cependant, l'annulation peut provoquer l'application de sanctions que la modification directe ne déclencherait pas. Dis autrement, peut-être que le partenaire peut accepter une modification (éventuellement tout à fait justifiée) alors qu'il refusera l'annulation dans la séquence annulation création. C'est la raison pour laquelle, nous avons introduit les jeux de modification en action et propositionnel $Update.ActionC$ et $Update.PropC$. Ils permettent de réaliser plus simplement des modifications en les considérant comme des actions dialogiques primitives (c'est-à-dire un jeu à part entière).

Dans le jeu $Update.ActionC$, l'agent initiateur x demande à son partenaire y s'il accepte de modifier un engagement du type $C(deb, cre, \alpha, t)$ en $C(deb, cre, \alpha', t)$, où (1) si l'initiateur x est le créditeur, alors $cre = x$ and $deb = y$, tandis que (2) si x est le débiteur $cre = y$ et $deb = x$. L'agent y peut alors accepter ou rejeter la modification. Si l'agent y accepte la modification, l'engagement $C(deb, cre, \alpha, t)$ est annulé et un nouvel engagement $C(deb, cre, \alpha', t)$ est créé (rendu actif et accepté). Les conditions et règles du jeu ***Update.ActionC*** sont donc les suivantes :

E_{uag}	$\exists\, t_i,\ t_i < t_j : C(deb, cre, \alpha, t_i)$
S_{uag}	$\neg C(deb, cre, \alpha, t_i)$ and $C(deb, cre, \alpha', t_f)$
F_{uag}	$C(deb, cre, \alpha, t_i)$
R_{uag}	1) $C_g(x, y, update_{d_1}(x, y, (\alpha, t_i), \alpha'), t_j)$
	2) $C_g(y, x, update_{d_1}(x, y, (\alpha, t_i), \alpha') \Rightarrow$
	$C_g(y, x, agree_{d_2}(y, x, update_{d_1}((\alpha, t_i), \alpha'))\vert$
	$disagree_{d_3}(y, x, update_{d_1}((\alpha, t_i), \alpha')), t_k), t_j)$
	3) $C_g(x, y, agree_{d_2}(y, x, update_{d_1}((\alpha, t_i), \alpha')) \Rightarrow$
	$C(deb, cre, \alpha', t_f), t_j)$
	4) $C_g(x, y, agree_{d_2}(y, x, update_{d_1}((\alpha, t_i), \alpha')) \Rightarrow$
	$\neg C(deb, cre, \alpha, t_i), t_j)$
	5) $C_g(x, y, disagree_{d_3}(y, x, update_{d_1}((\alpha, t_i), \alpha')) \Rightarrow$
	$C(deb, cre, \alpha, t_i), t_j)$

Le jeu de modification propositionnel : $Update.PropC\ Game$ **(**upg**)**

Ce jeu est le pendant du jeu de modification d'engagement en action $Update.ActionCGame$ présenté ci dessus. Les conditions et règles, mise à part le fait qu'elles concernent un contenu propositionnel, en sont identiques :

E_{upg}	$\exists\, t_i,\ t_i < t_j : C(deb, cre, p, t_i)$
S_{upg}	$\neg C(deb, cre, p, t_i)$ and $C(deb, cre, p', t_f)$
F_{upg}	$C(deb, cre, p, t_i)$
R_{upg}	1) $C_g(x, y, update_{d_1}(x, y, (\alpha, t_i), p'), t_j)$
	2) $C_g(y, x, update_{d_1}(x, y, (p, t_i), p') \Rightarrow$
	$C_g(y, x, agree_{d_2}(y, x, update_{d_1}((p, t_i), p'))\|$
	$disagree_{d_3}(y, x, update_{d_1}((p, t_i), p')), t_k), t_j)$
	3) $C_g(x, y, agree_{d_2}(y, x, update_{d_1}((p, t_i), p')) \Rightarrow$
	$C(deb, cre, p, t_f), t_j)$
	4) $C_g(x, y, agree_{d_2}(y, x, update_{d_1}((p, t_i), p')) \Rightarrow$
	$\neg C(deb, cre, p, t_i), t_j)$
	5) $C_g(x, y, disagree_{d_3}(y, x, update_{d_1}((p, t_i), p')) \Rightarrow$
	$C(deb, cre, p, t_i), t_j)$

Le jeu de décharge d'engagement satisfait : $Discharge.Satisfied\ Game$ **(**dsg**)**

Lorsqu'un engagement est satisfait, le débiteur souhaite que cela soit établi socialement, notamment pour éviter les éventuelles sanctions qui s'appliqueraient si l'engagement n'avait pas été satisfait. Pour ce faire, il joue le jeu de décharge d'engagement satisfait. Dans ce jeu, le débiteur propose la décharge de l'engagement tandis que le créditeur accepte ou refuse cette décharge. Ce jeu pourrait poser problème puisqu'on comprend aisément que le créditeur peut avoir intérêt à refuser la décharge bénéficiant ainsi de l'engagement satisfait et des sanctions attachées à sa violation. La section 5.4.3 indique comment un tel risque est contourné dans le modèle d'implantation proposé.

E_{dsg}	$\exists\, t_i,\ t_i < t_j : C(deb, cre, \alpha, t_i)$
S_{dsg}	$\neg C(deb, cre, \alpha, t_i)$
F_{dsg}	$C(deb, cre, \alpha, t_i)$
R_{dsg}	1) $C_g(x, y, discharge_{d_1}(x, y, C(deb, cre, \alpha, t_i), t_j)$
	2) $C_g(y, x, discharge_{d_1}(x, y, C(deb, cre, \alpha, t_i) \Rightarrow$
	$\quad C_g(y, x, accept_{d_2} \vert refuse_{d_3}, t_k), t_j)$
	3) $C_g(x, y, accept_{d_2}(y, x, \alpha) \Rightarrow apply(s_x)$ and $\neg C(x, y, \alpha, t_i), t_j)$
	4) $C_g(x, y, refuse_{d_3}(y, x, \alpha) \Rightarrow C(x, y, \alpha, t_i), t_j)$

Le jeu de décharge d'engagement violé : $Discharge.Violated\ Game$ **(**dvg**)**

Lorsqu'un engagement est violé, le créditeur souhaite que cela soit établi socialement, notamment pour que les éventuelles sanctions associées à ce cas s'appliquent. Pour ce faire, il joue le jeu de décharge d'engagement violé. Dans ce jeu, le créditeur propose la décharge de l'engagement tandis que le débiteur accepte ou refuse cette décharge. Ce jeu pourrait poser problème puisqu'on comprend aisément que le débiteur peut avoir intérêt à refuser la décharge évitant ainsi les éventuelles sanctions attachées à la violation. La section 5.4.3 indique comment un tel risque est contourné dans le modèle d'implantation proposé.

E_{dvg}	$\exists\, t_i,\ t_i < t_j : C(deb, cre, \alpha, t_i)$
S_{dvg}	$\neg C(deb, cre, \alpha, t_i)$
F_{dvg}	$C(deb, cre, \alpha, t_i)$
R_{dvg}	1) $C_g(x, y, discharge_{d_1}(x, y, C(deb, cre, \alpha, t_i), t_j)$
	2) $C_g(y, x, discharge_{d_1}(x, y, C(deb, cre, \alpha, t_i) \Rightarrow$
	$\quad C_g(y, x, accept_{d_2} \vert refuse_{d_3}, t_{j+1}), t_j)$
	3) $C_g(x, y, accept_{d_2}(y, x, \alpha) \Rightarrow apply(s_x)$ and $\neg C(x, y, \alpha, t_i), t_j)$
	4) $C_g(x, y, refuse_{d_3}(y, x, \alpha) \Rightarrow C(x, y, \alpha, t_i), t_j)$

Notons que les jeux pour manipuler les engagements en action et les engagements propositionnels sont semblables et que c'est essentiellement la nature des objets manipulés - actions ou proposition - qui diffère. Cette redondance, qui existe également dans le langage naturel, nous semble utile dès lors qu'elle permet de définir quels types d'engagements les agents pourront manipuler en choisissant quels sont les jeux mis à leur disposition. Pour autant de multiples autres jeux seraient envisageables, dont certains sont utiles dans certaines applications (jeu de question ouverte, jeu de négociation, ...). La section suivante indique pourquoi ces douze jeux plutôt que d'autres ont été proposés.

Complétude et adéquation pour les dialogues séquentiels

La sémantique du langage DIAGAL s'exprime en termes des manipulations effectuées sur le couche des engagements sociaux. Selon notre typologie des engagements, différenciant engagements en action et engagements propositionnels et compte tenu que les engagements sont orientés (du débiteur envers son créditeur), il existe quatre types d'engagements extra-dialogiques pouvant être acceptés entre deux agents x et y :

– L'engagement en action de x envers y ;

– L'engagement en action de y envers x ;

– L'engagement propositionnel de x envers y ;

– L'engagement propositionnel de y envers x ;

En outre, notre modèle d'engagement (figure 4.1) considère 5 opérations possibles sur ces engagements : création, annulation, modification, décharge d'un engagement satisfait et décharge d'un engagement violé. Si bien d'autres jeux sont envisageables, nous avons implanté l'ensemble des 12 jeux présentés précédemment car sous l'hypothèse que les actions dialogiques sont effectuées de manière séquentielle au sein d'un même dialogue, cet ensemble de jeux est *complet* et *adéquat* vis à vis de notre modèle de l'engagement social.

Rappelons que la *complétude* d'un système formel[4] est acquise dès lors que tout ce qui est défini sémantiquement dans le modèle est possible syntaxiquement. Dit autrement, la complétude signifie que tout ce que la sémantique décrit est effectivement modélisé par le système (ici le langage DIAGAL), que tous les états sémantiquement possibles sont accessibles. Ainsi, cette propriété assure que toute séquence de manipulations de la couche sociale impliquant les opérations définies dans notre modèle est possible. Dans notre cas le système ne serait pas complet si, par exemple, il existait pour l'un des quatres types d'engagement isolés ci-dessus une transition à laquelle aucun jeu de dialogue ne correspond. Or étant donnée que les jeux de décharges d'engagements violés ou satisfaits consomment les transitions 3 et 5 et 4 et 7 respectivement, ce n'est pas le cas. Le tableau 5.2 indique d'ailleurs à quelle transition correspond chacun des jeux.

L'*adéquation*, quand à elle, désigne le fait que seulement ce qui est permis sémantiquement est possible, c'est-à-dire que rien d'autre que ce que permet le modèle sémantique n'est

[4] Complétude et adéquation sont des propriétés issues des systèmes logiques dans lesquels elles servent à lier les théories de la preuve (syntaxe) et les théories sémantiques (voir [Alliot et Schiex, 1993] pour une introduction aux formalismes sous-jacents à l'intelligence artificielle et à l'informatique théorique).

possible. Dans notre cas, le système aurait été inadéquat, si par exemple les agents pouvaient annuler, modifier ou décharger un engagement qui n'a pas été préalablement créé. Si la complétude signifie que n'importe laquelle des opérations/transitions permises par notre modèle d'engagement peut être effectuée (ou en tout cas tentée) par un des jeux défini dans cette section, l'adéquation indique que rien d'autre n'est possible. Dans notre système ce sont les conditions d'entrée dans les jeux qui assurent le respect du cycle de vie des engagements, c'est-à-dire l'adéquation entre DIAGAL et notre modèle de l'engagement social.

Plus précisément, voici comment les douze jeux précédents réifient les 7 transitions de notre modèle :

- pour tenter de faire accepter un engagement en action de y envers x, l'agent x peut utiliser le jeu de requête ($Request\ Game$, rg) ;
- pour tenter de faire accepter un engagement en action de x envers y, l'agent x peut utiliser le jeu d'offre ($Offer\ Game$, og) ;
- pour tenter de faire accepter un engagement propositionnel de x envers y, l'agent x peut utiliser le jeu d'information ($Inform\ Game$, ig) ;
- pour tenter de faire accepter un engagement propositionnel de y envers x, l'agent x peut utiliser le jeu de question fermée ($Ask\ Game$, ag) ;
- pour tenter d'annuler un engagement en action de x envers y, l'agent x peut utiliser le jeu de d'annulation d'action ($Cancel.ActionC\ Game$, cag) ;
- pour tenter d'annuler un engagement en action de y envers x, l'agent x peut utiliser le jeu de de libération d'action ($Release.ActionC\ Game$, rag) ;
- pour tenter d'annuler un engagement propositionnel de x envers y, l'agent x peut utiliser le jeu de d'annulation d'engagement propositionnel ($Cancel.PropC\ Game$, cpg) ;
- pour tenter d'annuler un engagement propositionnel de y envers x, l'agent x peut utiliser le jeu de libération de proposition ($Release.PropC\ Game$, rpg) ;
- pour tenter de modifier un engagement en action, les agents x ou y peuvent utiliser (ce jeu est symétrique) le jeu de modification d'action ($Update.ActionC\ Game$, uag) ;
- pour tenter de modifier un engagement propositionnel, les agents x ou y peuvent utiliser (ce jeu est s$y$métrique) le jeu de modification de proposition ($Update.PropC\ Game$, upg) ;
- pour décharger un engagement violé, les agents x ou y peuvent utiliser le jeu de décharge d'engagement violé ($Discharge.Violated\ Game$, dvg) ;

– pour décharger un engagement satisfait, les agents x ou y peuvent utiliser le jeu de décharge d'engagement satisfait ($Discharge.Fullfilled\ Game$, dfg).

Ainsi défini, le langage DIAGAL peut être conçu comme un *système d'actions conjointes* pour la manipulation des engagements sociaux. Sa formulation générique et ses bonnes propriétés, rendent DIAGAL utilisable d'une multitude de manières, la section 5.4 présente les éléments saillants concernant l'utilisation de DIAGAL dans les SMAs. Un exemple de dialogue y est développé. Avant cela, les sections suivantes présentent brièvement les principaux modules logiciels du simulateur de dialogue DGS qui a été conçu pour implanter et valider différents usages de DIAGAL.

5.3 Le simulateur de dialogue DGS et l'implantation de DIAGAL

Un environnement de travail, le DGS [Dialogue Game Simulator] a été développé[5] afin de valider le langage DIAGAL ainsi que nos différents travaux sur les communications agents. Le DGS permet de tester, et de visualiser l'utilisation du langage DIAGAL.

L'interface principale du DGS permet de gérer les agents connectés au système, de charger les jeux de dialogue DIAGAL disponibles dans ledit système (ils seront connus de tous) et de visualiser les informations relatives aux conversations des agents présentées de manière synthétique et conviviale. Le DGS a été développé en JAVA à l'aide de la plateforme de développement orienté agent JACK^{TM} [Howden et al., 2001]. Dans cette section, on présente les différentes composantes logicielles permettant d'utiliser le langage DIAGAL dans un système multi-agents qui font également partie du DGS.

5.3.1 Les fichiers de jeux

Chaque jeu de dialogue du langage DIAGAL est encodé dans un fichier XML [Extended Markup Language]. L'utilisation d'XML comme langage de description a l'avantage de faciliter l'implantation des jeux. En outre, XML est très bien supporté par l'environnement

[5] Le DGS, initié par Brahim Chaib-draa, Nicolas Maudet et l'auteur, a été initialement conçu et développé avec David Bourget. Le code a ensuite été réutilisé par Marc-André Labrie qui est le principal responsable de la version actuelle. Le tronc commun du DGS est désormais maintenu et étendu par Mathieu Bergeron dans le cadre de sa Maîtrise.

de développement JAVA puisque de nombreux outils y existent qui facilitent son intégration. Le DTD [Document Type Definition] associé aux fichiers XML, décrit de manière précise l'ensemble des balises qui peuvent être utilisées pour encoder un jeu de dialogue. Cela offre une solution puissante et simple aux concepteurs/designers pour encoder les jeux qu'ils souhaitent utiliser. Les jeux sont chargés par l'utilisateur du DGS et la liste des jeux chargés est transmise aux agents lors de leur connexion au système.

5.3.2 L'agenda, la pile des jeux et le gestionnaire de dialogue

L'agenda et le gestionnaire de dialogue sont les deux outils principaux qu'un agent doit inclure pour pouvoir utiliser DIAGAL. Un agent doit inclure ces deux composantes pour pouvoir utiliser les jeux de dialogue de DIAGAL.

Le gestionnaire de dialogue (ce qui tend à justifier son nom) gère le bon déroulement des interactions de l'agent. En particulier, le gestionnaire de dialogue charge les jeux de dialogue disponibles dans le système (via son gestionnaire de jeux), incluant le jeu de contextualisation. Le gestionnaire de dialogue filtre toutes les actions posées par l'agent (via son gestionnaire d'action). C'est lui qui vérifie que les actions communicatives de l'agent sont licites. C'est donc lui qui assure le respect des règles du jeu de contextualisation et des jeux ouverts. C'est lui qui vérifie également que les conditions d'entrée des jeux proposés sont vérifiées en les instanciant avec les paramètres désirés (côté initiateur). C'est également lui qui formate les messages et assigne, le cas échéant, les degrés d'intensité des forces illocutoire des messages produits (cet aspects, qui justifie la présence d'un gestionnaire de degré d'intensité, sera discuté section 5.4.2). Une grande partie de l'activité du gestionnaire de dialogue consiste à tenir à jour l'agenda de l'agent.

L'agenda (tel qu'introduit en sections 2.3.3 et 2.3.3) est un tableau d'engagements (commitment store) individuel, c'est-à-dire local à l'agent. L'ensemble des agendas individuels constituent une version distribuée du tableau d'engagement traditionnel. Cette distribution permet d'éviter le modèle en étoile habituel, n agents et un tableau de conversation, et ainsi de réduire la complexité du système en minimisant le nombre d'interactions tout en évitant qu'un goulet d'étranglement se forme au niveau du tableau centralisé. Avec les agendas, le tableau de conversation de deux agents est simplement l'intersection de leurs deux agendas.

Seuls les engagements concernant l'agent (c'est-à-dire ceux dont il est débiteur ou créditeur) sont inscrits dans l'agenda, triés en fonction de la date de leur création (le paramètre t dans notre modèle de l'engagement social). Ces engagements sont autant les engagements extra-dialogiques stockés par l'agent (propositionnels ou en action) que les engagements dia-

logiques contractés lors d'un dialogue. Chaque engagement est stocké avec son état actuel (cet état résulte naturellement des dialogues tenus à son propos).

L'agenda est géré par le gestionnaire de dialogue, qui y ajoute les engagements lors de leur création et en change l'état courant en fonction des événements perçus par l'agent (incluant ses dialogues). Par exemple, un engagement en action est satisfait lorsque que l'action correspondante a été effectuée. Aussi, lorsqu'un agent transmet une action α à son gestionnaire de dialogue, celui-ci identifie si un engagement (dialogique ou extra-dialogique) est satisfait par cette action. Pour satisfaire un engagement, l'action perçue doit être exactement similaire à celle attendue et elle doit réussir. Si un engagement dialogique est satisfait, le gestionnaire de dialogue se contente de le noter dans l'agenda tandis que si un engagement extra-dialogique est satisfait par cette action, le gestionnaire de dialogue va initier un jeu de décharge auprès du créditeur et ce n'est que si le jeu de décharge est joué avec succès, qu'il va mettre l'agenda à jour. C'est également le gestionnaire de dialogue qui initie les jeux de décharge des engagements violés.

Pour chaque dialogue, le gestionnaire de dialogue maintient une pile informatique des jeux ouverts. La *pile des jeux* mémorise l'ensemble des jeux ouverts dans un dialogue particulier. Chaque fois qu'un jeu est ouvert, il est placé au sommet de la pile. C'est cette pile qui permet de savoir quel jeu devient actif lorsque le jeu courant est fermé et retiré de la pile. Les règles de priorité entre les jeux ouverts sont donc prises en compte via cette structure.

Finalement, c'est le gestionnaire de dialogue qui renseigne l'interface du DGS pour l'affichage des différentes informations concernant les dialogues en cours.

5.3.3 Espace de dialogue (Dialogue Workspace) et visualisation

L'espace de dialogue regroupe pour l'utilisateur du simulateur toutes les informations pertinentes à un dialogue. Un espace de dialogue est créé dans le DGS pour chaque paire d'agents dialoguant. Chaque espace de dialogue récapitule la pile de jeux du dialogue considéré, le diagramme des échanges dialogiques effectués jusque là (ce diagramme est stocké en format UML [Unified Markup Language]) ainsi que le contexte social relatif des agents conversant.

L'interface principale du DGS permet de visualiser graphiquement l'ensemble des espaces de dialogue actifs. La liste des dialogues chargés (il est possible d'en afficher la définition XML si nécessaire), la liste des agents connectés au système et pour chacun d'entre eux, le contenu de son agenda. Un module de métrique permet de suivre pour chaque agent,

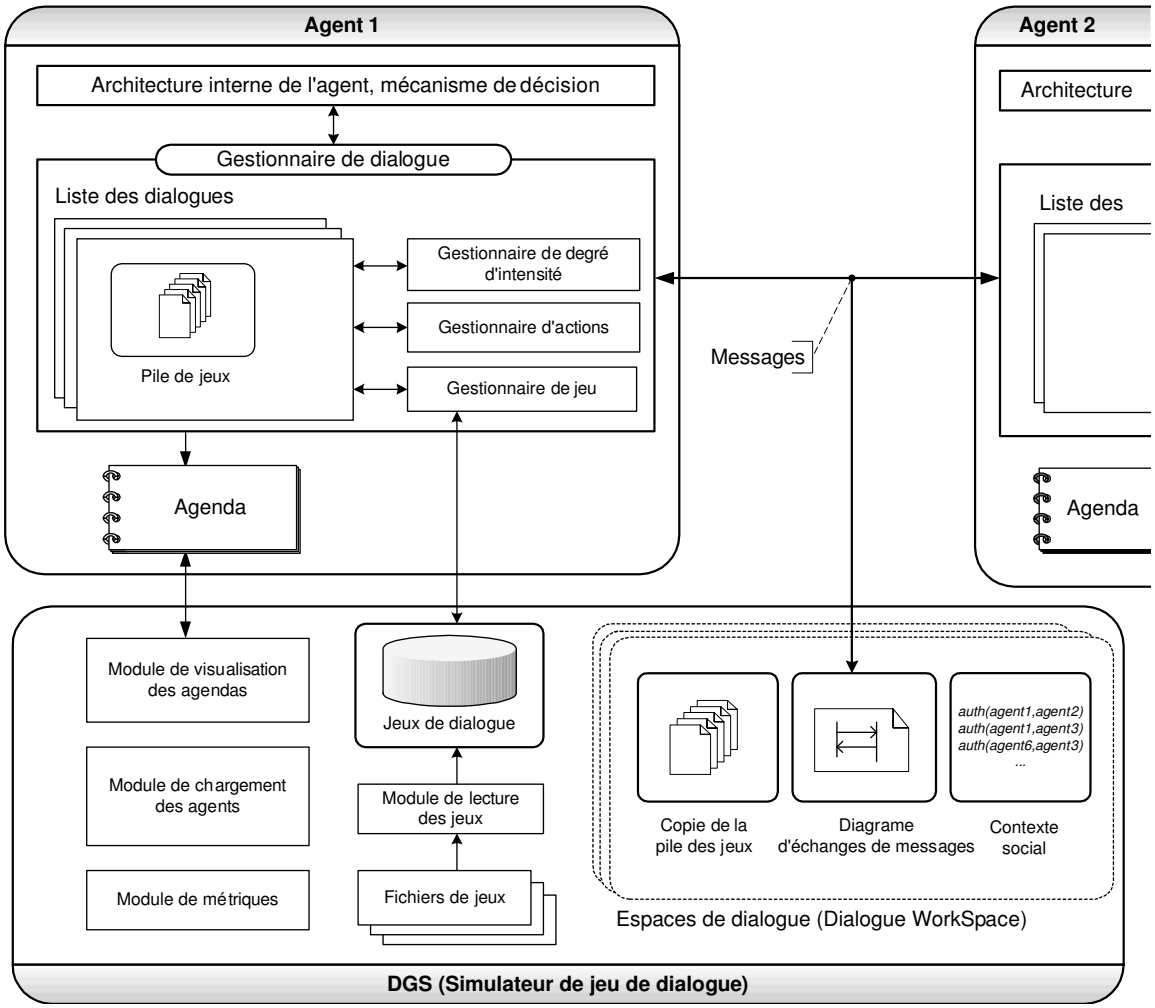

FIG. 5.1 – Vue d'ensemble des composantes logicielles associées à l'utilisation de DIAGAL et du DGS.

dialogue ou groupe d'agents : le nombre d'engagements pris, le nombre de jeux ouverts, le nombre d'engagement respectés, le nombre d'engagement violés, ... Ce module, développé par Mathieu Bergeron dans le cadre de sa maîtrise, est détaillé dans son mémoire [Bergeron, 2005].

La figure 5.1 présente une vue d'ensemble (simplifiée pour une meilleure lisibilité) des composantes logicielles associées à l'utilisation de DIAGAL dans le DGS.

Le détail de l'historique de cette architecture ainsi que les différents algorithmes et détails d'implémentation du DGS et de ses composantes sont disponibles dans le mémoire de maîtrise de Marc-André Labrie [Labrie, 2003].

5.4 Les diverses utilisations de DIAGAL

5.4.1 DIAGAL et actes de langage

La spécification via XML des jeux de dialogue offre une grande souplesse dans leur formulation. En particulier, les règles des jeux définissent les actions dialogiques que les agents doivent effectuer pour jouer le jeu. La spécification précise de ces actions dialogiques peut être raffinée. Ces actes dialogiques peuvent être de simples messages ad hoc, par exemple le message $update_{d_1}(x, y, (\alpha, t_i), \alpha')$ dans le jeu $Update.ActionC\ Game$, où bien ils peuvent être formatés selon un format de message plus générique, tel que spécifié par l'un des nombreux ACLs disponibles pour les systèmes multi-agents (et introduits section 1.2.2). Dans ce cas, seule la syntaxe de l'ACL sera utilisée, puisque la sémantique des messages est déjà spécifiée (en termes d'engagements) par les règles des jeux. On aurait plutôt $propose_{d_1}(update(x, y, (\alpha, t_i), \alpha'), t_j)$, où $propose$ tient pour un directif (ou un commissif) neutre. Ainsi, il est tout à fait possible d'utiliser DIAGAL comme sur-couche d'un ACL standard comme KQML ou FIPA-ACL pour profiter des facilités syntaxiques associées à ces ACLs et à leurs implantations déjà existantes. Malheureusement, à notre connaissance, aucun des ACLs rencontrés ne permet de spécifier le degré d'intensité des forces illocutoires utilisées (ce qui est une des options de DIAGAL discutée section 5.4.2).

Ainsi, les jeux de dialogue de DIAGAL, par leur caractère dialogique plutôt que monologique, permettent d'encapsuler les actes de langage traditionnellement utilisés pour la communication agent. Ce lien nous permettra (en section 5.5.1) d'étendre les conditions de succès et de satisfaction associés aux actes de langage dans la théorie philosophique.

Finalement, seuls des actes de langage assertifs, directifs et commissifs sont utilisés dans DIAGAL. Cela se comprend aisément du fait que les actes expressifs, auxquels sont associés la direction d'ajustement vide, ne sont pas liés au monde et donc ne sont pas traduisibles en terme de manipulation de la couche des engagements sociaux. Les actes déclaratifs, auxquels sont associés la double direction d'ajustement des mots aux choses, seraient modélisés par un jeu susceptible de créer un engagement satisfait, ce qui ne nous semble pas utile dans le contexte des systèmes multi-agents.

5.4.2 Degré d'intensité

Dans les spécifications des jeux présentés précédemment, les actes de langage imbriqués dans les jeux (voir section précédente) sont indicés avec une étiquette (d_1, d_2, ...) qui prend

valeur dans les entiers relatifs et qui indique le degré d'intensité de la force illocutoire de l'acte de langage correspondant (la notion de degré d'intensité des forces illocutoires est introduite section 1.2.1). Ce degré est indiqué de manière relative au degré d'intensité neutre en accord avec la classification de Vandervecken [Vanderveken, 1990] pour les verbes performatifs anglais :

- pour les assertifs : $suggest = -1$, $assert = 0$, $tell = +1$, $inform = +2$, $reveal = +3$, $divulge = +4$, ...
- pour les commissifs : $commit = 0$, $accept = 1$, $promise = +2$, ...
- pour les directifs : $suggest = -2$, $direct = -1$, $request = 0$, $demand = +1$, $order = +2$, ...

Par exemple, dans le jeu de requête pour l'action, l'acte de requête tient pour la catégorie des directifs toute entière et c'est le degré d'intensité de la force illocutoire qui détermine l'acte de langage réellement produit. C'est à l'agent de décider dynamiquement de ce degré, ce sans quoi le degré neutre sera sélectionné par défaut.

Si elle est cruciale pour les agents dialogiques destinés à interagir avec des humains ou dans des communautés mixtes, cette possibilité de choix du degré d'intensité des actes utilisés peut également être importante pour les communications entre agents artificiels puisque certaines caractéristiques sociales d'architectures d'agents développés dans le cadre de la modélisation cognitive permettent d'utiliser ces degrés. En particulier, notre approche de la pragmatique des communications agents tirera profit de cette possibilité. À notre connaissance, aucun autre ACL n'inclut ce raffinement.

5.4.3 Le problème de la décharge des engagements

Le problème de l'établissement de la satisfaction ou de la violation d'un engagement est délicat. Si - comme nous l'avons indiqué auparavant - ce problème reste théoriquement ouvert, il faut bien, pour pouvoir utiliser concrètement une approche reposant sur les engagements sociaux comme DIAGAL, le résoudre.

Pour ce faire, par simplification et pour les raisons indiquées en section 5.5.3, nous ne considérons que les engagements en action. On fait l'hypothèse qu'à chaque action est associée une date de butée. La satisfaction d'un engagement en action est acquise lorsque l'action spécifiée est effectuée par le débiteur avant la-dite date butoir.

La solution retenue pour l'implantation de l'établissement de la satisfaction consiste à donner aux gestionnaires de dialogue des agents conversant, le contrôle de la satisfaction des engagements pris. Concrètement, dès qu'une action satisfaisant un engagement est posée par un agent, son gestionnaire de dialogue engage un dialogue de décharge d'engagement satisfait ($Discharge.Full filled\ Game$) avec le gestionnaire de dialogue du créditeur et les éventuelles récompenses s'appliquent.

Concernant la violation, lorsque la date de butée d'un engagement est atteinte sans qu'il ait été satisfait, l'engagement est considéré comme violé et le gestionnaire de dialogue du créditeur entame avec le gestionnaire de dialogue un jeu de décharge d'engagement violé ($Discharge.Violated\ Game$) durant lequel les éventuelles sanctions sont appliquées.

Cette solution à l'avantage que la satisfaction d'un engagement est connue dès qu'elle survient, et la violation est définie comme la non-satisfaction ce qui est une définition complète de la notion de satisfaction. C'est ainsi qu'est résolu, concrètement le problème de la décharge des engagements. Puisque cette solution permet d'assurer que les engagements violés seront découverts et que nous avons fait l'hypothèse (lors de la définition de notre modèle de l'engagement social, section 4.5.2) que les sanctions applicables sont appliquées, l'hypothèse de responsabilité sûre tient, assurant le bon fonctionnement du système de contrôle social choisi.

5.4.4 DIAGAL comme langage de communication agent (ACL)

Le langage DIAGAL peut évidemment être utilisé comme un langage de communication agent traditionnel dont chaque primitive est un jeu de dialogue, c'est d'ailleurs de cette manière que nous l'utiliserons dans notre validation informatique (chapitre 7). L'idée de base, introduite dans [Pasquier et Chaib-draa, 2003b], en ce qui concerne l'utilisation de DIAGAL est qu'un agent peut tenter une opération donnée sur la couche des engagements sociaux en choisissant le jeu DIAGAL dont les conditions d'entrée et la condition de succès s'unifient avec la situation initiale et le résultat désiré (exprimé en termes d'engagements), respectivement. Le tableau 5.2 présente les différents jeux à utiliser selon le type d'engagement que l'agent initiateur souhaite manipuler et la transition qu'il souhaite consommer. La section 5.4.9 présente un exemple d'utilisation de DIAGAL dans cette optique.

Engagement	**Opération**	**Transition**	**Jeu DIAGAL à utiliser**
$C(x, y, \alpha)$	création	1	Jeu d'offre ($Offer\ Game$)
	anulation	2	Jeu d'annulation ($Cancel.ActionC\ Game$)
	modification	6	Jeu de modification ($Update.ActionC\ Game$)
	décharge violation	3+5	Jeu de décharge d'engagement violé ($Discharge.Violated\ Game$)
	décharge satisfaction	4+7	Jeu de décharge d'engagement satisfait ($Discharge.Fulfilled\ Game$)
$C(y, x, \alpha)$	création	1	Jeu de requête ($Request\ Game$)
	anulation	2	Jeu d'annulation ($Release.ActionC\ Game$)
	modification	6	Jeu de modification ($Update.ActionC\ Game$)
	décharge violation	3+5	Jeu de décharge d'engagement violé ($Discharge.Violated\ Game$)
	décharge satisfaction	4+7	Jeu de décharge d'engagement satisfait ($Discharge.Fulfilled\ Game$)
$C(x, y, p)$	création	1	Jeu d'information ($Inform\ Game$)
	anulation	2	Jeu d'annulation ($Cancel.PropC\ Game$)
	modification	6	Jeu de modification ($Update.PropC\ Game$)
	décharge violation	3+5	Jeu de décharge d'engagement violé ($Discharge.Violated\ Game$)
	décharge satisfaction	4+7	Jeu de décharge d'engagement satisfait ($Discharge.Fulfilled\ Game$)
$C(y, x, p)$	création	1	Jeu de question ($Ask\ Game$)
	annulation	2	Jeu d'annulation ($Release.PropC\ Game$)
	modification	6	Jeu de modification ($Update.PropC\ Game$)
	décharge violation	3+5	Jeu de décharge d'engagement violé ($Discharge.Violated\ Game$)
	décharge satisfaction	4+7	Jeu de décharge d'engagement satisfait ($Discharge.Fulfilled\ Game$)

TAB. 5.2 – Liens entre notre modèle de l'engagement social et les jeux de dialogue de DIAGAL (se reporter à la Figure 4.1 pour les numéros de transition). Pour chaque type d'engagement extra-dialogique pouvant tenir entre un locuteur x et son partenaire y, le tableau indique les jeux de dialogue DIAGAL correspondant aux différentes transitions (opérations) autorisées dans notre modèle de l'engagement social.

5.4.5 Variante déontique

L'utilisation de notre modèle d'engagement implique que les engagements extra-dialogiques, qui capturent les effets extra-dialogiques du dialogue, persistent après le dialogue et tiennent pour l'interprétation commune des conséquences du dialogue. Aussi, en accord avec l'objectif de flexibilité des engagements que nous nous sommes fixé en section 4.3, nous avons permis l'annulation unilatérale et la modification des engagements. Puisque des anticipations sont nécessaires à la coordination des agents, plus les agents anticipent vers un futur lointain, plus ces anticipations seront sujettes à rectification.

Cependant, cette manipulabilité des interprétations communes permise par les engagements sociaux, que nous avons nommée flexibilité sémantique, ajoute à la complexité du système (notamment la complexité des mécanismes de décision des agents susceptibles d'utiliser DIAGAL). Pour certains systèmes, il est souhaitable que la gestion des engagements soit plus simple. En particulier, par simplification, les engagements sociaux peuvent être considérés comme de simples obligations dirigées. Nous avons présentée et critiquée cette possibilité en section 4.3. Aussi, si ce n'est pas souhaitable dans le cas général, cela peut être nécessaire dans certains cas particuliers. Aussi, pour adapter DIAGAL à ce type de système, il suffit de n'utiliser que les jeux de création et de décharge, ce qui revient à supprimer les transitions 2 et 6 de l'automate à état fini qu'est notre modèle d'engagement de la Figure 4.1.

Dans ce cadre, si on omet la décharge qui a un statut particulier car elle conclut le cycle de vie des engagements, le système dialogique peut être considéré comme *monotone*. C'est à dire que le dialogue n'est utile qu'à ajouter des engagements (qui pourront être satisfaits ou violés) et les agents seront incapables de revenir sur les conséquences des dialogues tenus.

5.4.6 Prise en compte des relations d'autorité

Un grand nombre d'applications des systèmes multi-agents reposent sur un niveau organisationnel qui facilite le contrôle social et assure une meilleure efficacité du système. Ce niveau d'organisation est généralement capturé à l'aide d'une relation d'autorité qui reflète l'organigramme du groupe ou de l'organisation considérée [Royakkers et Dignum, 2000]. $Auth(X) := \{auth(a,b); a,b \in X\}$, où $auth(a,b)$ indique que l'agent a a autorité sur b et $Auth(X)$ désigne l'ensemble des relations d'autorité du groupe X.

Dans ces systèmes, il se peut, par exemple, qu'un agent ne puisse pas refuser les requêtes lorsqu'elles sont issues d'un supérieur hiérarchique. Il y a plusieurs manières pour prendre en compte ce type d'éléments du contexte social. Il se peut que le comportement à suivre soit

explicitement indiqué dans la spécification des rôles (en termes d'engagements, par exemple un engagement à ne pas refuser les requêtes des supérieurs hiérarchiques, ...), mais il peut être également nécessaire de contraindre le langage en fonction des relations d'autorité pour que cela ne soit plus possible de refuser les requêtes provenant d'un supérieur hiérarchique. Le gestionnaire de dialogue DIAGAL, dans son implantation actuelle, prend cette possibilité en compte en discriminant trois possibilités à l'ouverture d'un dialogue. Selon que le partenaire est supérieur, égal ou inférieur hiérarchiquement à l'agent initiateur (c'est-à-dire, selon qu'une relation d'autorité $auth(x, y) \in Auth(X)$ liant les deux agents existe ou non) les jeux seront contraints de la manière souhaitée.

Au moment de leur utilisation, les jeux sont instanciés (côté initiateur) avec les valeurs désirées. C'est ce qui permet, entre autres, au gestionnaire de dialogue de vérifier si les conditions d'entrée sont respectées avant même que de proposer le jeu. Lors de cette étape, le gestionnaire de dialogue est capable de prendre en compte les éventuelles relations d'autorité. Par exemple, si un agent $Boss$ communique avec l'agent Bob dont il est le supérieur hiérarchique et sur lequel il a autorité ($auth(Boss, Bob)$) à l'aide d'un jeu de requête ($Request\ Game$), les gestionnaires de dialogue vont instancier le jeu en prenant en compte cet aspect hiérarchique organisationnel en utilisant une variante déontique du jeu de requête dont les règles sont définies de la manière suivante :

$$R_{rg} \left| \begin{array}{l} 1)\ C_g(x, y, request_{d_1}(x, y, \alpha), t_j) \\ 2)\ C_g(y, x, request_{d_1}(x, y, \alpha) \Rightarrow \\ \quad C_g(y, x, accept_{d_2}(y, x, \alpha) \\ 3)\ C_g(y, x, accept_{d_2}(y, x, \alpha) \Rightarrow C(y, x, \alpha, t_f), t_j) \end{array} \right.$$

Avec ces règles, l'interlocuteur (Bob) ne peut refuser la requête de $Boss$. C'est ainsi que l'idée qu'une requête d'un supérieur est un ordre et ne peut être refusée est modélisée. Cela dit, même si cette situation est rendue rare par l'utilisation d'un système de sanction adéquate, Bob pourra toujours violer l'engagement ainsi contracté (volontairement ou involontairement).

5.4.7 DIAGAL pour la spécification de protocoles

DIAGAL peut être et a été utilisé pour spécifier des protocoles comme une structuration pre-définie de jeux de dialogue. Par exemple, Chaib-draa et al. [2002] ont fourni une spécification du protocole de demande pour l'action de Winograd et Flores (présenté section 1.3.2) avec DIAGAL. Il serait possible d'aller plus loin. Puisque DIAGAL est complet en ce qui concerne la couche sociale des engagements, il est possible de construire des algorithmes qui

construisent la séquence des jeux de dialogue nécessaires pour atteindre un état donné de la couche d'engagement. C'est l'une de nos perspectives.

5.4.8 DIAGAL pour les réseaux d'engagements

Une autre manière d'utiliser DIAGAL dans les systèmes multi-agents est de spécifier une tâche et sa dynamique en terme d'un réseau d'engagements indiquant la dynamique de celui-ci en spécifiant quelles sont les interdépendances entre les différents états des différents engagements extra-dialogiques liés à la tâche à accomplir. Bergeron et Chaib-draa [2004] ont développé cette approche, validée dans le cadre d'un exemple de tâche d'organisation d'un festival [Bergeron et Chaib-draa, 2005].

5.4.9 Exemple de dialogue

Dans cette section, nous donnons un exemple d'utilisation de DIAGAL comme ACL. On suppose, qu'un agent x souhaite qu'un agent y écrive un rapport (ayant pour identifiant $u19$) avant une date précise (le $17/03/2008$ à $17h$). Dans cette optique, l'agent souhaite donc que y s'engage en action envers lui à réaliser cette action avant la date voulue, c'est-à-dire qu'il aimerait faire accepter l'engagement :

$$C(x, y, SendReport(ident = u19, date = 17/03/08/17h), t, s_x, s_y)$$

L'agent x va donc tenter d'initier un jeu de requête ($Request\ Game$) avec y pour lui soumettre sa demande. Le choix de ce jeu de requête s'impose car c'est le seul jeu DIAGAL dont les conditions d'entrées sont respectées (y n'est pas déjà engagé envers lui à faire ou ne pas faire cette action précise) et dont la condition de succès est susceptible de s'unifier avec le type d'engagement voulu. L'agent x va donc utiliser le jeu de contextualisation et proposer à l'agent y d'entrer dans un jeu de requête ($Request\ Game$, aussi noté rg). Quand l'agent y reçoit le message de contextualisation $prop.in(x, y, rg)$, son gestionnaire de dialogue ajoute l'engagement dialogique suivant (donné par le jeu de contextualisation présenté section 5.2.2) dans son agenda :

$$C_g(y, x, acc.in(y, x, rg)|ref.in(y, x, rg)|prop.in(y, x, g'))$$

Pour satisfaire cet engagement dialogique (ce qui est une obligation stricte, pour les raisons discutées section 4.5.2), y peut soit accepter d'entrer dans le jeu de requête, refuser d'entrer dans le jeu proposé ou encore proposer un autre jeu. La seconde possibilité, signifiant que y n'est pas dialogiquement coopératif puisque cela revient à refuser de dialoguer, n'est pas

à exclure et il est possible que y ait de bonnes raisons pour reffuser d'entrer dans un jeu : il peut être déjà occupé à d'autres tâches qui monopolisent l'essentiel de ses ressources, il peut également être indisposé à répondre à x du fait de leurs échanges passés, cela peut-être l'expression d'une sanction sociale, ... Cette possibilité permet de prendre en compte le niveau attentionel du dialogue (voir les niveaux du dialogue section 1.3.2). Notons qu'il est possible d'implanter un jeu de contextualisation ne considérant pas cette possibilité de refus, dans ce cas, la contextualisation sera réduite à l'établissement du jeu à jouer (coordination au niveau du jeu courant) et sera utile pour structurer le dialogue, ce qui est de toute manière indispensable.

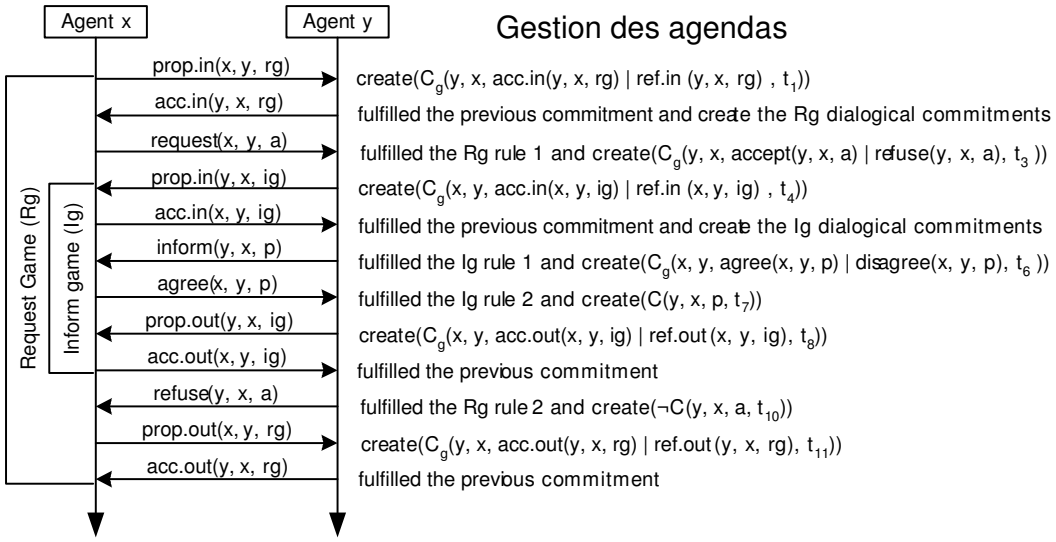

FIG. 5.2 – Exemple de conversation entre agents avec la gestion des agendas.

Pour développer notre exemple, nous supposons que y accepte de jouer le jeu de requête. Les gestionnaires de dialogue de x et y ajoutent alors le jeu de requête sur la pile des jeux ouverts et inscrivent les engagements dialogiques constituant les règles du jeu de requête (présentées section 5.2.3) dans leurs agendas respectifs. La première règle indique que x est engagé envers y à effectuer une requête :

$$C_g(x, y, request_{d_1}(x, y, \alpha), t_j)$$

L'agent x produit donc un acte directif avec le degré d'intensité et le contenu souhaité. La seconde règle du jeu est un engagement conditionnel (c'est-à-dire dont le contenu utilise l'opérateur conditionnel) et indique que y est engagé envers x à ce que si x produit la requête sus-mentionnée, alors il doit l'accepter ou la refuser. Supposons que le mécanisme de décision de y considère que le rapport peut et doit être fait, mais que la date proposée est trop proche. Avant de répondre par la négative à x, y va en informer x en imbriquant un

jeu d'assertion ($Inform\ Game$)... La conversation se poursuit ainsi en fonction des règles du cadre interactionnel et des décisions des agents. Finalement, après la conversation, seuls deux engagements extra-dialogiques subsistent $C(y, x, p)$ et $\neg C(y, x, a)$ où p est la proposition $DateToClose$ et a l'action $SendReport(ident = u19, date = 17/03/08/17h)$[6].

La figure 5.2 présente la conversation complète sous forme de diagramme de séquence (tel que produit par l'interface graphique du DGS) dans sa partie gauche et les différentes opérations sur le contenu des agendas des agents. Notons que le traitement des deux agendas est identique. C'est tout l'enjeu de l'établissement et de la mise en commun dans le dialogue qui sont nécessaires à la co-construction d'interprétations communes.

5.5 Discussion : avantages, limites et comparaisons

5.5.1 Succès et satisfaction dans DIAGAL

Dans la théorie Searlienne des actes de langage (présentée section 1.2.1), les conditions de succès et de satisfaction jouent un rôle central. Cette section indique comment ces notions sont étendues et raffinées dans le cadre du langage DIAGAL.

Les *conditions de succès* d'un acte de langage sont définies (en section 1.2.1) comme étant les conditions qui doivent tenir dans le contexte de l'énonciation afin que le locuteur réussisse son accomplissement. Dans le cas d'une promesse, par exemple, le locuteur doit être prêt à s'engager sur le contenu de sa promesse, il doit réussir à énoncer sa promesse et il doit être écouté et entendu par les agents appropriés [Searle et Vanderveken, 1985]. En fournissant plus qu'une approche monologique des primitives du langage, les jeux de dialogue de DIAGAL permettent d'étendre et de raffiner ces conditions de succès en deux catégories : les conditions de succès dialogique et les conditions de succès extra-dialogique.

Tout d'abord, chacun des jeux de DIAGAL est joué avec *succès dialogique* (qui est la notion la plus proche des conditions de succès de la théorie des actes de langage), si et seulement si tous les engagements dialogiques qui en constituent les règles sont satisfaits, c'est-à-dire que les conditions de succès ou d'échec du jeu sont atteintes. Cela signifie que l'interlocuteur accepte de jouer le jeu, c'est-à-dire accepte d'y entrer (ce sans quoi les conditions de succès

[6] Cette action et cette proposition font partie du langage de contenu que l'on suppose formel et bien spécifié. Aussi, c'est au niveau du langage de contenu que ces deux éléments ont pu être liés par les agents lors de leur raisonnement

ne pourront être réunies puisque le jeu ne sera pas joué) et que le jeu est joué complètement (après d'éventuelles interruptions dues aux structurations du dialogue).

Ensuite, les conditions de succès et d'échec des jeux rendent elles-mêmes compte d'une seconde dimension du succès, le *succès extra-dialogique*. Cette notion de succès extra-dialogique permet d'incorporer la dimension sociale du dialogue en prenant en compte l'interlocuteur. En effet, du point de vue de l'interlocuteur, un acte de langage assertif n'est réussi (indépendamment de sa véracité qui est liée à sa satisfaction) que s'il en accepte le contenu, une requête ou une offre n'est réussie que s'il l'accepte, ... Ainsi, un jeu de dialogue DIAGAL n'est joué avec succès extra-dialogique que si ses conditions de succès sont atteintes (ce qui signifie également qu'il a été joué avec succès dialogique), c'est-à-dire si la manipulation de la couche sociale proposée dans le jeu a été acceptée socialement.

Dans ce contexte, il y a plusieurs raisons pour lesquelles, le dialogue peut échouer : (1) l'interlocuteur peut refuser d'entrer dans le jeu ce qui est un cas d'échec dialogique, (2) les règles du jeu de dialogue peuvent ne pas être respectées (en particulier les conditions de succès, au sens de la théorie des actes de langage cette fois, d'un des actes de langage produit au cour du jeu de dialogue peuvent ne pas être réunies), ce qui est un autre cas d'échec dialogique ou encore (3) les conditions d'échec du jeu sont atteintes, ce qui signifie un échec extra-dialogique du jeu de dialogue.

Les *conditions de satisfaction* d'un acte de langage (section 1.2.1) rassemblent les conditions sous lesquelles les effets perlocutoires de l'acte en question sont atteints. Par exemple, une assertion n'est satisfaite que si son contenu est vrai, une requête n'est satisfaite que si elle est remplie, une promesse est satisfaite si elle est tenue, ... Les conditions de satisfaction lient l'acte de langage et le monde selon l'une de quatre directions d'ajustement (introduites section 1.2.1). C'est cette notion de satisfaction, qui introduit les aspects extra-dialogiques dans la théorie des actes de langage, en permettant de tenir compte des interactions idéales entre le monde et les actes de langage.

Tandis que, dans notre cadre dialogique, le succès extra-dialogique (que l'on aurait aussi bien pu nommer satisfaction dialogique) d'un jeu survient lorsque la condition de succès est atteinte, signifiant que l'opération de l'initiateur sur la couche sociale a été socialement acceptée, la notion de satisfaction des actes de langage imbriqués dans les jeux est capturée par celle de satisfaction des engagements extra-dialogiques. Ainsi, une requête sera satisfaite, si elle a été accomplie avec succès dialogique et extra-dialogique et que l'engagement extra-dialogique en découlant a lui-même été satisfait.

Dans notre approche, les engagements extra-dialogiques gardent donc une trace du dialogue entre son succès (dialogique et extra-dialogique) et son éventuelle satisfaction. No-

tons que de manière analogue à la théorie des actes de langage, les notions de succès et de satisfaction sont liées entre elles par des relations de dépendance fortes : pas de succès extra-dialogique sans succès dialogique préalable, et pas de satisfaction sans succès extra-dialogique préalable.

Les conditions de succès et de satisfaction telles que définies dans la théorie des actes de langage classique ont toujours été problématiques à implanter et à vérifier dans l'environnement distribué propre aux systèmes multi-agents ouverts. Il nous semble important, dans le cadre d'une formalisation informatique, d'expliciter cette zone floue qui sépare succès et satisfaction dans la théorie des actes de langage et les approches de la communication agent qui en découle. Cela participe à opérationnaliser la théorie.

5.5.2 Autres avantages de DIAGAL

Le cadre interactionnel DIAGAL, tel que décrit dans ce chapitre, cumule les avantages des approches sociales et conventionnelles présentées en section 2.3.7 avec certains avantages propres. On les récapitule ici :

- *L'hypothèse de sincérité n'est plus nécessaire* : les engagements sociaux n'étant pas nécessairement sincères, l'hypothèse de sincérité n'est plus nécessaire au bon fonctionnement du cadre interactionnel ;
- *Le problème de vérification est résolu* : comme nous l'avons montré dans ce chapitre, la vérification de la conformité des comportements agents par rapport aux dialogues tenus est rendue possible dans cette approche sans accéder aux états internes des agents, ce qui autorise de considérer cette vérification dans des systèmes ouverts[7] ;
- *DIAGAL supporte les systèmes ouverts* : muni du gestionnaire de dialogue de DIAGAL, des agents d'architecture interne hétérogènes et issues de développeurs différents, peuvent se communiquer des messages mutuellement compréhensibles. L'application des éléments de contrôle social (le système de sanction) est assuré par les gestionnaires de dialogue des agents ;
- *DIAGAL propose un cadre élégant et uniforme* : les engagements sociaux sont utilisés tant au niveau dialogique, qu'extra-dialogique via les structures de jeu. Dans le même esprit que la distinction entre le plan du dialogue et le plan de la tâche dans les approches intentionnelles basées sur la planification, cela permet par exemple de distinguer un engagement de répondre à une question de l'engagement qui résultera

[7] Notons tout de même que la solution proposée ne s'applique qu'aux engagements en action.

éventuellement de la réponse. DIAGAL permet donc de prendre en compte les aspects conventionnels de la pragmatique tels que les obligations dialogiques et la gestion des tours de parole. L'utilisation des engagements dialogiques et extra-dialogiques assure donc la possibilité de méta-dialogues (c'est-à-dire d'un dialogue à propos d'engagements dialogiques) et l'uniformité du formalisme ;

– *DIAGAL a une sémantique explicite* : le modèle de l'engagement social fournit (si on le considère comme une machine à états abstraite) une sémantique opérationnelle des jeux de dialogue. Autrement dit, chaque jeu de dialogue peut être vu comme une action conjointe dont l'effet en cas de succès extra-dialogique est de modifier l'état d'un engagement extra-dialogique ;

– *DIAGAL offre un cadre flexible et complet* : la flexibilité offerte est maximale par rapport à ce qui est possible sémantiquement, c'est-à-dire avec des engagements flexibles tels que définis en 4.5.1. C'est une conséquence de la complétude évoquée en section 5.2.3. La complétude par rapport au modèle d'engagement découle du fait que toutes les transitions du modèle d'engagement sous-jacent (Figure 4.1) peuvent être consommées par un jeu DIAGAL. Les transitions 3 et 4, qui font exceptions, sont elles socialement établies comme ayant été consommées lorsque les transitions 5 ou 7 sont consommées (respectivement). Notons au passage que la décharge des engagements, étape cruciale qui vient clore le cycle de vie d'un engagement, n'est pas prise en compte dans les autres approches des langages de communication agents rencontrées jusqu'alors ;

– *DIAGAL assure l'établissement sur différents niveaux* : DIAGAL permet de garantir l'établissement du fond commun et ce de manière particulièrement complète. En effet, chaque jeu de dialogue permet la négociation et l'établissement (présentation et acceptation, ou refus) d'une modification de la couche sociale des engagements (réifiée dans les agendas des agents conversant). L'établissement assure l'interprétation commune (en terme de manipulation des engagements) des dialogues. En outre, le jeu de contextualisation permet de garantir l'établissement de chacun des jeux de dialogue et de leurs structurations.

– *DIAGAL capture le niveau attentionnel du dialogue* : L'établissement, via le jeu de contextualisation, des jeux de dialogue permet de capturer le niveau attentionnel du dialogue (voir les niveaux de dialogue, section 1.3.2). Ainsi, un agent peut, par exemple, refuser d'entrer dans un jeu de requête (ce qui est différent que de refuser la requête elle-même) soit parce qu'il n'est pas disposé à dialoguer (comme conséquence des interactions passés, par sanction sociale ou par choix stratégique) ou encore parce que ses ressources sont déjà engagées ailleurs.

– *DIAGAL a été implémenté et validé* : à ce sujet, le cadre DIAGAL et le simulateur de jeux de dialogue du laboratoire DAMAS est à notre connaissance le plus avancé des

cadres interactionnels reposant sur les jeux de dialogue proposé dans la communauté de recherche sur les systèmes multi-agents.

5.5.3 Discussion des problèmes courants

Les approches de la communication agent reposant sur les engagements sociaux sont assez récentes. Aussi de nombreux aspects restent à approfondir et à éclaircir. Les sous-sections suivantes discutent les problèmes courants sur lesquels nous travaillons présentement. Ces problèmes touchent essentiellement la modélisation de la notion d'engagement social, centrale pour les modèles utilisant les jeux de dialogue. Cette notion est une notion de second ordre (ce qui en complexifie considérablement la formalisation) dont la modélisation pose de nombreux challenges.

Le problème du langage de contenu

De manière générale, bien peu de choses ont été dites à propos du contenu des engagements. Aussi les engagements sont proposés sans que le langage formel régissant leur contenu soit indiqué. Cela rend la tâche de développement d'un modèle d'engagement formel quasi impossible tant les formalismes de représentation des contenus proportionnels et des actions sont nombreux et hétérogènes.

Cette dimension des engagements n'a que très peu été discutée et aucune formalisation précise d'un langage de contenu « générique » adapté aux engagements sociaux n'a été produite. Les langages de contenu associés aux ACLs classiques (KIF, Prolog, ...) peuvent sans doute être mis à profit dans ce cadre. Pour autant, comme l'indiquent Hulstijn et al. [2005], les liens logiques entre le contenu des énoncés (qui entraîne celui des engagements créés) et le contexte d'énonciation doivent être pris en compte plus finement que dans les ACLs classiques. En outre, les liens entre le langage de contenu retenu et le système de sanction devront être éclaircis.

Le choix d'un langage de contenu nous semble une des conditions importantes pour progresser dans la modélisation des engagements. En effet, des phénomènes plus complexes liés à la satisfaction des engagements en dépendent. C'est par exemple le cas de la satisfaction par agrégation : si A s'est engagé à deux reprises à payer $5\$$ à B, ces deux engagements peuvent être satisfaits simultanément par un versement de $10\$$. Sans un modèle de la structure (ici algébrique) du langage de contenu et des éléments qu'il comporte, il n'est pas possible de déterminer qu'un paiement de $10\$$ équivaut à deux versements de $5\$$. De la même manière,

un versement de $3\$$ satisfait partiellement un engagement de $5\$$. Impossible de le déduire sans un modèle précis du langage de contenu. Nous avons vu (section 5.4.3) que dans la modélisation présentée, nous supposons simplement que l'identité est définie dans le langage de contenu et nous exigeons, pour qu'une action satisfasse un engagement qu'elle soit identique à celle sur laquelle l'engagement a été pris.

Engagements propositionnels et engagements en action

La distinction entre engagement propositionnel et engagement en action pose également problème. En effet, s'engager sur une action, c'est généralement s'engager sur le résultat de cette action.

A *: je vais fermer la porte avant minuit.* (action α)

A *: la porte sera fermée à minuit.* (proposition p)

Symétriquement, de manière générale, Walton et Krabbe [1995] indiquent que les engagements propositionnels contraignent les actions subséquentes. Par exemple et selon les contextes, A devra être capable d'argumenter que p, de fournir des preuves empiriques que p, de prouver p, de faire en sorte que p soit vrai,.... Finalement, A se trouve engagé sur un certain nombre de stratégies partielles, ce qui tend à réduire l'engagement propositionnel à l'engagement en action. Cependant, tous ces engagements subséquents sont centrés sur la proposition p. Les engagements propositionnels sont donc utilisés comme des moyens de s'engager sur un ensemble de stratégies partielles centrées sur p.

Cependant, Walton et krabbe (*ibid.*), qui ne sont pas informaticiens, ne fournissent pas tous les aspects formels nécessaires pour déduire une implantation informatique de leurs propos. En outre, on sent bien que les relations entre propositions et actions sont dépendantes du langage de contenu considéré. Aussi, concrètement et dans l'état actuel de nos travaux il est raisonnable à l'instar d'autres chercheurs, comme Flores [2002], de se cantonner aux engagements en actions. C'est ce que nous avons fait section 5.4.3 pour régler le problème de la responsabilité sure (introduit section 4.4.3).

Gestion des aspects temporels des engagements

Même si quelques travaux se sont déjà intéressés aux aspects temporels des engagements [Mallya et al., 2004; Verdicchio et Colombetti, 2005], beaucoup de travail reste à faire

dans ce domaine. En particulier le cas des engagements se référant au passé doit être éclairci. Pour l'heure, nous n'autorisons pas la création d'engagement se référant au passé. Aussi les énoncés suivants :

1.A *: J'ai fermé la porte hier à minuit.* (jeu de décharge d'engagement satisfait).

2.A *: Hier, à minuit, la porte était fermée.* (jeu de décharge d'engagement satisfait).

seront interprétés comme des tentatives de décharge d'engagement satisfait plutôt que comme des tentatives de création d'engagement référant au passé. Cet aspect est connexe à celui du choix d'un langage de contenu formel.

5.6 Conclusion

Dans ce chapitre, nous avons présenté le langage de communication agent DIAGAL, qui repose sur le modèle de l'engagement social flexible présenté au chapitre précédent. Le langage DIAGAL opérationnalise notre modèle de l'engagement social, tandis que le DGS en propose un cadre d'implémentation. Finalement, nous avons discuté les avantages et problèmes attachés à ce cadre de communication agent parmi les plus complets.

Ce chapitre, avec le précédent, nous a donc permis de présenter les aspects syntaxiques, structurels, sémantiques et conventionnels sur lesquels nous avons travaillé et qui seront le cadre général de validation de notre apport théorique concernant les aspects cognitifs de la pragmatique des communications agents[8]. Le chapitre suivant traite plus spécifiquement de notre problématique (présentée au chapitre 3) et introduit un cadre théorique pour le traitement des aspects cognitifs de la pragmatique des communications entre agents.

[8] Il en résulte un modèle en couche des aspects syntaxiques et sémantiques des conversations. Ce modèle, plus général a été élaboré en collaboration avec Roberto Flores et est présenté de manière formelle dans le langage de spécification Z dans des publications communes [Flores et al., 2005a,b].

Chapitre 6

Approche de la pragmatique des communications agents par la cohérence cognitive

6.1 Introduction

Dans les deux précédents chapitres, nous avons proposé un cadre interactionnel agent complet (tant que faire ce peut) et opérationnel. Si cela doit être considéré comme une contribution en soi, cela ne répond en rien à notre problématique, introduite au chapitre 3. Pour autant, cette étape était indispensable, puisque pour pouvoir parler de l'utilisation d'un langage de communication, ledit langage doit être défini.

Avec ce chapitre[1], nous dépassons donc ce niveau et attaquons notre problématique à proprement dite. En effet, on présente dans ce chapitre le corps théorique de notre approche de la pragmatique des communications agent. L'annexe D introduit un certain nombre de définitions et de résultats de sciences cognitives (psychologie cognitive et psychologie sociale, principalement) qui sous-tendent notre modèle[2].

Dans ce chapitre, on introduit les approches motivationnelles (section 6.2) telles qu'elles existent en sciences cognitives, puis on présente la théorie de dissonance cognitive sur laquelle nous nous sommes basés (section 6.3). Ensuite, la formalisation que l'on en propose

[1] Ce chapitre reprend les éléments publiés et présentés dans le cadre des Journées Francophones d'Intelligence Artificielle Distribuée et des Systèmes Multi-Agents [Pasquier et Chaib-draa, 2002] et de la conférence internationale Autonomous Agents and Multi-Agent Systems (AAMAS, 2003) [Pasquier et Chaib-draa, 2003a].

[2] Elle devra être lue comme préalable à ce chapitre ou au suivant en cas de difficultés de compréhension.

(section 6.4) est présentée et la notion de changement d'attitude, qui est fondamentale dans notre approche, est introduite (section 6.5).

Le reste du chapitre présente et exemplifie notre extension de ce modèle motivationnel générique à la pragmatique de la communication agent qui constitue notre principal apport (section 6.6). En particulier, les différentes dimensions de l'usage des communications capturées sont discutées. Ce modèle, reposant sur un principe motivationnel unique, formel et générique, permet de couvrir l'ensemble des dimensions de notre problématique, c'est-à-dire de répondre (au moins partiellement) aux questions :

- Quand un agent prend-il l'initiative d'une conversation, à quel sujet et pourquoi (voir section 6.6.3) ?
- Avec qui (voir section 6.6.3) ?
- Par quel type de dialogue (voir section 6.6.4) ?
- Quelle intensité donner aux forces illocutoires des actes de langage utilisés (voir section 6.6.8) ?
- Comment définir et mesurer l'utilité d'une conversation (voir section 6.6.7) ?
- Quand arrêter le dialogue ou le cas échéant comment le poursuivre (voir section 6.6.7) ?
- Quels sont les impacts du dialogue sur les attitudes de l'agent (voir section 6.5) ?
- Quels sont les impacts du dialogue sur l'humeur de l'agent (voir section 6.6.8) ?
- Quelles sont les conséquences du dialogue sur les accointances de l'agent (voir section 6.6.7) ?

Le chapitre suivant présente la validation informatique dans le cadre des communications entre agents du modèle présenté ici. Cette validation repose sur le cadre interactionnel introduit dans les deux chapitres précédents et viendra faire le lien entre nos contributions.

6.2 Définitions et éléments préliminaires

6.2.1 Intentionnalité, cognitions et attitudes

En sciences cognitives (et donc en intelligence artificielle[3]), l'intentionnalité (à ne pas confondre avec la notion d'intention individuelle introduite section 2.2.1) est la propriété qu'ont les états mentaux de représenter des états de choses du monde [Houdé et al., 1998] :

- *réalisées* : c'est le cas des croyances, dont le contenu fixe l'état du monde représenté[4] ;
- *à réaliser* : c'est le cas des désirs, buts et aspirations dont le contenu fixe l'état[5] que le monde devrait atteindre au sens de l'agent[6].

Les cas paradigmatiques d'états intentionnels sont les attitudes propositionnelles : « On regroupe sous le nom d'attitudes propositionnelles les croyances, buts, désirs, intentions, obligations, craintes, espoirs, souhaits, attentes, . . . , qui ont en commun d'être identifiées par leur contenu propositionnel. » [Houdé et al., 1998]. Les cognitions regroupent tous les éléments cognitifs : les perceptions, les attitudes propositionnelles (composante cognitive), les émotions (composante affective). Les engagements sociaux sont des cognitions sociales, un type particulier de cognition. De l'ensemble des cognitions résultent les *attitudes* qui sont des dispositions psychologiques positives ou négatives en rapport à un objet concret, abstrait ou à un comportement. Pour les psychologues contemporains, les attitudes sont les principales composantes de la cognition humaine. Elles sont le préliminaire subjectif à l'action rationnelle [Erwin, 2001].

6.2.2 Approches motivationnelles

En psychologie cognitive, en psychologie sociale et plus généralement en sciences cognitives, on regroupe sous l'appellation d'*approches motivationnelles*, une famille de théories qui postulent un système automatique visant à maintenir un état interne d'harmonie ou d'équilibre pour le système cognitif.

[3] Rappelons que l'intelligence artificielle est une des sciences cognitives, telle que présenté dans l'introduction de cette thèse.

[4] On parle également d'états mentaux informationnels.

[5] Notons que les états représentés ne sont pas nécessairement des états existants ni même des états possibles du monde. Par exemple, on peut désirer rencontrer le père Noël et on peut croire que $4 + 3 = 9$.

[6] On parle également d'états mentaux motivationnels, cependant nous utiliserons ce terme à d'autres fins.

Toutes les théories motivationnelles, en particulier les théories de la cohérence cognitive, font donc appel au concept d'*homéostasie*, c'est-à-dire à la faculté qu'ont les êtres vivants de maintenir ou de rétablir certaines constantes physiologiques ou psychologiques qu'elles que soient les variations du milieu extérieur.

Ces théories partagent en prémisse le *principe de cohérence* qui pose la cohérence comme mécanisme organisateur premier :

L'individu est plus satisfait avec la cohérence qu'avec l'incohérence.

L'individu forme un système ouvert dont le but est de maintenir la cohérence autant que possible (on parle aussi de balance ou d'équilibre). Ce principe d'application globale est la clé de voûte des théories de la cohérence qui sont elles-mêmes une espèce des théories dites motivationnelles.

Les théories de la cohérence cognitive : (a) décrivent les conditions d'équilibre et de déséquilibre au sein du système cognitif, (b) considèrent que le déséquilibre motive l'individu à restaurer l'équilibre et (c) décrivent les procédures par lesquelles cet équilibre peut être de nouveau atteint.

Par exemple, supposons un employé qui est contre les centrales nucléaires du fait d'un certain nombre de croyances à leur sujet, se retrouve engagé socialement (par exemple, parce qu'un supérieur hiérarchique le lui aura ordonné) à défendre un projet de construction d'une centrale nucléaire. Cet employé se trouve engagé sur un comportement anti-attitudinal. Autrement dit, il y a incohérence entre sa cognition et cet engagement social. L'hypothèse commune des théories motivationnelles est que cette incohérence *motive* un changement, une réaction. Ce changement peut s'exprimer par des comportements visant à réaffirmer les attitudes (actions externes : tentative de persuasion du supérieur, démission, désobéissance au supérieur hiérarchique, ...) ou bien par un changement d'attitude (actions internes : reconsidérer l'attitude envers les centrales nucléaires, plus ou moins profondément par exemple en justifiant son action par des contraintes ou des avantages matériels externes, ...).

Pour fonder nos travaux, nous avons retenu la théorie de la dissonance cognitive qui est la plus étudiée des approches motivationnelles. Nous l'avons choisie tant pour son caractère complet et général que pour sa proéminence historique et l'abondance des travaux et des formalisations qu'elle a engendrés[7]. Nous détaillons celle-ci dans la prochaine section, tan-

[7] Avant de faire ce choix, on a étudié de nombreuses théories du changement d'attitude, parmi lesquelles : la théorie du renforcement [Miller et Dollar, 1941], la théorie du traitement de l'information [Hovland et al., 1953], la théorie du jugement social [Sherif et Hovland, 1961], la théorie de la balance [Newcomb, 1953], la théorie

dis que l'annexe B dresse le portrait du riche paysage théorique environnant ce jalon de la psychologie sociale.

6.3 Généralités sur la théorie de la dissonance cognitive

La théorie de la dissonance cognitive, initialement formulée par Festinger [1957], est l'une des plus importantes théories de psychologie sociale. Elle a généré des centaines d'études et d'extrapolations sur les attitudes, les comportements et les croyances humaines, l'internalisation (sic) des valeurs, les motivations et les conséquences des prises de décisions, les désaccords inter-personnels, la persuasion et autres phénomènes psychologiques importants [Harmon-Jones et Mills, 1999]. Ceci s'explique en partie par la formulation très générale et abstraite de cette théorie qui la rend facile à manipuler. Dans les théories de la communication, elle apparaît comme l'une des principales théories de réception et de traitement des messages [Littlejohn, 2002].

Dans sa version originale, cette théorie considère que deux éléments de cognition (perceptions, attitudes propositionnelles ou comportements) sont en rapport ou pas (un lien pertinent les relie ou non). Deux cognitions liées sont soit consonantes (ou cohérentes) soit dissonantes (incohérentes). Elles sont dites consonantes si l'une entraîne ou supporte l'autre. À l'inverse, deux cognitions sont dites dissonantes si l'une entraîne ou supporte le contraire de l'autre.

L'hypothèse de base de cette théorie est que la dissonance produit chez le sujet une *tension* qui l'incite au changement. L'existence d'une dissonance plonge le sujet dans un état inconfortable de sorte que cela le *motive* à réduire cette dissonance. Plus la dissonance est intense, plus ce « malaise » psychologique est fort et plus la pression pour réduire la dissonance l'est aussi. Une dissonance peut être réduite en : (1) supprimant ou réduisant l'importance des cognitions dissonantes, (2) ajoutant ou augmentant l'importance des cognitions consonantes.

La seconde hypothèse de Festinger est qu'en cas de dissonance, l'individu ne va pas seulement changer ses cognitions ou essayer de changer celles des autres pour essayer de la réduire, il va aussi éviter toutes les situations qui risquent de l'accroître. Ces deux hypothèses ont été vérifiées par de nombreuses expériences de psychologie [Wickland et Brehm, 1976].

Un des intérêts majeurs de la théorie de la dissonance cognitive est de fournir une mesure de la dissonance, c'est-à-dire une métrique de la cohérence cognitive. Initialement, Festinger

de la congruence [Osgood, 1963], et la théorie de la consistance cognitive [McGuire, 1960]. Cette étude a été menée sous la direction du professeur Paquette, chercheur en psychologie sociale à l'Université Laval [Pasquier, 2003]. L'annexe B présente brièvement ces autres théories motivationnelles de psychologie sociale.

définissait l'intensité (la magnitude) de la dissonance introduite par une cognition X avec une mesure du taux de dissonance[8] définie de manière informelle comme suit :

$$dissonance_X = \frac{produit\ des\ importances\ des\ cognitions\ dissonantes\ avec\ X}{produit\ des\ importances\ de\ toutes\ les\ cognitions\ en\ rapport\ avec\ X}$$

La dissonance globale d'un agent peut, elle aussi, être calculée à partir des dissonances introduites par ses différentes cognitions. Toutes ces mesures ont été raffinées par la suite, donnant lieu à de nombreuses formalisations de la dissonance cognitive [Shultz et Lepper, 1999; Sakai, 2001].

On peut se demander dans quelles circonstances la dissonance survient. En fait, il y a différentes situations dans lesquelles la dissonance est presque inévitable :

1. *contact direct initial avec une situation* : une situation entièrement nouvelle est susceptible d'introduire un certain nombre de nouveaux éléments de cognition dissonants avec ceux qui pré-existent ;
2. *un changement dans la situation* : de la même façon, un changement dans la situation peut amener des éléments de cognition jusqu'alors consonants à devenir dissonants ;
3. *communication* : la communication avec les autres est susceptible d'introduire de nouveaux éléments qui sont dissonants avec ceux de l'agent ;
4. *existence simultanée de différentes cognitions dont certaines sont consonantes et d'autres dissonantes* : dans le cas général, une cognition est liée à plusieurs autres dont certaines sont consonantes et d'autres dissonantes.

Un état de consonance est un état d'équilibre et aucune force n'agit pour changer les relations entre les différentes cognitions de cet état. Au contraire, un état dissonant fait naître une motivation pour rétablir la consonance, c'est-à-dire retourner à l'équilibre. Pour éliminer cette tension, l'agent peut agir de différentes façons :

– *changer la situation afin qu'elle soit consonante avec ses cognitions* : un agent peut agir sur l'environnement pour l'amener dans un état où ses cognitions sont de nouveau consonantes avec la situation en question.

[8] On peut tout aussi bien mesurer la notion duale de la dissonance : la consonance.

– *changer ses cognitions afin qu'elles soient consonantes avec la situation* : c'est le principe de base du changement d'attitude qui permet l'apprentissage et l'adaptation dans la théorie de Festinger.

Si une consonance existe, l'agent va éviter les changements d'attitudes ou de comportements susceptibles d'introduire de la dissonance. De même, si une dissonance est présente, l'agent va éviter les changements de cognition susceptibles d'augmenter la magnitude de la dissonance et s'orienter vers des changements susceptibles de la diminuer. En l'absence d'équilibre, ces tendances se manifestent donc par des changements d'attitudes ou de comportements. Pour rendre la théorie plus utilisable, il faut pouvoir déterminer quelles sont les conditions qui décident si ce sont les attitudes ou les comportements qu'il faut changer. Cela dépend de rapports d'importance entre ces tendances et de la résistance au changement des différentes cognitions et comportements en jeu.

En effet, les cognitions ne sont pas toutes également manipulables. La probabilité qu'une cognition soit modifiée pour réduire la dissonance dépend de ce que Festinger nomme sa *résistance au changement*. La résistance au changement d'une cognition est directement fonction du type de cognition considérée, du nombre et de l'importance des éléments avec lesquels elle est consonante ou dissonante, de son ancienneté ainsi que de la manière dont elle a été acquise : perception, raisonnement ou communication. En effet, il est commun que la résistance au changement des cognitions acquises par perception soit supérieure à celle des cognitions issues du raisonnement ou de la communication.

Festinger a également écrit sur le lien entre dissonance cognitive et communication en se restreignant au problème de l'acquisition des cognitions. Pour Festinger [1954], un individu a deux sources majeures d'informations : sa propre expérience et la communication avec les autres. L'impact de l'expérience directe est plus grand en ce qu'elle exerce une forte pression cognitive pour s'y conformer. En effet, la communication peut être vue comme une source d'expérience indirecte. L'intensité de l'impact des communications dépend de la relation entre ceux qui communiquent. Cette relation peut être analysée en termes de rôle, de confiance/réputation, d'attraction (intérêt mutuels, relation sociale, ...) et de passif (historique des interactions, ...). Plus cette relation est « forte », plus la communication aura de l'impact sur la cognition des agents communicants. Souvent, ces deux sources, directe et indirecte, sont utilisées simultanément : un enfant peut apprendre de sa mère que le feu est dangereux et également se brûler au toucher.

6.4 Formalisation de la dissonance cognitive en terme d'éléments et de contraintes

De nombreuses formalisations et modélisations des phénomènes de dissonance cognitive ont été proposées par le passé en psychologie [Harmon-Jones et Mills, 1999, Chapitre 3], sans que celles-ci ne soient adaptées au cadre conceptuel et technique des systèmes multi-agents. Notre reformulation de la théorie de la dissonance cognitive emprunte à la théorie de la cohérence du philosophe computationnel Thagard et Verbeurgt [1998] qui nous permet de faire directement le lien entre la théorie de la dissonance cognitive et les notions, plus communes en informatique, d'éléments et de contraintes. Dans notre théorie, les *éléments* sont les cognitions privées et publiques des agents : croyances, désirs, intentions et engagements. Les éléments sont partitionnés en deux ensembles : l'ensemble A des éléments acceptés (qui sont interprétés comme crus vrais, activés ou valides selon le type des éléments) et l'ensemble R des éléments rejetés (qui sont interprétés comme crus faux, inactivés ou non valides selon le type des éléments). Tous les éléments qui ne sont pas explicitement acceptés sont rejetés. Les *contraintes* binaires[9] sur ces éléments sont induites des relations qui existent entre ces éléments dans le modèle cognitif de l'agent :

- *Les contraintes positives* : des contraintes positives correspondent aux relations de cohérence ou de consonance que sont les relations d'explication, les relations de déduction, les relations de facilitation et toutes les associations jugées positives.
- *Les contraintes négatives* : des contraintes négatives sont induites des relations d'incohérence ou de dissonance comme l'exclusion mutuelle, l'incompatibilité, l'inconsistance et toutes les relations jugées négatives.

À chacune de ces contraintes est attribué un poids reflétant l'importance et le degré de validité de la relation sous-jacente. Ces contraintes peuvent être satisfaites ou pas : une contrainte positive est satisfaite, si et seulement si, les deux éléments qu'elle lie sont soit tous les deux acceptés soit tous les deux rejetés. À l'inverse, une contrainte négative est satisfaite si et seulement si un des deux éléments qu'elle lie est accepté et l'autre rejeté. Ainsi, deux éléments sont dits *cohérents* s'ils sont liés par une relation à laquelle correspond une contrainte satisfaite. Et inversement, deux éléments sont dits *incohérents* si et seulement si ils sont liés par une relation à laquelle correspond une contrainte non-satisfaite.

Étant donné une partition des éléments entre A et R, on peut mesurer le *degré de cohérence* d'un élément en calculant la somme des poids des contraintes afférentes à celui-ci qui

[9] Rappelons qu'une contrainte binaire est une contrainte qui lie deux variables, ici deux éléments.

sont satisfaites divisée par le nombre des contraintes afférentes[10]. Et symétriquement, on peut mesurer le *degré d'incohérence* d'un élément comme la somme des poids des contraintes non satisfaites divisée par le nombre total de contraintes afférentes. De la même façon, on peut mesurer la cohérence d'un ensemble d'éléments comme la somme des poids des contraintes afférentes à cet ensemble (les contraintes dont au moins un pôle est un élément de l'ensemble considéré) qui sont satisfaites divisée par le nombre des contraintes afférentes total. Symétriquement, l'incohérence d'un ensemble de cognitions peut être mesurée comme étant la somme des poids des contraintes non-satisfaites afférentes à cet ensemble divisée par le nombre de contraintes afférentes total.

Un des intérêts majeurs de la théorie de la dissonance cognitive capturé par notre formulation est de fournir une mesure de l'incohérence (la dissonance), c'est-à-dire une métrique de la cohérence cognitive. Si on considère que le poids des contraintes rend compte de l'importance des cognitions les unes pour les autres, la mesure d'incohérence définies précédemment correspond précisément à la mesure utilisée par Festinger (définie à la section 6.3). Dans notre formalisation, une contrainte non-satisfaite est une relation de dissonance tandis qu'une contrainte satisfaite est une relation de consonance.

Ainsi, les hypothèses de Festinger s'appliquent également dans notre cadre. L'incohérence (ce que Festinger nomme la dissonance) produit chez le sujet une tension qui l'incite au changement. Plus l'incohérence est intense, plus l'insatisfaction est forte et plus la motivation pour la réduire l'est aussi. La seconde hypothèse de Festinger reste qu'en cas d'incohérence, l'individu ne va pas seulement changer ses cognitions ou essayer de modifier l'environnement (en particulier les cognitions d'autres agents) pour essayer de la réduire, il va aussi éviter toutes les situations qui risquent de l'accroître.

6.5 Dissonance, changement d'attitude et influence sociale

Dans les systèmes multi-agents, la question de savoir quand l'agent doit essayer de modifier l'environnement (entre autres, la couche des engagements sociaux publics) pour satisfaire ses intentions et quand l'agent doit modifier ses états mentaux pour être cohérent avec son environnement est cruciale. Dans notre modèle, l'agent cherche à maximiser sa cohérence, c'est-à-dire qu'il cherche à réduire ses incohérences en commençant par la plus intense. Pour réduire une incohérence, l'agent doit accepter ou rejeter certaines cognitions de manière à satisfaire au mieux les contraintes qui les lient. Ces cognitions peuvent être privées (les états mentaux) ou publiques (les engagements). Mais toutes les cognitions ne sont pas également

[10] Pour une mesure normalisée, on peut également calculer la somme des poids des contraintes afférentes à celui-ci qui sont satisfaites divisée par la somme de l'ensemble des poids de contraintes afférentes.

modifiables : c'est la notion de résistance au changement d'une cognition définie à la section 6.3. Pour pouvoir intégrer la communication dans notre modèle, il faut maintenant introduire le lien fondamental qui existe entre notre formulation de la théorie de la dissonance cognitive et la notion d'engagement.

Les engagements sociaux sont des cognitions particulières qui ne sont pas modifiables individuellement, mais doivent être socialement établis (c'est l'objet des jeux de dialogue que de fournir des outils pour la manipulation conjointe des engagements). C'est-à-dire que pour modifier, faire rejeter en l'annulant ou faire accepter un nouvel engagement afin de réduire une incohérence, un agent doit dialoguer. C'est par le dialogue que les agents vont essayer de faire établir les engagements sociaux cohérents avec leurs autres cognitions. Cependant, à l'issue de ces dialogues, certains engagements peuvent rester incohérents tout en étant plus difficilement modifiables (en particulier du fait des sanctions qui sont éventuellement attachées à ces modifications ultérieures). Ce sont alors des « obligations sociales » et cela fixe un des pôles des contraintes qui leur sont liées. Pour réduire d'éventuelles incohérences tout en se conformant aux engagements pris, c'est alors ses états mentaux que l'agent devra changer pour rétablir la cohérence. C'est le ressort du changement d'attitude dans notre système et cela formalise la vision des psychologues Brehm et Cohen [1962] à ce sujet, soutenue par des myriades d'expériences [Brehm et Cohen, 1962; Zimbardo et al., 1965; Zimbardo, 1977].

Pour ces psychologues, l'individu va chercher (par le dialogue) les engagements sociaux qui sont cohérents avec ses attitudes. Dans un débat d'idées, par exemple, chacun va poser des engagements propositionnels correspondant à ses croyances et être prêt à s'en justifier. Néanmoins, il n'est pas rare que l'agent accepte (ou se retrouve engagé sur) des engagements qui sont incohérents avec ses attitudes. Prenons l'exemple de l'argumentation : si l'individu est confronté à des éléments incohérents avec son point de vue, il peut argumenter pour défendre la cohérence de son point de vue. Supposons que de cette argumentation ressorte que le point de vue de l'interlocuteur est plus convainquant, plus acceptable que le sien (au sens du système d'argumentation choisi), il va alors devoir accepter/adopter les engagements posés par son interlocuteur. Ces engagements étant incohérents avec ses attitudes et les tentatives de modifier ces engagements ayant déjà échouées, il lui faut effectuer les changements d'attitudes rétablissant la cohérence. En effet, lorsque des engagements non modifiables sont incohérents avec ses attitudes, un des pôles de l'incohérence est fixé. Pour rétablir la cohérence, l'agent ne peut alors qu'essayer de revenir sur ces engagements dissonants (ce qui n'est pas toujours possible et porte généralement à conséquence via l'application de sanctions . . .) ou bien changer ses attitudes.

De nombreuses expériences, ainsi qu'une vaste littérature viennent préciser les modalités du changement d'attitude. Pour autant, la théorie de la dissonance cognitive ne fait pas de prédiction concernant le fait que le locuteur s'engage ou non sur des éléments dissonants,

c'est-à-dire incohérents. Cependant, dans notre théorie de la cohérence cognitive, les outils fournis peuvent être utilisés à cette fin. Lorsque cela arrive, c'est-à-dire un agent s'engage sur une voie incohérente avec ses attitudes (généralement par soumission sociale), il se peut alors que ces attitudes soient modifiées pour réduire l'incohérence. Des expériences montrent cela. Par exemple, Brehm a montré au travers des goûts culinaires que moins une personne aime un met, plus le fait de s'être engagé à en manger va susciter un changement d'attitude envers ce produit[11]. Évidemment, ladite personne aura, a priori, tendance à éviter ce type d'engagements. C'est-à-dire que ce phénomène ne survient que si l'individu est obligé de maintenir cet engagement. Sinon, toutes les autres tentatives de réduction sont observées : tentative de décrédibiliser ce pour quoi ou celui pour qui il s'est engagé, tentative de désengagement, rejet de la communication, rejet de l'information, ... Il faut donc que toutes ces possibilités soient rendues difficiles par la situation (par exemple, du fait des sanctions sociales ou autres qui pèsent sur le désengagement).

Le chapitre 7 illustre ce mécanisme de changement d'attitude et la théorie de la cohérence cognitive dans une application à l'automatisation de l'utilisation des jeux de dialogue de DIAGAL par des agents BDI.

6.6 Extension à la communication agent

6.6.1 Application aux systèmes multi-agents

Dans le cadre des systèmes multi-agents, les mesures de cohérence définies précédemment définissent une métrique de la cohérence cognitive qui peut être appliquée à l'agent comme aux groupes d'agents. Ainsi, le modèle de cohérence cognitive que nous proposons s'articule autour de deux axes :

- *la cohérence cognitive comme moteur cognitif et comportemental* : un agent travaille à maintenir une cohérence interne la plus élevée possible. Une incohérence qui ne peut être réduite par l'agent seul fera l'objet de communications ou de changements d'attitudes.
- *la cohérence cognitive comme moteur social* : dans un cadre coopératif (au sens de Camps [1998]), les agents travaillent à maintenir la cohérence du groupe la plus élevée possible. Dans le cas général, la notion mérite également d'être étudiée.

[11] Ceci n'est vrai que dans la mesure ou le changement d'attitude n'est pas justifié autrement (voir, à ce propos, le raffinement du modèle introduit ici présenté à la section 8.2).

Pour ces deux dimensions, celle de l'agent individuel et celle du groupe d'agents (au moins dans les situations coopératives), la cohérence cognitive proposent un système de régulation (une homéostasie). Puisque les agents cognitifs sont des systèmes complexes et que cette complexité est accrue par leur multiplicité dans le groupe, de tels systèmes de régulation sont nécessaires. En effet, se pose pour les systèmes multi-agents les mêmes questions que pour les systèmes complexes : comment garantir la convergence vers des états émergents qui soient satisfaisants en rapport à ce pourquoi ces systèmes ont été développés ? La théorie de la cohérence cognitive propose un système de régulation de la cohérence générique qui mérite d'être exploré dans ce sens. Il nous semble en outre que cette théorie serait apte à fournir de nombreuses indications pour résoudre des problèmes de conception des systèmes multi-agents.

Cependant, pour des raisons de temps et d'espace, nous n'explorerons en fait - dans cette thèse - qu'une partie du premier axe, c'est-à-dire que nous considérerons la théorie comme un moteur comportemental pour la cognition liée à la communication. Le chapitre 8 présente nos perspectives en discutant celles qui dépassent le cadre des aspects cognitifs de la pragmatique qui nous intéresse ici.

Le corps de notre théorie de la cohérence cognitive pour la pragmatique des communications agent ce construit en deux étapes :

1. *l'unification de la théorie de la dissonance cognitive de Festinger avec celle de la cohérence cognitive de Thagard* : cette unification permet d'obtenir un modèle de la dissonance cognitive qui soit formel et adapté aux systèmes multi-agents. Ce modèle a été présenté dans la section précédente. Il propose une théorie cognitive motivationnelle très générale.

2. *une extension à la communication* : le modèle précédent, s'il est d'application globale ne donne aucune information spécifique sur la pragmatique des communications agents. Dans cette section, nous étendons ce modèle au cas distribué pour y inclure le traitement de la communication. Cette extension embrasse la théorie de Festinger comme celle de Thagard.

Dans les sections suivantes, nous présentons cette extension à la communication. Nous indiquons, ensuite, différents problèmes théoriques, spécifiques ou pas aux communications dans les systèmes multi-agents, que la théorie de la cohérence cognitive permet de modéliser ou pour lesquels elle peut être utile.

6.6.2 Typologie des incohérences

Cette sous-section propose une typologie des incohérences qui adapte notre formulation de la théorie de la dissonance cognitive à un cadre explicitement distribué comme celui des systèmes multi-agents. La typologie que nous avançons a pour but d'introduire un vocabulaire simple, mais utile pour traiter des problèmes de cohérence. L'incohérence étant conceptuellement très proche de la notion de conflit, la typologie suivante est empruntée à nos travaux passés sur le concept de conflit [Pasquier et Dehais, 2000] :

- *incohérences internes et externes* : une incohérence est *interne* quand toutes les cognitions impliquées sont relatives à un même agent et *externe* quand les cognitions incohérentes impliquent au moins deux agents. Plus concrètement, une incohérence est externe pour un agent si c'est une incohérence entre des éléments de son modèle du monde et des éléments de ce qu'il connaît des modèles des autres.
- *incohérences explicites et implicites* : nous définissons *explicite* par le fait d'être dans « l'état d'avoir connaissance de » et *implicite* par celui d'être dans « l'état de ne pas avoir connaissance de ». On peut avoir connaissance de quelque chose sans être dans « l'état d'avoir connaissance ». C'est le cas de l'oubli par exemple (on a la connaissance que pour conduire la nuit, il faut allumer les phares, mais il peut arriver que l'on oublie). Une incohérence est *explicite* pour un agent si toutes les cognitions qui y participent sont explicites pour l'agent. Une dissonance est *implicite* si au moins une des cognitions incohérentes est implicite pour au moins une des parties concernées. Il est à noter qu'une incohérence implicite est une incohérence explicite potentielle. Notons également que dans un cadre multi-agent l'incohérence interne sera sans doute toujours explicite. La figure 6.1 de la section 6.6.4 détaille cette typologie et la section 6.6.9 fournit un exemple de chaque type de d'incohérence.

6.6.3 Lien cohérence - initiative, sujet et pertinence

Un problème particulièrement délicat en intelligence artificielle comme dans le reste des sciences cognitives est celui de l'ouverture du dialogue. Pourquoi un agent prend-il l'initiative d'un dialogue ? Quand doit-il la prendre et à quel sujet ? La réponse fournie par notre cadre de cohérence est qu'un agent prend l'*initiative* d'un dialogue s'il a une incohérence à réduire et qu'il ne peut la réduire seul. Soit parce qu'il sait que c'est une incohérence externe qui implique d'autres agents, soit parce qu'il s'agit d'une incohérence interne et qu'il n'a pas les capacités de la résoudre seul (il doit alors compter sur la coopération d'autres agents). Notons, en outre, que l'incohérence qui initie le dialogue en donne également *le sujet*.

On peut régler le seuil de l'intensité de l'incohérence à partir duquel un agent prend l'initiative du dialogue, cela définit une partie de son « caractère ». Le « réglage » de ce paramètre est un sujet de recherche ouvert. Ce « réglage » est en fait un problème de choix de modélisation comparable à celui de l'affectation des priorités entre attitudes propositionnelles qui définissent le caractère d'un agent dans l'architecture d'agent BOID [Broersen et al., 2001]. Aussi si un agent fait passer ces croyances avant toute autre cognition, il sera dit réaliste, tandis que s'il fait passer ses désirs avant toute autre chose, il sera dit hédoniste.... De la même manière si l'agent prend l'initiative d'un dialogue à la moindre incohérence rencontrée, il sera dit bavard, tandis que s'il ne s'exprime que lors d'incohérences de fortes magnitudes, il sera dit renfermé....

Finalement, la notion de *pertinence* joue un rôle central en pragmatique. Avec la théorie de la pertinence, Sperber et Wilson [1986] ont avancé l'idée que le locuteur choisit ce qu'il va dire en évaluant dynamiquement la pertinence de ses idées. Cette pertinence varie pour chaque élément de cognition au cours de la conversation. Le locuteur ne s'engage dans un acte de langage que lorsque la pertinence en est maximale. Avec notre approche, un agent qui prend l'initiative va s'attaquer à l'incohérence qui a la plus grande magnitude, car c'est ce qui est cognitivement le plus pertinent pour lui (en vertu du principe de cohérence). La section suivante indique comment le cadre de cohérence permet à l'agent de choisir quel type de dialogue engager.

6.6.4 Lien avec les types de dialogues

Dans cette sous-section, nous analysons comment les types de dialogues observés en dialectique peuvent être liés à la cohérence cognitive. Un certain nombre de travaux récents utilisent pour les dialogues la typologie de Walton et Krabbe [1995]. Ils y distinguent six types de dialogues définis par leur but premier (auquel les interlocuteurs souscrivent) et les buts/intentions propres à chacun (qui peuvent être incompatibles, c'est-à-dire incohérents).

1. *la persuasion* : la situation initiale est une incohérence externe de point de vue et le but global est de la résoudre. Chaque participant essaye de ne pas changer ses croyances (c'est la résistance au changement) et de faire changer celles des autres. Pour ce faire, les agents ont typiquement recours à l'argumentation [Keefe, 1991]. La persuasion est donc une technique de réduction d'incohérence externe. Ceci étant dit, il s'agit en fait pour un agent qui souhaite persuader, de montrer aux autres que son point de vue est plus cohérent que le leur. Alors qu'argumenter se résume, pour un agent, à exhiber des éléments cohérents avec sa proposition (c'est-à-dire liés à elle par des contraintes satis-

faites). À l'inverse, attaquer/réfuter l'autre, consiste à exhiber des éléments incohérents avec sa proposition (c'est-à-dire, liés à elle par des contraintes non satisfaites).

2. *la négociation* : à partir d'un conflit d'intérêts (un type d'incohérence externe), le but global est de conclure un contrat, d'arriver à un accord. Chaque agent a son propre but et veut maximiser son profit, ses intérêts. La résolution du conflit se fait habituellement par un échange d'offres et de contre-offres. Il est fréquent que des dialogues de persuasion soient imbriqués dans une négociation, les offres étant ainsi argumentées. La négociation est donc une technique de réduction d'incohérence externe explicite.

3. *l'investigation* : les participants de ce type de dialogues sont dans une situation d'incohérence interne partagée (ils souffrent tous d'une incohérence interne qui a été reconnue comme commune). Le but commun coïncide avec les buts individuels. Il s'agit de prouver un fait pour renforcer la cohérence. C'est une réduction d'incohérence interne partagée (c'est-à-dire que les agents impliqués dans le dialogue ont reconnu qu'ils ont la même incohérence à résoudre et ils s'entraident)

4. *la délibération* : chacun des agents participant au dialogue a ses préférences et tous doivent choisir parmi les offres de chacun. Les participants ont pour but commun de prendre une décision (choisir un plan ou une action). Leur but individuel est d'influencer la décision dans leur intérêt (les préférences des uns et des autres ayant généralement été reconnues différentes, pour qu'il y ait délibération) ou dans ce qu'ils considèrent être l'intérêt commun (afin d'accroître leur cohérence). C'est une technique de réduction d'incohérence externe explicite.

5. *la recherche d'informations* : c'est le seul type de dialogue qui est toujours asymétrique. Un agent cherche à obtenir de l'information des autres. Il s'agit d'une technique de réduction d'incohérence interne. Dans ce type de dialogue, seul l'agent demandeur d'information est dans un état d'incohérence. La réduction est asymétrique, mais pour la faciliter, il est fréquent que le demandeur explicite son incohérence aux autres agents (en indiquant pourquoi il cherche de l'information, c'est-à-dire en explicitant son incohérence). Cette technique de réduction peut prendre la forme d'un dialogue, mais peut aussi prendre la forme d'autres actions (par exemple : lire un livre de référence, chercher sur Internet, ...) pour autant que l'incohérence soit réduite.

6. *la dispute* : la situation initiale est conflictuelle et incohérente. Contrairement aux autres types de dialogues, la dispute n'est pas rationnelle, elle fait généralement plus appel aux sentiments, émotions et pulsions qu'à la raison et à la cohérence. C'est pourquoi nous ne la détaillerons pas ici.

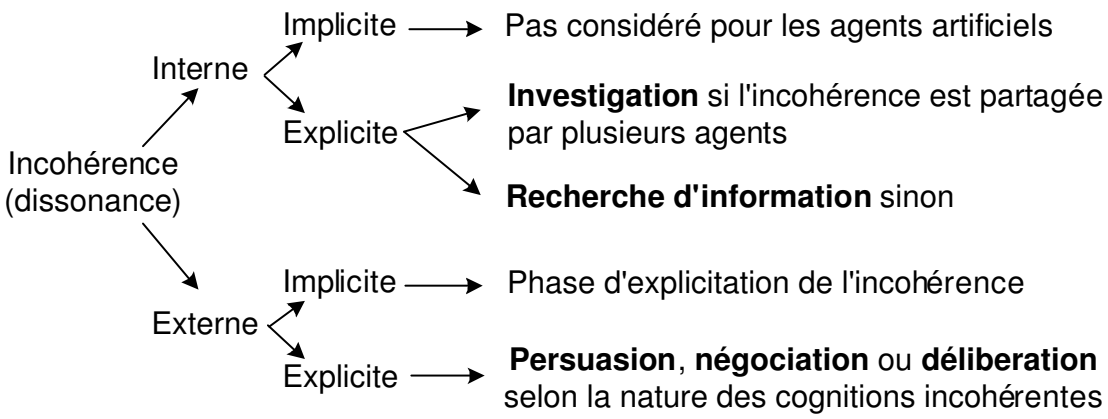

FIG. 6.1 – Typologie des incohérences/dissonances cognitives et lien avec les types de dialogue.

Comme on peut le constater, tous les dialogues naissent d'une situation d'incohérence. Notons que la réciproque n'est pas vraie, c'est-à-dire toutes les incohérences ne sont pas traitées par le dialogue.

> *Notre conjecture est donc que les agents communiquent si une incohérence les y contraint.*

La figure 6.1 résume les différents types d'incohérence ainsi que les types de dialogue qu'elles peuvent engendrer dans le cas général. Le chapitre suivant raffine ce choix dans le cadre plus spécifiquement multi-agents des jeux de dialogue du langage DIAGAL (chapitre 5).

Dès lors, la conversation est vue, outre ses caractéristiques déjà connues, comme une procédure générique d'explicitation et de tentative de réduction ou d'évitement d'incohérence. De par la proximité conceptuelle entre les notions de dissonance, d'incohérence et celle de conflit, cette conjecture est proche de la position classique de la dialectique : tout dialogue naît d'un conflit [Hamblin, 1970]. Depuis quelques années, plusieurs auteurs insistent de nouveau sur le rôle des conflits dans le processus conversationnel :

- pour Walton et Krabbe [1995], la question « Y a-t-il un conflit ? » est à la base de l'analyse des situations initiales du dialogue (voir figure 1.2) ;
- pour Reed et Long [1997],« nombre de dialogues ont pour situation initiale un conflit » ;

– pour Dessalles [1998b], « un grand nombre de dialogues trouvent leurs origines dans les conflits cognitifs entre désirs et/ou croyances » ;

– pour Baker [1991], « les dialogues résultent d'une opposition entre des buts conflictuels ».

6.6.5 Lien cohérence - explicitation

Pour qu'un agent envisage de réduire une incohérence, il faut qu'il ait connaissance de cette incohérence. Au sens de notre typologie, cela signifie que l'incohérence attaquée doit être explicite. Nous avons fait l'hypothèse que l'incohérence interne est toujours explicite en cela que nous ne traitons pas de niveau implicite dans les modèles des agents cognitifs[12]. Par contre, nous avons vu que les incohérences externes sont implicites pour un agent s'il n'est pas dans l'état d'avoir connaissance de toutes les cognitions impliquées dans l'incohérence. Cela signifie que dans la plupart des cas l'incohérence doit être explicitée avant qu'un dialogue concernant sa réduction soit entamé.

Il existe de nombreuses méthodes par lesquelles une incohérence externe implicite (potentielle) peut être explicitée. Tout d'abord, notons que l'incohérence n'est pas nécessairement implicite pour tous les agents impliqués. Dans ce cas, c'est généralement un agent pour qui l'incohérence est explicite qui prend l'initiative du dialogue visant à en entamer la réduction. Ce dialogue comportera nécessairement une phase d'explicitation de ladite incohérence (cette phase est indiquée figure 6.1).

Par exemple, avant qu'un acheteur et un vendeur négocient un prix, au moins l'un des deux doit informer l'autre de son intention. Dans ce cas, une convention (dans le monde occidental) veut que le vendeur indique le prix qu'il souhaite vendre l'item. L'acheteur est en accord avec le prix proposé ou pas. Si le prix proposé n'entre pas en incohérence avec les vues de l'acheteur, celui-ci va essayer de réduire son incohérence interne entre son désir de posséder l'item et sa croyance de ne pas être en possession de celui-ci en faisant une tentative pour acheter le produit. Par contre, s'il n'est pas d'accord avec le prix proposé, disons $p1$, cela va expliciter une incohérence externe entre son intention d'acquérir l'item pour $p2$, le prix souhaité, et l'intention du vendeur de s'en séparer pour $p1$. Plusieurs dialogues, entre autres comportements, peuvent alors faire suite. L'agent peut initier un dialogue de négocia-

[12] Cette hypothèse peut être levée si nos agents sont des humains et l'on peut alors étudier le rôle d'éléments de cognition implicites mais actifs. En particulier, on peut observer leurs expressions dans le comportement manifeste (par l'intermédiaire de lapsus, par exemple). Cet aspect sur lequel notre théorie reste ouverte a été présenté et discuté avec les membres de l'école Lacanienne de psychanalyse de Montréal lors du colloque « État de la psychanalyse en 2004 ».

tion dans lequel une des premières phases sera sans doute d'expliciter l'incohérence externe qui le préoccupe à l'agent vendeur. D'éventuels sous dialogues de persuasion seront ensuite imbriqués durant lesquels les agents essayeront de justifier et de convaincre du bien-fondé de leurs propositions. Finalement, un dialogue de délibération visant à prendre une décision en ce qui concerne l'échange peut également être entamé, ... Dans tous les cas, aucun de ces dialogues ne peut survenir tant que l'incohérence n'aura pas été explicitée.

Dans de très nombreux cas similaires, l'explicitation a lieu via une convention sociale de diffusion d'information qui assure qu'au moins un agent est dans l'état d'avoir connaissance des potentielles incohérences externes. On pense au tableau de réservation des cours de tennis ainsi qu'à d'innombrables dispositifs de coordinations dont le sens pourrait être revisité à la lumière de notre typologie des incohérences. D'autres cas nécessitent des processus d'explicitation ad hoc.

Finalement, notons que même dans le cas d'une incohérence interne, l'agent devra généralement expliciter son incohérence pour permettre à son interlocuteur de mieux l'aider à la réduire.

6.6.6 Lien cohérence - projet conjoint

La communication entre agents se fait par conversation, séquence d'actes de langage dont la somme des sens isolés ne rend pas compte de la signification. C'est pourquoi, de nombreux chercheurs, influencés par les idées de Searle [1990] et Clark [1996], voient la conversation comme une activité conjointe et rendent aux conversations leur dimension sociale et tentent d'analyser le dialogue du niveau conversationnel vers le niveau des actes de langage plutôt que le contraire (voir section 1.3.2). Nous proposons de voir le dialogue comme *un projet conjoint de réduction d'incohérence*.

À notre sens, le modèle de cohérence cognitive peut fournir des éléments de compréhension sur la nature du caractère conjoint des conversations. Selon notre hypothèse et en accord avec le principe de cohérence, *les agents communiquent pour réduire, éviter ou empêcher l'accroissement d'une incohérence cognitive*.

Dès lors que le dialogue est entamé parce qu'une incohérence ne peut être réduite par un agent seul, cette réduction (le dialogue) est un projet commun, une activité conjointe. On y reconnaît alors une intention commune quant à un but commun, même si cette intention reste

généralement implicite dans les dialogues entre humains[13]. Ajoutons que les étapes de mise en commun (*établissement* par présentation et acceptation, voir section 2.3.3) introduites dans le cadre interactionnel sont indispensables pour que cette réduction commune soit un succès. Sans mise en commun, la réduction risquerait d'être unilatérale c'est-à-dire un leurre, un malentendu, une incompréhension, un quiproquo, ...

Nous différencions deux aspects dans cette mise en commun. Il y a une mise en commun structurelle capturée par les jeux de dialogue. Par exemple, il faut s'assurer que lorsqu'un joueur propose une négociation, l'autre accepte de négocier avant que celle-ci ne débute. Les agents doivent être d'accord sur le type de dialogue dans lequel ils s'engagent. Mais comme les dialogues sont des tentatives de réduction de dissonance, une mise en commun cognitive de cette dissonance est également nécessaire à leur bon déroulement. Cela signifie que le sujet du dialogue, la dissonance à laquelle le dialogue s'attaque doit être explicitée pour tous les agents qui participent au dialogue. C'est-à-dire qu'en plus d'être prêts à négocier, il faut que les agents soient d'accord sur le problème, c'est-à-dire la dissonance à régler par la négociation.

En différenciant les incohérences externes implicites des dissonances externes explicites, nous avons souhaité montrer en quoi la phase d'explicitation de la dissonance est cruciale. En effet, pourquoi un agent accepterait-il de s'engager dans un type de dialogue si cela ne règle pas un problème dans lequel il se sait impliqué ? Plus précisément, un agent peut être amené à communiquer à cause d'une incohérence interne ou externe (incohérence entre son modèle du monde et ce qu'il connaît des modèles d'autrui). Dans les deux cas, le dialogue devra passer par une phase d'explicitation de la dissonance. L'incohérence devra être mise en commun, l'exemple 6.1, présenté ci-bas, illustre cette nécessité.

En effet, les dialogues de type négociation, persuasion et délibération sont tous initiés dans des cas d'incohérences externes (voir section 6.6.4). Les agents qui participent au dialogue sont généralement les agents impliqués dans ladite incohérence et sa résolution constitue leur projet conjoint. Dès lors, il est indispensable que chaque étape soit validée, acceptée par chacun des agents (établissement) ce sans quoi la résolution ne sera pas commune et la dissonance risque de subsister.

Dans le cas de l'investigation, l'incohérence n'est pas externe mais interne partagée. Les agents doivent alors expliciter cette incohérence pour s'assurer qu'elle est bien commune (et que les incohérences de chacun ne diffèrent pas trop, de sorte que le problème à résoudre soit

[13] Les expériences montrent bien que malgré les processus physiologiques qui y sont attachés, les humains agissent sans connaître ni reconnaître la théorie de la dissonance cognitive. Comme le font remarquer Dessalles et Ghadakpour [1999], il y a néanmoins des domaines ou le processus est effectué consciemment. C'est le cas, par exemple, avec la recherche scientifique où les chercheurs explicitent une problématique (une dissonance, un conflit, une incohérence) avant d'en débattre.

adopté comme étant commun). Ils vont ensuite entamer la résolution/réduction comme projet conjoint.

Finalement, dans le cas de la recherche d'informations, le projet qui doit être adopté comme commun vise à réduire l'incohérence interne du demandeur d'information. Généralement, celui-ci partage son incohérence avec un agent choisi pour sa bonne volonté coopérative, ses compétences ou la qualité de leur relation dans l'espoir que ce dernier l'aide à résoudre son incohérence.

Exemple 6.1

L'agent A prévoit d'utiliser la ressource non-partageable R pendant un intervalle de temps futur T et l'agent B aussi. Tant que A et B n'ont pas communiqué, l'incohérence reste implicite : seul un observateur averti peut avoir conscience du conflit de ressource qui attend nos deux agents. En l'absence d'un système de gestion de la ressource R, pour qu'un agent divulgue spontanément son intention quant à la ressource R, il faut :

- qu'il sache que R est une ressource non partageable et qu'il y a donc un risque de conflit potentiel ;
- qu'il sache que l'incohérence (ou le conflit) sera d'autant plus intense qu'elle adviendra au dernier moment (le moment de l'utilisation de la ressource), ce qui l'amène à désirer anticiper et à expliciter l'incohérence potentielle sur le champ ;
- qu'il sache qui pourrait vouloir utiliser la ressource (ici B) ;

Muni de ces connaissances contextuelles, A sait quoi communiquer et à qui le communiquer tout comme il sait pourquoi cela peut être utile. Il saura a posteriori si la conversation aura été utile ou pas selon qu'une incohérence externe aura été explicitée ou pas, puis réduite ou pas.

6.6.7 Lien cohérence - utilité et dynamique du dialogue

Utilité des dialogues

En théorie de la décision, comme en micro-économie, la notion d'utilité est une propriété de certaines fonctions de valuation. Une fonction de valuation est une fonction d'utilité si et

seulement si elle rend compte des préférences des agents. Dans la théorie de la cohérence cognitive définie ci-dessus, en vertu du principe de cohérence, les agents préfèrent la cohérence à l'incohérence. Ainsi, les mesures de cohérence sont interprétables directement comme des mesures d'utilité. Il s'en suit que le gain d'utilité d'une action est simplement défini comme une différence entre la cohérence avant et après l'occurrence de l'action. Cela signifie en particulier que les agents peuvent calculer l'*espérance d'utilité* des dialogues qu'ils envisagent. Cette espérance d'utilité est égale à l'intensité de l'incohérence de l'élément auquel le dialogue s'attaque de laquelle on retranche l'intensité de l'incohérence de cet élément après le dialogue, si celui-ci réussit en faveur de l'agent[14]. Les agents peuvent également calculer l'*utilité* d'une conversation dynamiquement en recalculant l'incohérence de l'élément en question au cours du dialogue. Lorsqu'une unité de dialogue est terminée, soit l'incohérence est réduite et le dialogue se termine, soit l'agent peut continuer (par un enchaînement dialogique ou par un changement d'attitude) à essayer de la réduire.

Dynamique intra-dialogue

Un agent sélectionne un type de dialogue en fonction du type de l'incohérence qu'il souhaite réduire, c'est-à-dire du type de problème qu'il souhaite régler. Mais au cours de cette résolution, d'autres incohérences peuvent apparaître. Celles-ci peuvent parfois devoir être réduites pour que la réduction principale puisse se poursuivre. C'est ce qui amène les agents à imbriquer un sous-dialogue pour réduire la nouvelle incohérence avant de reprendre le dialogue principal concernant l'incohérence première. Dans d'autres cas, l'incohérence peut se déplacer, amenant les agents à enchaîner deux types de dialogues. On a donc également un outil pour gérer l'imbrication et l'enchaînement des unités de dialogue au cours d'une conversation.

Par exemple, il est habituel de délibérer sur les modalités de paiement ainsi que des conditions de réception de la marchandise lors de la négociation d'un produit. Les agents souhaitent que cet aspect du contrat recherché soit cohérent avec leurs attentes (probablement divergentes) à ce sujet. Le fait que cela ne soit pas le cas est une incohérence (au sens de notre théorie). C'est l'irruption de cette nouvelle incohérence liée au reste de la négociation du prix qui va amener les agents à imbriquer une délibération. En effet, les modalités de paiement

[14] Le qualificatif espéré renvoie ici à : dans le cas où le dialogue, comme tentative de modification de la couche sociale, serait un succès et non à l'espérance probabiliste habituellement associée aux mesures d'utilité dans la théorie de la décision classique. À cet égard, les probabilités n'interviennent pas dans notre formalisation car nous nous situons sous une hypothèse d'équiprobabilité des alternatives qui représente bien l'incertitude des agents. L'introduction des probabilités et de leur apprentissage dans notre formalisme est envisagée comme l'une de nos perspectives, présentée section 8.4.3.

sont un élément important de la négociation d'un produit et son imbrication dans cette dernière garantit de pouvoir tenir compte de cette sous-résolution pour la suite de la négociation.

Pour faire le lien avec la théorie du jugement de l'école de Yale[15], il n'est pas rare que les engagements pris en définitive dans un dialogue soient des compromis entre les attitudes défendues par les différents interlocuteurs (c'est par exemple le cas typique dans les dialogues de négociation). De sorte que, pour réduire l'incohérence externe qui les occupait, chacun des agents accepte de réduire une incohérence interne plus faible en changeant ses attitudes. Ainsi, des changements d'attitudes peuvent avoir lieu durant un dialogue, en modifiant ainsi la dynamique.

Dynamique inter-dialogues

Dans notre théorie, les dialogues sont considérés comme des *tentatives de réduction d'incohérence*. Bien entendu, de telles tentatives peuvent échouer. La mesure d'utilité des dialogues définie dans le cadre de consonance permet de guider l'agent dans sa conduite communicationnelle. En effet, suite à un dialogue pas ou peu utile, c'est-à-dire lorsque l'incohérence n'est pas réduite, l'agent doit décider comment réagir. L'agent va probablement persévérer dans sa tentative de réduction en prenant en compte cet échec : il proposera un type de dialogue différent ou une proposition différente du même type de dialogue, il prendra note de cet échec qui pourra être utile pour le guider lors des dialogues/tentatives suivants. C'est, par exemple, cette mesure d'utilité qui peut lui permettre de ne pas tenir plusieurs fois de suite le même dialogue infructueux avec le même agent.

En particulier, dans un cadre ouvert et hétérogène, un agent est amené à communiquer avec des agents inconnus. Il lui faut alors se faire une idée des dialogues tenus avec ceux-ci. L'agent pourra tenir compte de l'utilité des dialogues tenus par le passé pour sélectionner ses interlocuteurs. Un agent aura intérêt à renforcer ses échanges avec les agents avec lesquels les dialogues sont utiles et de nombreuses incohérences (c'est-à-dire problèmes) ont été résolues. À l'inverse, il pourra tenir compte des dialogues inutiles en affaiblissant ses liens sociaux avec les interlocuteurs concernés. C'est-à-dire que la mesure de l'utilité des dialogues fournit une information précieuse qui pourra être utilisée par un outil de gestion des accointances (qui pourrait être opérationnalisé via l'apprentissage par renforcement, ce qui n'est pas notre objet ici).

Par exemple, en commerce électronique, un agent n'a pas intérêt à continuer de tenter des négociations qui n'aboutissent jamais (peut-être ses conditions de paiement et celles de

[15] Celle-ci est très brièvement introduite en section B.5.

l'interlocuteur sont incompatibles et les sous-dialogues de négociation de celles-ci n'aboutissent pas non plus). Au contraire, il aura intérêt à procéder avec des agents avec qui il tient des dialogues utiles. Dans ce cadre de commerce électronique, le cadre de dissonance (on suppose la modélisation réussie) pourrait être utilisé pour garantir que les dialogues les plus utiles correspondront aux contrats les plus intéressants, les plus satisfaisants.

6.6.8 Lien cohérence - humeur, intensité

Ces dernières années, le besoin d'intégrer les émotions aux agents artificiels s'est fait sentir [Bates, 1994; Velasquez, 1997]. Le modèle de consonance proposé permet de faire un lien direct entre les mesures de cohérence et l'humeur de l'agent. Notre théorie fournit un système de valeurs dans lequel un gain de cohérence est un soulagement, une joie, un réconfort. Un état de cohérence est un état de bien-être (sourire, aspect relaxé, ...). À l'inverse, un agent peut avoir peur d'une incohérence perçue comme potentiellement future, être préoccupé ou malheureux d'un état d'incohérence ou encore déçu d'une tentative de réduction échouée (par exemple une conversation). On pense aux interfaces hommes/machines et aux systèmes de tutoriels intelligents, entre autres. En outre, certains cadres interactionnels autorisent l'utilisation de différents degrés d'intensité des forces illocutoires des actes de langage[16]. Pourtant, aucune théorie d'agents n'indique comment la sélection de ce degré d'intensité s'effectue. Les mesures quantitatives définies par la théorie de la cohérence cognitive fournissent selon nous le moyen de guider l'agent dans le choix du degré d'intensité approprié.

Dès lors qu'une conversation est engagée pour réduire une incohérence, il semble légitime que la magnitude de celle-ci influence de manière directe le choix des degrés d'intensité des actes à utiliser. Par exemple, un agent qui a besoin d'une information pour réduire une incohérence interne va se lancer dans un dialogue de recherche d'informations et produire un acte directif pour essayer d'obtenir cette information. Le degré d'intensité de la force illocutoire va dépendre du degré d'intensité de ladite incohérence : (1) une invitation ou un conseil, si l'incohérence est très légère, (2) une recommandation ou une demande, si elle est un peu plus intense et (3) une supplication, une imploration ou un ordre, si la magnitude de l'incohérence est très intense et donc sa réduction cruciale.

Si ces paramètres d'émotions, d'humeur et d'intensité semblent moins importants pour les systèmes multi-agents entièrement artificiels, cette piste est intéressante pour les systèmes homme-machine. Évidemment, ce facteur de sélection du degré d'intensité n'est pas unique, car il y a un certain nombre d'autres facteurs susceptibles d'intervenir dans ce choix : les

[16] Cette idée est valable pour les actes de langage classiques comme pour les autres actes de conversation : actes d'établissement, actes de dialogue, ...

conventions sociales (il est généralement interdit de donner un ordre à un supérieur hiérarchique. . .) et les relations entre les agents (proximité, confiance, passif de la relation, . . .) sont aussi importantes pour sélectionner les degrés d'intensité des forces illocutoires des actes posés.

6.6.9 Exemples supplémentaires

Cette sous-section présente deux exemples informels qui couvrent tous les types d'incohérence que l'on peut rencontrer dans les systèmes multi-agents ainsi que tous les types de dialogues tels que définis par Walton et Krabbe [1995]. On suppose que les agents utilisent notre modèle de la cohérence cognitive pour le choix des actes de langages, des actes conversationnels ou des coups du jeu de dialogue selon le cadre interactionnel utilisé.

Incohérence externe implicite et explicite (persuasion, argumentation, délibération)

Pour illustrer ce type d'incohérence, considérons deux agents A et B qui voyagent ensemble en voiture vers un lieu L. A est le chauffeur et B le passager. À un moment donné, ils s'arrêtent à un carrefour où ils peuvent tourner à droite ou à gauche. L'agent A a l'intention de passer par la droite alors que B a l'intention de passer par la gauche. Tant que ces informations n'ont pas été communiquées, on est en présence d'une incohérence externe implicite (en effet, les deux intentions sont incompatibles et les partenaire n'en ont pas connaissance mutuelle).

> 1.A *: On prend à droite ?* < l'agent A cherche un support pour renforcer son état de cohérence >

L'impact de cette information sur l'agent B dépend de la relation qu'entretiennent A et B. Pour l'agent B qui a l'intention contraire, cette intervention de A explicite une incohérence externe. B peut calculer sa réponse selon la résistance au changement de son intention de passer par la gauche et l'impact de l'information donnée par A qui influence la magnitude de l'incohérence, c'est-à-dire l'importance qu'il lui accorde.

Si la résistance au changement de cette intention est faible et/ou l'impact des communications avec A fort, il peut changer d'attitude pour réduire l'incohérence (qui est maintenant explicite pour lui et implicite pour A) et répondre sans plus discuter :

$2.B$ *: OK.* < fin de la conversation, incohérence réduite >

Si la résistance au changement de cette intention est suffisamment élevée et/ou l'impact des communications avec A faible, il peut expliciter l'incohérence pour ouvrir une conversation plus complète :

$2'.B$ *: Moi, je prendrais plutôt à gauche.* < B explicite pour A l'incohérence externe qu'il a perçue >

Les agents ayant mis leur modèle d'autrui à jour, l'incohérence est maintenant une incohérence externe explicite. À ce moment, si la résistance au changement de l'intention de A de tourner à droite est faible et/ou que sa relation avec B est forte, celui-ci peut abandonner son intention et réduire l'incohérence en concluant :

$3.A$ *: OK, on passe par la gauche.* < fin de la conversation, incohérence réduite >

Si au contraire, la résistance au changement de son intention est forte et/ou l'impact de la communication de B est faible, il peut se lancer dans un dialogue de persuasion en réaffirmant ce qui est cohérent avec son intention, c'est-à-dire en montrant pourquoi son intention est cohérente. Cela peut être :

$3'.A$ *: Par la droite, c'est plus court et nous sommes pressés.*

Selon ses états mentaux, B peut s'incliner ($4.B$) ou essayer de persuader A à son tour ($4'.B$) :

$4.B$ *: D'accord, allons-y.* < fin de la conversation, incohérence réduite >

$4'.B$: *Oui, mais il y a des embouteillages à cette heure-ci, ce sera plus rapide par la gauche.* < . . . , jusqu'à ce que l'un des points de vue l'emporte ou que d'autres impératifs l'emportent sur la conversation (le feu passe au vert ou l'automobiliste suivant klaxonne, . . .) >

Ce faisant, la magnitude de l'incohérence augmente. Finalement, A peut s'incliner ou bien surenchérir selon la résistance au changement de son intention de passer par la droite et l'impact des communications avec B, . . .

On note qu'il est difficile de savoir si les dialogues ci-dessus relèvent de la persuasion (chaque agent persuade l'autre d'adopter son point de vue), de la négociation (les deux agents négocient un chemin) ou de la délibération (chacun exprime ses préférences et le groupe doit choisir). Ce qui est sûr, c'est qu'il s'agit d'une réduction d'incohérence externe comme c'est le cas dans chacun de ces trois types de dialogues (voir section 6.6.4). On peut constater, en outre, que lorsque l'incohérence est réduite, le dialogue s'arrête. Le contraire n'est pas vrai, car il se peut que l'incohérence persiste, mais que le dialogue cesse, car le feu passe au vert et A devant agir, tourne à droite.

Incohérence interne (explicite) et incohérence interne partagée (recherche d'informations et investigation)

On rappelle l'exemple 3.1 de la section 3.2 :

1. A *: Est-ce que je peux te poser une question ?*
2. B *: oui, vas-y.*
3. A *: Est-ce que tu as l'heure ?*
4. B *: Non, j'aimerais bien la connaître.*

Dans cet exemple, l'agent A commence un dialogue de type recherche d'information pour obtenir l'heure (énoncé $1.A$ à $3.A$) mais B explicite une incohérence interne commune en répondant qu'il a la même incohérence interne que l'agent A à résoudre, car il n'a pas l'heure et il souhaiterait la connaître aussi (énoncé $4.B$). Dans cette situation d'incohérence interne commune explicite, le seul type de dialogue (si les agents décident de poursuivre le dialogue pour réduire l'incohérence plutôt que toute autre action) possible est l'investigation (voir section 6.6.4).

6.7 Conclusion

Dans ce chapitre, les théories motivationnelles reposant sur l'homéostasie et le principe de cohérence ont été introduites et la théorie de la dissonance cognitive a été détaillée. Notre apport théorique qui consiste en une formalisation et extension à la communication de ce cadre motivationnel générique a été présenté. En présentant les différentes dimensions de la communication agents capturée par cette approche, nous avons souhaité insister sur la vaste

couverture de la théorie proposée. En particulier, cette approche permet de traiter, même partiellement, les questions suivantes qui sont autant de dimensions des aspects cognitifs de la pragmatique :

- Quand un agent prend-il l'initiative d'une conversation, à quel sujet et pourquoi (voir section 6.6.3) ?
- Avec qui (voir section 6.6.3) ?
- Par quel type de dialogue (voir section 6.6.4) ?
- Quelle intensité donner aux forces illocutoires des actes de langage utilisés (voir section 6.6.8) ?
- Comment définir et mesurer l'utilité d'une conversation (voir section 6.6.7) ?
- Quand arrêter le dialogue ou le cas échéant comment le poursuivre (voir section 6.6.7) ?
- Quels sont les impacts du dialogue sur les attitudes de l'agent (voir section 6.5) ?
- Quels sont les impacts du dialogue sur l'humeur de l'agent (voir section 6.6.8) ?
- Quelles sont les conséquences du dialogue sur les accointances de l'agent (voir section 6.6.7) ?

La théorie de la cohérence cognitive présentée dans les sections précédentes doit être envisagée comme une sur-couche des architectures agents existantes. Son intégration sera réalisée via la reformulation des réseaux de cognitions en termes d'éléments et de contraintes de sorte que les différentes mesures de cohérence et d'utilité définies ci-dessus s'appliquent.

Le prochain chapitre (chapitre 7) détaille notre validation informatique de cette théorie dans un cadre strictement multi-agent et en utilisant notre modèle de l'engagement social (présenté au chapitre 4) et le langage DIAGAL (présenté au chapitre 5) comme cadre interactionnel. Le chapitre 8 propose et discute un raffinement en ce qui concerne la modélisation du changement d'attitude et la prise en compte des sanctions. Ce dernier chapitre présente également les perspectives ouvertes par cet apport théorique original ainsi que les usages que d'autres équipes de recherche ont faits de nos idées.

Chapitre 7

Application à l'automatisation de la communication agent avec DIAGAL

7.1 Introduction

Dans ce chapitre[1], nous présentons notre validation informatique du modèle pour le traitement des aspects cognitifs de la pragmatique des communications entre agents proposé au chapitre précédent (chapitre 6). Dans la mesure où nous ne proposons pas une approche entièrement cohérentiste de la modélisation agent (ce qui est l'une de nos perspectives, présentées à la section 8.4), nous avons validé notre approche en étendant un modèle d'agent déjà existant : le modèle d'agent BDI [Beliefs, Desires and Intentions] (section 7.2). Le modèle BDI est le modèle d'agent cognitif le plus répandu et le plus étudié, c'est pourquoi il constitue un bon cadre de validation pour notre approche. L'intégration de notre approche au modèle BDI ne pouvant se faire au niveau de sa formulation théorique qui est exprimée en terme de logiques modales incompatibles avec notre formalisme, nous la réaliserons directement au niveau de l'implantation du système.

Sans vouloir suppléer à l'importante littérature qui couvre cette approche et ses nombreux développements, la section suivante présente succinctement une vision procédurale du modèle BDI classique. Ensuite, nous présenterons notre étude des liens entre les états mentaux privés utilisés dans le modèle BDI et la notion publique d'engagement social, centrale à notre approche de la communication agent (section 7.3). Les éléments théoriques et algorithmiques

[1] Ce chapitre reprend en l'étendant le contenu de notre présentation au workshop international sur la communication agent associé à la conférence AAMAS 2003 qui eu lieu à Melbourne en Australie [Pasquier et al., 2003]. Certaines parties de notre article à paraître dans le journal Cognitive System ont également été utilisées [Pasquier et Chaib-draa, 2004a].

nécessaires à l'intégration de notre approche à l'architecture BDI sont ensuite présentés et discutés (sections 7.4 à 7.9). Un exemple d'exécution du système résultant est détaillé (section 7.10) et le modèle proposé est synthétisé (section 7.11) et discuté (section 7.12).

7.2 Le modèle BDI [Beliefs, Desires and Intentions]

De nombreuses caractérisations du modèle BDI ont été proposées. On en trouve des formulations logiques, telles que celles produites par Rao et Georgeff [1991, 1995] (reprise dans le plus récent [Singh et al., 1999]) autant que des spécifications formelles [Inverno et al., 1998, 2004] ou procédurales [Wooldridge, 2001a, chapitre 4]. C'est ce type de vision procédurale, plus proche de l'implémentation, qui nous sera utile dans le cadre de notre validation informatique.

L'architecture BDI repose sur deux processus principaux : la délibération et le raisonnement fin-moyen (*means-end reasoning*). Pour un agent, la délibération consiste à déterminer ses intentions, à partir de ses croyances et de ses désirs, tandis que le raisonnement fin-moyen consiste en l'élaboration d'une suite d'actions, un plan, que l'agent va exécuter comme tentative pour réaliser ses intentions courantes. L'algorithme de contrôle de l'agent BDI effectue un compromis entre la délibération et le raisonnement fin moyen. En particulier, le fonction de reconsidération, centrale dans cette approche indique au vu des croyances et intentions actuelles de l'agent si celui-ci doit délibérer de nouveau (activité cognitive coûteuse) ou bien agir (via la planification et l'exécution de plan). L'algorithme BDICycle() (figure 7.1) présente le pseudo-code[2] de la procédure de contrôle d'un agent BDI classique.

À chaque cycle, l'agent BDI met à jour ses croyances en fonction de ses perceptions (ligne 6 et 7). Si c'est nécessaire (ce qu'indique la fonction booléenne Reconsider() de la ligne 8), l'agent (re)délibère pour mettre à jour ses désirs et ses intentions (ligne 9 et 10). Ensuite (ligne 12), l'agent (re)planifie une suite d'actions à réaliser si nécessaire, c'est-à-dire si le plan courant est vide ou que l'intention poursuivie est déjà réalisée ou encore qu'elle est devenue impossible ou que le plan courant n'est plus valide (ligne 11). Sinon (c'est-à-dire si toutes ces conditions sont fausses), l'agent exécute une action du plan courant (lignes 14-16). Notons que cette action peut elle-même être une recette complexe.

[2] Dans cet algorithme, comme dans les suivants, les variables et valeurs apparaissent en italique, les éléments du langage algorithmique sont en gras et les noms de procédures, fonctions et méthodes qui commencent par une majuscule sont en caractères normaux. Les algorithmes sont présentés dans une syntaxe la plus proche possible de l'implémentation JAVA qui en a été faite. La notation pointée, propre à la programmation orientée objet y est parfois utilisée. Comme pour tous code informatique développé dans un contexte international (la recherche scientifique en est un), les algorithmes sont présentés en anglais.

Procedure BDICycle(B_0,I_0)

1: **Inputs :** B_0, set of initial beliefs ;
   I_0, set of initially accepted intentions ;
2: **Outputs :** *none*, this is not a function !
3: **Local :** B := B_0, object that store the agent's beliefs ;
   I := I_0, object that store the agent's intentions ;
   D, objet that store the agent's desires ;
   List ρ, store both internal and external percepts ;
   List π := $null$, current plan, sequence of (possibly complex) actions ;
4: **Body :**
5: **while** *true* **do**
6: Get new percepts ρ ;
7: Update B on the basis of ρ ;
8: **if** Recondider(I,B) **then**
9: D := Options(B,I) ;
10: I := Deliberate(B,D,I) ; // deliberate if necessary
11: **end if**
12: **if** Empty(π) or Succeeded(I,B) or Impossible(I,B)) **then**
13: π := Plan(B,I) ; // replan if necessary
14: **else**
15: α := Head(π) ;
16: Execute(α) ; // execute an action
17: π := Tail(π) ;
18: **end if**
19: **end while**

FIG. 7.1 – Algorithme de contrôle d'un agent BDI (repris et adapté de [Wooldridge, 2001a, chapitre 4] et [Schut et Wooldridge, 2001]).

Cohen et Levesque [1990a], entre autres, ont insisté sur le fait que l'intention est un choix sur lequel l'agent s'engage individuellement. Ce type d'engagement individuel [Bratman, 1990; von Wright, 1980] ne devra pas être confondu avec les engagements sociaux. C'est la fonction Reconsider() (ligne 8) qui assure la persistance temporelle des intentions, rendant ainsi compte de l'engagement individuel qui leur est associé. Le mécanisme utilisé par l'agent pour déterminer quand et comment rejeter une intention préalablement acceptée est appelé *stratégie d'engagement individuelle*. On utilise ici la terminologie introduite par Rao et Georgeff [1991] et utilisée depuis par d'autres [Wooldridge, 2001a; Singh et al., 1999]. Il est classique de distinguer trois familles de stratégies d'engagement individuel :

- *l'engagement aveugle ou stratégie fanatique (blind commitment)* : une intention à laquelle une telle stratégie est associée sera poursuivie jusqu'à ce que l'agent croit que l'intention a été réalisée ;
- *l'engagement mixte (single-minded)* : une intention à laquelle une telle stratégie est associée sera poursuivie jusqu'à ce que l'agent croit que l'intention a été réalisée ou bien qu'il n'est plus possible de la réaliser ;
- *l'engagement ouvert (open-minded)* : une intention à laquelle une telle stratégie est associée sera poursuivie jusqu'à ce que l'agent croit qu'il n'est plus possible de réaliser cette intention.

Notons que ces trois familles de stratégies d'engagement individuel, introduites par Rao et Georgeff (ibid.) correspondent en fait à trois variantes de leur axiomatique logique, dont – de leur propre aveu – il est difficile de rendre compte en langage naturel. Nous verrons comment cette notion de stratégie d'engagement individuel est prise en compte dans notre approche à la section 7.8.

Avant de présenter notre approche de la pragmatique comme un nouveau module de traitement cognitif du modèle BDI, encore faut-il que celui-ci s'y prête. En l'occurrence, les modèles BDI classiques n'incluent pas le traitement des engagements sociaux. Or, utiliser un cadre conventionnel pour la communication agent, reposant sur les engagements sociaux, implique un changement de paradigme au niveau du raisonnement pratique des agents. En effet, dans un tel contexte : *un agent cognitif ne doit plus simplement raisonner sur ses intentions et celles qu'il a cru reconnaître des autres agents, mais il doit également raisonner sur les engagements potentiels ou déjà existants*. Ainsi, pour pouvoir appliquer notre approche à la communication entre agents BDI tout en utilisant le cadre DIAGAL comme langage de communication, il nous faut étudier les liens entre les cognitions privées de l'agent (telles que considérées dans le modèle BDI classique) et la notion d'engagement social (centrale dans notre modélisation des communications). De tels liens sont proposés dans la section suivante.

7.3 Lier les cognitions privées aux cognitions publiques

Puisque les agents raisonnent avec leurs états mentaux privés, mais à propos des engagements publics, il est nécessaire d'établir un lien entre ces types de cognitions. Cela nous semble être une condition nécessaire pour permettre à des agents cognitifs - tels que ceux reposant sur le modèle BDI - d'utiliser les cadres interactionnels conventionnels tels que DIAGAL (décrit au chapitre 5).

Quand on dit d'un agent cognitif de type BDI qu'il est en interaction avec son environnement, cela signifie que la cognition de l'agent est couplée à son environnement extérieur par ses perceptions et ses comportements. Dans le modèle BDI, les comportements d'un agent résultent de ses intentions. En effet, ce sont les intentions (voir section 2.2.1), issues de la délibération, qui par l'entremise du comportement manifeste, font le « lien » avec le milieu extérieur, c'est-à-dire son environnement physique (éventuellement logiciel) et social. Pour établir le lien entre les cognitions publiques et les cognitions privées, on ne considère, parmi les cognitions privées des agents, que les intentions.

Parmi les intentions personnelles, nous distinguons les intentions sociales et les intentions individuelles. Les *intentions sociales* sont des intentions personnelles qui requièrent le concours d'autres agents pour être accomplies. Plus généralement, toute intention incluse dans une activité collective, distribuée, nécessitant la coordination des agents sera considérée comme une intention sociale[3]. Ces intentions sociales concernent donc une activité (même indirectement) collective, c'est-à-dire requérant l'action, la permission ou l'opinion d'autres agents. Ces intentions sociales sont très répandues dans les environnements multi-agents et interviennent dans la plupart des activités distribuées, qu'elles soient de résolution de problèmes ou de coordination. Dans le cas d'une délégation, par exemple, l'agent A peut avoir l'intention sociale que l'agent B exécute une action α donnée. Ces intentions sociales, on l'aura compris, sont celles qui sont potentiellement impliquées dans les incohérences externes.

Parmi les intentions individuelles, nous appelons *intentions échouées*, les intentions individuelles pour lesquelles l'agent n'a aucun plan ou encore pour lesquelles tous les plans individuels disponibles ont échoués. Dans le premier cas, cela signifie que le raisonnement fin-moyen a échoué, tandis que dans le second, c'est la compétence de l'agent qui est en cause (manque de ressources, manque de savoir-faire ou simple état de fait). Ce type d'intention échouée rend compte des intentions impliquées dans une incohérence interne que l'agent ne parvient pas à résoudre seul.

[3] Nous ne considérons pas les cas particuliers comme les intentions faisant partie d'un plan collectif déjà socialement accepté, ...

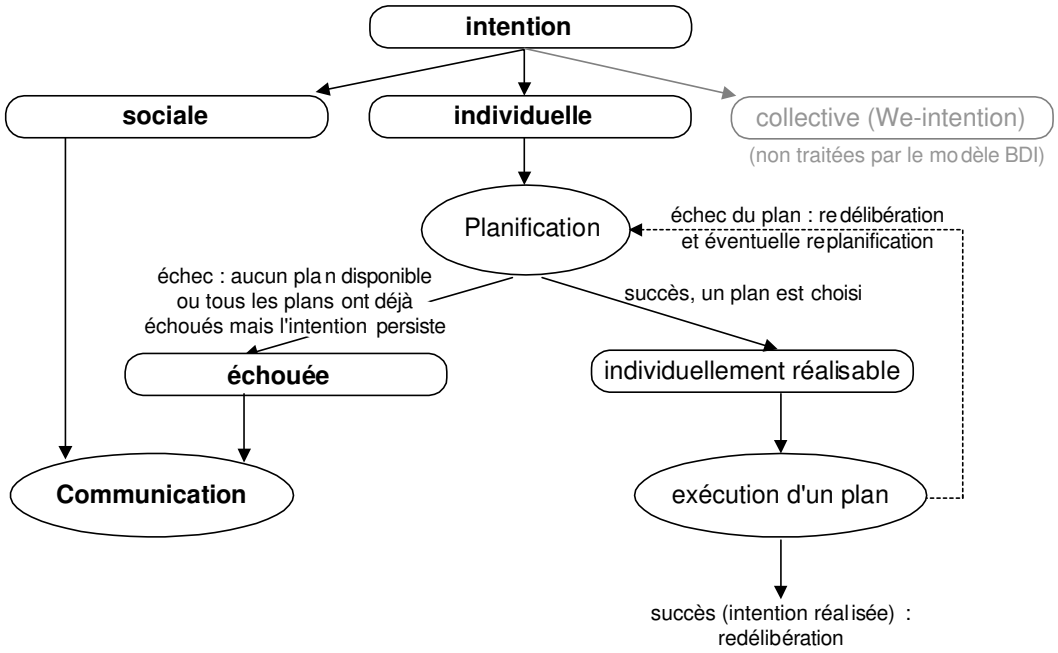

FIG. 7.2 – Typologie des intentions.

Cette distinction entre les intentions impliquant de fait une dimension sociale et les autres (les intentions individuelles qui ne sont ni sociales, ni échouées) nous semble cruciale pour l'intégration des approches conventionnelles et sociales de la communication avec les théories d'agent existantes. La figure 7.2 illustre cette typologie. Nous indiquerons section 7.9 comment ce filtrage est effectué de manière algorithmique.

Parmi ces intentions sociales ou échouées, nous faisons encore la distinction classique entre *intention de* (faire quelque chose ou faire faire quelque chose) et *intention que* (une certaine proposition tienne, c'est-à-dire, soit vraie ou considérée comme vraie) [Bratman, 1990], selon que l'intention porte sur une action ou un contenu propositionnel.

Les cognitions sociales/publiques considérées sont les engagements sociaux et on distingue les *engagements propositionnels* des *engagements en action* (voir section 2.3.3) [Bell, 1995; Walton et Krabbe, 1995]. Comme tous les autres éléments, dans notre modèle, les engagements sociaux peuvent être acceptés ou rejetés (on utilise notre modèle de l'engagement social flexible, introduit en section 4.5.1).

Dans ce contexte, nous pouvons retourner à la question générale : quels sont les liens entre les engagements sociaux et les états mentaux des agents ? Comme réponse, nous proposons les liens suivants. On considère qu'idéalement *un engagement en action est le pendant so-*

cialement accepté d'une « intention de » et qu'un engagement propositionnel est le pendant socialement accepté d'une « intention que ».

Dans notre approche, ces liens sont capturés par un ensemble de contraintes positives ou négatives liant les intentions et les engagements possibles. Des contraintes positives reflètent ces relations de correspondances tandis que des contraintes négatives rendent compte des éventuelles incompatibilités. Du point de vue de la représentation des connaissances et la formalisation du raisonnement, ces contraintes indiquent des liens idéaux, mais elles ne sont pas nécessairement satisfaites, elles sont à satisfaire.

Par exemple, un agent A qui a l'intention (sociale) que l'agent B accomplisse l'action $TournerGauche$ devra réussir à engager B sur cette action. Ainsi une correspondance, capturée par une contrainte positive existe entre l'intention de A que B effectue l'action (intention que l'on notera $I_A(TournerGauche_B)$) et l'engagement de B envers A sur cette réalisation (noté $C(B, A, TournerGauche_B, t, s_B, s_A)$). Pour que cette contrainte soit satisfaite, ces deux éléments devront être soit tous les deux acceptés soit tous les deux rejetés. Des contraintes négatives sont également inférées des relations d'incompatibilités qui peuvent exister entre intentions et engagements. Par exemple, dans la situation décrite précédemment, des incompatibilités existent entre l'intention de A, $I_A(TournerGauche_B)$ et des engagements tels que $C(B, A, TournerDroite_B, s_B, s_A)$ ou $C(B, A, AllerToutDroit_B, s_B, s_A)$[4].

Si, pour l'heure, cette liaison entre cognitions privées et cognitions publiques n'est proposée dans aucun des cadres interactionnels conventionnels agents rencontrés (approches basées sur les engagements ou jeux de dialogue), c'est sans doute qu'elle relève plus de notre problématique, c'est-à-dire de l'utilisation des cadres interactionnels plutôt que de leur définition. Pour autant, ces relations entre cognitions publiques et privées ne sont pas complètement neuves, puisque nous avons vu section 7.2 qu'il est classique d'associer l'intention individuelle à un type particulier d'engagement individuel. Nos liens étendent cela au niveau social dans les cas appropriés, c'est-à-dire pour les intentions sociales ou échouées, telles que définies ci-haut.

Par ailleurs, des liens similaires ont été introduits dans les modèles d'agents cognitifs de manière générale, sans références spécifiques à la communication. En effet, le besoin d'intégrer le traitement des conventions sociales, des normes, des obligations et dans une certaine mesure des engagements sociaux dans les architectures d'agents délibératifs se fait sentir. Dans la littérature multi-agent récente, on parle d'architecture hybride *normative-*

[4] On ne donne pas ici d'algorithme pour ce qui est de l'inférence de ces contraintes, car elles reflètent la structure logique du domaine considéré qui est capturée par le langage de contenu et son ontologie qui n'ont pas été fixés dans notre cas.

délibérative [Castelfranchi et al., 1999; Broersen et al., 2001; Boella et Lesmo, 2001][5] pour rendre compte de ces tentatives. Parmi les travaux qui traitent explicitement d'engagements sociaux dans ce domaine, on note l'effort de Royakkers et Dignum [2000] qui ont proposés, dans la lignée de Castelfranchi [2004, 1995] les axiomes logiques suivants :

$$S\text{-}COMM(i,j,\tau_i) \rightarrow I_j(\tau_i), [6] et \tag{7.1}$$

$$S\text{-}COMM(i,j,\tau_i) \rightarrow I_i(\tau_i) \tag{7.2}$$

Ces axiomes qui proposent une autre modélisation des liens entre cognitions privées et publiques nous semblent d'une rigidité excessive. Le premier (7.1) indique que lorsque i est engagé envers j à réaliser l'action τ, j a l'intention que i réalise cette action, tandis que le second (7.2) indique que i a également cette intention. On peut les lire sous la forme développée, c'est-à-dire sous la forme du théorème suivant (la démonstration est triviale) : $\vdash \neg S\text{-}COMM(i,j,\tau_i) \vee (I_i(\tau_i) \wedge I_j(\tau_i))$. Ce dernier indique que dans tous les cas, soit l'agent i n'est pas engagé envers l'agent j à réaliser τ, soit i et j ont tous les deux l'intention que i réalise τ. L'association via des contraintes plutôt que ces liens logiques trouvés dans la littérature est plus souple. En effet, tandis que (et c'est leur définition) les axiomes ne peuvent être mis à mal, les contraintes peuvent être satisfaites ou non et tous les cas sont envisageables (et utiles à envisager).

En particulier, les axiomes précédemment évoqués sont incompatibles avec l'objectif de flexibilité sémantique qui sous-tend notre modèle de l'engagement social. Par exemple, nous pensons qu'il est utile de considérer les cas ou un agent viole un engagement volontairement, ce qui signifie qu'il n'a pas l'intention prétendue. Aussi, si l'agent i viole volontairement l'engagement $S\text{-}COMM(i,j,\tau_i)$, c'est bien qu'il n'a pas accepté l'intention correspondante $I_i(\tau_i)$. Dans ce cas, le second des axiomes précédemment cités (7.2) est contredit, ce qui invalide le système logique proposé. Autre exemple, si l'agent j souhaite annuler l'engagement sus-mentionné, c'est bien qu'il n'accepte pas l'intention sociale $I_j(\tau_i))$ et dans ce cas, c'est le premier des axiomes cités précédemment (7.1) qui est contredit. Ainsi, ces deux axiomes supposent plus que la sincérité des agents, ils supposent une très forte normativité, au détriment de la flexibilité. Castelfranchi [1997] va d'ailleurs plus loin en affirmant que lorsque

[5] Dont il faut bien reconnaître à leur décharge qu'elles cherchent plutôt à intégrer les normes et les obligations que les engagements sociaux tels qu'ils apparaissent dans les cadres de communications récents.

[6] Qui est parfois formulé : $S\text{-}COMM(i,j,\tau_i) \rightarrow Goal_j(Does_i(\tau))$

$C(x, y, \tau)$ est accepté « x and y mutually know that x intend to do τ and that's y goal ». Rien n'est dit sur la méthode par laquelle une telle connaissance[7] commune peut être établie.

Il existe d'autres contre-exemples, à cette famille d'axiomes par trop réducteurs. Ainsi, si nous souscrivons aux analyses et approches de la communauté telles qu'introduites dans les différentes architectures délibératives-normatives, nous mettons en cause la rigidité des termes logiques utilisés et insistons sur la prévalence de la souplesse offerte par les contraintes qui nous permettent une modélisation plus réaliste et plus complète. En particulier, les axiomes ci-haut sont orientés et tentent de réduire les engagements à des attitudes privées de manière *unidirectionnelle* : ils indiquent que l'engagement implique l'intention, mais l'intention n'implique pas l'engagement (ce sans quoi on aurait équivalence). Les contraintes que nous utilisons sont symétriques, indiquant que le lien est *bidirectionnel*[8], ce dont rend compte la possibilité du changement d'attitude que nous allons redéfinir section 7.5. Mais avant tout, introduisons le cadre interactionnel retenu pour notre validation informatique.

7.4 Cadre interactionnel utilisé

Pour notre validation, on suppose que dans le système considéré, le langage DIAGAL (tel que défini en section 5.2) est utilisé comme cadre interactionnel. Par simplification, nous n'utiliserons pas le jeu de modification. Mais pour éviter la monotonie associée à la variante déontique stricte du langage DIAGAL (introduite section 5.4.5), nous autorisons le désengagement unilatéral, c'est-à-dire l'annulation. Ainsi, dans cette variante, quatre jeux de dialogue peuvent être utilisés par les agents dans l'espoir d'ajouter un engagement dans la couche publique : $offer$, $request$, $inform$ et ask. Deux jeux d'annulation et deux jeux de décharge viennent compléter la panoplie des unités dialogiques disponibles aux agents[9].

Dans ce cadre, les engagements ne sont pas modifiables et impliquent des sanctions s'ils ne sont pas respectés. L'annulation, c'est-à-dire le désengagement unilatéral est incluse dans ce non-respect. Le modèle d'engagement utilisé est donc celui présenté au chapitre 5, dans lequel un engagement accepté[10] de x envers y pris au temps t sur la proposition p (ou l'action α) entraîne les sanctions s_x s'il est violé ou annulé par x, et s_y s'il est annulé par y. Un

[7] La connaissance est plus forte que la croyance et est généralement définie comme la croyance vraie, ce qui est capturé par l'axiome suivant : $Know_i(p) = p \wedge Bel_i(p)$ (tiré de [Cohen et Levesque, 1995]).

[8] Thagard [2000a] développe un argumentaire pour justifier la bidirectionnalité des contraintes cognitives dans les approches cohérentistes.

[9] Concrètement, seulement ces jeux seront chargés dans l'interface du DGS par l'utilisateur.

[10] C'est-à-dire qui appartient à l'ensemble des éléments acceptés de tous les agents ayant participé à son établissement. Concrètement, l'engagement a été noté comme accepté (actif, violé ou satisfait) dans l'agenda de chacun des agents par leurs gestionnaires de dialogue respectifs, selon la systématique décrite au chapitre 5.

tel engagement est noté : $C(x, y, p, t, s_x, s_y)$. Dans un premier temps, par souci de simplification, nous ne discuterons pas les sanctions (c'est-à-dire que nous supposerons toutes les sanctions et récompenses comme étant nulles). Par contre, puisque les éléments nécessaires au traitement des sanctions ont été introduits dans notre modélisation de l'engagement social (chapitre 4) ainsi que dans l'implantation du langage DIAGAL (chapitre 5), nous reviendrons sur le traitement des sanctions en section 8.2.4.

7.5 Formulation BDI du changement d'attitude

Dans notre modèle (présenté au chapitre 6), chaque agent cherche à maximiser sa cohérence cognitive, c'est-à-dire essaie de réduire ses incohérences en commençant par la plus intense. Pour réduire l'incohérence, l'agent doit accepter ou rejeter des éléments pour mieux satisfaire les contraintes qui les lient. Ces éléments peuvent être les représentants de cognitions privées ou publiques. Nous avons vu qu'en vertu de leur nature et de leur résistance au changement, toutes les cognitions ne sont pas également modifiables. En particulier, les engagements sociaux *doivent être socialement établis* et dans le cas qui nous concerne, ce sont les jeux de dialogue qui permettent de manipuler la couche sociale. Ainsi, pour changer l'état d'acceptation d'un engagement social, les agents doivent dialoguer. Le dialogue est le seul médium par lequel les agents peuvent essayer d'obtenir des engagements sociaux cohérents avec leurs cognitions privées. Cependant, certains engagements incohérents avec leurs cognitions privées peuvent également résulter de ces dialogues. Plus précisément, cela signifie qu'un engagement peut être accepté sans que l'intention correspondante le soit ou encore un engagement peut être rejeté tandis que l'intention correspondante est acceptée. Dans ces deux cas, la contrainte qui lie l'intention et l'engagement n'est pas satisfaite.

Conformément à notre modèle de l'engagement social, l'annulation ou la violation d'un engagement accepté implique des sanctions. En outre, on considère qu'un engagement qui est rejeté après un dialogue visant à le faire accepter gagne en résistance au changement (cela dépend cependant de la stratégie d'engagement individuelle retenue par l'agent, tel que discuté section 7.8). Cela interdit aux agents de faire indéfiniment des tentatives pour faire accepter les modifications qu'ils souhaitent (à l'exception des agents fanatiques qui ne tiennent pas compte des échecs de leurs tentatives en n'augmentant pas la résistance au changement des engagements poursuivis, comme nous le verrons section 7.8). Aussi, cette résistance au changement des engagements rejetés ainsi que les sanctions associées aux engagements acceptés rendent coûteuses les tentatives de modifications ultérieures. Si les engagements sont plus flexibles que de simples obligations sociales, l'idée n'en reste pas moins que l'un des pôles des contraintes afférentes à un tel engagement est « fixé » ou en tout cas plus difficile à modifier. Dès lors, pour réduire une éventuelle incohérence impliquant cet engagement

comme élément, les agents auront tendance à modifier leurs cognitions privées, c'est-à-dire leurs intentions. C'est le ressort du changement d'attitude tel qu'introduit en section 6.5.

Dans le cadre BDI qui nous intéresse ici, les seules cognitions privées impliquées dans notre calcul de cohérence sont les intentions des agents, mais l'on suppose que les éventuels changements d'attitudes seront répercutés sur les autres cognitions via les fonctions de de mise à jour des croyances et des désirs du système BDI. Celles-ci pourraient prendre en compte le changement d'attitude dans la mise à jour des différentes cognitions privées. Cette possibilité n'a pas été implantée pour l'heure et nous verrons comment différents types de changement d'attitude peuvent être envisagés en section 8.2. Nous nous sommes contentés ici de présenter le changement d'attitude comme résultant en l'adoption ou l'abandon d'une intention. Un exemple de ce type de changement d'attitude dans notre cadre de validation est fourni à la section 7.10.

La figure 7.3 schématise (quelque peu grossièrement) les liens entre intentions, engagements et jeux de dialogue que nous proposons. De la gauche à la droite, nous avons les intentions échouées et les intentions sociales, classées selon les deux catégories introduites section 7.3, qui sont liées aux quatre types d'engagements susceptibles de leur correspondre. Notons que tant que ces engagements n'ont pas été discutés, ils sont de simples possibilités envisagées et générées par l'agent dans son raisonnement[11], tandis que tous les engagements non explicitement acceptés sont rejetés (notre hypothèse du monde clos pour les engagements). Afin d'assurer la cohérence avec ses intentions, l'agent va, par l'entremise de son calcul de cohérence (introduit au chapitre 6 et présenté plus en détail dans le cadre de cette validation à la section suivante), faire un certain nombre de tentatives pour modifier les engagements correspondants afin de mieux satisfaire les contraintes qui lient les cognitions privées aux cognitions publiques. Pour effectuer ces tentatives, l'agent n'aura qu'à sélectionner le jeu DIAGAL (seuls les jeux de création d'engagement sont représentés sur la droite de la figure 7.3) dont les conditions d'entrée s'unifient (au sens de la logique mathématique) avec la situation initiale et la condition de succès s'unifie avec l'état désiré[12]. Selon l'issue des dialogues, le chemin inverse (de droite à gauche) pourra être emprunté et le changement d'attitude éventuellement initié.

[11] L'état réel des engagements est conservé dans l'agenda de l'agent tenu à jour par son gestionnaire de dialogue.

[12] Le tableau 5.2 présenté à la section 5.4.4 peut être utilisé à cet égard.

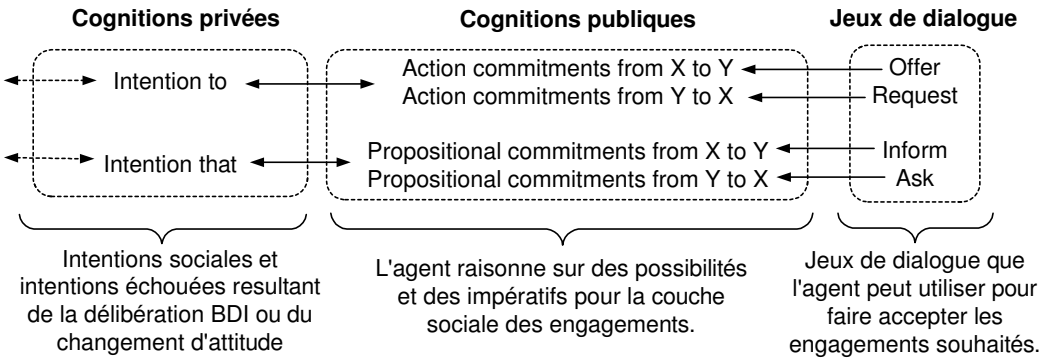

FIG. 7.3 – Schématisation simplifiée des liens entre les cognitions privées, publiques et les jeux de dialogue DIAGAL.

7.6 La fonction d'utilité espérée

7.6.1 Résolution du problème de maximisation de la cohérence

Comme nous l'avons vu à la section 6.4, la cohérence cognitive d'un ensemble d'éléments de cognition est exprimée comme étant la somme des poids des contraintes satisfaites divisée par le nombre de contraintes considérées au sein de cet ensemble. Le problème de cohérence se formule comme suit : étant donné un ensemble d'éléments et les contraintes positives et négatives qui les lient, il s'agit de trouver une partition des éléments entre l'ensemble des éléments acceptés A et l'ensemble des éléments rejetés R qui maximise la cohérence. Notons qu'il ne s'agit pas d'un problème de satisfaction de contraintes classique (CSP[Constraint Solving Problem]) mais d'un problème d'optimisation sous contraintes puisqu'on ne peut présupposer qu'une solution complète (au sens de la satisfaction de contraintes, c'est-à-dire dans laquelle toutes les contraintes sont satisfaites) existe[13]. Thagard et Verbeurgt [1998] ont isolé cinq algorithmes de résolution de ce problème de maximisation de la cohérence :

1. *un algorithme exhaustif* qui considère toutes les solutions possibles pour retenir la meilleure ;
2. *un algorithme incrémental* qui construit une solution en considérant les éléments dans un ordre aléatoire ;

[13] On pourra se reporter à [Alliot et Schiex, 1993] ou [Kumar, 1992] pour une introduction aux techniques algorithmiques de satisfaction de contraintes.

3. *un algorithme connexioniste* qui utilise un algorithme de propagation de l'activation (similaire à ceux développés pour les réseaux de neurones artificiels) pour converger vers une solution ;

4. *un algorithme glouton* qui utilise un choix localement optimal pour approximer une solution globalement optimale ;

5. *un algorithme de programmation semi-définie* qui garanti qu'une forte proportion des contraintes satisfaites dans la solution optimale le sont également après son exécution.

La première solution offre un algorithme de résolution exact qui renvoie toujours la solution optimale. Cependant, pour un réseau de n éléments, il y a 2^n solutions possibles à considérer. La complexité de l'algorithme exhaustif explose donc combinatoirement en fonction de la taille des données et celui-ci n'est donc applicable que dans des cas triviaux. Notons au passage que le problème général de la maximisation de la cohérence est NP-complet, comme l'ont formellement montré Thagard et Verbeurgt [1998]. Le second algorithme n'est pas bon du point de vue de la qualité de la solution renvoyée puisque celle-ci dépend essentiellement de l'ordre dans lequel les éléments sont considérés lors de la construction incrémentale de la solution. Cet algorithme a été introduit par la communauté de modélisation cognitive pour discuter certains aspects de la rationalité limitée humaine et mettant en avant la complexité des traitements non monotones qui sont nécessaires pour la « rationalité parfaite ». Les trois derniers algorithmes proposés sont satisfaisants, et ce, aussi bien en ce qui concerne la qualité de la solution retournée que du point de vue de leur complexité informatique.

Dans notre cas, une restriction inhérente à notre approche nous aide à trancher entre ces trois solutions. En effet, puisque chercher une solution consiste à accepter et rejeter des éléments pour mieux satisfaire les contraintes et donc augmenter la cohérence du réseau, les éléments doivent être modifiables indifféremment. Or, et nous avons déjà insisté sur ce point à plusieurs reprises, dans notre modèle appliqué à la communication, tous les éléments ne sont pas modifiables individuellement. En particulier l'agent ne peut modifier seul l'état d'acceptation d'un engagement social. Aussi l'approche connexionniste qui met à jour de manière parallèle les seuils d'activations[14] des éléments indifféremment de leur nature est difficilement applicable dans notre cas. Il en va de même avec l'approche par programmation semi-définie. Ainsi, dans notre approche, le calcul de cohérence se fait via un algorithme glouton modifié que nous nommons algorithme de recherche locale et qui est présenté dans la sous-section suivante.

[14] Dans cette approche, l'état d'acceptation (discret) d'un élément est déduit de son seuil d'acceptation (continu) qui peut être positif (élément accepté) ou négatif (élément rejeté).

7.6.2 Algorithme de recherche locale

Ainsi, à partir de l'état initial, à chaque étape de son raisonnement pragmatique sur le réseau formé par les différentes intentions et engagements considérés, l'agent va chercher le changement d'état d'acceptation d'un élément qui maximise le gain de cohérence, en prenant en compte la résistance au changement de cet élément. Techniquement, c'est ce que l'on appelle, dans le jargon de la satisfaction de contraintes, un changement « 1-optimal ». Si cet élément s'avère être un engagement, l'agent accomplira sa tentative de changement par le dialogue tandis que s'il s'agit d'une intention, le changement sera consommé directement et l'algorithme poursuivra son exécution après avoir mis à jour la résistance au changement de l'élément modifié.

Dans notre implantation, l'agent détermine qu'elle est l'élément dont le changement d'acceptation maximise le gain de cohérence en explorant les n états accessibles à partir de l'état courant en modifiant l'état d'acceptation d'un des n éléments du réseau considéré. Comme, à nature égale, toutes les cognitions ne sont pas identiquement modifiables, nous avons introduit la notion de coût pour prendre en compte les résistances au changement et les éventuelles sanctions. Tous les états accessibles sont donc explorés et évalués à l'aide de la fonction d'utilité espérée[15] suivante :

$$\begin{aligned} g(ExploredState) \;=\; & coherence(ExploredState) - coherence(CurrentState) \\ & - cost(CognitionChanged) \end{aligned}$$

où $coherence()$ calcule la cohérence cognitive d'un état, $CognitionChanged$ est la cognition dont on considère le changement d'état d'acceptation et $ExploredState$ est l'état évalué, c'est-à-dire l'état courant dans lequel le changement d'état considéré est réussi. Enfin, $cost$ est une fonction de coût calculée comme suit :

– Si $CognitionChanged$ est une intention, son coût de modification est sa résistance au changement (telle qu'établie par l'architecture sous-jacente ou telle que mise à jour lors de la dernière modification d'état de celle-ci) ;

[15] Le qualificatif « espéré » renvoie ici à : dans le cas où le dialogue, comme tentative de modification de la couche sociale, serait un succès et non à l'espérance probabiliste habituellement associée aux mesures d'utilité dans la théorie de la décision classique. À cet égard, les probabilités n'interviennent pas dans notre formalisation car nous nous situons sous une hypothèse d'équiprobabilité des alternatives qui représente bien l'incertitude des agents. L'introduction des probabilités et de leur apprentissage dans notre formalisme est envisagée comme l'une de nos perspectives, présentée section 8.4.3.

– Si $CognitionChanged$ est un engagement rejeté, son coût de modification est sa résistance au changement subjective (qui est généralement initialement faible mais qui est susceptible d'augmenter à chaque tentative de changement d'état infructueuse) ;

– Si $CognitionChanged$ est un engagement accepté, son coût de modification correspond à sa résistance au changement à laquelle s'additionnent les éventuelles sanctions attachées à son annulation ou à sa violation ainsi que les récompenses attachées à son respect (dans le cas du créditeur les récompenses sont soustraites des sanctions puisqu'il en aurait été le débiteur).

Le changement « 1-optimal » recherché par cette procédure n'existe pas toujours. Il se peut qu'aucune des modifications envisageables n'améliore la cohérence. Pour autant, si la cohérence n'est pas maximale, l'agent sait qu'il est peut-être dans un état sub-optimal donc un maximum local. Dans ce cas, il est utile de développer un second niveau de l'arbre de recherche, en particulier à partir des états cognitifs internes (les intentions dans notre cas) qui sont individuellement modifiables. Une heuristique de moins pire d'abord est utilisée pour ce faire tandis qu'une borne sur la profondeur de la recherche limite cette exploration supplémentaire à quelques pas. À chaque pas, cet algorithme développe donc un niveau de l'arbre de recherche, le meilleur, même si celui-ci est d'utilité négative tant que la cohérence n'est pas optimale et que la borne de profondeur n'a pas été atteinte. Si des branches d'utilité négative peuvent être ainsi explorées, une branche ne sera entérinée que si son utilité totale est positive.

Techniquement, l'algorithme de recherche locale est un algorithme de recherche de type profondeur d'abord. C'est une procédure de recherche informée utilisant une stratégie gloutonne avec une heuristique de moins pire d'abord[16]. C'est une approche itérative avec une contrainte sur le type des éléments modifiés (seules les intentions sont individuellement modifiables). Si ce type d'algorithme a une complexité d'ordre $O(b^p)$, où b est le facteur de branchement et p la profondeur de recherche maximale, l'heuristique et la contrainte évoquée ci-dessus réduisent considérablement cette complexité, puisque le processus dialogique ainsi que d'autres tâches cognitives ou comportementales de l'agent viennent interrompre le processus de satisfaction de contraintes. En outre, le nombre d'éléments considérés dans les simulations réalisées pour l'heure est restreint à quelques dizaines dans les cas les plus complexes.

Nous ne fournissons pas de preuve de convergence de cette procédure de recherche pour la maximisation de la cohérence, mais comme Thagard et Verbeurgt [1998] pour la leur, elle nous a semblé optimale sur les exemples testés. L'algorithme de recherche local et la

[16] Le lecteur pourra au besoin consulter l'ouvrage de Russell et Norvig [2003, chapitres 3 et 4] qui clarifie la terminologie utilisée ici.

Procedure CommunicationPragmatics(*initiate*)

```
1:  Inputs : initiate, boolean variable set to true when the underlying
                 BDI architecture call the Procedure, set to false otherwise
2:  Outputs : none, this is not a function !
3:  Global : agenda, objet that store the agent's agenda
4:  Local :
5:  List commitments := agenda.GetCommitments() ;
6:  List dialogCommitments := agenda.GetDialogCommitments() ;
7:  Body :
8:  TreatCommitments(commitments) ;
9:  if dialogCommitments.IsEmpty() and initiate=true then
10:    initiate := false ;
11:    InitiateDialogue() ; // initiate a dialogue
12: else
13:    if dialogCommitments.IsEmpty() and initiate=false then
14:       ModifiedBDICycle() ; // dialog finished
15:    else
16:       TreatDialogCommitments(dialogCommitments) ; // pursue a dialogue
17:    end if
18: end if
```

FIG. 7.4 – Algorithme de traitement pragmatique.

fonction d'utilité espérée sont exemplifiés en section 7.10. Cet algorithme est appelé par les procédures de traitement pragmatique décrites dans la section suivante (par l'intermédiaire des procédures d'initiative du dialogue, InitiateDialogue() et de traitement des engagements dialogiques TreatDialogCommitments()).

7.7 L'algorithme de traitement pragmatique

Le comportement de nos agents est guidé par leur cohérence cognitive, celle-ci impliquant autant leurs cognitions privées que les engagements sociaux. Ces derniers sont inscrits avec leur état courant dans l'agenda de l'agent par son gestionnaire de dialogue. Lorsqu'un agent doit déterminer quelle action dialogique entreprendre, il exécute l'algorithme de traitement pragmatique présenté figure 7.4.

Comme nous l'avons vu en section 5.2.1, nous distinguons les engagements extra-dialogiques (assignés à un objet de la classe List en ligne 5) des engagements dialogiques (assignés à un autre en ligne 6). Les engagements dialogiques regroupent les engagements issus des règles des jeux de dialogue autant que de celles du jeu de contextualisation. La procédure pour traiter les engagements extra-dialogiques TreatCommitments(*commitments*) (ligne 8) consiste à mettre à jour le modèle cognitif de l'agent en parcourant les modifications de l'agenda réalisées par le gestionnaire de dialogue. Ces modifications résultent autant des actions dialogiques qu'extra-dialogiques des agents (dans le cas de la satisfaction ou de la violation d'engagements).

Ensuite, l'algorithme distingue trois cas :

1. *l'initiative d'un dialogue* : il n'y a pas d'engagement dialogique actif dans l'agenda et la variable booléenne *initiate* est à vrai (test, ligne 9), signifiant que l'architecture BDI sous-jacente vient d'appeler la procédure de traitement pragmatique. La variable *initiate* est mise à faux et la procédure pour initier un nouveau dialogue est appelée (InitiateDialogue(), ligne 11) ;

2. *la terminaison d'un dialogue* : il n'y a plus d'engagements dialogiques actifs dans l'agenda et la variable booléenne *initiate* est à faux (teste, ligne 13), signifiant que le segment de dialogue est terminé. L'architecture BDI sous-jacente est appelée de nouveau (ModifiedBDICycle(), ligne 14) ;

3. *la continuation d'un dialogue* : il y a des engagements dialogiques à traiter, la procédure TreatDialogCommitment(*dialogCommitments*) (ligne 16) est appelée.

La fonction pour initier un dialogue, InitiateDialogue() (ligne 11) appelle une sous procédure qui génère le réseau de contraintes qui lient les intentions considérées et les engagements correspondants[17]. Ensuite, toujours dans cette procédure pour initier un dialogue, la procédure de recherche locale (présentée à la section 7.6.2) est appelée et effectue des modifications d'éléments (afin d'accroître la cohérence cognitive) dans le modèle cognitif maintenu par l'agent pour son raisonnement pragmatique[18]. La procédure de recherche locale est appelée jusqu'à ce qu'un engagement soit rencontré et que l'agent initie un dialogue pour consommer la transition désirée[19]. Le jeu DIAGAL à utiliser pour ce faire dépend sim-

[17] On renvoie le lecteur intéressé au rapport d'Andrillon [2003], rédigé sous notre direction, pour le détail de cet algorithme.

[18] Rappelons que ce modèle cognitif est le réseau d'éléments acceptés ou rejetés et de contraintes formé par les intentions sociales et échouées, les engagements extra-dialogiques et les contraintes positives et négatives qui les lient selon les principes de représentation des connaissances indiqués dans les sections précédentes.

[19] Notons, qu'il se peut que la procédure de recherche locale ne renvoie aucun élément, notamment si la cohérence est déjà maximale.

plement de l'état actuel de l'engagement et de l'état souhaité, tandis que les différents champs de l'engagement à modifier indiquent le partenaire et le sujet du dialogue.

La procédure de poursuite du dialogue, TreatDialogCommitments(*dialogCommitments*) (ligne 16), consiste à effectuer le traitement des engagements dialogiques pendants en évaluant l'utilité de toutes les actions possibles, permises par les règles du jeu, et choisir celle qui a les meilleures conséquences sur la cohérence cognitive à l'aide de la fonction d'utilité introduite section 7.6. Ce choix est mis en regard de celui indiqué par l'algorithme de recherche locale. Si la modification sur l'engagement extra-dialogique proposée par le jeu courant n'est pas celle indiquée par la fonction de recherche locale, l'agent va proposer un sous-dialogue subjectivement plus utile en proposant l'imbrication d'un autre jeu de dialogue approprié.

Enfin, dans le cas où un segment de dialogue se termine, c'est la procédure ModifiedBDI-Cycle() (ligne 14) qui est appelée, signifiant que l'agent va redonner la main à l'architecture délibérative sous-jacente. Celle-ci va procéder aux propagations des éventuels changements d'attitudes de l'agent, délibérer et fournir d'éventuelles nouvelles intentions qui seront filtrées pour différencier : (1) les intentions individuelles classiques et agir en conséquence (2) les intentions sociales ou échouées, pour la satisfaction desquelles le dialogue est souvent nécessaire (ce n'est pas systématique puisque la cohérence cognitive peut être déjà maximale). Dans ce dernier cas, la procédure de traitement pragmatique décrite ci-dessus sera appelée de nouveau.

Finalement, l'algorithme CommunicationPragmatics(*initiate*) (figure 7.4) est appelé :

– chaque fois que l'architecture BDI sous-jacente termine une délibération (ou une re-délibération, après un appel de notre algorithme pour un changement d'attitude) et que les intentions produites sont soit des intentions sociales soit des intentions échouées, signifiant que l'agent ne peut les satisfaire de lui-même et doit communiquer ;

– lorsque que le gestionnaire de dialogue modifie l'agenda et que cette modification n'est pas une violation ni une satisfaction d'engagement extra-dialogique. Cela assure (1) que l'agent ré-exécute l'algorithme tant que tous les dialogues en cours ne sont pas terminés et (2) que l'agent traite les dialogues initiés par d'autres. Par exemple, lorsqu'un agent reçoit un $prop.in$ pour entrer dans un jeu DIAGAL particulier, le gestionnaire de dialogue ajoute l'engagement dialogique correspondant dans l'agenda conformément au jeu de contextualisation courant. C'est au sein de la procédure TreatDialogCommitments() que sera décidé si l'agent accepte ou refuse d'entrer dans le jeu proposé.

Dans le second cas, il est à noter que la gestion des tours de parole est assurée par les règles des jeux et répercutée via le gestionnaire de dialogue qui indique lorsqu'un engagement dialogique est satisfait (signe que l'interlocuteur termine son tour de parole) et cela appelle la procédure de traitement pragmatique. La figure 7.5 présente l'algorithme de contrôle BDI modifié, ModifiedBDICycle().

Notons que si l'imbrication de ces deux algorithmes (ModifiedBDICycle() et CommunicationPragmatics()) implique un traitement séquentiel, leur grain de traitement, relativement fin, assure un certain « parallélisme » qui permet à l'agent de tenir en parallèle certaines activités dialogiques et l'exécution pas à pas de ses plans individuels.

7.8 Résistance au changement et stratégie d'engagement individuel

Dans la théorie BDI classique, les intentions sont associées à des engagements individuels (également appelés engagements psychologiques) visant à en assurer la persistance. Les différentes méthodes de gestion de ces engagements individuels appelées *stratégies d'engagement individuel* tiennent pour les mécanismes par lesquels un agent détermine quand rejeter une intention préalablement acceptée (tel qu'introduit section 7.2). Notre modèle, grâce à l'introduction du changement d'attitude étend cela en indiquant également quand adopter des intentions préalablement rejetées sans passer par la délibération du système sous-jacent. Le changement d'attitude permet donc d'étendre la notion de reconsidération d'intention de manière significative et de capturer des aspects normatifs d'adaptation sociale, jusqu'ici ignorés dans les modèles d'agents cognitifs.

Nous avons vu section 7.2 les différentes stratégies d'engagement disponibles dans la caractérisation logique du cadre BDI classique. Ces stratégies ont cours dans le modèle sous-jacent et notre modèle permet de les étendre grace à la notion de résistance au changement. À chaque modification d'un élément dans le modèle cognitif de l'agent, la résistance au changement de celui-ci est mise à jour. La manière dont ces mises à jour sont effectuées implante différentes stratégies d'engagement individuel sur les différentes cognitions. Par exemple, lorsqu'une intention sociale a une très forte résistance au changement, cela correspond à une stratégie d'engagement individuel de type fanatique (blinded). Dans notre modèle, cette notion de résistance au changement ne s'applique pas qu'aux engagements psychologiques, elle s'applique également aux engagements sociaux. Par exemple, lorsqu'un engagement est explicitement refusé suite à une tentative visant à le faire accepter, l'agent à l'initiative de la

Procedure ModifiedBDICycle(B_0, I_0)

1: **Inputs :** B_0, set of initial beliefs ;
   I_0, set of initial intentions ;
   Those inputs are optional (used for the first call)
2: **Outputs :** *none*, this is not a function !
3: **Global :** $B := B_0$, object that store the agent's beliefs ;
   $I := I_0$, object that store the agent's accepted intentions ;
   I_s, object that store the agent's social or failed intentions ;
   D, objet that store the agent's desires ;
   List ρ, store both internal and external percepts ;
   List $\pi := null$, current plan, sequence of (possibly complex) actions ;
4: **Body :**
5: **while** *true* **do**
6: ρ.GetNewPercepts() ; // get new percepts ρ
7: B.Update(ρ) ; // update B on the basis of ρ
8: **if** Recondider(I,B) **then**
9: D := Options(B,I) ;
10: I := Deliberate(B,D,I) ; // deliberate if necessary
11: **end if**
12: **if** Empty(π) or Succeeded(I,B) or Impossible(I,B)) **then**
13: π := Plan(B,I) ; // replan if necessary
14: I_s := Filter(B,I) ; // assigne failed or social intentions
15: **else**
16: α := Head(π) ;
17: Execute(α) ; // execute an action
18: π := Tail(π) ;
19: **end if**
20: **if** *agenda*.Modified()=*true* **then**
21: CommunicationPragmatics(*false*) ; // pursue a dialogue or answer a new dialogue offer
22: **end if**
23: **if** not Empty(I_s) **then**
24: CommunicationPragmatics(*true*) ; // initiate a dialogue
25: **end if**
26: **end while**

FIG. 7.5 – Algorithme de contrôle d'un agent BDI modifié pour prendre en compte notre approche de la communication entre agents.

tentative doit mettre à jour la résistance au changement de cet engagement dans son modèle cognitif. Trois cas sont à distinguer dans la mise à jour de cette résistance au changement :

1. si l'agent ne change pas cette résistance au changement (augmentation nulle), toutes choses étant égales par ailleurs, (en particulier la résistance au changement de l'intention correspondante, acceptée) l'agent va réitérer ses efforts pour le faire accepter, implantant ainsi une stratégie d'engagement fanatique (blindly committed) ;

2. si l'agent augmente cette résistance au changement de manière drastique, (à résistance au changement de l'intention correspondante, acceptée, égale) cela va favoriser le changement d'attitude en ce qui concerne l'intention correspondante (et les cognitions sous-jacentes liées), on obtient alors un agent influencable/faible (open-minded) dont l'intention ne persistera pas ;

3. entre les deux (à résistance au changement de l'intention correspondante, acceptée, égale) on a un continuum de stratégies d'engagements mixtes (single minded) impliquant divers degrés de persévérance pour l'agent.

Ajoutons que la frontière entre (1), (2) et (3) dépend également de la résistance au changement adoptée pour l'intention correspondante (qui rend compte de la stratégie d'engagement individuelle issue de la délibération par l'architecture BDI sous-jacente) et de celles des cognitions environnantes dans un calcul de cohérence plus global.

7.9 Implémentation et implantation

Notre implémentation, réalisée avec l'aide de Nicolas Andrillon [Andrillon, 2003] (que nous avons supervisé lors de son stage de fin d'étude d'ingénieur) repose sur celle du DGS, elle-même dûe à un effort collectif et introduite en section 5.3. Nos agents BDI ont été implantés à l'aide de la plateforme de développement orientée agent JACK™. JACK™est un environnement de programmation agent commercial, distribué par la firme Australienne Agent Oriented Systems (AOS). JACK™étend le langage objet JAVA et génère des agents logiciels en JAVA qui implémentent les concepts BDI [Howden et al., 2001], en particulier ceux de PRS [Ingrand et al., 1992] (Procedural Reasoning System) et dMars [Inverno et al., 1998, 2004] (Distributed Multi Agent Reasoning System). Les agents que nous avons développés étendent cette architecture BDI de manière générique grâce à un module qui encapsule les différents algorithmes présentés dans ce chapitre. Aussi, les agents développés utilisent notre approche de la pragmatique pour leurs communications. Les communications se font à l'aide

du langage DIAGAL au sein du simulateur DGS. L'utilisateur créé, initialise, exécute et visualise les agents via l'interface graphique du DGS, étendue avec Nicolas Andrillon pour l'intégration des agents du type de ceux décrits dans ce chapitre.

À la section 7.5, il a été question de la prise en compte du changement d'attitude par l'architecture BDI sous-jacente. Dans le cadre de notre implantation, cette prise en compte n'inclut que la prise en compte du changement d'intention. Rappelons que le changement d'attitude, tel qu'étudié en psychologie sociale (voir annexe C) peut impliquer d'autres cognitions privées. Notre implantation de la prise en compte du changement d'attitude par les agents JACK™, qui agissent conformément aux interpréteurs BDI classiques, tels que celui présenté en section 7.2, consiste simplement à mettre à jour la structure intentionnelle de l'agent. Lorsqu'une nouvelle intention est acceptée suite à un changement d'attitude (c'est également vrai lorsque c'est suite à une délibération), un événement interne de type intention est déclenché ($BDIGoalEvent$ dans la terminologie JACK™). Si cet événement est consommé par un plan, l'agent ajoute ce plan à sa pile d'actions courantes. Lorsqu'une intention précédemment acceptée est rejetée, les éventuels plans visant à la satisfaire sont abandonnés.

À la section 7.3, il a été question d'un filtre pour sélectionner les intentions sociales et les intentions échouées. Dans l'algorithme de contrôle BDI présenté figure 7.5, ce filtre est implanté par la fonction Filter(B,I). Dans le cadre des agents JACK™, ce filtre s'implante de manière simple. Le raisonnement fin-moyen des agents JACK™ s'effectue grâce à un système d'appareillage des intentions ($BDIGoalEvent$, dans le cadre JACK™) et des plans disponibles. Dans ce cadre, notre filtre consiste simplement à récupérer les intentions pour lesquelles cet appareillage échoue. Il s'agit bien des cas correspondants aux intentions sociales (pour lesquels aucun plan individuel n'est généralement défini) et les intentions échouées (pour lesquelles tous les plans disponibles, s'il y en avait, ont échoués).

Comme il serait fastidieux de détailler notre implémentation JACK™ et Java et son implantation complète ici, nous renvoyons le lecteur intéressé aux documents techniques produits à cet effet [Andrillon, 2003][20]. La section suivante fournit un exemple détaillé d'une exécution du système résultant.

7.10 Exemple détaillé

Supposons que deux agents, Paul et Peter, se sont entendus sur un plan commun pour aller ensemble au concert de leur groupe préféré tout en partageant les frais. Une sous-tâche de ce

[20] Le code, ainsi que des démonstrations du système peuvent également être présentés sur demande.

plan commun est d'aller acheter les billets. Paul s'est vu assigner cette sous-tâche et délibère sur la méthode par laquelle il va se rendre au magasin où se vendent les billets. Il doit choisir entre deux intentions mutuellement exclusives : y aller en taxi ($I_{Paul}(Aller\_en\_taxi_{Paul})$, noté $I_{Paul}(T)$ dans la suite) ou y aller à pied ($I_{Paul}(Aller\_a\_pied_{Paul})$, noté $I_{Paul}(P)$ dans la suite). On suppose que l'architecture BDI de Paul a acceptée la première et refusée la seconde (peut-être afin d'économiser du temps sachant que le taxi est plus rapide). Ces intentions ont une résistance au changement qui dépend de la stratégie d'engagement individuel retenue à la délibération ou déterminée à la conception de l'agent. Des résistances au changement faibles[21] ont été affectées à ces intentions pour nous permettre d'exemplifier le changement d'attitude. Les intentions acceptées se sont vu affecter une résistance au changement de 0.1 tandis que l'on assigne aux intentions rejetées (elles le sont par défaut) une résistance au changement de 0.05.

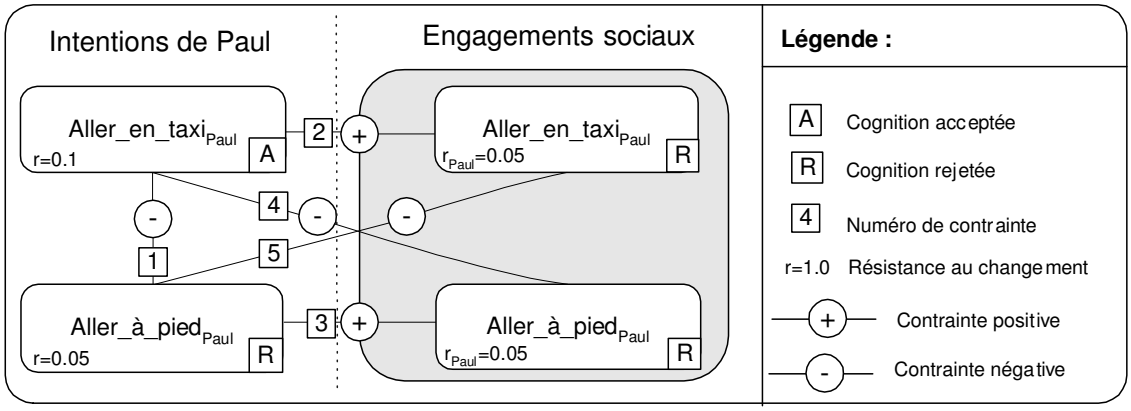

FIG. 7.6 – Modèle cognitif de Paul à l'état initial.

Ces intentions sociales peuvent être associées aux deux engagements correspondants (en accord avec les liens établis en section 7.3) : l'engagement social de Paul envers Peter d'aller à pied ($C(Paul, Peter, Aller\_a\_pied_{Paul})$, noté $C(P)$ dans la suite) et celui de prendre un taxi ($C(Paul, Peter, Aller\_en\_taxi_{Paul})$, noté $C(T)$ dans la suite). En outre, l'engagement de Paul envers Peter d'y aller à pied et l'intention d'y aller en taxi sont incompatibles de même que l'engagement d'y aller en taxi et l'intention d'y aller à pied. À partir de cet état initial, et conformément à notre modèle (Chapitre 6), des contraintes positives sont associées aux relations de correspondance et des contraintes négatives sont associées aux relations d'exclusion mutuelle et d'incompatibilité décrites ci-dessus. La Figure 7.6 présente le réseau d'intentions et d'engagements de Paul, tel que décrit ci-dessus. Notez que les engagements sont simplement les représentations que l'agent (ici Paul) s'en fait pour pouvoir raisonner dessus. À cette étape, ce ne sont pas de réels engagements puisqu'ils n'ont pas été établis ni rejetés par le dialogue. Ils sont toutefois rejetés, car c'est l'état par défaut d'un engagement inactif.

[21] Indiquant que l'agent ne va pas poursuivre une stratégie fanatique.

L'établissement, c'est-à-dire la mise en commun, de cet état initial pour les engagements est assurée par le fait que notre hypothèse de monde clos est partagée (tous les engagements non explicitement acceptés sont rejetés). Les engagements rejetés ont par défaut une très faible résistance au changement initial (0.05 dans notre exemple). Si l'état courant de l'engagement (accepté ou rejeté) est maintenu commun par les mécanismes décrits au chapitre 5, chacun des agents maintien une représentation subjective de ces engagements et de la résistance au changement qui leur est attribuée subjectivement. Celle-ci est donc susceptible de varier selon l'agent considéré. C'est pourquoi, dans la figure 7.6 comme dans les suivantes, la résistance au changement d'un engagement est indicée par l'identificateur de l'agent qui la maintient. Finalement, dans cet exemple, un poids unitaire a été affecté à chaque contrainte par simplification[22].

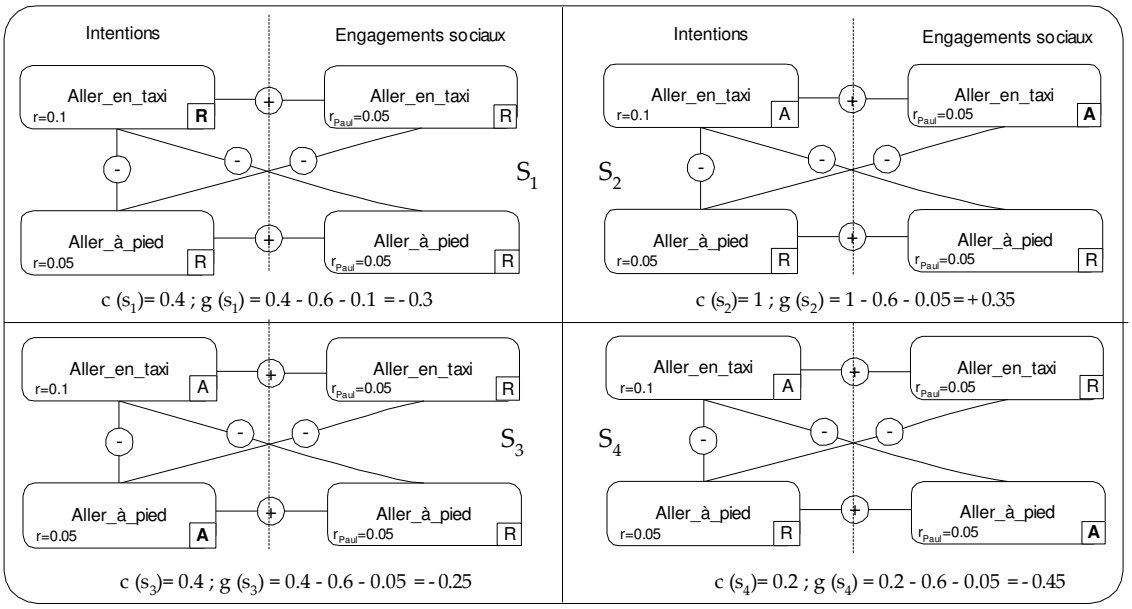

FIG. 7.7 – États explorés par l'algorithme de recherche local de Paul à partir de l'état initial.

Le simulateur DGS permet de choisir quel agent a l'initiative. Supposons que l'on donne l'initiative à Paul. Initialement, comme l'indique la figure 7.6, Paul a trois de ses contraintes satisfaites (numérotées 1, 3 et 4 sur la figure 7.6) sur les cinq contraintes le concernant dans cet exemple. Cela fait une cohérence de 0.6 ($3/5$, c'est-à-dire trois contraintes satisfaites sur un total de cinq) pour une cohérence maximum de 1 ($5/5$, c'est-à-dire que toutes les contraintes sont satisfaites). Paul va donc essayer d'augmenter cette cohérence en localisant le changement d'état (d'un élément de cognition) qui en cas de succès serait le plus utile.

[22] Les techniques de représentation des connaissances dans les formalismes hybrides symboliques connexionistes ne sont pas le sujet de cette thèse. Par contre, la section 8.3.3 montre comment elles sont introduites par la présente approche dans le champ des systèmes multi-agents et quels en sont les enjeux et perspectives.

La figure 7.7 montre les différents états accessibles à partir de l'état courant ainsi que leurs valeurs de cohérence (c) et d'utilité espérée (g).

En accord avec ces valeurs, Paul va essayer de faire accepter l'engagement $C(T)$. Puisqu'il s'agit d'un engagement social, Paul va utiliser l'un des jeux de dialogue DIAGAL à sa disposition (voir section 7.4). Peter sera le partenaire du dialogue puisque c'est lui qui est concerné par l'incohérence attaquée et par l'engagement convoité. Paul choisit donc parmi les jeux de dialogue celui dont les conditions de succès s'unifient avec l'engagement désiré et dont les conditions d'entrées s'unifient avec la situation actuelle. Le seul jeu DIAGAL à réunir ces conditions est le jeu d'offre, $Offer\ Game$.

Conformément au jeu de contextualisation de DIAGAL, Paul va alors proposer à Peter de jouer ce jeu. On suppose que Peter est dialogiquement coopératif et accepte de jouer le jeu. Pour respecter les règles du jeu d'offre, Paul va alors produire un acte de langage commissif[23].

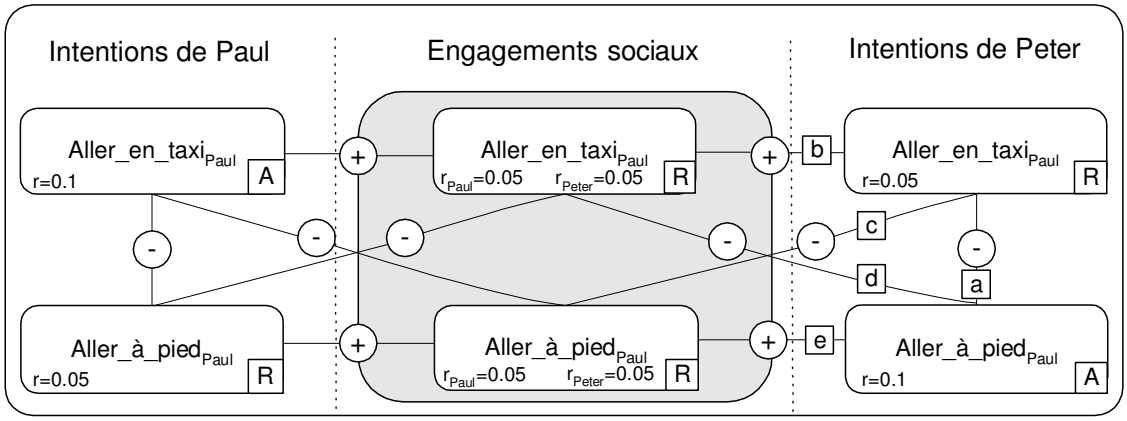

FIG. 7.8 – Modèles cognitifs de Paul et Peter avant la réponse de Peter.

À la réception de cette proposition, Paul est engagé à y répondre, conformément à la seconde règle du jeu d'offre DIAGAL. Comme les deux agents partagent les frais et que le taxi coûte plus cher que la marche à pied, on suppose que suite à sa délibération (menée par l'architecture sous-jacente), Peter a accepté l'intention que Paul aille à pied ($I_{Peter}(P)$) et rejeté l'intention que Paul aille en taxi (ce que l'on notera $\neg I_{Peter}(P)$). La Figure 7.8 rend compte de manière synthétique du modèle cognitif de Paul (sur la gauche) et Peter (sur la droite) à ce moment. La cohérence initiale de Peter est de 0.6, puisque les trois contraintes

[23] Et ce, éventuellement avec le degré de force illocutoire approprié. Nous ne considérons pas les degrés de force illocutoire dans cette application, mais il est très simple de remédier à ce manque. Une association telle que [0 ;0.2]= -2 ,[0.2 ;0.4]=-1 ,[0.4 ;0.6]=0 , [0.6 ;0.8]=1, [0.8 ;1]=2, permet de mettre concrètement en application le relation théorique introduite section 6.6.8 (le langage DIAGAL le permet, voir section 5.4.2). La même remarque pourrait s'appliquer au traitement des émotions des agents. Cependant pour ne pas alourdir le texte et puisque ces points méritent d'être approfondis, nous reléguons ces aspects au rang de perspectives.

nommées a, b et c sur la figure 7.8 sont satisfaites sur les cinq contraintes de poids unitaire considérées. Avant de répondre en fonction du gain ou de la perte de cohérence entraînée par ce changement pour lui, Peter va chercher s'il n'a pas une incohérence de plus grande magnitude à traiter et isole l'engagement de Paul envers lui-même d'y aller à pied $C(P)$, comme indiqué par la figure 7.9. Peter va donc imbriquer un jeu de requête ($Request\ Game$) concernant cet engagement. Paul accepte de jouer le jeu et Peter produit un acte directif conformément aux règles du jeu. Paul répond à la requête de Peter par la négative, guidé par les effets de sa réponse sur sa cohérence cognitive, qui diminuerait en cas d'acceptation, tandis qu'elle stagne en cas de refus. Le jeu de requête est alors fermé puisque ses conditions d'échec ont été atteintes.

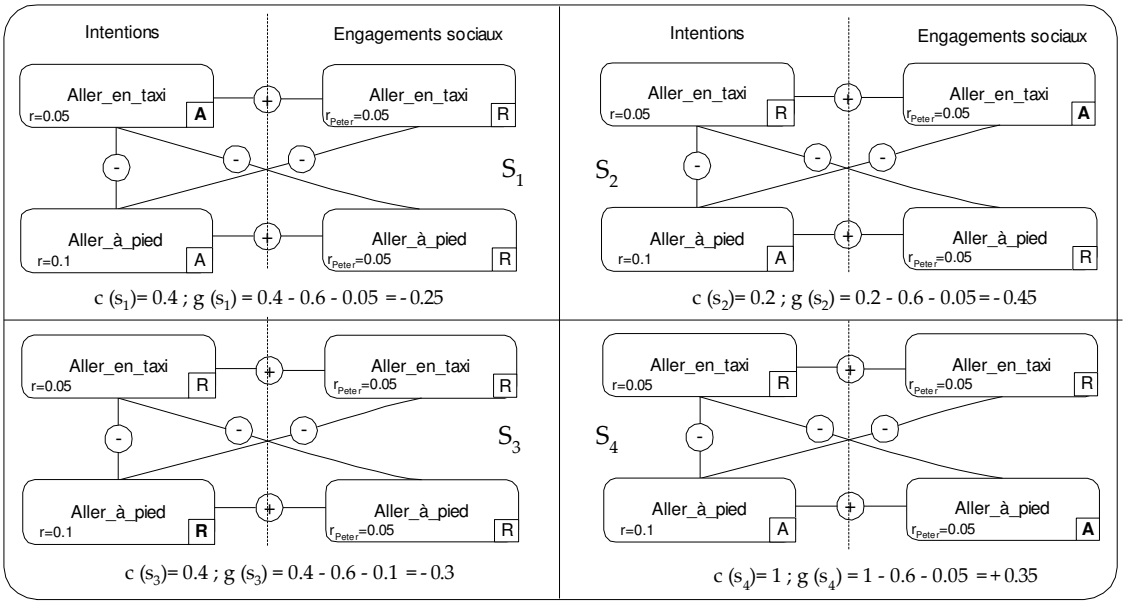

FIG. 7.9 – États explorés par Peter lors de sa recherche locale initiale.

Pour Paul, comme pour Peter, et de manière indépendante, la résistance au changement de cet engagement explicitement rejeté est mise à jour en fonction de leurs stratégies d'engagement individuel (telles qu'introduites en section 7.8). Afin d'illustrer le changement d'attitude, considérons que Peter augmente cette résistance au changement de manière importante, prenant ainsi la réponse de Paul de manière ferme, implémentant ainsi une stratégie de type influençable. Supposons que la résistance au changement de l'engagement rejeté $C(P)$ pour Peter est désormais de 1.

Le calcul de cohérence de Peter (via l'algorithme de recherche locale) indique alors un changement d'attitude. La figure 7.10 présente sous la forme d'un arbre de décision les différentes itérations de notre algorithme de recherche locale de Peter à partir de l'état décrit par la figure 7.8 dans lequel la résistance au changement de $C(P)$ pour Peter est désormais de 1.

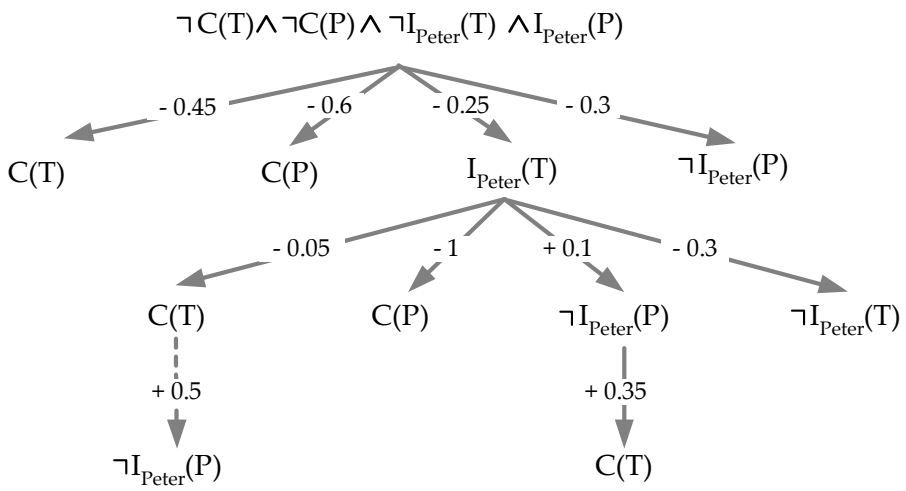

FIG. 7.10 – Arbre de décision de Peter (dans le cas où $r_{\neg I_{Peter}(T)} < r_{I_{Peter}(P)}$, 0.05 et 0.1 respectivement). La flèche en pointillés indique le chemin retenu par l'algorithme de recherche locale dans le cas où $r_{PeterC(T)} < r_{I_{Peter}(P)}$.

Les arcs sont étiquetés avec la valeur de la fonction d'utilité espérée (g) pour le changement considéré. La première itération de la recherche locale indique que Peter devrait accepter l'intention que Paul aille en taxi ($I_{Peter}(T)$). Une seconde itération est déclenchée par le fait que le résultat de la première indique une cognition privée sur laquelle Peter a le contrôle. L'augmentation (de 0.5) de la résistance au changement de l'intention acceptée lors de la première itération découle du fait que chaque élément explicitement accepté ou rejeté voit sa résistance au changement augmenter conformément à la stratégie d'engagement de l'agent. Techniquement, cela a pour effet de bord d'éviter une oscillation. En effet, si cette résistance au changement n'était pas mise à jour, alors au tour suivant, l'algorithme refuserait de nouveau cette intention et cela mènerait vers le maximum local dont on sort.

La seconde itération de la recherche locale indique que Peter devrait en outre rejeter l'intention que Paul aille à pied ($I_{Peter}(P)$). La résistance au changement de cette intention rejetée est alors mise à jour (augmentée de 0.5). La troisième itération indique que Peter devrait finalement essayer de faire accepter l'engagement de Paul d'y aller en taxi. C'est justement ce que lui a proposé Paul et il est engagé à accepter ou à refuser cette proposition.

Puisque le choix des valeurs des différentes résistances au changement peut sembler discutable au lecteur attentif, indiquons que si initialement la résistance au changement de l'intention $I_{Peter}(T)$ est inférieure à celle de l'engagement refusé que Paul aille en Taxi $\neg C(T)$, la seconde itération de l'algorithme de recherche locale de Peter va déboucher sur l'acceptation de cet engagement (la recherche locale sera alors stoppée pour laisser place au traitement dialogique) et à la prochaine itération, $I_{Peter}(P)$ sera rejetée, comme l'indique l'arc

en pointillé sur la figure 7.10. Notons que dans les deux cas la cohérence de Peter finie par être optimale ($c = 1$, les 5 contraintes sont satisfaites) et l'utilité totale est identique ($-0.25 + 0.1 + 0.35 = -0.25 - 0.05 + 0.5 = +0.2$).

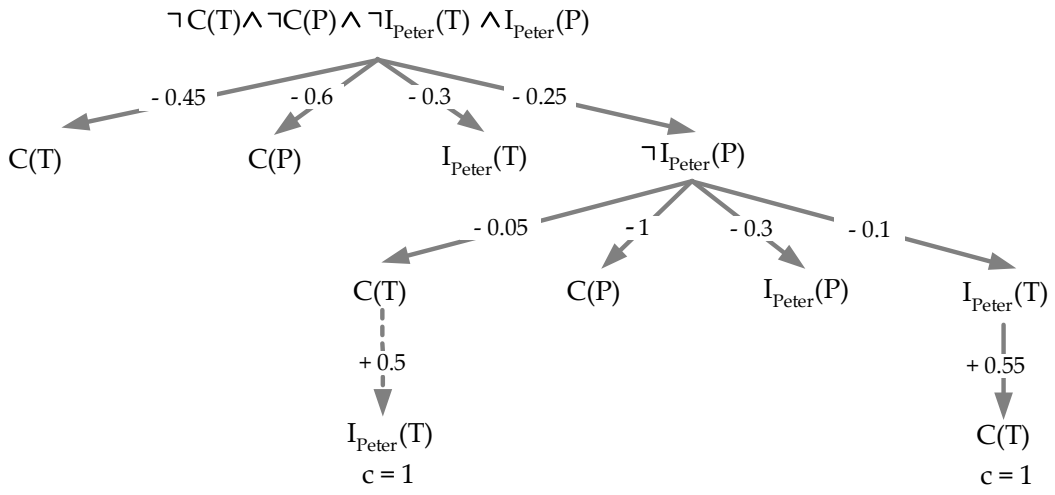

FIG. 7.11 – Arbre de décision de Peter dans le cas où $r_{\neg I_{Peter}(T)} > r_{I_{Peter}(P)}$. La flèche en pointillés indique le chemin retenu dans le cas où $r_{PeterC(T)} < r_{I_{Peter}(P)}$.

En outre, on peut se demander ce qui serait advenu si la résistance au changement de l'intention (initialement acceptée par Peter) que Paul aille à pied avait été inférieure à celle que Paul prenne un taxi (qui est rejeté par défaut). Aussi, au lieu de mettre la première à $+0.1$ et la seconde à $+0.05$ en argumentant que les intentions explicitement acceptées par un processus délibératif sont plus difficilement modifiables que celles refusées par défaut on pourrait mettre la première à $+0.05$ et la seconde à $+0.1$ en argumentant qu'après tout, ce qui importe le plus à Peter ce n'est pas que Paul aille à pied mais bien qu'il n'y aille pas en taxi. Dans ce cas, les états explorés par la recherche locale de Peter sont alors ceux indiqués par la figure 7.11. Sans détailler de nouveau les calculs qui mènent à cette arborescence, notons que le résultat est équivalent puisque dans tous les cas le changement d'attitude a lieu et Peter accepte l'engagement de Paul d'aller en taxi. Dans tous les cas, cela mène à une cohérence optimale pour Peter avec une utilité totale du dialogue de $+0.2$.

À la suite de ces calculs, la procédure $TreatCommitments()$ de l'algorithme de gestion pragmatique de Peter appelle l'architecture sous-jacente pour propager le changement d'attitude et redélibérer. Cette ré-considération devra inclure (pour respecter les calculs précédents) (1) le rejet de son « intention de » que Paul aille à pied et (2) l'acceptation de l'intention « l'intention de » que Paul aille en taxi[24] ainsi (3) qu'une révision des autres états

[24] L'architecture sous-jacente peut également produire une redélibération importante et finalement rejeter la nouvelle intention et générer des intentions alternatives, cette possibilité permet de rendre compte de la vision

mentaux de l'agent. La propagation du changement d'attitude (qui est en soi une forme d'apprentissage, dont indiquer comment elle est opérationnalisée est un sujet de recherche ouvert autant en intelligence artificielle qu'en sciences cognitives) et la re-délibération qui seraient normalement effectuées par l'architecture sous-jacente sont simplement simulées dans notre implémentation : on suppose que le résultat de nos algorithmes est systématiquement accepté et que les modifications correspondantes sont répercutées sur les autres états mentaux. L'état des modèles cognitifs de nos agents après ce dialogue est indiqué par la figure 7.12.

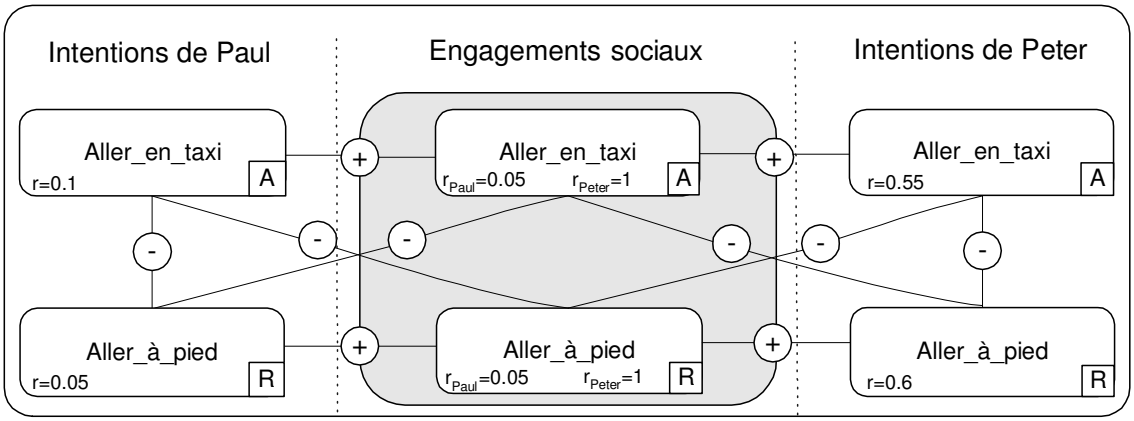

FIG. 7.12 – Modèles cognitifs de Paul et Peter après le changement d'attitude de Peter.

Peter va alors satisfaire son engagement de répondre à l'offre de Paul, en accord avec son calcul d'utilité espérée sur son nouvel ensemble de cognitions (et conformément à l'arbre de décision 7.10). Il va donc accepter l'offre de Paul, ce qui lui permet d'atteindre une cohérence optimale. Le jeu d'offre va être fermé et le dialogue terminé puisque la cohérence de Paul est également optimale.

Ainsi, à la suite de ce dialogue, Paul est engagé envers Peter à aller acheter les places en taxi. L'utilité globale du dialogue peut être calculée pour chacun des agents. Elle est de $+0.35$ pour Paul et de $+0.2$ pour Peter. Il est bien normal que cette utilité soit positive puisque les deux agents sont parvenus à un accord (l'incohérence externe qui les a occupés a été réduite avec succès). En outre, il est logique que cette utilité soit plus grande pour Paul puisque celui-ci n'a pas eu à faire l'effort d'un changement d'attitude comme Peter.

Un dialogue de décharge aura lieu après la date butoir de l'engagement (précisée dans l'action qui en est le contenu) si aucun des agents n'annule cet engagement entre temps,

plus large que permet la prise en compte des désirs et des croyances y compris concernant les accointances dont seule l'architecture sous-jacente dispose dans notre modèle. Dans ce cas, le changement d'attitude effectué par nos algorithmes peut être envisagé comme une option supplémentaire qui sera retenue ou non par l'architecture sous-jacente.

il sera alors déterminé si l'engagement a été respecté ou non et les éventuelles sanctions s'appliqueront en accord avec les mécanismes décrits au chapitre 5.

Dialogue résultant

Le diagramme de séquence de la Figure 7.13 illustre l'échange de messages entre Paul et Peter dans le cadre de l'exécution détaillée ci-dessus. Ce diagramme est une reproduction du diagramme d'échange de messages tel qu'affiché par le DGS et qui permet à l'utilisateur de suivre pas à pas le déroulement de l'interaction (voir section 5.3.3). On y visualise les actes de contextualisation aussi bien que les actes de langage (ou actes communicationnels) utilisés par les agents et résultant du respect des règles du jeu de contextualisation et des jeux de dialogue utilisés (exprimées en termes d'engagements dialogiques). Notons, car c'est l'un des principaux apports de cette thèse, que l'ensemble de ces interactions ont été tenues automatiquement par les agents via l'implémentation de notre approche des aspects cognitifs de la pragmatique.

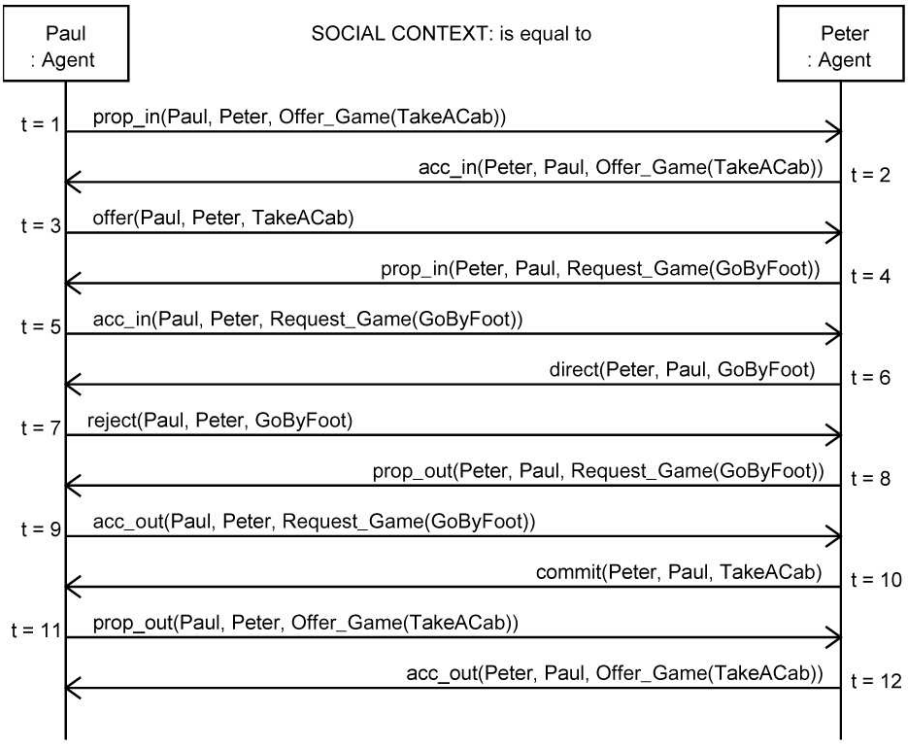

FIG. 7.13 – Diagramme de séquence du dialogue entre Paul et Peter.

Dans le cas où Peter aurait eu l'initiative au début du dialogue, un dialogue symétrique aurait eu lieu. Peter aurait alors essayé d'établir l'engagement que Paul y aille à pied, Paul

aurait imbriqué un jeu d'offre pour y aller en taxi, sans succès, et aurait finalement accepté la requête de Peter après avoir changé d'attitude. C'est-à-dire que le dialogue aurait mené à la situation inverse. Aussi surprenant que cela puisse paraître, c'est tout à fait normal puisque nous avons supposé que la résistance au changement d'un engagement socialement rejeté augmente considérablement, c'est-à-dire que nos agents ne persistent pas du tout dans leurs tentatives. Avec ce type de stratégie d'engagement individuel, les agents sont extrêmement adaptatifs, c'est-à-dire influençables. Tout refus social entraîne presque immédiatement un changement d'attitude. Le moindre échec les amène à abandonner leurs intentions. Dans ce cas particulier (choisi parce qu'il permet d'illustrer simplement le changement d'attitude), l'initiative joue un rôle crucial (on pourra méditer sur ce résultat).

Finalement, nos deux agents ont tenu une conversation à laquelle pourrait correspondre le dialogue en langage naturel suivant (même si les phases d'établissement, d'acceptation et de mise en commun sont généralement plus implicites et pas forcément linguistiques avec la communication naturelle) :

Paul : Écoute ce que je te propose. <proposition d'entrer dans un jeu d'offre (le jeu DIAGAL $offer$)>

Peter : Ok, vas-y. <proposition acceptée>

Paul : Je vais aller acheter les billets en taxi. <offre>

Peter : Justement, je voulais te demander quelque chose à ce propos ? <proposition d'imbriquer un jeu de requête (le jeu DIAGAL $request$)>

Paul : Oui. <proposition acceptée>

Peter : Tu ne préfères pas y aller à pied ? <requête>

Paul : Non. <requête rejetée>

Peter : Ha bon, ... <proposition de fermer le jeu imbriqué>

Paul : Bah, oui. <proposition acceptée>

Peter : (soupir) bon, donc tu va y aller en taxi. <offre acceptée>

Paul : C'est d'accord ? <proposition de fermer le jeu d'offre>

Peter : Oui (long soupir). <proposition acceptée>

7.11 Résumé/synthèse du modèle

La figure 7.14 synthétise et résume le comportement des agents BDI, pour les aspects cognitifs de la pragmatique des communications, c'est-à-dire, les aspects concernant son uti-

lisation et ses effets. Ce comportement est dirigé par leur cohérence cognitive, comme nous l'avons détaillé dans ce chapitre ainsi que dans le précédent. La partie gauche (*part A*) de la figure schématise le modèle d'agent décrit dans ce chapitre tandis que la partie droite (*part B*) synthétise, sous la forme d'un arbre de décision, le traitement pragmatique de chacun des agents.

À chaque étape de son raisonnement pragmatique, l'agent cherche l'élément dont le changement maximiserait sa fonction d'utilité espérée. Trois cas sont alors à distinguer :

1. *si cet élément est une intention*, cela déclenche un changement d'attitude. S'il s'agit d'une intention acceptée, l'agent va la rejeter et à la fin du segment de dialogue, l'architecture sous-jacente sera appelée pour propager le changement d'attitude et entamer une re-délibération. S'il s'agit d'une intention rejetée, l'agent va l'accepter. L'architecture sous-jacente propagera ce changement d'attitude (une adoption d'intention) lorsque le segment de dialogue sera terminé. C'est le cas où modifier le monde extérieur (la couche des engagements sociaux dans notre cas) est trop coûteux pour l'agent. Dans ce cas, l'agent va changer ses propres cognitions pour restaurer la cohérence avec son environnement. C'est le changement d'attitude.

2. *si cet élément est un engagement accepté*, l'agent va devoir l'annuler ou le violer[25]. Dans ce cas, l'agent devra faire face aux éventuelles sanctions attachées. Notons que ces sanctions ont été prises en compte dans le calcul d'utilité espéré qui mène l'agent à cette décision.

3. *si cet élément est un engagement rejeté*, trois sous-cas sont à considérer :

   – *si aucun dialogue n'est ouvert avec l'agent concerné par l'engagement* : l'agent va initier un dialogue en proposant le jeu de dialogue susceptible de permettre la modification désirée, c'est-à-dire dont les conditions d'entrée et les conditions de succès s'unifient avec l'état présent et l'état souhaité respectivement ;

   – *si un dialogue est déjà ouvert avec l'agent concerné* : l'agent va satisfaire ses engagements dialogiques en fonction de son calcul d'utilité et va imbriquer ou pré-séquencer un jeu adapté, susceptible de permettre la modification désirée ;

   – *si un dialogue est déjà en cours à propos l'engagement considéré pour la modification considérée* : l'agent va simplement traiter ses engagements dialogiques en accord avec son calcul de cohérence cognitive.

[25] Notons que l'annulation est généralement moins coûteuse que la violation. En outre, par définition, seul son débiteur peut violer un engagement.

Finalement, chaque jeu de dialogue se conclut par l'atteinte des conditions d'échec ou de succès signifiant respectivement le succès ou l'échec de l'initiateur dans sa tentative dialogique. En cas d'échec, il revient aux agents de mettre à jour la résistance au changement des engagements discutés pour prendre en compte cet échec. Cette mise à jour influence la possibilité de changement d'attitude subséquent de la manière décrite à la section 7.8.

Le modèle proposé dans ce chapitre est est à mettre en regard avec le l'approche intentionnelle de la communication agent apliquée au modèle BDI traditionnel et représenté sur la figure 7.15. Dans notre approche, les éléments grisés sur la figure 7.15 ont été substitués par ceux indiqués sur la partie gauche (*part A*) de la figure 7.14. Ce faisant, c'est une nouvelle architecture délibérative-normative que nous avons proposée. Celle-ci nous semble complémentaire des propositions déjà effectuées dans la communauté puisque qu'aucune ne traite avec la même précision les aspects communicatifs en isolant le raisonnement pragmatique comme nous l'avons fait.

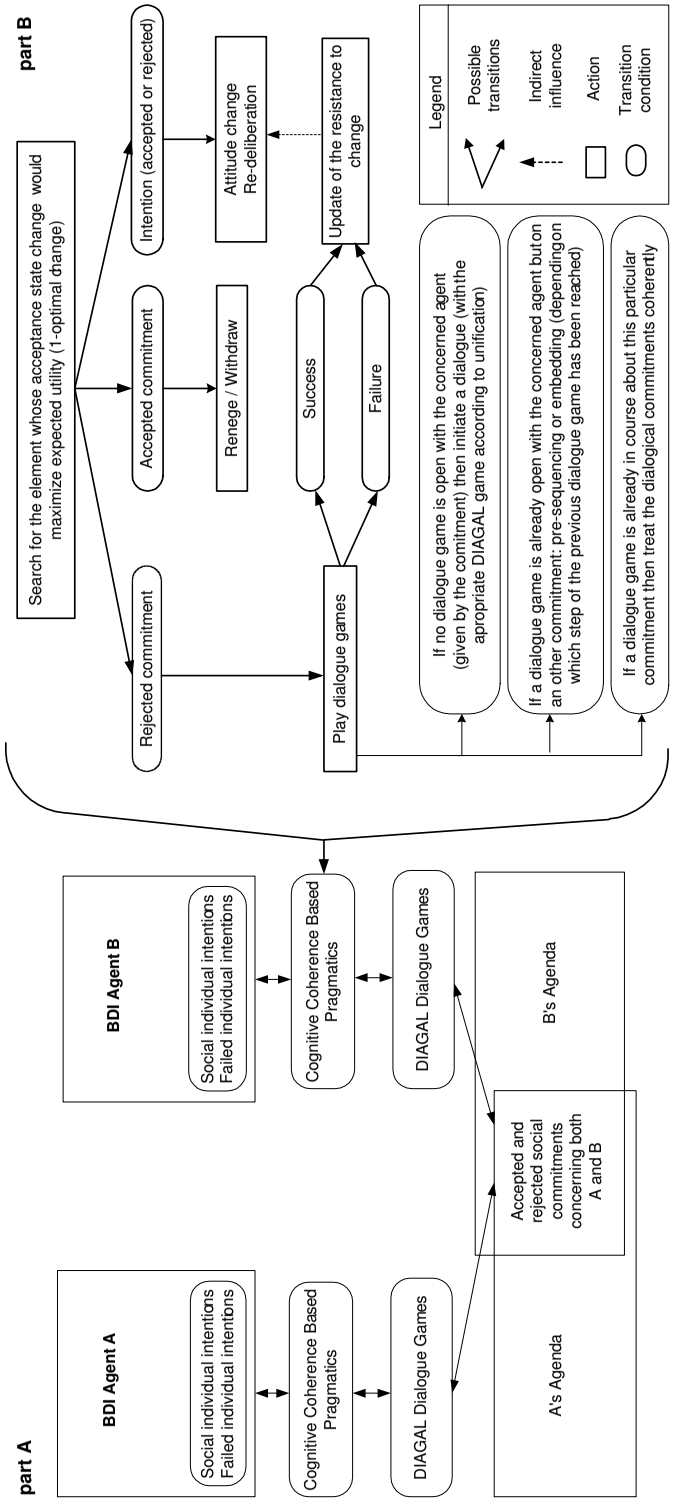

FIG. 7.14 – Schématisation du modèle développé pour notre validation informatique.

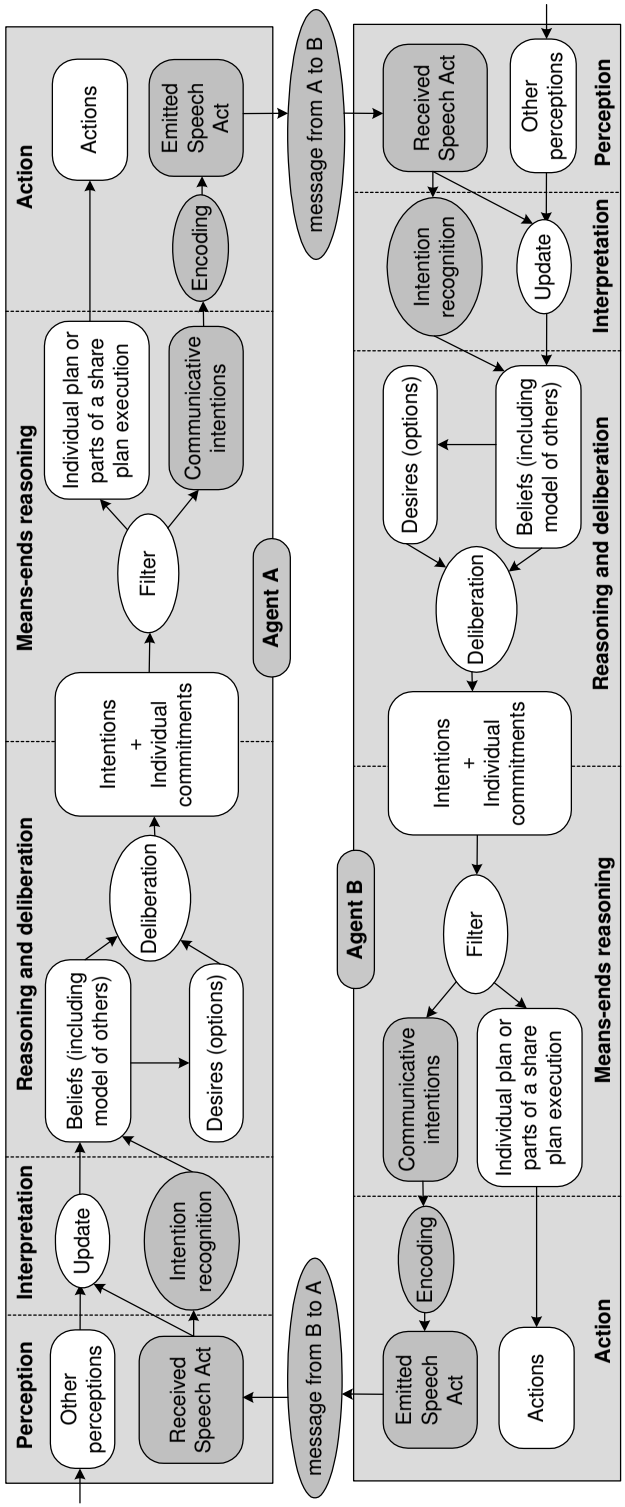

FIG. 7.15 – Modèle de la communication entre deux agents BDI.

7.12 Conclusion

Dans ce chapitre, nous avons présenté notre validation informatique, dans le cadre des systèmes multi-agents, de la théorie élaborée au chapitre précédent. Cette validation s'appuie sur notre cadre interactionnel (le langage DIAGAL et le simulateur DGS, présentés au chapitre 5), lui-même construit à partir de notre modèle d'engagement (présenté au chapitre 4). Cette validation nous a permis de mettre en oeuvre un cadre technique pour l'émergence de conversation entre agents cognitifs de type BDI (section 7.2) dans un cadre conventionnel reposant sur les engagements sociaux. Pour ce faire, nous avons :

- introduit et opérationnalisé les liens théoriques existants entre intentions et engagements sociaux dans les cas appropriés (section 7.3) ;
- opérationnalisé le calcul de cohérence via notre algorithme de recherche locale et la fonction d'utilité espérée qui guide l'agent (section 7.6) ;
- introduit un algorithme de traitement pragmatique (section 7.7) et discuté les interrelations existantes entre cette surcouche communicationnelle et l'architecture BDI sous-jacente ;
- discuté la prise en compte des stratégies d'engagement individuel et la manière dont la notion de reconsidération d'intention est raffinée et étendue par notre approche (section 7.8) ;
- détaillé un exemple d'exécution du système résultant[26] (section 7.10).

Cette architecture logicielle, synthétisée section 7.11 répond dans la pratique aux six premières questions de notre liste de questions concernant les aspects cognitifs de la communication :

1. Quand un agent prend-il l'initiative d'une conversation, à quel sujet et pourquoi ?
2. Avec qui ?
3. Par quel type de dialogue ?
4. Comment définir et mesurer l'utilité d'une conversation ?
5. Quand arrêter le dialogue ou le cas échéant comment le poursuivre ?

[26] D'autres exemples de simulations sont disponibles dans [Pasquier, 2002] et [Andrillon, 2003].

6. Quels sont les impacts du dialogue sur les attitudes de l'agent ?
7. Quelle intensité donner aux forces illocutoires des actes de langage utilisés ?
8. Quels sont les impacts du dialogue sur l'humeur de l'agent ?
9. Quelles sont les conséquences du dialogue sur les accointances de l'agent ?

Dans ce travail d'implémentation et de validation, nous n'avons pas tenu compte des questions 7 et 8, car (1) elles sont théoriquement résolues par les éléments introduits section 6.6.8 et (2) elles ne sont pas pertinentes pour les dialogues entre agents artificiels de type BDI puisque cette architecture n'inclut pas de traitement émotionnel. La question 9, quant à elle, mériterait plus de travail puisque, comme indiqué en section 6.6.7, l'idée sous-jacente est d'exploiter l'utilité des dialogues tenus pour la gestion des accointances de l'agent. Ce type de traitement est reporté comme l'une de nos perspectives (section 8.4.3).

Si cette validation informatique est en elle-même une contribution, celle-ci mériterait d'être approfondie, raffinée et explorée plus avant. On envisage notamment de fournir (par démonstration ou par simulation) des résultats théoriques supplémentaires (preuve de convergence, ...). En ce qui concerne les comparaisons, c'est à notre connaissance la première tentative visant à automatiser la communication entre agents artificiels BDI dans un cadre conventionnel reposant sur les engagements sociaux. Le chapitre suivant discute les perspectives et les travaux connexes à notre approche.

Chapitre 8

Perspectives, travaux connexes et discussion

8.1 Introduction

Dans les chapitres précédents, nous avons présenté : (1) un modèle de l'engagement social (chapitre 4), (2) le langage de communication agent DIAGAL, reposant sur ce modèle (chapitre 5), (3) un modèle général des aspects cognitifs de la pragmatique des communications entre agents (chapitre 6) et (4) une implémentation de ce modèle qui étend et adapte le modèle BDI aux cadres de communication tel que celui présenté dans les trois chapitres précédemment cités (chapitre 7).

Dans le présent chapitre et avant de présenter notre conclusion, nous proposons et discutons quelques pistes de raffinement et d'extension de notre modèle qui en explicitent les forces et les faiblesses. Après avoir discuté l'importance du changement d'attitude dans les systèmes multi-agents (section 8.2.2), nous commencerons par introduire un raffinement de notre modélisation du changement d'attitude (section 8.2). Cela nous amènera à discuter des difficultés concernant la prise en compte des sanctions dans la prise de décision (section 8.2.4).

Ensuite, nous discuterons des enjeux théoriques et pratiques des approches cohérentistes comme la nôtre et des formalismes hybrides symboliques connexionistes qui y sont attachés (section 8.3). Quelques perspectives d'extensions de ce travail sont ensuite présentées (section 8.4).

Finalement, nous introduirons brièvement les travaux qui nous semblent connexes aux nôtres et nous discuterons les utilisations qui sont faites de notre approche dans la communauté de recherche en intelligence artificielle sur les systèmes multi-agents et leurs applications.

8.2 Discussion et raffinements de notre modèle des aspects cognitifs de la pragmatique

8.2.1 Hypothèse de coopération : des approches intentionnelles à la cohérence cognitive

Dans la vision inaugurée par Austin [1962], qui est à la base des approches intentionnelles détaillées au chapitre 2, la communication agent consiste en une suite d'actions individuelles (les actes de langage) dont les effets attendus sont des changements dans les états mentaux des interlocuteurs (avec les conséquences ultérieures que cela peut avoir sur l'environnement, via les actions individuelles). Dans ce cadre, une forte hypothèse de coopération peut seule assurer le bon fonctionnement du système (c'est l'une des limites des approches intentionnelles que nous avons discutées à la section 2.2.6). Pour autant, dans les systèmes multi-agents, la communication est utilisée par les agents pour se coordonner, que ces agents soient coopératifs ou non [Reed et al., 2002].

Dans notre modèle, lorsqu'un agent en vient à essayer de réduire une incohérence interne (explicite) et si aucune récompense ne tend à justifier cet effort, son interlocuteur doit être coopératif pour accepter de l'aider. C'est le cas avec les types de dialogues qui traitent des incohérences internes, comme la recherche d'informations et l'investigation. Pour autant, l'hypothèse de coopération posée dans le cas général par les approches intentionnelles (voir section 2.2.1) est trop forte. Elle a été introduite pour pallier au fait que ces approches ne donnent pas une caractérisation satisfaisante des aspects motivationnels du dialogue. Aussi, dans la plupart des systèmes reposant sur cette approche l'initiative dialogique n'est pas réellement prise en compte et les agents ne font que réagir coopérativement à l'intention reconnue chez l'allocataire. Les systèmes de question-réponse, qui sont en fait des systèmes de réponse en ce qu'ils ne prennent généralement pas l'initiative dans le dialogue, sont un exemple parfait de ce cas.

Dans notre modèle, la communication n'est pas simplement utilisée pour réduire des incohérences internes, mais également pour expliciter et éventuellement réduire des incohérences

externes. En effet, lorsqu'ils communiquent, les agents d'un système ouvert peuvent être en situation d'accord ou de désaccord préalable (cela est généralement implicite). Dans le premier cas, la communication peut se réduire à la dissémination des informations susceptibles d'expliciter un accord préalable jusque-là implicite tandis que dans le second cas elle sera également utilisée pour traiter le désaccord explicité. En effet, la notion d'explicitation, que nous avons introduite section 6.6.5, est inhérente à la nature distribuée des SMAs dans lesquels chaque agent n'a qu'une perception partielle de l'environnement. Cette explicitation peut être celle d'une situation d'incohérence comme d'une situation cohérente. Par exemple, si dans l'exemple présenté au chapitre précédant, l'architecture interne de l'agent Peter avait, dès le départ, accepté l'intention que Paul aille en taxi et rejetée celle qu'il aille à pied, le dialogue aurait été plus simple. En effet, lorsque les deux agents sont a priori implicitement d'accord, le premier qui fait une proposition ou une requête la voit acceptée par l'interlocuteur et la cohérence est simplement établie. La communication sert alors à partager, expliciter et établir cet accord préalable.

Si dans le cas de l'accord préalable implicite, la communication peut se résumer à une dissémination d'information révélant un accord préalablement implicite, la communication est également l'outil par lequel les agents se mettent d'accord, ou font des tentatives dans ce sens, en cas d'incohérence. C'est le cas avec les types de dialogues étudiés en dialectique qui présupposent tous une forme de conflit de position ou d'intérêt en préalable, comme nous l'avons indiqué section 6.6.4.

Aussi, si la coopération dialogique est systématiquement requise pour que le cadre de communication puisse être utile, l'hypothèse de coopération extra-dialogique n'est pas nécessaire dans le cas des incohérences externes. Le principe de cohérence nous semble suffisant pour rendre compte des aspects motivationnels du dialogue. Si nous indiquons cela comme une piste de réflexion quant au dépassement éventuel de l'hypothèse de coopération habituelle, nous laissons cependant les développements de cette application du principe de cohérence au groupe à d'ultérieures études.

8.2.2 Le changement d'attitude dans les communications agents

Dans le cas du désaccord préalable, c'est-à-dire lorsque qu'une incohérence externe est explicitée, le changement d'attitude d'au moins une des parties est attendu. Plus généralement, le changement d'attitude est au coeur de tous les dialogues impliquant une dimension persuasive[1]. En effet, il est désormais accepté en psychologie sociale que le but premier de

[1] Les notions de crédibilité et de confiance, omniprésentes dans les communications humaines, amènent la plupart des auteurs à attribuer une valeur persuasive à toute action communicationnelle [Petty et Cacioppo, 1996].

toute communication persuasive est le changement d'attitude [Erwin, 2001]. Cette notion a été largement ignorée dans la littérature concernant la communication agent. Pourtant, dans les théories d'agents cognitifs, comme le modèle BDI que nous avons utilisé jusqu'alors, il y a la notion de *reconsidération d'intention* qui est proche de celle du changement d'attitude.

Dans les travaux passés de la communauté sur les agents BDI, il a été accepté que les agents doivent de temps à autre reconsidérer leurs intentions, soit parce qu'elles ne sont plus réalisables ou encore parce que d'autres opportunités ont pu apparaître [Wooldridge, 2001a]. Ce sujet est d'ailleurs toujours en discussion et de nombreux raffinements ont été apportés au sein de la communauté [Schut et Wooldridge, 2001; Parsons et al., 2000]. Cependant, dans la liste des motifs susceptibles de déclencher la reconsidération d'intention, invoqués dans la littérature spécialisée [Rao et Georgeff, 1995], il est fait peu de cas de la communication. Notre modèle permet de combler ce manque en tenant compte des effets potentiels de la communication sur les intentions de l'agent (sans hypothèse de coopérativité). Une distinction importante à cet égard reste à faire. La distinction entre changement d'attitude partiel et changement d'attitude complet. La sous-section suivante raffine notre contribution à l'égard du changement d'attitude dans cette direction, en se basant sur notre étude des résultats de sciences cognitives à ce sujet. La formalisation et l'expérimentation sur ces derniers aspects restent à faire.

8.2.3 Raffiner notre modèle du changement d'attitude

Notre modèle du changement d'attitude peut être raffiné en prenant en compte des résultats plus fins de psychologie sociale. La notion d'attitude est une notion de méta-niveau qui regroupe, comme nous l'avons vu, l'ensemble des cognitions se rapportant à un même objet. Si cette notion s'accompagne naturellement d'un besoin de cohérence des différentes cognitions impliquées dans une attitude, les psychologues ont été plus loin et ont montré la validité du principe de cohérence cognitive : *l'individu préfère la cohérence à l'incohérence*. C'est ce principe motivationnel que nous avons exploité dans notre approche.

Les liens entre d'une part, les attitudes (leurs composantes cognitives et affectives) et les intentions (aspects volitifs des attitudes) et d'autre part entre les intentions et le comportement manifeste de l'agent ont été approfondis et raffinés tout au long du dernier demi-siècle d'investigation en psychologie cognitive et sociale. L'annexe C présente certains éléments constitutifs préalables tels qu'on les trouve dans les états de l'art classique en psychologie sociale comme ceux de Petty et Cacioppo [1996] ou de Erwin [2001]. Le schéma classique

qui ressort de ces études est[2] que les croyances (cognitions informationnelles) et les désirs (pulsions et affect) mènent aux intentions qui peuvent elle-même mener aux comportements manifestes ou aux tentatives dialogiques pour obtenir les engagements sociaux correspondants selon leur nature et en accord avec les liens introduits en section 7.3.

En retour, il arrive (du fait des hiérarchies sociales, des relations de pouvoir, de la nature des dialogues de négociations, d'argumentation ou de persuasion, . . .) qu'un agent se retrouve socialement engagé sur un objet contre-attitudinal. Dans ce cas, le changement d'attitude peut survenir.

Dans notre approche, les liens entre cognitions privées et publiques établis en section 7.3 permettent de définir le changement d'attitude de la manière décrite dans l'étude, classique en psychologie sociale, de Brehm et Cohen [1962]. Des études plus récentes permettent cependant de raffiner ces résultats et de distinguer deux types de changement d'attitudes. En effet, le changement d'attitude peut être (1) *partiel*, c'est-à-dire que l'agent peut se contenter d'adopter l'intention correspondante en supportant le fait qu'elle est incohérente avec le reste de sa cognition ou (2) *complet*, auquel cas, le changement d'attitude est propagé parmi les autres cognitions de manière à rétablir la cohérence interne.

Des résultats complémentaires de psychologie sociale [Myers et Lamarche, 2000], indiquent comment distinguer (1) de (2). Le changement d'attitude sera partiel quand il est externalement justifié[3] et que ces justifications externes sont suffisamment importantes pour compenser le coût de l'adoption et de la réalisation d'une intention anti-attitudinale. En revanche, lorsqu'aucune justification externe n'est trouvée, l'agent va procéder à un changement d'attitude complet de manière à rétablir la cohérence entre ses cognitions privées et les cognitions publiques (les intentions et les engagements, dans notre cas) ainsi qu'au sein de ses cognitions privées (les croyances, désirs et intentions, dans notre cas).

Cette distinction est un résultat connu en psychologie sociale comme *le paradigme de justification insuffisante* [Erwin, 2001]. Il est quelque peu contre-intuitif dans la mesure ou il signifie que les comportements anti-attitudinaux non externalement justifiés mèneront à un changement d'attitude plus complet, plus profond, que ceux qui sont justifiés. Cela signifie simplement qu'une fois l'agent engagé sur une action (a course of action) contre-attitudinal, il va justifier l'intention adoptée par un changement d'attitude complet dans la mesure ou cela n'est pas déjà justifié externalement. En d'autres termes, si l'agent a accepté une intention pour la récompense associée à l'engagement accepté correspondant ou pour éviter les

[2] Notons que celui-ci est conforme aux analyses des philosophes de l'esprit, prises comme références pour les modèles d'agents cognitifs actuels.

[3] Ces justifications externes sont, dans notre cas, les récompenses associées à la satisfaction des engagements ou encore le fait d'échapper aux sanctions associées au non-respect des engagements acceptés.

sanctions associées à son non-respect, il ne va pas procéder à un changement d'attitude complet. Ces résultats qui sont de grande valeur pour l'étude de l'apprentissage humain et pour les théories pédagogiques laissent entrevoir de nombreux développements dans le domaine des SMAs. Cependant, comme nous allons le voir dans la prochaine sous-section, le raffinement présenté dans cette section nous amène à lever le problème de la prise en compte des sanctions.

8.2.4 Prise en compte des sanctions

Pour pouvoir implanter les deux types de changements d'attitude – partiel ou complet – définis à la section précédente, les sanctions doivent être précisément prises en compte dans les calculs motivationnels décrits auparavant. Rappelons que dans la fonction d'utilité individuelle g définie en section 7.6, $cost$ est une fonction de coût calculée comme suit :

- *si la cognition à changer ($cognitionChanged$) est une intention*, son coût de modification est sa résistance au changement (telle qu'établie par l'architecture sous-jacente ou telle que mise à jour lors de la dernière modification d'état de la cognition considérée) ;
- *si la cognition à changer est un engagement rejeté*, son coût de modification est sa résistance au changement subjective (qui est généralement initialement faible, mais qui est susceptible d'augmenter à chaque tentative de changement d'état infructueuse) ;
- *si la cognition à changer est un engagement accepté*, son coût de modification correspond à sa résistance au changement à laquelle s'additionne les éventuelles sanctions attachées à son annulation ou à sa violation ainsi que les récompenses attachées à son respect (dans le cas du créditeur les récompenses sont soustraites des sanctions puisqu'il en aurait été le débiteur).

Ce dernier point mérite d'être clarifié au moyen d'un exemple. Supposons que l'engagement $C(A, B, \alpha_A, t, s_A, s_B)$, noté ζ dans la suite de cet exemple, ai été établi comme accepté lors d'un dialogue passé entre les agents A et B. Supposons qu'en vertu de la stratégie de punition en cours dans le système considéré, $s_A = \{4, 2, 1\}$ signifiant pour A une sanction de 4 pour la violation, 2 pour l'annulation, 1 pour la modification et $s_B = \{3, 2, 4\}$ signifiant pour B une sanction de 3 pour l'annulation, de 2 pour la modification et une récompense de 4 (à débourser par B au crédit de A) pour la satisfaction. Soit S_A, l'état du réseau pour l'agent A ($S_A = \{A, R, C+, C-\}$, c'est à dire l'ensemble des éléments acceptés (A) ou rejetés (R) et les différentes contraintes positives ($C+$) ou négatives ($C-$) les liants) et S'_A l'état dans lequel ζ est rejeté, toutes choses étant égales par ailleurs ($S'_A = \{A - \zeta, R + \zeta, C+, C-\}$).

S_B et S'_B sont les états représentant ces situations pour l'agent B. Dans ces conditions, on a alors :

$$g_A(S'_A) = coherence(S'_A) - coherence(S_A) - cost(reject(C(A, B, \alpha_A, t, s_A, s_B)), \text{ soit :}$$

1. $g_A(S'_A) = coherence(S'_A) - coherence(S_A) - (2 + 4)$, pour l'annulation par A (et la perte de la récompense subséquente) ;
2. $g_A(S'_A) = coherence(S'_A) - coherence(S_A) - (4 + 4)$, pour la violation par A (et la perte de la récompense subséquente) ;
3. $g_B(S'_B) = coherence(S'_B) - coherence(S_B) - (3 - 4)$, pour l'annulation par B.

Dans ces formules, les quantités retournées par la fonction de mesure de cohérence et celles utilisées pour traduire les sanctions doivent être de même nature, ce qui n'est pas simple à assurer. Cela met en évidence le fait que la prise en compte des sanctions pose un problème théorique classique pour lequel nous n'avons pas de solution. Ce problème est celui de l'intégration de différents critères dans la prise de décision multicritère et s'il est souvent formulé dans le cadre des théories de la décision classique, il est également pertinent pour les sciences cognitives. En effet, nous sommes en présence d'éléments « psychiques » subjectifs (cohérence cognitive, résistance au changement perçue) et d'éléments matériels (sanctions matérielles explicites) qui ne sont pas exprimés selon les mêmes métriques. Or rien ne semble nous indiquer une manière de rapprocher ces grandeurs dont il reste à savoir comment les intégrer dans un seul et même calcul motivationnel. Il en résulte nombre de difficultés théoriques et techniques pour lesquelles bien peu de solutions satisfaisantes ont été proposées [Kast, 1993].

Certains individus feraient pour 10$ ce que d'autres ne feraient pas pour 100 000$. Théoriquement, il en va de même pour les agents logiciels qui les représentent. Dés lors, il parait difficile d'avancer sur le sujet avec les schémas analytiques traditionnels. Par ailleurs, le savoir empirique en psychologie progresse rapidement et de nombreux facteurs ont déjà été isolés qui permettent d'envisager une progression sur ces questions d'intégration de différents critères.

Nous allons maintenant décrire de manière plus générale les enjeux théoriques et pratiques des approches cohérentistes, telles que la nôtre.

8.3 Enjeux théoriques et pratiques des approches cohérentistes

8.3.1 Approche cohérentiste en philosophie de l'esprit

Dans cette section, nous souhaitons rendre clair le fait que notre approche, quoique originale, n'est pas aussi isolée que cela pourrait sembler de prime abord, mais qu'elle s'inscrit dans un courant d'idée qui l'englobe. Étonnamment, ce courant n'est pas représenté dans la littérature agent tandis qu'il est parmi les plus importants dans le domaine, plus vaste, de la modélisation cognitive.

En effet, notre formulation théorique doit beaucoup aux approches cohérentistes telles qu'on les trouve en philosophie de l'esprit. Nombre de philosophes de l'esprit travaillent désormais de concert avec les autres sciences cognitives, refusant l'isolement et l'anti-psychologisme, des principaux mouvements philosophiques du XXe siècle qu'étaient la philosophie analytique et la phénoménologie. En effet, bien que ces dernières soient encore pratiquées et enseignées, les sciences cognitives se sont imposées avec leur bagage d'investigations empiriques et expérimentales à nombre de philosophes de l'esprit. Aussi, au sein de ces philosophes, les approches cohérentistes font leur chemin et permettent si ce n'est de résoudre, au moins de donner un éclairage nouveau et progresser sur des problèmes philosophiques anciens.

Pour autant, la philosophie de l'esprit diffère des autres sciences cognitives en ce qu'elle n'est pas simplement descriptive (comment pense-t-on ?) mais également normative (comment doit-on penser ?[4]). Au centre de cette préoccupation, il y a la notion de justification[5] : est-ce justifié de croire ce que l'on croit ? et comment peut-on justifier l'acquisition de nouvelles croyances ? Pour beaucoup de philosophes, la justification consiste à isoler un ensemble de croyances indubitables à partir desquelles d'autres croyances pourront être inférées. Deux sources de certitude ont été explorées jusqu'alors : les rationalistes comme Platon et Descartes ont mis de l'avant la raison, tandis que les empiristes comme Locke, Berkeley et Hume ont préféré l'expérience empirique sensorielle comme fondation du savoir. Aujourd'hui, il est généralement reconnu que ces deux approches fondatrices sont erronées. Il n'y a pas de vérité de la raison indubitable (ou s'il y en a, elles ne sont pas suffisantes pour rendre compte du reste de ce que l'on croit savoir), de même qu'il n'y a pas de vérité indubitable de l'expérience sensorielle. L'échec de l'épistémologie fondationnaliste a conduit

[4] Cette question se raffine pour l'intelligence artificielle : Comment un agent doit raisonner ?

[5] Notons que ce sujet est d'intérêt pour ce qui est de la représentation des connaissances et de la formalisation du raisonnement qui sont les éléments centraux des études en intelligence artificielle.

de nombreux philosophes (parmi lesquels : Hegel [1807], Bradley [1914], Bosanquet [1920], Neurath [1959], Quine [1963], BonJour [1985], Harman [1986] et Thagard [2000b]) à rendre compte de la notion de justification en terme de cohérence. Selon ceux-ci, une croyance n'est pas justifiée comme dérivant de vérités indubitables, mais plutôt parce qu'elle est cohérente avec d'autres croyances qui se supportent les unes les autres. La cognition consiste alors à effectuer des ajustements jusqu'à ce qu'un certain équilibre réflectif (selon l'expression de Rawls [1971]), du type de celui proposé par la cohérence cognitive, soit atteint.

La justification cohérentiste, formulée en terme de satisfaction de contraintes permet de rendre compte de nombreuses dimensions psychologiques et philosophiques de manière unifée :

- la justification logique et l'inférence inductive et déductive [Goodman, 1965] ;
- la justification des principes éthiques [Rawls, 1971] ;
- la délibération et le raisonnement fin-moyens [Hurley, 1989; Thagard et Millgram, 1995] ;
- la formation d'impression [Read et Marcus-Newhall, 1993] ;
- la perception : la vision stéréoscopique [Marr et Poggio, 1976; Feldman, 1981], la reconnaissance de formes [McClelland et Rumelhart, 1981] ;
- la théorie de la vérité et de la connaissance [Davidson, 1986] ;
- la justification épistémique [Haack, 1993; Lehrer, 1990] ;
- la justification légale [Raz, 1992] ;
- la justification éthique [Rawls, 1971] ;
- le raisonnement mathématique [Kitcher, 1983] ;
- le raisonnement analogique [Holyoak et Thagard, 1989].

Dans son ouvrage « Coherence in Thought and Action », Paul Thagard [Thagard, 2000a] fait systématiquement référence aux résultats de psychologie, de neurosciences, de linguistique et d'intelligence artificielle. Il y définit cinq types de cohérences qui permettent de résoudre tous les problèmes ci-dessus lorsqu'ils sont formulés comme des problèmes de cohérence. C'est sur cette dernière approche que nous nous sommes basés pour fonder notre théorie de la pragmatique des communications entre agents cognitifs. Paul Thagard, est un philosophe computationnel de l'esprit et dirige le département de sciences cognitives de

l'Université de Waterloo au Canada. La section suivante présente brièvement les réponses aux différentes objections que l'on pourrait formuler à l'égard d'une approche cohérentiste comme la nôtre.

8.3.2 Objections classiques aux approches cohérentistes

Dans cette section, on présente brièvement les principales objections théoriques à l'encontre des approches cohérentistes en discutant chacune d'elles.

Sous spécification

Les approches cohérentistes sont parfois qualifiées de vagues, sous spécifiées. Dans notre cas, il n'en est rien puisque qu'en indiquant clairement le type des éléments et des contraintes en présence (comme il est traditionnel de le faire en représentation des connaissances et formalisation du raisonnement) et en fournissant un algorithme de calcul de la cohérence on parvient à opérationnaliser complètement le concept de cohérence[6].

Indiscrimination

Une autre critique commune est celle de l'indiscrimination, qui considère que les approches cohérentistes ne permettent pas de discriminer les éléments selon différents degrés d'importance ou de crédibilité. Par exemple, les croyances ou cognitions issues de la perception directe doivent généralement être traitées comme étant plus importantes que les autres croyances (acquises indirectement par raisonnement ou communication) dans les déductions. Dans notre approche, les résistances aux changements des cognitions permettent de discriminer les différents éléments. Dans ce cas, une cognition issue de la perception directe se verra attribuer une résistance au changement plus importante (pour rendre compte de la fiabilité de ce mode d'établissement de l'acceptation, le cas échéant).

En outre, le fait de favoriser certains éléments n'est pas une garantie qu'ils soient finalement acceptés, ce sans quoi notre approche ne serait plus cohérentiste, mais fondationnaliste et ces éléments nécessairement acceptés seraient alors les fondations du système (comme

[6] C'est l'un des avantages méthodologiques inhérents à l'informatique : un algorithme, pour être bien défini, ne peut être sous spécifié.

c'est le cas en logique avec les axiomes ou les hypothèses). Une certaine flexibilité et relativité sont donc prises en compte dans l'inférence cohérentiste.

Circularité

La circularité est une autre des objections classiques des approches cohérentistes. Les logiciens en particulier mettent en garde contre le caractère fallacieux des raisonnements circulaires du type : *dieu existe parce c'est écrit dans la bible et que l'on peut faire confiance à la bible puisqu'elle est d'inspiration divine*. Aussi, le raisonnement cohérentiste dans lequel les éléments se supportent les uns les autres de manière symétrique peut sembler circulaire à première vue. Pourtant, les algorithmes introduits par Thagard et Verbeurgt [1998] et développés dans cette thèse montrent que l'inférence cohérentiste est bien différente de l'inférence déductive logique traditionnelle avec laquelle les propositions sont dérivées les unes des autres de manière linéaire à partir des axiomes et autres hypothèses ou encore de la déduction probabiliste utilisant les probabilités conditionnelles [Thagard, 2000b]. L'inférence cohérentiste est issue d'un calcul global pour lequel la notion même de circularité ne s'applique pas. En effet, les liens bidirectionnels capturés par des contraintes ne sont pas des implications logiques ou probabilistes qui s'enchaînent, mais simplement des inter-dépendances que l'agent essaye de satisfaire au mieux.

Vérité

Le raisonnement cohérentiste n'est donc pas circulaire, mais on peut se demander s'il est juste ? Autrement dit, est-ce que les inférences cohérentistes sont justes ? Il y a deux grandes familles de théories de la vérité. Les idéalistes, rejetant l'utilité de considérer l'existence d'une réalité extérieure à la pensé ont introduit une notion purement cohérentiste de la vérité qui revient à considérer que la vérité d'une proposition se réduit à sa cohérence avec les autres propositions considérées. La seconde famille de théories de la vérité, moins radicale, considère la vérité comme une correspondance entre les propositions et une réalité objective indépendante des propositions. Comme l'argumente Thagard [2000b], ces deux familles peuvent être capturées par l'inférence cohérentiste selon que l'on utilise des éléments discriminatoires ou non. Par exemple, la cohérence explicatoire de Thagard donne priorité (sans en garantir complètement l'acceptation) aux éléments issus de l'observation empirique. De la même manière, dans le cadre de notre validation informatique (chapitre 7) différents degrés de priorité sont capturés grace aux résistances aux changements et ainsi priorité est donnée tantôt aux intentions individuelles échouées ou sociales issues de la délibération, tantôt aux engagements issus des dialogues.

8.3.3 Les approches hybrides symboliques-connexionistes

L'incarnation formelle des approches cohérentistes, discutées et décrites ci-haut, sont les approches hybrides symboliques-connexionistes. Si celles-ci ne se limitent pas aux approches cohérentistes, elles leur sont particulièrement adaptées. Dans cette section, on introduit brièvement ces approches de manière générale afin de pouvoir situer la notre dans ce champ formel, considéré comme le plus prometteur au sein des sciences cognitives pour la représentation des connaissances et la formalisation du raisonnement [Varela, 1996; Vignaux, 1991].

Depuis un peu plus d'une décennie, de nombreuses recherches pour intégrer les traitements connexionistes et symboliques ont amenées à un consensus sur le caractère prometteur de ces modèles qui s'orientent vers des architectures plus robustes, plus puissantes et plus souples pour la modélisation cognitive ou les systèmes intelligents [Sun, 1997]. Ces avancées théoriques, se sont concrétisées par un nombre croissant d'applications dans des domaines aussi variés que l'interprétation du langage naturel parlé, la robotique, le diagnostic médical et les applications financières.

Du point de vue des sciences cognitives (dont l'intelligence artificielle est partie prenante) et des neurosciences, un connexionnisme pur dont une interprétation symbolique est possible semble la voie la plus prometteuse et réaliste qui soit à ce moment. En effet, force est de constater que le cerveau (dont le fonctionnement exact reste par ailleurs encore grandement méconnu) permet le traitement symbolique via le substrat neuronal, de nature connexionniste. Ainsi, l'hétérogénéité des processus cognitifs, dont certains sont symboliques (mémorisation, traitement du langage naturel, raisonnement explicite, ...) et d'autres pas (vision, traitement du signal acoustique, intégration perceptuelle, ...) invite à utiliser des structures elles-aussi hétérogènes. L'idée des approches hybrides symboliques connexionistes est d'unifier ces processus en terme du type d'outils – « connexionnistes » – mis en jeu.

Les approches connexionnistes regroupent l'ensemble des approches qui s'articulent autour de représentations connexionnistes, c'est-à-dire qui mettent l'accent sur le réseau formé par des éléments et leurs connexions, dont l'exemple typique est celui des réseaux de neurones artificiels. Aussi, si la supériorité des approches connexionistes n'est plus discutée pour ce qui est des traitements non explicitement symboliques, il restait à montrer que les traitements symboliques peuvent être pris en compte par les approches connexionistes. Ce processus est en marche et de nombreux apports couvrant les domaines « traditionnellement » traités par les approches symboliques ont été proposés. Ainsi, de nombreux modèles de la causalité et de l'inférence ont été proposés [Sun et Browne, 2001]. Les problèmes liés à la quantification, à l'unification et à l'instantiation de variables ont été résolus (dans leur état de l'art, Sun et Browne [2001] présentent les différentes solutions proposées à cet égard). À titre

d'exemple, Sun [1994] a formellement démontré l'équivalence d'une version simplifiée[7] de son modèle connexioniste FEL [Fuzzy Evidential Logic] avec la logique des clauses de Horn et avec CT, la théorie causale de Shoham [1987, 1990b], basée sur une logique modale temporelle avec procédure d'inférence implémentable. Ainsi, Sun établit clairement le lien entre le raisonnement dans les réseaux connexionistes et les systèmes à base de règles. Il montre ensuite comment les modèles connexionistes vont plus loin que les approches logiques classiques.

En effet, du point de vue des systèmes à base de connaissance en intelligence artificielle, les formalismes hybrides symboliques connexionistes ont montré de nombreux avantages sur leurs prédécesseurs purement symboliques. La double nature de ces approches permet l'intégration de qualités complémentaires des deux facettes impliquées. Le caractère symbolique de ces approches permet de conserver les avantages des approches symboliques : interprétation simple, contrôle explicite, encodage initial rapide, possibilités de quantification et d'instanciation de variables dynamiques et abstraction des connaissances. À ceux-ci viennent s'ajouter les avantages des approches connexionistes, qui permettent de prendre en compte de nombreuses caractéristiques du raisonnement de sens commun, souvent ignorées par les modèles symboliques classiques : la gradualité des concepts, les connexions causales partielles, la gestion de l'incertain, l'intégration de l'apprentissage facilitée, la robustesse et la tolérance aux interférences et aux erreurs, l'adaptativité, ...

On note que les modèles connexionistes sont plus simplement implémentables que les modèles probabilistes ou logiques, par ailleurs peu plausibles au niveau des mécanismes cognitifs impliqués. En outre, les algorithmes connexionistes permettent de distribuer le calcul généralement massivement parallèle et d'aboutir à des implantations efficaces pour des problèmes qui sont généralement dans des classes de complexité qui requièrent ce type d'optimisation. Plus généralement, ils offrent une nouvelle perspective sur le raisonnement comme processus dynamique, complexe et continue.

De nombreuses classifications de ces approches ont été proposées et nous ne les détaillerons pas ici. Nous nous contenterons d'introduire les éléments nécessaires pour positionner notre approche du traitement des aspects cognitifs de la pragmatique des communications entre agents dans ce champ dynamique et prometteur. On distingue généralement quatre grandes classes de systèmes qui couvrent le continuum du tout symbolique au tout connexioniste pour la modélisation du traitement symbolique :

1. *les approches symboliques* : les approches symboliques telles que la logique sont celles qui ont été classiquement développées dans le domaine de la représentation des connais-

[7] Les éléments sont restreints à des valeurs d'activation bipolaires pour simuler la valuation sémantique logique.

sances et de la formalisation du raisonnement. Le paradigme de celles-ci est le système à base de règles que l'on retrouve dans la plupart des disciplines concernées par la problématique du traitement symbolique. L'architecture BDI classique, présentée à la section 7.2, est un exemple typique d'approche symbolique ;

2. *les approches par transformation* : les approches par transformations visent, comme leur nom l'indique, à transformer une représentation symbolique en représentation connexioniste et vice-versa. Un algorithme de transformation des éléments de cognitions présents dans l'architecture BDI classique en un réseau d'éléments et de contraintes adapté à notre modèle tomberait dans cette catégorie ;

3. *les approches connexionnistes unifiées* : les approches connexionistes unifiées (parfois qualifiées d'approches neuronales unifiées) regroupent l'ensemble des approches connexionnistes dont une interprétation symbolique est possible. Dans cette catégorie, on distingue généralement les architectures neuronales localistes qui utilisent un élément connexionniste pour chaque concept, des architectures neuronales distribuées dans lesquelles un concept est représenté par un ensemble d'éléments connexionnistes. Cet ensemble est éventuellement variable et ses éléments ne sont pas nécessairement distincts de ceux utilisés pour d'autres concepts. Notre approche du traitement des aspects cognitifs de la pragmatique tombe dans la catégorie des architectures connexionistes localistes ;

4. *les approches hybrides* : les approches hybrides sont des approches modulaires mêlant des modules de traitement utilisant les approches symboliques et des modules utilisant des traitements connexionistes. L'architecture d'agent proposée dans le cadre de notre validation informatique (présentés au chapitre 7) entre dans cette catégorie puisqu'elle mêle les modules de l'architecture BDI sous-jacente (approche symbolique) et notre module de raisonnement pragmatique (approche connexioniste).

La section suivante, présente les principales perspectives ouvertes par les contributions introduites aux chapitres 6 et 7.

8.4 Perspectives

L'architecture de communication entre agents et le modèle de la pragmatique présentés dans cette thèse offrent de nombreuses perspectives. Cette section détaille celles qui nous semblent les plus prometteuses du point de vue des systèmes multi-agents.

8.4.1 Extension de la puissance et de l'expressivité du modèle

Afin d'accroître la puissance expressive du formalisme sous-jacent à notre approche, nous devrons étendre le formalisme actuel au premier ordre (variables, quantification) avec les techniques disponibles pour ce faire dans le champ des formalismes hybrides symboliques-connexionistes [Sun et Browne, 1999]. On pourra également étendre la nature des contraintes, puisque c'est la notion de satisfaction (de contraintes) qui est centrale au modèle. Ainsi des contraintes n-aires pourront être considérées. Ces aspects techniques de notre travail ne devront pas être négligés puisqu'ils permettront d'en élargir l'applicabilité. Notons que notre implantation et validation, présentée au chapitre 7, démontre une bonne applicabilité de notre approche si on la compare à d'autres approches moins procédurales (pour lesquels aucun algorithme n'est fourni) et plus descriptives comme les approches logiques qui, lorsqu'elles ont une bonne expressivité, restent généralement très éloignées d'une éventuelle implémentation.

8.4.2 Une architecture d'agent cohérentiste

Dans cette thèse, nous avons présenté notre approche des aspects cognitifs de la communication entre agents comme un module à intégrer aux modèles d'agents existants. En particulier, nous avons montré comment celle-ci peut s'intégrer à l'architecture BDI. Cependant, une de nos perspectives à long terme est de développer une architecture entièrement cohérentiste. Celle-ci serait dirigée par des principes motivationnels semblables à ceux présentés dans cette thèse et intégrerait les différentes avancés pratiques et théoriques décrites précédemment (section 8.3). Nos propres travaux passés[8], ainsi que les éléments déjà existants dans le champs des approches hybrides symboliques-connexionistes, nous porte à croire en la faisabilité d'un tel projet. Notons que les formalismes hybrides symboliques-connexionistes sont peu utilisés par la communauté de recherche sur les systèmes multi-agents. Une partie de notre travail sera alors de convaincre celle-ci de leur utilité pour le domaine.

8.4.3 Apprentissage de la communication

De nombreux aspects de notre théorie restent sous-exploités par les travaux réalisés jusqu'ici. En particulier, la notion d'utilité du dialogue a été introduite et de nombreuses ex-

[8] Avec Rivaland, nous avons développé et implanté une architecture de gestion des croyances reposant sur le modèle cohérentiste de Shultz et al. [2001]. Cette architecture a été étendue pour rendre compte des dialogues d'investigation [Pasquier et Rivaland, 2002]. Il s'agit d'une architecture purement connexioniste et dont les aspects motivationnels sont strictement cohérentistes.

ploitations de celle-ci restent à explorer. Au sein des communautés de recherche sur l'apprentissage artificiel et sur les systèmes multi-agents, peu de travaux se sont intéressés aux problématiques de l'apprentissage de l'usage de la communication[9]. Un état de l'art de ces tentatives est présenté dans [Weiss et Co, 2001]. Dans cette perspective, les mesures d'utilité des dialogues tenus fournissent une information précieuse, susceptible d'être utilisée pour nourrir : un modèle des accointances (ou relations sociales), une théorie de l'esprit (ou modèle des autres), un modèle des sanctions sociales, un modèle de la confiance et de la réputation, ... L'utilité, positive ou négative, d'un dialogue peut être envisagée comme une récompense et des algorithmes d'apprentissage automatique modernes comme ceux développés dans le cadre de l'apprentissage par renforcement [Sutton et Barto, 1998] pourraient être appliqués aux problématiques de l'apprentissage de la communication ou de l'apprentissage via la communication.

Dans notre approche, un certain nombre d'hypothèses rendent l'effet des actes de langage *déterministe* et publiquement établis. C'est d'ailleurs ce qui permet de progresser sur le front de la vérifiabilité. En effet, avec les approches mentalistes, ce déterminisme était obtenu à l'aide des hypothèses de sincérité et de coopérativité (bénévolat). Pour autant, pour un agent donné, les jeux de dialogue (qui résultent en une série d'actes de langage), envisagés comme des actions communes, sont non-déterministes[10]. En effet, un agent qui initie un jeu de dialogue dans l'espoir de réaliser une modification de la couche sociale des engagements ne sait pas, au moment où le jeu est ouvert, si cette tentative va être un succès (extra-dialogique, voir section 5.5.1) ou non.

Dans ce cadre, l'agent prend ses décisions en fonction de la valeur des actions entreprises, c'est-à-dire en utilisant la fonction d'utilité espérée définie sections 6.6.7 et 7.6. Le qualificatif « espéré » renvoie ici à : dans le cas où le dialogue, comme tentative de modification de la couche sociale, serait un succès (extra-dialogique) et non à l'espérance probabiliste, habituellement associée aux mesures d'utilité dans la théorie de la décision classique [Kast, 1993]. À cet égard, les probabilités n'interviennent pas dans notre formalisation car nous nous situons sous une hypothèse d'équiprobabilité des alternatives. Cette hypothèse capture l'ignorance des agents avec la notion d'équiprobabilité et représente bien leur incertitude.

L'introduction des probabilités et de leur apprentissage dans notre formalisme est envisagée comme l'une de nos perspectives principales. La prise en compte des probabilités permet de prendre en compte les chances de réalisation des alternatives pour pondérer la valeur de celles-ci. Cette approche, qui est l'approche classique en théorie de la décision, est tout à fait commune dans les systèmes multi-agents modernes [Parsons et Wooldridge, 2002]. De

[9] À ne pas confondre avec la construction et l'apprentissage de la communication au sens de l'établissement d'un vocabulaire et d'une grammaire commune, comme d'autres l'étudie Steels et al. [2002].

[10] Quoique les relations d'autorité peuvent les rendre déterministes, comme indiqué section 5.4.6.)

nombreux outils formels permettant d'acquérir et d'exploiter ces probabilités sont déjà disponibles à cet égard. Pour autant, très peu d'applications concernent la prise de décision quant au comportement communicationnel des agents. En outre, le traitement stochastique vient avec ses limites. En particulier, il reste à savoir dans quelles conditions les hypothèses de stationnarité, généralement nécessaires à la convergence des algorithmes proposés, sont vérifiées dans le contexte de la communication entre agents cognitifs.

De nombreux formalismes de traitement de l'incertitude, autres que les approches stochastiques, sont également envisageables. Nous tirerons parti, au passage, de l'une de nos études passées qui traite des inter-relations de la résolution de conflit (notion proche de celle d'incohérence) et du traitement formel de l'incertitude en intelligence artificielle [Pasquier, 2000].

8.4.4 Simulation sociale

Sur la base du simulateur et de l'architecture agent présentés chapitres 5 et 7, de nombreuses extensions et raffinements sont envisageables à des fins de simulation sociale[11]. Dans un souci de transfert de connaissance similaire à celui qui nous anime dans le reste de nos contributions, des éléments de psychologie sociale nous semblent importants à modéliser :

1. Le paradigme de justification insuffisante, présenté section 8.2, sera utilisé pour raffiner le modèle de changement d'attitude ;
2. Les recherches sur la persuasion [Petty et Cacioppo, 1996] pour ce qui est de la prise en compte de la source, et du contenu du message ainsi que pour le choix du destinataire ;
3. La théorie des influences sociales pour ce qui est des aspects sociaux comme la réputation, la hiérarchie, la confiance, l'appréciation.

8.4.5 Explication, justification et argumentation

Les jeux de dialogue, issus des systèmes dialectiques présentés section 2.3.5, sont particulièrement adaptés pour traiter des dialogues d'explication, de justification et d'argumentation. Aussi, il est possible d'étendre les jeux de dialogue de DIAGAL avec des jeux spécifiquement dédiés aux dialogues argumentatifs. Dans cette thèse, nous nous sommes contenté de donner

[11] On pourra consulter l'état de l'art inclut dans la thèse de Amblard [2003] pour un survol de la simulation sociale informatique.

(au chapitre 5) un ensemble de jeux complet et adéquat par rapport au modèle de l'engagement social que nous souhaitions couvrir. Là encore, ce ne sont pas les aspects syntaxiques ou sémantiques des dialogues argumentatifs qui nous intéressent (de nombreuses contributions valables ont été faites à cet égard, aussi bien en dialectique formelle que dans le domaine des systèmes multi-agents). Ce sont de nouveau les aspects cognitifs de ces dialogues qui nous semblent avoir été largement négligés dans les formalismes rencontrés. Pourquoi argumenter (cherche à persuader, expliquer, justifier) ? Quand argumenter ? Avec qui ? A quelle fin ? Quels sont les processus cognitifs mis en jeu et comment les modéliser ? Quand une argumentation mène-t-elle à une persuasion, c'est-à-dire à un changement d'attitude ?

Aussi, l'un de nos objectifs à moyen terme est d'étendre notre approche des aspects cognitifs de la pragmatique pour considérer l'argumentation. En effet, l'argumentation (comme la justification ou l'explication) peut être intégrée dans notre approche comme propagation de contraintes cognitives ou sociales. Dans ce cadre, indiquer aux autres agents quelles sont les contraintes cognitives ou sociales qui sont en jeux dans son propre calcul de cohérence permet à un agent d'expliquer aux autres en quoi les actes posés sont cohérents (justification et explication) autant que de participer à les convaincre que son point de vue est plus cohérent que le leur (persuasion et argumentation). En utilisant la propagation de contraintes comme mécanisme d'explication, de justification et d'argumentation, on autorise ainsi les agents à prendre en compte dans leur raisonnement les contraintes cognitives ou sociales des autres, ce qui permet une coopération informée (en contraste avec une coopération de principe). Notons que le modèle proposé dans cette thèse ne prend pas en compte ce type de coopération dans le raisonnement pragmatique mais que notre validation le fait via le modèle BDI sous-jacent qui maintient (éventuellement) un modèle des autres et en tient compte dans la délibération.

Notons que du point de vue technique, Jung et al. [2001] ont déjà progressé sur le thème de l'argumentation comme propagation de contraintes entre agents réactifs. Leur formalisme pourrait être étendu aux agents cognitifs, de nouveaux résultats pourraient alors venir enrichir les leurs. Les sections suivantes présentent les travaux connexes à notre approche.

8.5 Travaux connexes

Comme nous l'avons souligné au chapitre 3, bien peu de travaux théoriques se sont intéressés aux aspects cognitifs de la pragmatique des communications en intelligence artificielle et dans la communauté multi-agent aussi bien que dans les domaines qui leur servent habituellement de fondations. Nous avons déjà mentionné les approches intentionnelles qui sont d'importance historique indéniable. De fait, comme ces approches ne sont pas adaptées aux

cadres de communication conventionnels reposant sur les engagements sociaux et les jeux de dialogue, une approche alternative était nécessaire. Cette alternative, nous l'avons vue, en dépasse grandement le champ d'application puisque de nouvelles dimensions de la pragmatique sont prises en compte.

Pour autant, plusieurs autres approches connexes à la nôtre ont été produites, quoique moins spécifiques aux systèmes multi-agents et pas forcément fondées sur des résultats de sciences cognitives comme la nôtre. En effet, comme nous l'avons mentionné à la section 6.6.4 d'autres chercheurs ont déjà fondé l'émergence des dialogues sur la résolution de conflit, une notion très proche de celle d'incohérence. Par exemple, Dessalles [1998b][12], dans son étude des contraintes logiques qui pèsent sur les conversations en langage naturel [Dessalles, 1998a], spécule que les conflits cognitifs entre les désirs ou croyances des agents permettent de rendre compte de l'émergence des conversations. Deux mécanismes simples sont proposés : l'abduction[13] et la propagation de coefficients de nécessité. Lorsqu'un conflit est identifié, les participants essaient de faire une abduction du terme le plus faible (en termes de son coefficient de nécessité). Si cela échoue, ils affirment la fausseté de ce dernier. Cela amène le conflit à se déplacer sur d'autres termes. Le nouveau conflit est traité de la même manière jusqu'à ce qu'une solution soit trouvée ou jusqu'à ce que les participants soient incapables de faire de nouvelles abductions. Notre approche, sans réfuter ce travail, l'étend en indiquant que le-dit conflit (l'incohérence dans notre cas) n'est pas nécessairement externe mais peut également être interne.

Dans la même veine, mais dans la communauté agent cette fois-ci, citons également les travaux de Fiorino [1998]; Fiorino et Maille [1998] qui ont fourni un modèle logique de l'établissement de conjecture commune reposant sur la résolution de conflit et dont notre approche pourrait fournir les fondements théoriques[14].

L'architecture BOID [Beliefs, Obligations, Intentions and Desirs] [Broersen et al., 2001], quoique que de portée plus générale que l'étude de la communication, est conceptuellement très proche de ce qu'une architecture entièrement cohérentiste, caractérisée de manière logique, serait. Les cognitions sont représentées par des règles de production logique. Dans BOID, quatre classes de cognitions sont considérées : les croyances (B), les obligations (O), les intentions (I), et les désirs (D). Ces quatres classes sont ordonnées (selon un ordre total) de sorte que : si la classe X est préférée à la classe Y, alors toutes les cognitions de la

[12] Notons que Dessalles [1993] s'est intéressé dès sa thèse de doctorat à la modélisation des aspects cognitifs des conversations spontanées.

[13] L'opération d'abduction désigne l'inférence d'hypothèses à partir d'une évidence, de raisons à partir d'un fait. L'abduction est souvent appelée « modus ponens inverse » et correspond (en logique classique) à la règle d'inférence suivante : $\frac{A \rightarrow B, B}{A}$.

[14] C'est ce que nous en a dit Humbert Fiorino lorsqu'on l'a rencontré lors de l'édition 2002 des journées francophones d'intelligence artificielle distribuée et des systèmes multi-agents (JFIADSMA'02).

classe X seront prioritaires sur celles de la classe Y. Dans notre approche, les résistances aux changements, affectés indépendamment à chaque cognition permettent le même type de caractérisation puisque la priorité sera donnée aux engagements sociaux ou aux intentions individuelles selon les résistances au changement qui leur auront été attribués. Un travail technique visant à montrer que l'architecture BOID est un cas particulier d'une architecture cohérentiste est envisagé comme l'une de nos perspectives.

Pour être complet, il nous faut mentionner quelques travaux à caractère plus technique (qui ne reposent pas sur des fondements théoriques clairs, issues des sciences cognitives), comme les expériences réalisées par Excelente-Toledo et al. [2001] et utilisant la théorie de la décision pour raisonner sur les engagements et les sanctions. Notons également que la théorie de la dissonance cognitive à déjà été introduite en intelligence artificielle, dans le cadre d'un système de maintien de la cohérence pour les bases de connaissances logiques (TMS, Truth Maintenance System) [Schwartz, 2001].

8.5.1 Utilisation de notre cadre

Du fait de sa généricité et de sa large couverture, notre approche est un bon candidat pour servir comme fondation d'autres systèmes. C'est dans cet esprit que Sansonnet et Valencia [2003b] de l'équipe du LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur de l'université d'Orsay, en France) ont étendus nos idées, reformulées dans un cadre logique du premier ordre implémantable et implémenté, pour automatiser le comportement communicationnel d'agents purement informationnels, c'est-à-dire non orientés tâche. Les auteurs ont également appliqué cette approche au domaine de la simulation sociale en considérant – en outre de l'utilisation de la dissonance cognitive comme principe motivationnel premier – les notions de pertinence et de confiance [Sansonnet et Valencia, 2003a].

Une autre application de notre approche est en cours de développement au GRAAL (Groupe Raisonnement, Action et Actes de Langage), un laboratoire de linguistique computationnelle attaché à l'IRIT (Institut de Recherche en Informatique de Toulouse, en France) et à l'ENSEEIHT (École Nationale Supérieure d'Électrotechnique, d'Électronique, d'Informatique, d'Hydraulique et des Télécommunications). Des chercheurs et étudiants y développent leur propre simulateur de jeux de dialogue adapté au traitement du langage naturel [Adam, 2003].

Nous renvoyons le lecteur intéressé par ces applications aux lectures indiquées ci-haut. La section suivante inscrit cette approche dans le champ plus général de la modélisation dialogique.

8.6 Discussion : la modélisation de l'activité dialogique

De par notre problématique, nous avons souhaité dépasser les aspects syntaxiques et sémantiques de la communication et mettre de l'avant l'usage de la communication. Ce que les linguistes nomment l'*interprétation pragmatique* est un processus qui n'est pas en lien avec le langage en particulier, mais plutôt avec la proactivité des agents cognitifs. Lorsqu'un agent cognitif rationnel agit, il a des raisons pour cela (par définition du terme rationnel). Fournir une interprétation pragmatique d'une action consiste à en déterminer les raisons.

L'école classique en intelligence artificielle, inspirée de la philosophie analytique de l'esprit, consiste à remonter des actions aux intentions puis aux désirs et croyances de l'agent (ou de son interlocuteur) pour expliquer les actions de l'agent, en particulier les actions communicatives. Les cognitions volitives premières sont donc les désirs des agents. Notons que la formalisation de la notion de désir avec les outils logiques habituels est difficile (voir à ce propos [Dignum et al., 2002]). Généralement, les désirs sont donnés à la conception plutôt que générés dynamiquement. Dans le cas des actions communicatives, notre approche, qui se distingue par l'utilisation des concepts et outils cohérentistes permet de compléter à ce niveau les analyses des philosophes avec des résultats de psychologie cognitive pour aboutir à un modèle de plus grande couverture, plus complet, plus générique. En particulier, et c'est ce que nous avons souhaité montrer dans cette thèse, la cohérence cognitive *motive* la communication comme elle en motive les effets (via le changement d'attitude). La cohérence cognitive est un moteur motivationnel[15] qui rend compte de la nécessite comme de la dynamique de la communication. En outre, via le changement d'attitude, l'aspect persuasif inhérent à toute communication est capturé.

La très grande quantité d'éléments du contexte susceptibles d'être inclus dans l'interprétation pragmatique est déroutante. Une singularité de l'interprétation pragmatique sur laquelle insiste les philosophes du langage [Recanati, 2001] est son caractère révisable (sa non-monotonie). La meilleure explication offerte à un moment donné peut être révisée lorsque de nouveaux éléments du contexte sont découverts. Aussi, dans notre modèle cohérentiste, le plus cohérent à un moment donné ne l'est plus forcément à un autre.

Comme Krauss et Morsella [2000], nous distinguons quatre grands paradigmes de modélisation du dialogue : encodage-décodage, le paradigme intentionnaliste, le paradigme de

[15] Dans les approches intentionnelles, on comprend que les intentions découlent d'une sélection des désirs par raisonnement sur les croyances (le processus de délibération), tout comme on comprend que les croyances sont issues du raisonnement sur les connaissances et les perceptions. Par contre, aucun mécanisme concernant la génération des désirs n'a été exhibé, de sorte que l'on a l'impression qu'il manque une composante motivationnelle de type homéostatique (comme la cohérence cognitive) à ce type de modèle. Les désirs sont supposés pré-exister et sont généralement incarnés par les buts de l'agent que son concepteur lui délègue.

prise de perspective et le paradigme dialogique. Par contre, et contrairement à Krauss et Morcela, plutôt que d'opposer ces modèles, nous souhaitons mettre en évidence leur complémentarité.

L'approche par *encodage-décodage* est la plus directe. La communication y est envisagée comme un transfert d'information par l'intermédiaire de codes. Le sens des codes doit être partagé et dans les cas les plus simples, il y a une correspondance un-à-un entre les codes et leurs sens. Si ce paradigme constitue l'approche dominante pour les communications informatiques traditionnelles[16], il ne permet pas de rendre compte de certaines dimensions du langage naturel qui sont pertinentes pour les systèmes multi-agents à base d'agents cognitifs.

Tout d'abord, la distinction entre sens littéral et sens de l'énoncé (notions introduites dans l'annexe A) via la prise en compte du contexte n'est pas considérée par ce paradigme. Par ailleurs, la distinction entre sens de l'énoncé et sens du locuteur ne peut être prise en compte de manière satisfaisante avec cette approche. Aussi un énoncé tel que : « avez-vous l'heure ? » devra être répondue comme une question fermée à laquelle on répond oui ou non (sens de l'énoncé) et non comme une requête exprimant l'intention du locuteur de satisfaire son intention d'avoir l'heure par la reconnaissance de celle-ci (sens du locuteur, en termes Gricéen). Le *paradigme intentionnel* permet de prendre en compte ces distinctions.

Cependant, si les énoncés qui sont compris non-littéralement sont courants dans les dialogues naturels, ils nous semblent moins utiles dans le cas des communications entre agents logiciels. En effet, la non-littéralité relève généralement du style et tout ce qui est dit de manière non littérale peut l'être de manière littérale. Aussi dans les systèmes multi-agents, c'est la tradition encodique qui a été appliquée aux approches intentionnelles. En effet, avec les sémantiques mentalistes, le message véhicule certains états mentaux (par la présence desquels il est déclenché) qui seront décodés par l'interlocuteur de manière non ambiguë (c'est-à-dire, conformément à la sémantique du type de message reçu).

Normalement, l'encodage et le décodage de ces messages fait également intervenir le fond commun (common ground). Puisque le fond commun qu'un locuteur doit considérer varie selon les interlocuteurs, le locuteur doit prendre en compte son interlocuteur dans la communication. Le *paradigme de prise de perspective* vise à tenir compte de cette variabilité en considérant comme l'énonce Brown [1965], que l'encodage effectif requiert que le point de vue de l'auditeur soit pris en compte de manière réaliste. Concrètement, les modèles représentatifs de ce paradigme reposant sur des approches mentalistes sont encore à l'élaboration et de nombreuses difficultés liées aux incertitudes concernant les autres et leurs

[16] À ce niveau, on prendra soin de ne pas confondre le paradigme encodage-décodage pour la communication avec les concepts d'encogage et décodage inhérents à l'informatique et à l'intelligence artificielle et qui sont nécessaires à toute tentative de représentation symbolique des connaissances et de formalisation du raisonnement.

modèles résistent. L'approche par les engagements sociaux (qui représentent une partie du fond commun) est prometteuse, quoique ne rendant pas suffisamment compte du point de vue de l'autre dès lors que les engagements le concernant ne rendent pas forcément compte de manière fidèle de son point de vue.

Finalement, *le paradigme dialogique* se fonde sur des bases tout à fait différentes en considérant la communication comme une activité collective. Avec la notion d'activité commune, le sens n'est plus une propriété des messages. Cela rompt avec les autres paradigmes, tous monologiques, qui conçoivent le sens comme une propriété du message (dans le paradigme encodage-décodage), comme résultant des intentions du locuteur (dans le paradigme intentionnel) ou comme découlant du point de vue de l'autre (selon le paradigme de la prise de perspective). Dans le paradigme dialogique, le sens de la communication est celui de l'activité commune qu'elle constitue. Le sens des énoncés est dit « socialement situé ». Les processus de mise en commun, c'est-à-dire d'établissement, sur lesquels nous avons insisté, permettent d'assurer que les interlocuteurs partagent le sens de chaque énoncé (en terme de leurs conséquences) avant de procéder aux suivants. C'est sur ce paradigme que repose le langage DIAGAL et avec lui, une partie de notre approche.

Aussi, avec le langage DIAGAL (chapitre 5), le sens de la communication est la manipulation des engagements sociaux. Chaque jeu de dialogue étant une activité commune pour l'avancement de l'état de la couche des engagements sociaux. Les engagements sociaux, qui lèvent l'attente d'actions, font le lien entre l'activité dialogique des agents et leurs activités extra-dialogiques. Cependant plutôt que de représenter explicitement ces activités comme c'est le cas avec les réseaux d'engagements ou avec la spécification de protocoles, nous avons repris la formulation mentaliste en l'étendant au raisonnement (cohérentiste) sur la couche sociale. C'est ce raisonnement, impliquant des éléments de cognition individuels et sociaux, qui motive la communication comme il motive l'action.

Finalement, décrit ainsi, on voit bien que ces quatre paradigmes de modélisation de l'activité dialogique sont essentiellement complémentaires et que tous doivent être idéalement couverts. Le chapitre suivant conclut cette thèse en en rappelant les principales contributions.

Chapitre 9

Conclusion

De notre état de l'art des modèles de communication entre agents cognitifs (chapitres 1 et 2), nous avons extrait une problématique originale. En effet, les aspects cognitifs de la pragmatique des communications agents n'ont pas été traités auparavant dans le cas des approches conventionnelles et sociales (chapitre 3). Pour répondre à cette problématique, raffinée en objectifs concrets, nous avons proposé un modèle complet de la communication agent, couvrant les quatre dimensions principales de la communication dialogique : syntaxe, structure, sémantique et pragmatique.

Du point de vue syntaxique, nous avons proposé le langage DIAGAL (chapitre 5). Les jeux de dialogues permettent de capturer les aspects conventionnels de la communication, tandis que le jeu de contextualisation permet de capturer les aspects structurels du dialogue. Au niveau sémantique, nous avons proposé un modèle de l'engagement social flexible et de son respect (chapitre 4). Pour ce qui est des aspects cognitifs de la pragmatique du dialogue entre agents, nous avons mis de l'avant la théorie de la cohérence cognitive (chapitre 6). Cette théorie a été validée informatiquement dans le cadre de l'architecture d'agent BDI (chapitre 7). Nous avons finalement explicité certains des enjeux pratiques et théoriques justifiant l'introduction de cette approche pour les systèmes multi-agents et l'intelligence artificielle et plus généralement pour les sciences cognitives (chapitre 8).

De l'ensemble de nos contributions syntaxiques et sémantiques résulte un modèle en couches. La figure 9.1 synthétise ce modèle, dont les différentes couches sont :

– *Couche attentionnelle et signalétique* : les signaux du jeux de contextualisation, c'est-à-dire les messages de contextualisation, permettent d'établir (*grounding*) les jeux de dialogue ainsi que de définir dynamiquement la structuration du dialogue ;

- *Couche actes de langage* : les messages (contenant des actes de langages) issus du respect des règles des jeux de dialogue ouverts permettent de remplir les engagements dialogiques pendants et font avancer l'état des jeux ouverts ;
- *Couche dialogique* : les jeux de dialogue permettent, lorsqu'ils sont joués avec succès extra-dialogique, de faire avancer l'état de la couche sociale des engagements extra-dialogiques ;
- *Couche sociale* : les engagements extra-dialogiques, s'ils sont respectés, amènent la production d'actions qui font avancer l'état des activités des agents ;
- *Couche activité* : les activités des agents avancent l'état de l'environnement. Elles sont l'objet des systèmes multi-agents et satisfont les agents ou leurs concepteurs.

Au niveau de cette dernière couche, nous avons proposé de considérer cette activité comme une réduction d'incohérence cognitive[1]. En nous inspirant de la psychologie cognitive et dans l'esprit des approches motivationnelles, nous avons caractérisé les activités des agents comme étant des activités cognitives (qui peuvent entraîner comme effet de bord des activités comportementales, sans que cela soit nécessaire puisqu'il existe également de nombreuses activités purement cognitives, c'est-à-dire non orientées tâches). Cette activité cognitive, qui subsume les autres, prend la forme du maintien et du rétablissement (lorsque c'est nécessaire) de la cohérence cognitive. Cette caractérisation des aspects cognitifs de la pragmatique des communication entre agent est indépendante du domaine d'application, ce qui produit une approche générique, c'est-à-dire indépendamment du domaine d'application.

La modélisation que nous proposons nous semble répondre correctement à la problématique développée au chapitre 3 ainsi qu'aux objectifs que nous nous y étions fixés. La théorie de la cohérence cognitive (chapitre 6), que nous proposons pour la pragmatique des communications agents, est épistémologiquement *fondée* sur les travaux de psychologie cognitive, de psychologie sociale et de philosophie computationnelle de l'esprit. Elle arbore la forme, classique en informatique, de raisonnements et de calculs sur des éléments et leurs contraintes associées. Ce faisant, elle constitue une théorie formelle de l'utilisation des cadres interactionnels conventionnels[2] autant qu'une théorie des effets des communications (changement d'attitude). Les mesures d'incohérence et d'utilité définies dans le cadre de cohérence proposé fournissent les mécanismes nécessaires pour répondre même partiellement aux problèmes

[1] D'autres possibilités sont envisageables à ce niveau et nous les avons brièvement présentées au chapitre 5, section 5.4 : construction de protocoles, spécification de l'activité commune en terme de réseau d'engagement, planification de conversations, ...

[2] Les cadres interactionnels considérés sont (1) le cadre interactionnel théorique défini par les types de dialogues isolés par Walton et Krabbe [1995] et (2) les jeux de dialogue : en particulier, mais de manière non restrictive, ceux du langage DIAGAL.

FIG. 9.1 – Modèle en couches de la communication entre agents résultant de nos contributions.

évoqués ci-dessous et qui sont autant de sujets peu traités dans la littérature concernant les systèmes multi-agents :

- Quand un agent prend-il l'initiative d'une conversation, avec qui, à quel sujet et pourquoi (voir sections 6.6.3, 7.7 et 7.11) ?
- Avec quel type de dialogue (voir section 6.6.4 et plus spécifiquement, pour les jeux de dialogue du langage DIAGAL, les chapitres 5 et 7) ?
- Quelle intensité donner aux forces illocutoires des actes de langage utilisés (voir section 6.6.8) ?
- Comment définir et mesurer l'utilité d'une conversation (voir sections 6.6.7 et 7.6) ?
- Quand arrêter le dialogue ou le cas échéant comment le poursuivre (voir sections 6.6.7 et 7.7) ?
- Quels sont les impacts du dialogue sur les attitudes de l'agent (voir sections 6.5, 7.5 et 8.2) ?
- Quels sont les impacts du dialogue sur l'humeur de l'agent (voir section 6.6.8) ?

– Quelles sont les conséquences du dialogue sur les accointances de l'agent (voir sections 6.6.7 et 8.4.3) ?

– Comment modéliser la dimension commune des conversations et capturer la notion de projet conjoint (voir section 6.6.6) ?

Évidemment, chacune de ces problématiques n'a pu être épuisée ici, mais l'objet de cette thèse était plutôt de donner une idée d'ensemble de cette nouvelle approche de la communication agent. En effet, l'un des intérêts de notre apport est la large couverture de la théorie proposée pour l'automatisation de la communication entre agents cognitifs. À notre connaissance, il n'existe pas d'autre approche de couverture comparable dans le champ des systèmes multi-agents. Pourtant, l'essentiel de notre apport n'est peut-être pas tant de répondre aux problématiques sus-nommées que d'y répondre à l'aide d'un seul et même principe : l'homéostasie de la cohérence cognitive.

La pragmatique (au sens de la théorie de l'usage) des cadres interactionnels développés pour les systèmes multi-agents est un aspect trop souvent négligé. Soulignons que ce n'est pas complètement un hasard ou un oubli puisque cet aspect est réputé difficile, et ce, dans tous les domaines d'étude de la communication (linguistique, philosophie du langage, théories de la communication, intelligence artificielle, ...). Selon Lochbaum [1994] : « les interlocuteurs s'engagent dans un dialogue parce qu'ils ont une raison pour cela. ». Notre approche propose un fondement pour les communications entre agents cognitifs en identifiant de manière générique qu'elle pourrait être cette raison : une incohérence cognitive. C'est de ce fondement, crédible au vu des connaissances actuelles en sciences cognitives que l'on a développé notre théorie de l'usage des approches conventionnelles et sociales de la communication entre agents cognitifs. À notre problématique (chapitre 3), on répond donc par une solution théorique générique, simple et élégante, reposant sur des bases théoriques et empiriques solides (philosophie de l'esprit, psychologie sociale, psychologie cognitive, socio-linguistique et intelligence artificielle). C'est donc une *contribution aux sciences cognitives*.

Finalement, le cadre de cohérence proposé fournit un système de valeurs aussi bien individuel que collectif qui nous semble être une étape nécessaire pour développer une plus grande autonomie des agents : le concepteur aura moins à s'occuper de l'utilité et de l'efficacité des comportements des agents dès lors que ceux-ci disposeront d'outils pour la mesurer eux-mêmes. C'est donc aussi un premier pas vers l'automatisation des communications agents et en cela une *contribution à l'informatique* en général et aux aspects théoriques et pratiques des systèmes multi-agents en particulier.

Bibliographie

Adam, C. (2003). Structure et simulation du dialogue. Mémoire de maîtrise, IRIT, Institut de Recherche en Informatique de Toulouse.

Airenti, G., Bara, B., et Colombetti, M. (1993). Conversation and behaviour games in the pragmatics of dialogue. *Cognitive science*, 17(7) :34–49.

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communication of the ACM*, 26 :832–843.

Allen, J. F. et Perrault, C. R. (1980). Analysing intention in dialogues. *Artificial Intelligence*, 15(3) :23–46.

Allen, J. F., Schubert, L. K., Ferguson, G., Heeman, P., Hwang, C. H., Kato, T., Light, M., Martin, N., Miller, B., Poesio, M., et Traum, D. R. (1995). The TRAINS project : A case study in building a conversational planning agent. *Journal of Experimental and Theorical AI*, 7(4) :7–48.

Alliot, J. M. et Schiex, T. (1993). *Intelligence Artificielle et Informatique Théorique*. Cépadues-éditions.

Allwood, J. (1976). *Linguistic communication as action and cooperation*. Thèse de doctorat, University of Goteborg, Sweden.

Allwood, J. (1994). Obligations and options in dialogue. *Think Quaterly*, 3(1) :12–34.

Amblard, F. (2003). *Comprendre le fonctionnement de simulation sociales individus-centrées*. Thèse de doctorat, Université Blaise Pascal, Clermont II, Clermont-Ferrant.

Amgoud, L., Maudet, N., et Parson, S. (2000). Modelling dialogue using argumentation. Dans *Proceedings of the 4th international conference on multi-agent systems (ICMAS'00)*.

Amgoud, L., Maudet, N., et Parsons, S. (2002). An argumentation-based semantics for agent communication languages. Dans *Proceedings of the 15th European Conference on Artificial Intelligence*, Lyon.

Andrillon, N. (2003). Pragmatique de la communication agent. Rapport de stage, Laboratoire DAMAS [Dialogue, Apprentissage et systèmes Multi-AgentS], Département d'Informatique et de Génie Logiciel, Faculté de Sciences et Génie, Université Laval, Québec, Canada.

Aronson, E. (1968). *Theories of Cognitive Consistency : a Sourcebook*, chapitre Dissonance theory : progress and problems. Chicago : rand-McNally.

Artikis, A., Pitt, J., et Sergot, M. (2002). Animated specifications of computational societies. Dans *Proceedings of the First Internationnal Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)*, pages 1053–1061, Bologna, Italy. ACM Press.

Austin, J. L. (1962). *How to Do Things With Words*. Oxford University Press : Oxford, England.

Bach, K. et Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. The MIT Press : Cambridge, MA, USA.

Baker, M. (1991). *Knowledge Acquisition in Physics and Learning Environments*, chapitre An Analysis of Cooperation and Conflict in Students' Collaborative Explanations for Phenomena in Mechanics. Springer-Verlag : Heidelberg, Germany.

Balkanski, C. et Hurault-Plantet, M. (2000). Cooperative requests and replies in a collaborative dialogue model. *Internationnal Journal of Human-Computer Studies*, 53 :915–968.

Barreau, H. (1995). *L'épistémologie*. Que sais-je ? Collection Encyclopédique. Presses Universitaires de France (PUF).

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, 37(7) :122–125.

Beccaria, C. (1963). *On Crimes and Punishments*. New Jersey : Prentice Hall.

Bell, J. (1995). Changing attitudes. Dans Wooldridge, M. et Jennings, N., rédacteurs, *Intelligent Agents : Theories, Architectures, and Languages (ATAL'94)*, volume 890, pages 40–55. Springer-Verlag : Heidelberg, Germany.

Bentahar, J. (2002). Communication et argumentation dans les systèmes multi-agents. Rapport technique, Département d'Informatique et de Génie Logiciel, Faculté de Sciences et Génie, Université Laval.

Bentahar, J., Moulin, B., Meyer, J.-J. C., et Chaib-draa, B. (2004). A logical model for commitment and argument network for agent communication. Dans *Proceedings of 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'04)*, pages 19–23, New York, USA. ACM Press.

Bentham, J. (1970). *An Introduction to the Principles of Morals and Legislation*. London : The Athlone Press.

Bergeron, M. (2005). Spécification, modélisation et analyse du dialogue entre agents par l'intermédiaire des engagements sociaux. Mémoire de maîtrise, Département d'Informatique et de Génie Logiciel, Faculté de Sciences et Génie, Université Laval, Québec, Canada.

Bergeron, M. et Chaib-draa, B. (2004). Les réseaux d'engagements. Dans Boissier, O. et Guessoum, Z., rédacteurs, *Systèmes multi-agents : défis scientifiques et nouveaux usages, actes de la conférence JFSMA'04*, pages 251–265. Hermes-science, Lavoisier.

Bergeron, M. et Chaib-draa, B. (2005). Acl : Specification, design and analysis all based on commitments. Dans *Proceedings of the Workshop on Agent Communication (AC2005), fourth International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS 2005)*, Utrecht, Netherlands.

Boella, G. et Lesmo, L. (2001). *Social Order in Multi-Agent Systems*, chapitre Deliberative normative agents, pages 85–110. Kluwer Academic.

Boissier, O. (2001). *Principes et architecture des systèmes multi-agents*, chapitre Modèles et architectures d'agents, pages 71–99. Hermes sciences publication, Lavoisier, Paris.

BonJour, L. (1985). *The structure of empirical knowledge*. Havard University Press, Cambridge.

Bosanquet, B. (1920). *Implication and linear inference*. MacMillan, London.

Bourget, D. (2002). The dialogue game simulator (dgs). Rapport de stage, Département d'Informatique et de Génie Logiciel, Faculté de Sciences et Génie, Université Laval, Québec, Canada.

Bradley, F. H. (1914). *Essays on truth and reality*. Clarendon Press, Oxford.

Brassac, C. (1994). Speech act and conversation sequencing. *Pragmatics and cognition*, 2(1) :191–205.

Bratman, M. E. (1987). *Intentions, Plans, and Practical Reason*. Harvard University Press : Cambridge, MA.

Bratman, M. E. (1990). What is intention ? Dans Cohen, P. R., Morgan, J. L., et Pollack, M. E., rédacteurs, *Intentions in Communication*, pages 15–32. The MIT Press : Cambridge, MA, USA.

Brehm, J. et Cohen, A. (1962). *Explorations in Cognitive Dissonance*. John Wiley and Sons, inc.

Brehm, S. S. et Brehm, J. W. (1981). *Psychological reactance : a theory of freedom and control*. New york : Academic Press.

Broersen, J., Dastani, M., Hulstijn, J., Huang, Z., et Van der Torre, L. (2001). The BOID architecture : Conflicts between beliefs, obligations, intention and desires. Dans *Proceedings of the fifth International Conference on Autonomous Agent*, pages 9–16, Montréal, Canada. ACM Press.

Brown, R. (1965). *Social Psychology*. The Free Press, New York.

Bruce, B. (1975). Generation as social action. *Theoretical issues in Natural Language Processing*, 1 :64–67.

Bunt, H. (1996). *The Structure of Multimodal Dialogue*, chapitre Dynamic Interpretation and Dialogue Theory. John Benjamin, Amsterdam.

Bunt, H. (2000). *Abduction, Belief and Context in Dialogue : Studies in Computational Pragmatics*, volume 1 de *Series Natural Language Processing*, chapitre Dialogue pragmatics and context specification, pages 81–150. John Benjamins, Amsterdam.

Bunt, H. et Black, W. (2000). *Abduction, Belief and Context in Dialogue : Studies in Computational Pragmatics*, volume 1 de *Series Natural Language Processing*, chapitre The ABC of Computational Pragmatics, pages 1–46. John Benjamins, Amsterdam.

Bylander, E. (1991). Complexity results for planning. Dans *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 274–279, Sydney, Australia.

Caelen, J. (1996). *Nouvelles interfaces homme-machine*, volume 18 de *OFTA, série ARAGO*. Lavoisier, Paris.

Camps, V. (1998). *Vers une théorie de l'auto-organisation dans les systèmes multi-agents basée sur la coopération : application à la recherche d'information dans un système d'information répartie*. Thèse de doctorat, Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France.

Carberry, S. (1990). *Plan Recognition in Natural Language Dialogue*. The MIT Press : Cambridge, MA, USA.

Cartwright, D. et Harary, F. (1956). Structural balance : a generalisation of heider's theory. *Psychological Review*, 63 :277–293.

Castelfranchi, C. (1995). Commitments : from individual intentions to groups and organizations. Dans *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 41–48, San Francisco, CA, USA.

Castelfranchi, C. (1997). Practical 'permission' : Dependence, power, and social commitment. Dans *Proceedings of the Second Workshop on Practical Reasoning and Rationality*. Research Publications, Dept. of Computer Science, Queen Mary and Westfield College, London.

Castelfranchi, C. (2000). Engineering social order. Dans *Engineering Societies in the Agents World (ESAW)*, volume 1972 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 1–18. Springer-Verlag : Heidelberg, Germany.

Castelfranchi, C. (2004). *Realism in action - Essays in the Philosophy of Social Sciences.*, chapitre Grounding We-intentions in Individual Social Attitudes, page in press. Kluwer Academic Press.

Castelfranchi, C., Dignum, F., Jonker, C., et Treur, J. (1999). Deliberative normative agents : Principles and architecture. Dans Jennings, N. R. et Lesperance, Y., rédacteurs, *Intelligent Agents V, Proceedings of the Internationnal Workshop on Agent Theories, Architectures, and Languages (ATAL)*, volume 1757 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 364–378. Springer-Verlag : Heidelberg, Germany.

Chaib-draa, B. et Dignum, F. (2002). Trends in agent communication language. *Computational Intelligence, Special Issue on Agent Communication Language*, 18(2) :89–101.

Chaib-draa, B., Maudet, N., et Labrie, M.-A. (2002). Request for action reconsidered as dialogue games based on commitments. Dans Huget, M.-P., rédacteur, *Internationnal Workshop on Agent Communication Language and Conversation Policies (AAMAS'02)*, volume 2650 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 284–299.

Chaib-draa, B., Maudet, N., et Labrie, M.-A. (2003). DIAGAL, a tool for analyzing and modelling commitment-based dialogues between agents. Dans *Proceedings of Canadian Artificial Intelligence Conference*, volume 2671 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 353–369. Springer-Verlag : Heidelberg, Germany.

Chaib-draa, B. et Vanderveken, D. (1998). Agent communication language : Toards a semantics based on success, satisfaction, and recursion. Dans *Intelligent Agents IV, Proceedings of the Internationnal Workshop on Agent Theories, Architectures, and Languages (ATAL)*, Paris.

Chaib-draa, B. et Vongkasem, L. (2000). ACL as a joint project between participants : A preliminary report. Dans Dignum, F. et Greaves, M., rédacteurs, *Issues in Agent Communication*, number 1916 in Lecture Notes in Artificial Intelligence (LNAI), pages 235–248. Springer-Verlag : Heidelberg, Germany.

Chaignaud, N. et El Fallah-Seghrouchni, A. (2001). Apport de la modélisation cognitive aux langages de communication entre agents. Dans El Fallah-Seghrouchni, A. et Magnin,

L., rédacteurs, *Fondements des systèmes multi-agent : nodèles, spécifications formelles et vérification, Actes des JFIADSMA'01*. Hermes : Lavoisier.

Chicoisne, G. (2002). *Dialogue antre agents naturels et agents artificiel : une application aux communautés virtuelles*. Thèse de doctorat, Institut National Polytechnique de Grenoble, Grenoble, France.

Chopra, A. et Singh, M. P. (2004). Nonmonotonic commitment machines. Dans *Proceedings of the International Workshop on Agent Communication Languages and Conversation Policies (ACL)*, Lecture Notes in Artificial Intelligence (LNAI), pages 183–200. Springer-Verlag : Heidelberg, Germany.

Clark, H. H. (1996). *Using Language*. Cambridge University Press.

Clark, H. H. et Schaeffer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, 2 :19–41.

Clark, H. H. et Wilkes-Gibbs, D. (1986). Refering as a collaborative process. *Cognition*, 1(22) :1–35.

Cohen, P. R. (1996). *Survey of the State of Art in Natural Language Technology*, chapitre Discourse and Dialogue, pages 234–241. Universitat des Saarlandes, Germany.

Cohen, P. R. et Levesque, H. J. (1990a). Intention is choice with commitment. *Artificial Intelligence*, 42 :213–261.

Cohen, P. R. et Levesque, H. J. (1990b). Performatives in a rationally based speech act theory. Dans *28th Annual Meeting of the Association for Computational Linguistics*, pages 79–88, Pittsburg. Springer-Verlag : Heidelberg, Germany.

Cohen, P. R. et Levesque, H. J. (1990c). Rational interaction as the basis for communication. Dans Cohen, P. R., Morgan, J., et Pollack, M. E., rédacteurs, *Intentions in Communication*, pages 221–256. The MIT Press : Cambridge, MA, USA.

Cohen, P. R. et Levesque, H. J. (1991). Teamwork. Technote 504, SRI International, Menlo Park, CA, USA.

Cohen, P. R. et Levesque, H. J. (1995). Communicative actions for artificial agents. Dans *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 65–72, San Francisco, CA, USA.

Cohen, P. R. et Perrault, C. R. (1979). Elements of a plan based theory of speech acts. *Cognitive Science*, 3 :177–212.

Colombetti, M. (1998). Different ways to have something in common. Dans Christiansen, H., Andreasen, T., et Larsen, H. L., rédacteurs, *Proceedings of the third international conference on flexible query answering systems*, pages 95–109. Springer-Verlag : Heidelberg, Germany.

Colombetti, M. (2000). Commitment-based semantic for agent communication languages. Dans *1st Workshop on the History and Philosophy of Logic, Mathematics and Computation*. Extended Abstract.

Conte, R. et Castelfranchi, C. (1995). *Cognitive and Social Action*. UCL Press, London.

Cost, R. S., Chen, Y., Finin, T., Labrou, Y., et Peng, Y. (1999). Modeling agent conversation with colored petri nets. Dans *Proceedings of The Agent Conversation Policies Workshop at Third International Conference on Autonomous Agents (Agents-99)*, Seattle, WA, USA.

Craig, R. T. (1983). *Conversational Coherence : Form, Structure and Strategy*. Sage, Beverly Hills, CA, USA.

Craig, R. T. (1993). Why are there so many communication theories ? *Journal of Communication*, 43 :26–33.

Dastani, M., Hulstijn, J., et der Torre, L. V. (2000). Negotiation protocols and dialogue games. Dans *Proceedings of the Belgium/Dutch AI Conference (BNAIC'2000)*, Kaatsheuvel, Holland.

Davidson, D. (1986). *A coherence theory of truth and knowledge*. Oxford : Blackwell.

Dessalles, J.-L. (1993). *Modèle cognitif de la communication spontanée, appliqué à l'apprentissage des concepts*. Thèse de doctorat, École Nationale Supérieure des Télécommunications (ENST), Paris, France.

Dessalles, J.-L. (1998a). Casual conversation as logical constraint satisfaction. Dans Allwood, J., rédacteur, *Workshop on Pragmatics and Logic at the 10th European Summer School in Logic, Language and Information (ESSLLI'98)*, pages 27–34, Saarbrücken, Germany.

Dessalles, J.-L. (1998b). *Formal Semantics and Pragmatics of Dialogue*, volume TWLT-13, chapitre The Interplay of Desire and Necessity in Dialogue, pages 89–97. Enschede : University of Twente, Enschede, The Netherlands.

Dessalles, J.-L. et Ghadakpour, L. (1999). L'activité scientifique en tant que comportement naturel ancré sur le conflit cognitif. Dans *Actes des Huitièmes Journées de Rochebrune : Conflits des Interprétations et Interprétation des Conflits*, volume 99-S-001, pages 87–98, Paris, France. École Nationale Supérieure des Télécommunications (ENST).

Dignum, F. et Greaves, M. (2000). Issues in agent communication : An introduction. Dans Dignum, F. et Greaves, M., rédacteurs, *Issues in Agent Communication*, volume 1916 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 1–16. Springer-Verlag : Heidelberg, Germany.

Dignum, F., Kinny, D., et Sonenberg, L. (2002). From desires, obligations and norms to goals. *Cognitive Science Quarterly*, 2(3-4) :407–430.

Dignum, F. et Vreeswick, G. A. W. (2003). Toward a test bed for multi-party dialogue. Dans Dignum, F. et Huget, M.-P., rédacteurs, *Agent Communication Languages and Conversation Policies Workshop (AAMAS'03)*, pages 56–66.

Elio, R., Haddadi, A., et Singh, A. (2000). Task models, intentions and agent conversation policies. Dans *Pacific Rim International Conference on Artificial Intelligence*, pages 394–403. Springer-Verlag : Heidelberg, Germany.

Endriss, U., Maudet, N., Sadri, F., et Toni, F. (2004). Logic-based agent communication protocols. Dans Dignum, F., rédacteur, *Advances in Agent Communication*, volume 2922 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 91–107. Springer-Verlag : Heidelberg, Germany.

Erwin, P. (2001). *Attitudes and Persuasion*. Psychology Press.

Excelente-Toledo, C. B., Bourne, R. A., et Jennings, N. R. (2001). Reasoning about commitments and penalties for coordination between autonomous agents. Dans Müller, J., Andre, E., Sen, S., et Frasson, C., rédacteurs, *Proceedings of the Fifth International Conference on Autonomous Agents*, pages 131–138, Montreal, Canada. ACM Press.

Feldman, J. A. (1981). *Parallel models of associative memory*, chapitre A connectionist model of visual memory, pages 49–81. Erlbaum.

Ferguson, G. (1995). *Knowledge Representation and Reasoning for Mixed-Initiative Planning*. Report tr-562, University of Rochester, Computer Sciences Department, Rochester, USA.

Festinger, L. (1954). *A Social Communication and Cognition : A very Preliminary and Highly Tentative Draft*, chapitre Appendix A. American Psychological Association, Washington DC, USA.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.

Festinger, L. et Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58 :203–210.

Fikes, R. et Nilsson, N. (1971). STRIPS : A new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2 :189–208.

Finin, T. et Fritzson, R. (1994). KQML– a language and protocol for knowledge and information exchange. Dans *Proceedings of the Thirteenth International Workshop on Distributed Artificial Intelligence*, pages 126–136, Lake Quinalt, WA, USA.

Finin, T., Labrou, Y., et Peng, Y. (1999). Agent communication languages : The current landscape. *IEEE Intelligent Systems*, 14(2) :45–52.

Fiorino, H. (1998). *Élaboration de Conjectures par des Agents Coopérants*. Thèse de doctorat, École Nationnale Suppérieure de l'Aéronautique et de l'Espace (ENSAE-SUPAÉRO), Toulouse, France.

Fiorino, H. et Maille, N. (1998). Conflict solving through common conjecture elaboration. Dans *Workshop on Conflicts among agents (ECAI'98)*, Brighton, UK.

FIPA (2000). Agent communication language, FIPA [Foundation for Intelligent Physical Agents] 2000 specification. http ://www.FIPA.org.

FIPA (2004). Fipa [Foundation for Intelligent Physical Agents]. http ://www.FIPA.org.

Flores, R. (2002). *Modelling Agent conversations for Actions*. Thèse de doctorat, University of Calgary, Calgary, Alberta, Canada.

Flores, R. et Kremer, R. (2001). Bringing coherence to agent conversation. Dans Wooldridge, M., Ciancarini, P., et Weiss, G., rédacteurs, *Agent-Oriented Software Engineering II*, volume 2222 de *Lecture Notes in Computer Science (LNCS)*, pages 50–67. Springer-Verlag : Heidelberg, Germany.

Flores, R. et Kremer, R. C. (2004). A principled modular approach to construct flexible conversation protocols. Dans Tawfik, A. Y. et Goodwin, S. D., rédacteurs, *Proceedings of the 17th Canadian Conference on Artificial Intelligence*, volume 3060 de *Lecture Notes in Computer Science (LNCS)*, London, Canada. Springer-Verlag : Heidelberg, Germany.

Flores, R., Pasquier, P., et Chaib-draa, B. (2005a). Conversational semantics with social commitments. Dans van Eijk, R., Huget, M.-P., et Dignum, F., rédacteurs, *Agent Communication : International Workshop on Agent Communication (AC 2004)*, number 3396 in Lecture Notes in Artificial Intelligence (LNAI), pages 18–33. Springer-Verlag : Heidelberg, Germany.

Flores, R., Pasquier, P., et Chaib-draa, B. (2005b). A layered model for message semantics using social commitments. *Journal of Autonomous Agent and Multiagent Systems*. À paraître.

Fornara, N. et Colombetti, C. (2002). Operational specification of a commitment-based agent communication language. Dans C., C. et Johnson, W. L., rédacteurs, *Proceeding of the First Autonomous Agents and Multi-Agents Systems Joint Conference (AAMAS'02)*, volume 2, pages 535–543. ACM Press.

Fornara, N. et Colombetti, M. (2003). Defining interaction protocols using a commitment-based agent communication language. Dans Rosenchein, J. S., Sandholm, T., Wooldridge, M., et Yokoo, M., rédacteurs, *Proceedings of the second Autonomous Agents and Multi-Agents Systems conference (AAMAS'03)*, pages 520–527, Melbourne, Australie. ACM Press.

Fornara, N., Vigano, F., et Colombetti, M. (2005). Agent communication and institutional reality. Dans van Eijk, R., Huget, M., et Dignum, F., rédacteurs, *Agent Communication : Internationnal Workshop on Agent Communication (AC 2004)*, volume 3396 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 1–17. Springer-Verlag : Heidelberg, Germany.

Goethals, G. R. et Cooper, J. (1975). When dissonance is reduced : The timing of self-justificatory attitude change. *Journal of Personnality and Social Psychology*, 32 :361–387.

Goffman, E. (1959). *The presentation of the self in everyday life*. Doubleday, New York, USA.

Goodman, N. (1965). *Fact, fiction, and forecast*. Bobbs-Merrill, Indiannapolis, USA, 2^{ime} édition.

Grice, H. P. (1957). Meaning. *Philosophical Review*, 66 :377–388.

Grice, H. P. (1969). Utterer's meaning and intentions. *Philosophical Review*, 78 :147–177.

Grice, H. P. (1975). *Syntax and Semantics : Speech acts*, volume 3, chapitre Logic and Conversation. Academic Press.

Grosz, B. J. et Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligence*, 86 :269–357.

Grosz, B. J. et Sidner, C. L. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics*, 12 :175–204.

Grosz, B. J. et Sidner, C. L. (1990). Plans for discourse. Dans Cohen, P. R., Morgan, J., et Pollack, M. E., rédacteurs, *Intentions in Communication*, pages 417–444. The MIT Press : Cambridge, MA, USA.

Haack, S. (1993). *Evidence and Inquiry : towards reconstruction in epistemology*. Oxford : blackwell.

Habermas, J. (1984). *The Theory of Communicative Action*. Polity Press, Cambridge, UK.

Hamblin, C. (1970). *Fallacies*. Methuen, London, UK.

Hanson, D. J. (1980). Relationship between methods and findings in attitude-behaviour research. *Psychology*, 17 :11–13.

Harman, G. (1986). *Change in View : Principles of Reasoning*. The MIT Press : Cambridge, MA, USA.

Harmon-Jones, E. et Mills, J., rédacteurs (1999). *Cognitive Dissonance : Progress on a Pivotal Theory in Social Psychology.* American Psychological Association.

Hechter, M. et Opp, K. D. (2001). Introduction. Dans Hechter, M. et Opp, K. D., rédacteurs, *Social Norms*, pages xi–xx. Russell Sage Foundation.

Hegel, G. (1967, originally published in 1807). *The phenomenology of mind*. Harper and Row, New York, USA. Trad par Bailly, J.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. John Wiley and sons, New York, USA.

Hewitt, C. (1991). Open information systems semantics for distributed artificial intelligence. *Artificial Intelligence*, 47 :76–106.

Higgins, E. T., Rhodewalt, F., et Zanna, M. (1979). Dissonance motivation : Its nature, persistence, and reinstatement. *Journal of Experimental Social Psychology*, 15 :16–34.

Hoare, C. A. R. (1972). Proof of correctness of data representations. *Acta Informatica*, 1(4) :271–281.

Holyoak, K. J. et Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, 13 :295–355.

Holzmann, G. (1991). *Design and Validation of Computer Protocols*. Prentice-Hall.

Houdé, O., Kayser, D., Koenig, O., Proust, J., et Rastier, F. (1998). *Vocabulaire des Sciences Cognitives*. Psychologie et sciences de la pensée. Presses Universitaires de France (PUF).

Hovland, C., Janis, I., et Kelley, H. (1953). *Communication and Persuasion : Psychological studies of opinion change*. Yale University Press, New Haven, CT, USA.

Howden, N., Rönnquist, R., Hodgson, A., et Lucas, A. (2001). Jack intelligent agents : summary of an agent infrastructure. Dans Müller, J.-P., Andre, E., Sen, S., et Frasson, C., rédacteurs, *Proceedings of the Fifth International Conference on Autonomous Agents*, Montréal, Canada. ACM Press.

Huget, M. (2001). *Une ingénierie des protocoles d'interaction pour les systèmes multiagents*. Thèse de doctorat, Université Paris IX, Paris, France.

Huget, M. et Demazeau, Y. (2005). First steps towards multi-party communication. Dans van Eijk, R., Huget, M., et Dignum, F., rédacteurs, *Agent Communication, International Workshop on Agent Communication Languages (AC 2004)*, volume 3396 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 65–75. Springer-Verlag : Heidelberg, Germany.

Huget, M. et Koning, J. (2003). Interaction protocol engineering. Dans Huget, M., rédacteur, *Agent Communication Languages and Conversation Policies*, volume 2650 de *Lecture notes in Artificial Intelligence (LNAI)*, pages 179–193. Springer-Verlag : Heidelberg, Germany.

Hulstijn, J., Dignum, F., et Dastani, M. (2005). Coherence constraints for agent interaction. Dans van Eijk, R., Huget, M., et Dignum, F., rédacteurs, *Agent Communication : Internationnal Workshop on Agent Communication (AC 2004)*, volume 3396 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 134–152. Springer-Verlag : Heidelberg, Germany.

Hulstiln, J. (2000). Dialogue games are recipe for joint action. Dans *Proceedings of the 4th Workshop on the Semantics and Pragmatics of Dialogue (GOTALOG'00)*, volume 00-5, Goteborg, Sweden. Gothenburg Papers in Computational Linguistics.

Hunter, J., Danes, J., et Cohen, S. (1984). *Mathematical Models of Attitude Change*, volume 1 de *Human Communication research series*. Academic Press, Inc.

Hurley, S. L. (1989). *Natural reasons : Personality and polity.* Oxford University Press, New York, USA.

Ingrand, F., Georgeff, M., et Rao, A. (1992). An architecture for real-time reasoning and system control. *IEEE Expert, Knowledge-Based Diagnosis in Process Engineering*, 7(6) :34–44.

Inverno, M., Kinny, D., Luck, M., et Wooldridge, M. (1998). A formal specification of dmars. Dans Singh, M., Rao, M. P., et Wooldridge, M., rédacteurs, *Intelligent Agents IV : Proceedings of the Fourth International Workshop on Agent Theories, Architectures and Languages (ATAL'98)*, volume 1365 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 155–176. Springer-Verlag : Heidelberg, Germany.

Inverno, M., Luck, M., Georgeff, M., Kinny, D., et Wooldridge, M. (2004). The dmars architechure : A specification of the distributed multi-agent reasoning system. *Journal of Autonomous Agents and Multi-Agent Systems*, 1-2(9) :5–53.

Jaccard, M. (1996). *La conclusion de contrats par ordinateur - aspects juridiques de l'échange de données informatisées*. Thèse de doctorat, Stämpfli Editions SA.

Johnson, M. W., McBurney, P., et Parsons, S. (2003). When are two protocols the same ? Dans Huget, M.-P., rédacteur, *Communication in Multiagent Systems*, volume 2650 de *Lecture*

Notes in Artificial Intelligence (LNAI), pages 253–268. Springer-Verlag : Heidelberg, Germany.

Jung, H., Tambe, M., et Kulkarni, S. (2001). Argumentation as distributed constraint satisfaction : Applications and results.

Kasper, G. (1997). Can pragmatic competence be taught ? Honolulu : University of Hawaï, Second Language Teaching and Curriculum Center. Retrieved 6.07.2004 from the World Wide Web : http ://www.nflrc.hawaii.edu/NetWorks/NW06/.

Kast, R. (1993). *La théorie de la décision*, volume 120 de *Repères*. La découverte, Paris, France.

Kayser, D. (2001). *Traitement automatique du langage naturel*, volume 20 de *Technique et science informatiques*. Hermes, Paris.

Keefe, J. O. (1991). *Cognitive Dissonance*, chapitre Persuasion : Theory and research, pages 61–78. Sage, Newbury Park, California, USA.

Kitcher, P. (1983). *The nature of mathematical knowledge*. Oxford University Press, New York, USA.

Kone, M. T., Shimazu, A., et Nakajima, T. (2000). The state of the art in agent communication languages. *Knowledge and Information Systems*, 2 :259–284.

Kowtko, J., Isard, S., et Doherty, G. (1991). Conversational games within dialogue. Dans *Proceedings of the ESPRIT Workshop on Discourse Coherence*, pages 169–180, Edinburgh, UK.

Koza, J., Andre, D., et Keane, M. (1999). *Genetic Programming III : Darwinian Invention and Problem Solving*. Morgan Kaufmann Publishers, Inc.

Krauss, R. et Morsella, E. (2000). *The Handbook of Conflict Resolution : Theory and Practice*, chapitre Communication and Conflict, pages 131–143. Jossey Bass, San Francisco, USA.

Krishman, H. S. et Smith, R. E. (1998). The relative endurance of attitudes, confidence and attitudes-behaviour consistency. *Journal of Consumer Psychology*, 7 :273–298.

Kumar, S., Huber, M. J., Cohen, P. R., et McGee, D. R. (2002). Toward a formalism for conversation protocols using joint intention theory. *Computational Intelligence Journal (Special Issue on Agent Communication Language)*, 18(2) :174–228.

Kumar, V. (1992). Algorithms for constraint-satisfaction problems : A survey. *AI Magazine*, 13(1) :32–44.

Kurzweil, R. (1999). *The Age of Spiritual Machines : When Computers Exceed Human Intelligence*. Penguin Books.

La Piere, R. T. (1934). Attitudes vs. actions. *Social Forces*, 13 :230–237.

Labrie, M.-A. (2003). Langage de communication agent basé sur les engagements par l'entremise des jeux de dialogue. Mémoire de maîtrise, Département d'Informatique et de Génie Logiciel, Faculté des Sciences et de Génie, Université Laval, Québec, Canada.

Labrou, Y. (1996). *Semantics for an Agent Communication Language.* Thèse de doctorat, Computer Science and Electrical Engineering Department, University of Maryland, Baltimore, USA.

Labrou, Y. et Finin, T. (1997a). A proposal for a new KQML specification. Rapport Technique TR CS-97-03, Computer Science and Electrical Engineering Department, University of Maryland Baltimore County, Baltimore, MD 21250, USA.

Labrou, Y. et Finin, T. (1997b). Semantics and conversations for an agent communication language. Dans Pollack, M. E., rédacteur, *Proceedings of the Fifteenth International Joint Conference on Artifical Intelligence (IJCAI-97)*, pages 584–591, Nagoya, Japan. Morgan Kaufmann publishers Inc. : San Mateo, CA, USA.

Labrou, Y., Finin, T., et Mayfield, J. (1995). *Software Agents*, chapitre KQML as an Agent Communication Language. The MIT Press : Cambridge, MA, USA.

Lambert, L. et Carberry, S. (1991). A tripartite plan-based model of dialogue. Dans *Proceedings of the 29th annual meeting of Association for Computational Linguistics (ACL)*, pages 47–54, Berkeley, USA. Association for Computational Linguistics.

Lamontagne, L. et Lapalme, G. (2004). Textual reuse for email response. Dans *Advances in Case-Based Reasoning, 7th European Conference, ECCBR 2004*, volume 3155 de *Lecture Notes in Computer Science*, pages 242–256. Springer-Verlag : Heidelberg, Germany.

Leech, G. (1983). *Principles of pragmatics*. London : Longman.

Lehrer, K. (1990). *Theory of Knowledge*. Boulder : Westview.

Lemeunier, T. (2000). *L'intentionnalité communicative dans le dialogue homme-machine en langue naturelle*. Thèse de doctorat, Université du Maine, France.

Lemeunier, T. (2003). De la modélisation de l'activité conversationnelle des systèmes de dialogue personne-machine. *In Cognito : Cahiers Romans de Sciences Cognitive*, 1(2) :23–52.

Levin, J. A. et Moore, J. A. (1978). Dialogue-games : Metacommunication structures for natural language understanding. *Cognitive Science*, 1(4) :384–420.

Levinson, D. J. (2003). Collective sanctions. Public law research paper no. 57, New York University, School of Law, Center for Law and Business Research.

Levinson, S. C. (1979). Activity type and language. *Linguistics*, 17 :365–399.

Levinson, S. C. (1983). *Pragmatics*. Cambridge University Press : Cambridge, UK.

Lewin, I. (2000). A formal model of conversation games theory. Dans *proceedings of the 4th workshop on the semantics and pragmatics of dialogue (GOTALOG'00)*, Göteborg University, Sweeden. Gothenburg Papers in Computational Linguistics 00-5.

Lienard, J. S. (1991). Communication homme machines ; rapport sur la définition, l'état de l'art et les perspectives scientifiques. Rapport du comité d'objectif scientifique et technique (COST), Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI), SPI-CNRS.

Lin, F., Norrie, D., Shen, W., et Kremer, R. (1999). A shema-based approach to specifying conversation policies. Dans Greaves, M. et Bradshaw, M., rédacteurs, *Workshop on Specifying and Implementing conversation Policies*, volume 1916 de *Lecture Notes in Computer Science (LNCS)*, pages 193–204. Springer-Verlag : Heidelberg, Germany.

Litman, D. et Allen, J. F. (1990). *Intentions in communication*, chapitre Discourse Processing and Common Sens Plans, pages 365–388. The MIT Press : Cambridge, MA, USA.

Littlejohn, S. W. (2002). *Theories of Human Communication*. Number seventh edition. Wadsworth Publishing Company.

Lochbaum, K. E. (1994). *Using Collaborative Plans to Model the Intentional Structure of Discourse.* Thèse de doctorat, Harvard University, Cambridge, MA., USA.

MacKenzie, J. (1979). Question-begging in non-cumulative systems. *Journal of philosophical logic*, 8 :117–133.

Maes, P. (1995). Intelligent software agent. *Scientific American*, 273(3) :66–68. Special Issue on Key Technologies for the 21st Century.

Mallya, A., Yolum, P., et Singh, M. P. (2004). Resolving commitments among autonomous agents. Dans *Proceedings of the International Workshop on Agent Communication Languages and Conversation Policies (ACL)*, Lecture Notes in Artificial Intelligence (LNAI), pages 166–182. Springer-Verlag : Heidelberg, Germany.

Mann, W. C. (1988). Dialogues games : conventions of human interaction. *Argumentation*, 4(2) :511–532.

Marconi, D. (1997). *La philosophie de langage au XXième siècle*. Éditions de l'éclat.

Marr, D. et Poggio, T. (1976). Cooperative computation of stereo disparity. *Science*, 194 :283–287.

Martindale, D. (1978). *Social Control for the 1980s : A Handbook for Order in a Democratic Society*, chapitre The theory of social control, pages 46–58. Westport, CT : Greenwood Press.

Maudet, N. (2001). *Modéliser les Conventions des Interactions Langagières : la Contribution des Jeux de Dialogue*. Thèse de doctorat, Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et des Télécommunications (ENSEEIHT), Université Paul Sabatier (UPS), Toulouse, France.

Maudet, N. et Chaib-draa, B. (2002). Commitment-based and dialogue-game based protocols - new trends in agent communication language. *Knowledge Engineering*, 17(2) :157–179.

McBurney, P. et Parson, S. (2001). Agent ludens : games for agent dialogues. Dans *Proceedings of the 2001 AAAI Spring Symposium on Game Theoretic and Decision Theoretic Agents*. Stanford, USA, AAAI Technical Report.

McBurney, P. et Parson, S. (2002). Games that agents play : A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11(3) :315–334.

McBurney, P. Parsons, S. et Wooldridge, M. (2002). Desiderata for agent argumentation protocols. Dans *Procceedings of the First International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'01)*, pages 402–409, Bologna, Italy. ACM Press.

McClelland, J. L. et Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. *Psychological Review*, 88 :375–407.

McGuire, W. J. (1960). Cognitive consistency and attitude change. *Journal of Abnormal Social Psychology*, 60 :345–353.

Miller, N. E. et Dollar, J. (1941). *Social Learning and Imitation*. Yale University Press, Hew Haven, Connecticut, USA.

Moore, D. J. (1990a). Dialogue game theory for explanation-giving systems. Dans *Alvet SIG Workshop on Explanation*, Manchester, UK.

Moore, D. J. (1993). *Dialogue Game Theory for Intelligent Tutoring Systems*. Thèse de doctorat, Leeds Metropolitan University, UK.

Moore, R. C. (1990b). A formal theory of knowledge and action. Dans Allen, J. F., Hendler, J., et Tate, A., rédacteurs, *Readings in Planning*, pages 480–519. Morgan Kaufmann Publishers : San Mateo, CA.

Morris, C. (1938). *International Encyclopedia of Unified Science*, volume 2, chapitre Foundations of the Theory of Signs. Chicago : The University of Chicago Press.

Morris, C. (1946). *Langage, Signs and Behavior*. New York : Braziller.

Moulin, B. (1997). *Agent and Multi-agent Systems*, volume 1441 de *Lecture Notes in Artificial Intelligence (LNAI)*, chapitre The Social Dimension of Interactions in Multi-agent Systems, pages 109–122. Springer-Verlag : Heidelberg, Germany.

Moulin, B., Irandoust, H., Bélanger, M., et Desbordes, G. (2002). Explanation and argumentation capabilities : Towards the creation of more persuasive agents. *Artificial Intelligence Review*, 17(3) :169–222.

Myers, D. et Lamarche, L. (2000). *Psychologie sociale*. McGraw-Hill, Montréal, Canada.

Neurath, O. (1959). *Logical Positivism*, chapitre Protocol Sentences, pages 199–208. Free Press, Glencoe, IL., USA.

Newcomb, T. (1953). An approach to the study of communicative acts. *Psychological Review*, 60(6) :393–404.

Nuyts, J. (1994). The intentional and socio-cultural in language use. *Pragmatics and Cognition*, 2(2) :237–268.

Osgood, C. (1963). On understanding and creating sentences. *American Psychologist*, 18 :735–751.

Parsons, S. et McBurney, P. (2003). Argumentation-based communication between agents. Dans Huget, M.-P., rédacteur, *Agent Communication Language*, volume 2650 de *Lecture Note in Artificial Intelligence (LNAI)*, pages 164–178.

Parsons, S., Petterson, O., Saffiotti, A., et Wooldridge, M. (2000). Intention reconsideration in theory and practice. Dans *Proceedings of the 14th European Conference on Artificial Intelligence*, Berlin.

Parsons, S. et Wooldridge, M. (2002). Game theory and decision theory in multi-agent systems. *Autonomous agents and Multi-Agent Systems*, 5 :243–254.

Parunak, H. (1996). Visualizing agent conversations : Using enhanced dooley graphs for agent design and analysis. Dans *Proceedings of the Second International Conference on Multi-agent Systems (ICMAS-96)*, pages 275–282.

Pasquier, P. (2000). Conflit et incertitude. Rapport de stage du DEA : Représentation des connaissances et formalisation du raisonnement, Institut de Recherche en informatique de Toulouse (IRIT) et Office Nationnal d'Étude et de Recherche en Aérospatial (ONERA), Centre d'Étude et de Recherche de Toulouse, Toulouse, France.

Pasquier, P. (2001a). Application de théories du langage naturel aux systèmes artificiels. Rapport de synthèse, Laboratoire DAMAS, Département d'informatique et de génie logiciel, Faculté de sciences et génie, Université Laval, Québec, Canada. "http ://www.damas.ift.ulaval.ca/˜pasquier/publications.html".

Pasquier, P. (2001b). Communication entre agents. Rapport de synthèse, Laboratoire DAMAS, Département d'informatique et de génie logiciel, Faculté de sciences et génie, Université Laval, Québec, Canada.

Pasquier, P. (2002). La cohérence cognitive comme fondement pour la pragmatique des communications agents. Proposition de thèse, Université Laval, Département d'informatique et de génie logiciel, Faculté de sciences et génie. http ://www.damas.ift.ulaval.ca/˜pasquier/publications.html.

Pasquier, P. (2003). Changements d'attitudes et systèmes multi-agents. Rapport de lecture dirigée, rédigé sous la direction de Guy Paquette, Professeur et chercheur en psychologie sociale, département de Communication, Université Laval), Département de communication de l'Université Laval.

Pasquier, P., Andrillon, N., et Chaib-draa, B. (2003). An exploration in using cognitive coherence theory to automate BDI agents' communicational behavior. Dans Dignum, F., rédacteur, *Advances in Agent Communication - International Workshop on Agent Communication Languages (ACL'03)*, volume 2922 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 37–58. Springer-Verlag : Heidelberg, Germany.

Pasquier, P., Bergeron, M., et Chaib-draa, B. (2004a). DIAGAL : a Generic ACL for Open Systems. Dans Gleizes, M.-P., Omicini, A., et Zambonelli, F., rédacteurs, *Proceedings of The Fifth International Workshop Engineering Societies in the Agents World (ESAW'04)*, volume 3451 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 139–152. Springer-Verlag : Heidelberg, Germany.

Pasquier, P. et Chaib-draa, B. (2002). Cohérence et conversations entre agents : vers un modèle basé sur la consonance cognitive. Dans Müller, J. et Mathieu, P., rédacteurs, *Systèmes multi-agents et systèmes complexes, Actes des 10ème journées francophones d'intelligence artificielle distribuée et des systèmes multi-agents (JFIADSMA'02)*, pages 188–203, Paris, France. Hermes Science Publication.

Pasquier, P. et Chaib-draa, B. (2003a). The cognitive coherence approach for agent communication pragmatics. Dans Rosenchein, J. S., Sandholm, T., Wooldridge, M., et Yokoo, M., rédacteurs, *Proceedings of The Second International Joint Conference on Autonomous Agent and Multi-Agents Sytems (AAMAS'03)*, pages 544–552, Melbourne, Australie. ACM Press.

Pasquier, P. et Chaib-draa, B. (2003b). Engagements, intentions et jeux de dialogue. Dans Herzig, A., Chaib-draa, B., et Mathieu, P., rédacteurs, *Modèles formels de l'interaction, Actes des Secondes Journées Francophones (MFI'03)*, pages 289–294. Cépaduès. Papier court.

Pasquier, P. et Chaib-draa, B. (2004a). Agent communication pragmatics : The cognitive coherence approach. *Cognitive Systems*, 6. À paraitre.

Pasquier, P. et Chaib-draa, B. (2004b). Modèles de dialogue entre agents cognitifs : un état de l'art. *In Cognito : Cahiers Romans de Sciences Cognitive*. À paraitre.

Pasquier, P. et Dehais, F. (2000). Approche Générique du Conflit. Dans Scapin, D. et Vergisson, E., rédacteurs, *ErgoIHM 2000*, Biarritz, France. ESTIA.

Pasquier, P., Flores, R., et Chaib-draa, B. (2004b). Modelling flexible social commitments and their enforcement. Dans Gleizes, M.-P., Omicini, A., et Zambonelli, F., rédacteurs, *Proceedings of the Fifth International Workshop Engineering Societies in the Agents World (ESAW'04)*, volume 3451 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 153–165. Springer-Verlag : Heidelberg, Germany.

Pasquier, P. et Rivaland, B. (2002). Émergence de dialogues entre agents par satisfaction de contraintes cognitives. Rapport de projet écrit pour le cours : Systèmes multi-agents et applications., Département d'Informatique et de Génie Logiciel, Université Laval.

Petty, R. et Cacioppo, J. (1996). *Attitudes and Persuasion : Classic and Contemporary Approaches*. Westview Press.

Philips, L. et Link, H. (1999). The role of conversation policy in carrying out agents conversations. Dans Greaves, M. et Bradshaw, M., rédacteurs, *Workshop on Specifying and Implementing Conversation Policies*.

Poesio, M. et Mikheev, A. (1998). The predictive power of game structure in dialogue act recognition : Experimental results using maximum entropy estimation. Dans *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP'98)*, Sydney, Australia.

Polinsky, M. et Shavel, S. (1998). *The New Palgrave Dictionary of Economics and The Law*, volume 3, chapitre Punitive Damages, pages 192–198. London : Macmillan Reference Limited.

Pollack, M. E. (1990). Plans as complex mental attitudes. Dans Cohen, P. R., Morgan, J., et Pollack, M. E., rédacteurs, *Intentions in Communication*, pages 77–104. The MIT Press : Cambridge, MA, USA.

Posner, R. A. et Rasmusen, E. B. (1999). Creating and enforcing norms, with special reference to sanctions. *International Review of Law and Economics*, 19(3) :369–382.

Quignard, M. (2000). *Modélisation cognitive de l'argumentation dialoguée : étude de dialogues d'élèves en résolution de problème de sciences physiques*. Thèse de doctorat, Université Joseph Fourier, Grenoble, France.

Quine, W. (1963). *From a logical point of view*. Harper Torchbooks, New York, 2 édition.

Ramshaw, L. (1989). A meta-plan model for problem solving discourse. Dans *Proceedings of the 4th Conference of the European Chapter of tha Association for Computational Linguistics (ACL)*, pages 35–42.

Rao, A. S. et Georgeff, M. (1995). BDI Agents : from theory to practice. Dans *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 312–319, San Francisco, CA, USA.

Rao, A. S. et Georgeff, M. P. (1991). Modeling rational agents within a BDI-architecture. Dans Fikes, R. et Sandewall, E., rédacteurs, *Proceedings of Knowledge Representation and Reasoning (KR&R-91)*, pages 473–484. Morgan Kaufmann Publishers : San Mateo, CA.

Rawls, J. (1971). *A theory of justice*. Harvard University Press, Cambridge.

Raz, J. (1992). The relevance of coherence. *Boston University Law Review*, 72 :273–321.

Read, S. et Marcus-Newhall, A. (1993). The role of explanatory coherence in the construction of social explanations. *Journal of Personality and Social Psychology*, 65 :429–447.

Recanati, F. (2001). What is said. *Synthese*, 128 :75–91.

Reed, C. (1998). Dialogue frames in agent communication. Dans *Proceedings of the Third International Conference on MultiAgent Systems (ICMAS'98)*, pages 246–253, Paris, France. IEEE Computer Society.

Reed, C. et Long, D. (1997). Collaboration, cooperation and dialogue classification. Dans *IJCAI 1997*, volume 2, Nagoya, Japan. Morgan Kaufmann.

Reed, C., Norman, T., et Jennings, N. (2002). Negotiating the semantics of agent communication languages. *Computational Intelligence*, 2(18) :229–252.

Riess, M. et Schlenker, B. (1977). Attitude change and responsability avoidance as modes of dilemma resolution in forced-compliance situations. *Journal of Personnality and Social Psychology*, 35 :21–30.

Rokeach, M. (1969). *Attitudes and Values : A Theory of Organisation and Change*. San Fransisco : Jossey-Bass.

Rosenberg, M. J. (1960). *Attitude Organisation and change*, chapitre An analysis of affective-cognitive consistency. Yale University Press, New Haven, CT, USA.

Rosenberg, M. J. et Abelson, R. P. (1960). *Attitude Organisation and change*, chapitre An analysis of cognitive balancing. Yale University Press, New Haven, CT, USA.

Royakkers, L. et Dignum, F. (2000). No organisation without obligation : How to formalise collective obligation ? Dans Ibrahim, M., Kung, J., et Revell, N., rédacteurs, *Proceedings of the* 11^{th} *Internationnal Conference on Databases and Expert Systems Applications*, volume 1873 de *Lecture Notes in Computer Science (LNCS)*, pages 302–311. Springer-Verlag : Heidelberg, Germany.

Russell, S. et Norvig, P. (2003). *Artificial Intelligence : A Modern Approach*. Prentice Hall Series in Artificial Intelligence, 2^{ime} édition.

Sadek, D. (1991a). *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication*. Thèse de doctorat, Université de Rennes 1, France.

Sadek, D., Bretier, P., et Panaget, F. (1997). ARTIMIS : Natural dialogue meets rational agency. Dans *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97)*, Yokohama, Japan.

Sadek, M. (1991b). Dialogue acts are rational plans. Dans *Proceedings of the ESCA/ETRW Workshop on the Structure of Multimodal Dialogue*, pages 1–29, Maratea, Italy.

Sakai, H. (2001). *A Multiplicative Power-Function Model of Cognitive Dissonance : Toward an Integrated Theory of Cognition, Emotion and Behavior After Leon Festinger*, chapitre Computer Simulation, pages 267–294. American Psychological Association.

Sandholm, T. W. et Lesser, V. R. (1995). Issues in automated negociation and electronic commerce : Extending the contract net framework. Dans *Proceedings of the First International Conference on Multiagent Systems (ICMAS-95)*, pages 328–335. AAAI Press.

Sandholm, T. W. et Lesser, V. R. (1996). Advantages of a leveled commitment contracting protocol. Dans *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, Portland, OR, USA.

Sansonnet, J.-P. et Valencia, E. (2003a). Agents informationnels pour l'étude expérimentale de concepts de socio-cognition : vers une approche agent de la socio informatique. Dans *Journées francophones des systèmes multiagents (JFSMA'03)*, Hamamet, Tunisia. http ://www.limsi.fr/Individu/jps/research/buzz/buzz.htm.

Sansonnet, J.-P. et Valencia, E. (2003b). Dialogue between non-task oriented agents. Dans *Proceedings of the 4th Workshop on Agent Based Simulation (ABS'04)*, Montpelier, France. http ://www.limsi.fr/Individu/jps/research/buzz/buzz.htm.

Schlenker, B. R. (1980). *Impression management : The self-concept, social identity, and interpersonal relations*. Brooks/Cole, Monterey, CA, USA.

Schut, M. et Wooldridge, M. (2001). Principles of intention reconsideration. Dans Muller, J., Andre, E., Sen, S., et Frasson, C., rédacteurs, *Proceedings of the Fifht International Conference on Autonomous Agents, AGENTS'01*, pages 340–347, Montreal, Canada. ACM Press.

Schwartz, P. J. (2001). Truth maintenance with cognitive dissonance. Rapport technique, University of Mariland at College Park, USA.

Searle, J. R. (1969). *Speech Acts : An Essay in the Philosophy of Language*. Cambridge University Press : Cambridge, UK.

Searle, J. R. (1979). *Expression and Meaning*. Cambridge University Press : Cambridge, UK.

Searle, J. R. (1983). *Intentionality : An Essay in the Philosophy of Mind*. Cambridge University Press : Cambridge, UK.

Searle, J. R. (1990). Collective intentions and actions. Dans Cohen, P. R., Morgan, J., et Pollack, M. E., rédacteurs, *Intentions in Communication*, pages 401–416. The MIT Press : Cambridge, MA, USA.

Searle, J. R. (1992a). *La Redécouverte de l'Esprit*. NRF Essais. Gallimard.

Searle, J. R. (1992b). *(On) Searle on Conversation*, chapitre Conversation, pages 7–29. Benjamins Pub, Philadelphia, USA.

Searle, J. R. et Vanderveken, D. (1985). *Foundations of Illocutionary Logic*. Cambridge University Press, NY, USA.

Shannon, C. E. et Weaver, C. (1975). *Théorie mathématique de la communication, traduit de l'anglais par Cosnier, Dahan et Economides*. Retz CEPL.

Sherif, M. et Hovland, C. (1961). *Social Judgement*. Yale University Press, New Haven, USA.

Shoham, Y. (1987). *Reasoning about change : Time and causation from the standpoint of artificial intelligence*. Thèse de doctorat, Computer science department, Yale University, USA.

Shoham, Y. (1990a). Agent-oriented programming. Rapport Technique STAN–CS–1335–90, Computer Science Department, Stanford University, Stanford, CA, USA.

Shoham, Y. (1990b). Nonmonotonic reasoning and causation. *Cognitive Science*, 2(14) :213–252.

Shultz, R. et Lepper, R. (1999). *Cognitive Dissonance : progress in a pivotal theory in social psychology*, chapitre Computer simulation of the cognitive dissonance reduction, pages 235–265. American Psychological Association.

Shultz, T. R., Katz, J., et Lepper, M. (2001). Clinging to belief : a constraint-satisfaction model. Dans Moore, J. D. et Stenning, K., rédacteurs, *Proceedings of the Twenty Third Annual Conference of the Cognitive Science Society*, University of Edinburgh, Scotland. Laurence Erlbaum Associates.

Simon, H. (1957). *Models of man : Social and Rational*. John Wiley, New York, USA.

Singh, M. P. (1994). *Multiagent Systems : A Theoretical Framework for Intentions, Know-How, and Communications*, volume 799 de *Lecture Notes in Artificial Intelligence (LNAI)*. Springer-Verlag : Heidelberg, Germany.

Singh, M. P. (1998). Agent communication languages : rethinking the principles. *IEEE Computer*, 12(31) :40–47.

Singh, M. P. (1999). An ontology for commitments in multiagent systems : Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7 :97–113.

Singh, M. P. (2000). A social semantics for agent communication languages. Dans Dignum, F. et Greaves, M., rédacteurs, *Issues in Agent Communication*, Lecture Notes in Artificial Intelligence (LNAI), pages 31–45. Springer-Verlag : Heidelberg, Germany.

Singh, M. P., Rao, S., et Georgeff, M. P. (1999). *Multiagent Systems : A Modern Approach to Distributed Artificial Intelligence*, chapitre Formal Methods in DAI : Logic-Based Representation and Reasoning, pages 331–376. The MIT Press : Cambridge, MA, USA.

Smith, R. G. (1977). The CONTRACT NET : A formalism for the control of distributed problem solving. Dans *Proceedings of the Fifth International Joint Conference on Artificial Intelligence (IJCAI-77)*, Cambridge, MA, USA.

Smith, R. G. (1980). The contract net protocol : High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, C-29(12) :1104–1113.

Smith, R. W., Hipp, D. R., et Biermann, A. W. (1995). An architecture for voice dialogue systems based on prolog-style theorem proving. *Computational Linguistics*, (21) :281–320.

Sperber, D. et Wilson, D. (1986). *Relevance*. Harvard University Press, Cambridge MA.

SRI International (1999). Open Agent Architecture. Internet, http ://www.ai.sri.com/ oaa.

Stalnaker, R. C. (1978). Assertion. *Syntax and Semantics*, (9) :315–322.

Steels, L., F., K., et Van Looveren, J. (2002). *The Transition to Language*, chapitre Crucial factors in the origins of word-meaning. Oxford University Press.

Sun, R. (1994). A neural network model of causality. *IEEE Transactions on Neural Networks*, 5(4) :604–611.

Sun, R. (1997). *Connectionist-Symbolic Integration*, chapitre An introduction to hybrid connectionist-symbolic models. Lawrence Erlbaum Associates.

Sun, R. (2001). Cognitive science meets multi-agent systems : a prolegomenon. *Philosophical Psychology*, 14(1) :5–28.

Sun, R. et Browne, A. (1999). Connectionist variable binding. *Expert Systems*, 16(3) :189–207.

Sun, R. et Browne, A. (2001). Connectionist inference models. *Neural Networks*, 14(10) :1331–1355.

Sutton, R. et Barto, A. (1998). *Reinforcement Learning : An Introduction*. The MIT Press : Cambridge, MA, USA.

Thagard, P. (2000a). *Coherence in Thought and Action*. The MIT Press : Cambridge, MA, USA.

Thagard, P. (2000b). Probabilistic network and explanatory coherence. *Cognitive science Quaterly*, (1) :91–114.

Thagard, P. et Millgram, E. (1995). *Goal driven learning*, chapitre Inference to the best plan : a coherence theory of decision, pages 439–454. The MIT Press : Cambridge, MA, USA.

Thagard, P. et Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science*, 22 :1–24.

Thaise, A. et al (1988). *Approche Logique de l'Intelligence Volumes 1,2 et 3*. Dunod Informatique.

Traum, D. (2004). Issues in multi-party dialogues. Dans Dignum, F., rédacteur, *Advances in agent Communication*, volume 2922 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 201–221. Springer-Verlag : Heidelberg, Germany.

Traum, D. R. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. Thèse de doctorat, Department of Computer Sciences, University of Rochester, Rochester, USA.

Traum, D. R. (1997). A reactive-deliberatie model of dialogue agency. Dans Müller, J.-P., Wooldridge, M., et Jennings, N., rédacteurs, *Intelligent Agent III*, volume 1193 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 157–171. Springer-Verlag : Heidelberg, Germany.

Traum, D. R. et Hinkelman, E. A. (1992). Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*. Special Issue on Non-literal Language.

Traum, D. R. et Poesio, M. (1997). Conversational actions and discourse situations. *Computational Intelligence*, 13(3) :309–347.

Trognon, A. et Brassac, C. (1992). L'enchaînement conversationnel. *Cahiers de Linguistique Française*, 13 :76–107.

Tuomela, R. et Miller, K. (1988). We-intentions. *Philosophical Studies*, 53 :367–389.

Vanderveken, D. (1990). *Meaning and Speech Acts : Principles of Language Use*. Cambridge University.

Vanderveken, D. (1999). *Analyse et simulation de conversations : de la théorie des actes de langage aux systèmes multi-agents*, chapitre La structure logique des dialogues intelligents, pages 61–100. L'interdiscilplinaire informatique. Paris, France.

Varela, F. (1996). *Invitation aux Sciences Cognitives*. Seuil.

Velasquez, J. (1997). Modeling emotions and other motivations in synthetic agents. Dans *Proceedings of the AAAI Conference 1997*, pages 10–15. Providence, RI, USA.

Verdicchio, M. et Colombetti, M. (2004). A logical model of social commitment for agent communication. Dans Dignum, F., rédacteur, *Advances in Agent Communication : International Workshop on Agent Communication Languages*, volume 2922 de *Lecture Notes in Artificial Intelligence (LNAI)*. Springer-Verlag : Heidelberg, Germany.

Verdicchio, M. et Colombetti, M. (2005). Dealing with time in content language expressions. Dans van Eijk, R., Huget, M., et Dignum, F., rédacteurs, *Agent Communication : Internationnal Workshop on Agent Communication (AC 2004)*, volume 3396 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 91–105. Springer-Verlag : Heidelberg, Germany.

Vignaux, G. (1991). *Les Sciences Cognitives : une Introduction*. le Livre de Poche : biblio essais, Paris, La Découverte édition.

Vitteau, B. et Huget, M. (2004). Modularity in interaction protocols. Dans Dignum, F., rédacteur, *Advances in Agent Communication, International Workshop on Agent Communication Languages (AC 2003)*, volume 2922 de *Lecture Notes in Artificial Intelligence (LNAI)*, pages 291–309. Springer-Verlag : Heidelberg, Germany.

Vold, G. B., Bernard, T. J., et Snipes, J. B. (2002). *Theoretical Criminology*. Oxford University Press, 5^{ime} édition.

von Wright, G. (1980). *Freedom and determination*. North Holland Publishing Co.

Walton, D. N. (1984). *Logical Dialogue Games and Fallacies*. University Press of America.

Walton, D. N. et Krabbe, E. (1995). *Commitment in Dialogue*. Suny Press.

Weiss, G. et Co (2001). *Multiagent Systems*, chapitre Glossary. The MIT Press : Cambridge, MA, USA.

Werner, E. (1992). The design of multi-agent systems. Dans Werner, E. et Demazeau, Y., rédacteurs, *Decentralized AI 3 — Proceedings of the Third European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW-91)*, pages 3–30. Elsevier Science Publishers B.V. : Amsterdam, The Netherlands.

Wicker, A. W. (1969). Attitudes versus actions : the relationship of verbal and overt behavioural responses to attitude objects. *Journal of Social Issues*, 25 :41–78.

Wickland, R. et Brehm, J. (1976). *Perspectives on Cognitive Dissonance*. NY : Halsted Press.

Winograd, T. et Flores, F. (1986). *Understanding Computers and Cognition : A New Foundation for Design*. Ablex, Norwood, NJ, USA.

Wittgenstein, L. (1953). *Philosophical Investigations*. MacMillan.

Wolper, P. (1991). *Introduction à la Calculabilité*. InterEditions.

Wooldridge, M. (2001a). *An Introduction to MultiAgent Systems*. Wiley.

Wooldridge, M. (2001b). *Multiagents Systems*, chapitre Intelligent Agents. The MIT Press : Cambridge, MA, USA.

Wooldridge, M. et Jennings, N. R. (1994). Formalizing the cooperative problem solving process. Dans *Proceedings of the Thirteenth International Workshop on Distributed Artificial Intelligence (IWDAI-94)*, pages 403–417, Lake Quinalt, WA, USA.

Wooldridge, M., Mac Burney, P., et Parsons, S. (2002). Desiderata for agent argumentation protocols. Dans *Proceedings of the Autonomous Agent and Multi-Agent Systems Conference(AAMAS'02)*, pages 402–409. ACM Press.

Wyer, R. et Golberg, L. (1970). A probabilistic analysis of the relationships among beliefs and attitudes. *Psychological Review*, (77) :100–120.

Zanna, M. (1975). The effect of distraction on resolving cognitive dilemmas. Dans *Meeting of American Psychology Association*, Chicago, USA.

Zanna, M. et Cooper, J. (1976). *New directions in attribution research (Vol.1)*, volume 1, chapitre Dissonance and attribution process. Hillsdale, N.J. : Erlbaum.

Zimbardo, P. (1977). *Influencing attitudes and changing behavior*. Reading. Addison-Wesley, MA, USA.

Zimbardo, P., Weisenberg, M., Firestone, I., et Levy, B. (1965). Communicator effectiveness in producing public conformity and private attitude change. *Journal of Personality*, (33) :233–255.

Annexe A

Sémantique linguistique et sémantique mathématique

Dans les travaux concernant la modélisation formelle du dialogue, deux notions de sémantiques sont en jeux : la sémantique linguistique et la sémantique mathématique. Ces deux notions entrent en conflit et il est prudent de bien indiquer de laquelle on traite.

En logique mathématique, la sémantique d'un langage formel est une relation entre le langage et un espace de structures mathématiques appelées modèles. Il est commun de définir plusieurs sémantiques mathématiques pour un langage, cela permet de démontrer plus de propriétés et les comparaisons de ces sémantiques peuvent également être enrichissantes ; on parle alors de théorie des modèles ou de méta-mathématiques.

En linguistique, la sémantique est la relation entre les signes et les objets auxquels ils s'appliquent. La sémantique linguistique devra être la plus générique et universelle possible dans l'esprit du projet linguistique lui-même. En linguistique, trois niveaux de sens sont distingués :

- *le sens de la phrase* (sentence meaning) : la caractéristique du sens de la phrase est qu'il est conventionnel et indépendant du contexte. Ainsi la phrase « J'ai grandi en Bretagne » a un sens invariable selon les énonciations que l'on peut composer à partir des références aux sens (en terme de dénotation) de ses composantes grammaticales ;
- *ce qui est dit* (what is said) : ce qui est dit fait référence à l'énonciation de la phrase en contexte, cela vient compléter le sens celle-ci en venant combler les variables restées libres. Par exemple, lorsque j'énonce la phrase « J'ai grandi en Bretagne », je suis alors le sujet auquel réfère le pronom personnel/démonstratif « J », ce qui ne serait pas le cas

si l'énoncé était de quelqu'un d'autre. On appelle « saturation » le processus par lequel le besoin d'intantiation contextuelle de ces « variables libres » est comblé. D'autres fois la référence est extra-linguistique ;

– *ce qui est impliqué* (what is implicated) : il y a des cas ou un énoncé signifie autre chose que son sens littéral dans le contexte de l'énonciation. Par exemple, si la phrase « Je suis Breton. » est énoncée en réponse à la question « Tu sais faire les crêpes ? », est en plus de son sens littéral une réponse positive à la question. Dans ces cas, le sens de l'énoncé inclut également ce qu'il implique dans son contexte d'énonciation. Dans d'autres cas, le contexte d'énonciation permet de déterminer un sens parmi plusieurs possibles. Par exemple, lorsque je réponds « J'ai déjà petit déjeuné » en réponse à la question « Est-ce que tu veux un bol de chocolat chaud ? », ce qui est impliqué c'est que j'ai déjà le ventre plein et non que j'ai déjà petit déjeuner au moins une fois par le passé ce qui serait le sens strict autorisé par la phrase.

Comme le remarque Recanati [2001], il y a encore débat pour savoir comment assigner ces trois niveaux de sens avec les catégories de *sens littéral* (literal meaning) et de *sens de l'énoncé* (speaker's meaning). La position la plus classique, « minimaliste », inclut les deux premières catégories dans le sens littéral.

Connaître un langage, c'est connaître sa syntaxe et sa sémantique linguistique suffisamment pour pouvoir construire le sens littéral d'un énoncé quelconque. L'*interprétation sémantique* est le processus par lequel un agent exploite sa connaissance d'un langage pour établir déductivement le sens littéral d'un énoncé.

L'*interprétation pragmatique* est un processus radicalement différent qui n'est pas en lien avec le langage en particulier, mais plutôt avec la proactivité des agents cognitifs. Lorsqu'un agent cognitif rationnel agit, il a des raisons pour cela (par définition du terme rationnel). Fournir une interprétation pragmatique d'une action consiste à en déterminer les raisons.

Dans le cadre des langages de communication agents (ACLs), ces deux notions de sémantique entrent en conflit et faute de précision, les discours sur la sémantique des ACLs sont souvent confus. En effet, les ACLs étant des langages formels, l'étude de leurs propriétés peut être réalisée par la définition de sémantiques mathématiques les concernant. En outre, comme ils sont un média de communication, de la même façon que le langage naturel, on aimerait s'assurer qu'ils ont une sémantique précise, garantissant à ceux qui l'utilisent de partager le sens des énoncés. Si, il est clair qu'un idéal visant à fournir une sémantique mathématique qui en soit également une linguistique est au moins implicitement poursuivi par de nombreux chercheurs, cela n'est pas évident. Nous nous contenterons pour l'heure de dresser une brève typologie des sémantiques mathématiques.

Une première manière de définir une sémantique mathématique d'un langage formel est de définir chaque élément du langage en terme des pré-conditions qui doivent tenir pour qu'il soit réalisé et des post-conditions qui doivent s'appliquer à son utilisation (comme pour les actions dans STRIPS). C'est ce que l'on nomme une *sémantique axiomatique*. Pour ce qui est de la communication agent, on peut faire la différence entre les sémantiques axiomatiques publiques qui n'utilisent que des éléments publics, accessibles à tous et les sémantiques axiomatiques privées qui font référence à des éléments internes aux agents. Pour illustrer cette idée de sémantique axiomatique, prenons l'exemple de la sémantique des langages de programmation. La méthode la plus fréquemment utilisée pour spécifier la sémantique d'un programme consiste à définir les pré- et post-conditions de chaque élément de code[1]. En guise d'exemple, considérons l'instruction « $x := 3.8$ ». La sémantique de cette instruction inclut la pré-condition : « x est le nom d'un emplacement mémoire pouvant recevoir la valeur d'un nombre de type Réel » et la post-condition : « A l'emplacement mémoire dénoté par x on trouve la valeur 3.8 »[2]. La pré-condition sert à imposer que l'instruction doit pouvoir être exécutée et la post-condition rend compte de l'effet minimal de l'instruction sur l'environnement du programme. La relation entre la pré et la post-condition correspond à la description de la compréhension intuitive que l'on a du fonctionnement de l'instruction, c'est pourquoi on parle de sémantique.

Les pré et post-conditions des instructions dans les langages de programmation séquentielle sont exprimées en termes de valeurs des variables, car la programmation informatique de ce niveau se limite à la manipulation de variables. Mais pour ce qui est des ACLs, on travaille à un niveau supérieur. En fait, les pré-conditions et post-conditions des ACLs devront être exprimées en termes d'attitudes mentales des agents en présence. En effet, les ACLs et leurs sémantiques doivent co-exister et s'aligner avec les théories mentales et comportementales des agents [Dignum et Greaves, 2000].

Les *sémantiques opérationnelles* sont un autre type de sémantiques mathématiques qui considèrent les énoncés comme des transitions dans une machine abstraite. Dans ce cadre, un énoncé est défini par le changement d'état qu'il opère dans la machine abstraite. Des sémantiques opérationnelles ont été définies pour certains ACLs. Une sémantique opérationnelle donne un sens aux messages en termes d'étapes de calcul (ou réécritures).

On trouve aussi des *sémantiques dénotationnelles*, dans lesquelles chaque élément du langage est mis en relation avec une entité mathématique abstraite appelée dénotation. La sémantique des mondes possibles est un exemple de sémantique dénotationnelle. Pour qu'une sémantique dénotationelle soit utile, il faut qu'il soit possible de déterminer le sens sémantique d'un énoncé en fonction du sens de ses éléments, une propriété appelée compositionna-

[1] Voir, par exemple, la méthode des triplets de Hoare [1972].

[2] Et ce, quelque soit la valeur dénotée par x auparavant.

lité. Cette propriété est absente dès qu'un énoncé a une dénotation qui ne correspond pas à la composition de celles de ses composants.

Finalement, les sémantiques de la théorie des jeux peuvent aussi s'avérer utiles dans certains cas. Dans celles-ci, chaque énoncé bien formé est associé à un jeu à deux joueurs, un protagoniste et un antagoniste. Un énoncé est alors considéré vrai, si et seulement si, il existe une stratégie gagnante (série de coups qui sont gagnants, quelque soit la réponse de l'antagoniste) pour le protagoniste.

Annexe B

Autres théories motivationnelles en psychologie sociale

On présente ici des théories motivationnelles (ou connexes) autres que celle de la dissonance cognitive. Ces théories sont vues ici comme des théories du changement d'attitude dans un contexte de communication passive (l'individu ne fait qu'écouter et percevoir les autres). Ces théories ont pour la plupart été formalisées [Hunter et al., 1984]. Les sous-sections suivantes décrivent brièvement : la théorie de la balance, la théorie de la congruence, la théorie du renforcement, la théorie du traitement de l'information, la théorie du jugement social, l'approche de Rokeach, la théorie de la consistance cognitive, la théorie du management de l'impression et la théorie de la réactance psychologique.

B.1 La théorie de la balance

Initialement pensée par Heider dans les années 40, c'est Newcomb [1953] qui rendra la théorie de la balance populaire. Selon cette théorie, il y a balance cognitive lorsque les éléments sont mutuellement cohérents entre eux. L'hypothèse principale en est : l'individu préfère les relations balancées entre les éléments aux relations non-balancées. Heider [1958] a étudié la cohérence entre deux ou trois éléments. Même si des réseaux de plus grandes importances ont été étudiés [Cartwright et Harary, 1956], nous ne détaillerons ici que les triplets d'éléments. La figure B.1 présente les huit configurations possibles entre deux individus P et O et un objet X. Si le nombre de relations négatives est impair, le triplet est dit non-balancé, tandis qu'il est dit balancé si le nombre de relations négatives est nul ou pair. La principale conséquence d'un système non-balancé est la motivation à restaurer la balance.

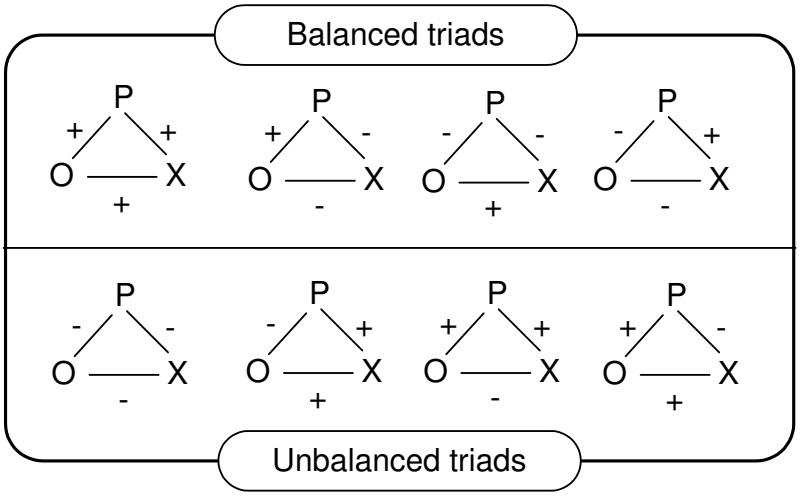

FIG. B.1 – Triplets balancés et non-balancés dans la théorie de la balance, d'après Newcomb [1953].

Rosenberg et Abelson [1960] indiquent que c'est la solution la plus simple qui sera sélectionnée. On retrouve ici le principe d'économie, central en psychologie. Cette simplicité est fonction de la solidité des relations considérées. Une relation plus solide, parce que plus ancienne ou plus importante ou vérifiée, sera moins facilement modifiable ou suppressible qu'une relation plus fragile et moins bien établie. Un grand nombre de termes ont été utilisés pour définir les procédures par lesquelles un tel rétablissement est possible (en anglais, on trouve : transcendance, bolstering, misconstrual, autism, ...). Elles se résument cependant à trois possibilités. Pour restaurer la balance l'individu peut : (1) changer le signe d'une des relations (2) supprimer une relation et ainsi changer la structure du système cognitif (3) différencier les attributs positifs et négatifs de l'objet ou de la personne. Par exemple, si l'individu considéré aime le café et fait confiance à son docteur mais que ce dernier lui déconseille le café car la caféine est nocive pour lui, nous avons un triplet non-balancé. L'individu peut alors réagir de diverses manières et, par exemple, dédoubler le triplet en deux triplets balancées : un pour le café décaféiné et un pour le café caféiné.

Les expérimentations ont confirmé et raffiné la théorie de la balance cognitive en ajoutant deux autres effets à celui de la désirabilité de la balance : (1) l'effet d'attraction indique que l'individu préfère les triades pour lesquels la relation $P-O$ est positive (indépendamment de leur relation à X) et (2) l'effet d'accord qui indique que l'individu préfère que $P-X$ et $O-X$ soit en accord (indépendamment de $P-O$). La principale explication proposée pour rendre compte de ces résultats ultérieurs est que par économie (cognitive, de temps ou d'intérêt), l'individu considère d'abord la simple relation $P-O$ puis le couple de relations $P-X$ et $O-X$ puis finalement le triade complet et la balance émerge ainsi de la complexification de

son raisonnement. Cette échelle a été vérifiée expérimentalement notamment en donnant un temps croissant aux sujets pour indiquer leurs préférences parmi différents triades.

B.2 La théorie de la congruence

La théorie de la congruence, reprend et approfondit la théorie de la balance décrite précédemment. La théorie de la congruence a été initiée par Osgood [1963] pour introduire de la variabilité dans les relations et rendre compte de degrés plus riches que le simple $+$ ou $-$. L'introduction de graduation dans les relations considérées permet à cette spécialisation de la théorie de la balance des prédictions plus précises et spécifiques que cette dernière. Comme la théorie de la balance, cette théorie est essentiellement une théorie de la cohérence émotionnelle et affective.

B.3 Théorie du renforcement

L'idée principale de la théorie de renforcement est que dans un cadre de communication passive, pour un individu, les messages cohérents avec ses attitudes vont renforcer celles-ci tandis que les messages inconsistants avec ses attitudes vont les affaiblir [Miller et Dollar, 1941]. Cette théorie repose sur les trois hypothèses suivantes : (1) une hypothèse forte d'influence sociale qui postule qu'un individu gère ses attitudes par imitation avec celles des autres, (2) la formation ou le changement d'attitude peut être vu comme un cas particulier d'apprentissage stimulus-réponse (3) la formation ou le changement d'attitude peuvent-être abordés sous l'angle des théories de la communication de masse (qui traitent des influences et des conditionnements de masse).

Le changement d'attitude comme réponse au stimulus peut-être envisagé de deux manières : comme changement d'attitude personnel ou comme changement d'attitude et réponse émotionnelle envers la source. Dans tous les cas, l'idée reste la même : un message aligné avec les attitudes renforce celles-ci ou entraîne une réponse émotionnelle et une attitude positive envers la source alors qu'un message contradictoire affaiblit les attitudes contredites ou entraîne une réponse émotionnelle et une attitude négative envers la source.

Dans ce modèle de base, tous les individus réagissent de la même façon aux messages ce qui est une hypothèse trop forte (contredite par bon nombre d'expériences ainsi que par le

quotidien de chacun). Ainsi, de nombreuses recherches ultérieures ont raffiné les différents types de réponses qui peuvent être envisagés.

B.4 Théorie du traitement de l'information

Pour Hovland et al. [1953], spécialisés dans la psychologie de la communication et les théories de la persuasion, le changement d'attitude découle de la comparaison du sujet entre sa position et la position défendue par les messages reçus. Le récepteur se pose des questions et compare ses réponses avec celles proposées dans les messages reçus. Soit m la valuation (évaluation qualitative) de l'attitude défendue par le message reçu et a la valuation de l'attitude présente chez le récepteur alors $\delta(a) = m - a$ indique si c'est le changement d'attitude (valeur positive) ou bien la tentative de persuasion (valeur négative) qui est décidée.

B.5 Théorie du jugement social

Cette théorie, initiée par Sherif et Hovland [1961] et leurs collègues de l'école de Yale, étudie la manière dont les humains jugent les messages. Ces recherches ont été développées sur la base de recherches en psychophysique. La psychophysique s'intéresse à la façon dont l'individu juge certaines valeurs physiques des objets perçus : le poids, la luminosité, la taille, ... Ces recherches montrent que les jugements sont établis sur la base de points de références, appelés ancres. Cette idée d'ancre est facilement vérifiable à l'aide de l'expérience suivante : plonger une main dans l'eau chaude et l'autre dans l'eau froide pendant un moment, puis plonger les deux mains dans l'eau tiède, les sensations des deux mains sont alors très différentes. Cela s'explique par le fait que les deux mains ont des ancres différentes. L'idée de la théorie du jugement sociale est que le jugement des messages s'effectue de manière similaire avec des ancres internes issues des expériences passées. En outre, plus une question est importante ou pertinente pour l'ego (lui aussi, issu des expériences passées), plus l'ancre va influencer l'interprétation.

Le Q-classement est l'expérience classique de cette théorie. On fournit à l'individu un ensemble d'affirmations sur un sujet donné que celui-ci doit classer en groupes, selon les similarités qu'il détecte. Le nombre de groupes est libre et une fois le classement fait, il doit ordonner les groupes du positif au négatif pour finalement indiquer lesquels lui semblent acceptables, inacceptables ou neutres. Cela définit un intervalle d'acceptabilité, un intervalle

de non-engagement et un intervalle d'inaceptabilité. Ces intervalles sont fondamentaux dans la vie quotidienne de chacun.

L'interprétation des messages reçus par l'individu peut être distordue de deux manières différentes :

- *par effet de contraste* : cet effet survient lorsque l'individu juge le message comme étant plus éloigné de son point de vue qu'il ne l'est en réalité ;
- *par effet d'assimilation* : cet effet survient lorsque l'individu juge le message plus proche de son point de vue qu'il ne l'est en réalité.

Les expériences montrent qu'un message sera traité par assimilation s'il est proche du point de vue de l'individu et par contraste sinon, le tout étant pondéré par l'importance que lui accorde l'ego. Concernant les changements d'attitudes, la théorie fait les prédictions suivantes :

1. Les messages qui « tombent » dans l'intervalle d'acceptabilité facilitent le changement d'attitude ;
2. Un message perçu comme étant dans l'intervalle d'inacceptabilité n'aura pas ou peu d'effets sur les attitudes, il sera ignoré ou rejeté. Pire, un effet « boomerang » peut au contraire se retourner contre le message en agissant dans le sens opposé à celui-ci dès lors qu'il est discrédité ;
3. Dans l'intervalle d'acceptabilité et l'intervalle neutre, plus un message est loin du point de vue du récepteur plus des changements d'attitudes sont prévisibles ;
4. Finalement, plus l'engagement de l'ego est fort sur le sujet du message, plus l'intervalle de rejet (d'inacceptabilité) est grand, plus l'intervalle d'acceptabilité et l'intervalle neutre sont petits et donc moins les changements d'attitudes sont probables. Ceci explique le fait que les individus dont l'ego est très engagé sur un point donné sont difficiles à persuader.

B.6 Approche de Rokeach : croyances, attitudes, valeurs

Dans la théorie de Rokeach [1969], les croyances sont les centaines de milliers d'assertions que l'individu fait sur le monde et sur lui-même. Les croyances sont organisées en terme

de centralité et d'importance pour l'ego. Le noyau central de ce système est constitué d'un ensemble de croyances bien établies qui sont relativement difficiles à modifier (on dit qu'elles ont une forte résistance au changement). Plus un élément de croyance est central, plus son changement porte à conséquences.

Les attitudes sont vues comme des dispositions à agir d'une certaine manière envers un objet. Chaque attitude consiste essentiellement dans un ensemble de croyances à propos de l'objet. Pour Rokeach, il y deux types d'attitudes qui doivent toujours être considérées ensemble : les attitudes envers l'objet et les attitudes envers la situation. Le comportement d'un individu dépend toujours d'une combinaison de ces deux types d'attitudes. Dans une situation donnée, si l'individu n'agit pas de manière consistante avec ses attitudes envers les objets, c'est sans doute que des attitudes envers la situation l'en empêchent. Par exemple, on mange des choses que l'on n'aime pas lorsque l'on est en position d'invité.

Des trois éléments explicatifs du comportement humain, les valeurs sont - pour Rokeach - le plus important. Les valeurs sont des types de croyances centrales particulières qui servent de guide de vie. Il en distingue deux types :

- *Valeurs instrumentales* : ce sont des guides de vie pour le quotidien (importance de travail, loyauté, ...) ;
- *Valeurs terminales* : ce sont les buts ultimes cherchés dans la vie (joie, santé, ...).

Rokeach donne aussi beaucoup d'importance à la notion d'estime personnelle (ensemble des croyances se rapportant à soi).

Finalement, il s'agit d'une théorie de la cohérence puisqu'elle affirme que l'individu est guidé par son besoin de consistance et que l'inconsistance crée une pression qui incite au changement. Cependant, la consistance définie par Rokeach est extrêmement complexe. Pour lui, les inconsistances les plus importantes sont celles qui impliquent les croyances sur soi du fait que le but premier du système est la maintenance d'une bonne estime de soi (cela passe par la consistance).

B.7 Théorie de la consistance cognitive

Cette théorie due à McGuire [1960] et complétée par Wyer et Golberg [1970] étend la logique formelle pour prendre en compte les probabilités subjectives autres que 0 ou 1. Ces

chercheurs, psychologues, ont établi que les croyances deviennent de plus en plus localement consistantes. Cette augmentation temporelle de consistance est due aux observations ayant rapport aux croyances concernées et aux croyances connexes. Cette théorie a été prévue comme une théorie statique (à l'instar des théories logiques formelles dont elle s'inspire), c'est-à-dire comme capable de prédire des croyances à partir d'autres croyances. Du fait de cette monotonie, elle ne prend pas en charge les changements d'attitudes.

B.8 La théorie du management de l'impression

En psychologie sociale, la théorie du management de l'impression s'intéresse à la manière dont l'individu présente une image de lui aux autres pour parvenir à ses buts. Goffman [1959] en est le père. La plupart des analyses du management de l'impression supposent que le but premier des interactions de l'individu avec d'autres est l'approbation sociale.

La théorie du management de l'impression conforte les résultats de la dissonance cognitive. Les travaux de Schlenker [1980], en particulier, indiquent qu'indépendamment du fait que l'individu ai un besoin psychologique de cohérence, l'apparence de cohérence entraîne une récompense sociale tandis que l'apparence d'incohérence entraîne une sanction sociale. Schlenker ajoute que faire preuve de cohérence dans les attitudes exprimées donne une impression désirable d'équilibre et de prédictibilité qui entraîne confiance, respect et appréciation positive de la part d'autrui.

B.9 La théorie de la réactance psychologique

Due à Brehm et Brehm [1981], la théorie de la réactance psychologique s'intéresse au mouvement de réaction d'un individu lorsqu'il se sent privé de liberté. La réactance psychologique est la force motivationnelle qui conduit l'individu a essayer de recouvrir une liberté menacée. Une expérience réalisée en 1977 par Petty et Cacioppo montre, par exemple, que l'effet de contre-argumentation est plus fort lorsque l'individu est prévenu que le discours entendu est à but persuasif qu'autrement. La réactance psychologique s'exprime alors par : (1) la réaffirmation du comportement ou de l'attitude menacée, (2) une contre-argumentation des raisons de la restriction et (3) une ré-évaluation positive des attitudes touchées par la menace. C'est une théorie de psychologie inverse, complémentaire des théories présentées dans les sections précédentes. En indiquant que la croyance explicite que le message est à but persuasif renforce les attitudes attaquées, cette approche fait le lien entre les théories de

la persuasion (qui mettent en garde contre l'effet boomerang de la persuasion explicite) et les théories motivationnelles. La persuasion est alors entendue au sens péjoratif du terme, c'est-à-dire vue comme une tentative de manipulation et donc de privation de liberté contre laquelle la cohérence interne doit être protégée.

B.10 Conclusion

Dans cette annexe, nous avons présenté certaines des théories connexes à celle de la dissonance cognitive sur laquelle nous nous basons pour notre approche des aspects cognitifs des communications entre agents. Sans être complète, tant s'en faut, cette annexe met en évidence qu'il s'agit en fait d'une famille de théories, dont la plupart donnent encore lieu à de nombreuses recherches. La convergence des résultats, ainsi que la richesse du support expérimental de ces théories ajoutent évidemment à la pertinence de choisir ce type d'approches, issues des sciences cognitives et en particulier de la psychologie sociale, comme fondation.

Annexe C

Transfert de connaissances : attitude et changement d'attitude en psychologie sociale

Dans notre approche pour la modélisation des agents artificiels cognitifs, nous utilisons des théories et résultats issus des recherches de psychologie sociale sur le changement d'attitude et la persuasion. Cette annexe[1], présente un certain nombre de pré-requis qui fondent notre modèle théorique de la pragmatique des communications présenté au chapitre 6 et sa validation informatique, présentée au chapitre 7. Nous pensons également que ce transfert de connaissance à un intérêt plus général comme source d'inspiration pour les modèles d'agents artificiels. En particulier, comme nous l'avons précisé en section 3.5, ces éléments issues de psychologie sociale et cognitive, validés expérimentalement, offrent une perspective complémentaire aux apports de la philosophie de l'esprit dont sont issues les modèles d'agents cognitifs actuels.

Dans cette annexe, nous synthétisons des aspects complémentaires concernant les concepts et mécanismes du changement d'attitude. Les principaux résultats de psychologie sociales concernant l'origine des attitudes et les liens entre attitudes et comportements manifestes sont détaillés.

[1] Les éléments de cette annexe sont issus de notre travail de recherche et de lecture dirigée Pasquier [2003], accomplie sous la direction du Pr. Paquette, chercheur et enseignant en psychologie sociale au département des sciences de la communication de l'université Laval.

C.1 La notion d'attitude

La modélisation classique des attitudes [Erwin, 2001] consiste en un modèle triadique, nommé *abc* pour *affect*, *behavior*, *and cognition*. L'affectif, le comportemental et le cognitif sont donc les trois composantes du modèle *abc*. L'*affectif* qui désigne le contenu émotionnel des attitudes peut être vu comme déterminé sur une échelle polarisée du négatif au positif. Le *cognitif* traduit la perception subjective des relations entre les objets psychologiques d'intérêts (cette notion d'intérêt rend compte du principe d'économie, omniprésent en psychologie). L'aspect *comportemental* consiste juste à agir de manière cohérente avec nos attitudes. En fait, et ce sera l'un des points développés dans la suite de cette annexe, il s'agit plus d'une prédisposition à agir dans le sens des attitudes, celle-ci pouvant succomber à de nombreux facteurs de situation, de contexte, d'influence et de contrôle social. Ces trois facettes des attitudes sont intriquées de manière souvent complexe, mais consistante.

La consistance interne joue un grand rôle dans le modèle *abc*. En particulier, les aspects émotionnels et cognitifs sont fortement corrélés. Les études de Rosenberg [1960], notamment, montrent bien que si la composante affective est modifiée, on observe les changements correspondants en terme de cognition et vive-versa. Les liens entre les comportements et les deux autres composantes sont moins clairs. Si théoriquement le comportement de l'agent est déterminé par ses attitudes, ce résultat a d'abord été mis à mal par les résultats contradictoires de nombreuses expériences.

C.2 Attitudes et comportement manifeste

L'étude des liens entre attitudes et comportement est le talon d'Achille du domaine. En effet, contrairement aux attentes, peu de résultats non contredits ont été proposés dans ce domaine jusqu'à récemment.

L'étude de La Piere [1934], une expérience célèbre et fondatrice de l'étude des liens entre attitudes et comportements, se déroule en deux étapes :

– un voyage (1930-1932) au travers des États Unis avec un couple d'étudiants chinois, visitant 67 hôtels et campings ainsi que 184 restaurants. Lors de ce voyage, seul un camping leur a refusé le service et les attentions supplémentaires (lorsqu'il y en avait) étaient plutôt favorables ;

– un questionnaire envoyé à chacun des lieux visités six mois après le voyage. Ce questionnaire incluait la question : accepteriez-vous une personne de race chinoise à séjourner dans votre établissement. Il ne récupérera les réponses que de 81 restaurants et 47 hôtels. 92% et 91%, respectivement, répondirent par la négative à la question cruciale. Toutes les autres réponses sauf une étaient : « incertain, cela dépend des circonstances ». Finalement, seul un propriétaire de camping répondit oui.

La Piere conclut que les questionnaires ne sont pas un moyen fiable de mesurer les attitudes. Cependant, l'étude fut généralement citée pour indiquer la faiblesse des liens qui unissent attitudes et comportements.

Bien que de nombreux théoriciens ont avancé l'idée d'une forte corrélation entre attitudes et comportements, les revues sur le sujet ont longtemps montré le contraire : en 1969, l'étude de Wicker [1969] indique que seulement 10% de la variabilité des comportements peut être prédite par les attitudes. La faiblesse de le corrélation fait penser à certains que la notion même d'attitude est inutile.

Mais les études et les méthodes se raffinant et en 1980, Hanson [1980] indique que 65% des 46 laboratoires dont il synthétise les résultats ont conclu une forte corrélation. Selon Erwin [2001], cette évolution est due aux progrès des méthodologies expérimentales en laboratoire (moins de biais de situation et meilleur contrôle des attitudes conflictuelles). Par exemple, rien n'indique dans l'étude de La Piere que les propriétaires répondant aux questionnaires soient ceux qui ont accueilli les trois voyageurs. En outre, les attitudes sont supposées et mesurées de manière très générale et isolée, tandis que les comportements sont toujours très spécifiques.

En dernière analyse, ces études montent bien qu'il n'existe pas de relation simple et régulière entre attitudes et comportement et que d'autres facteurs doivent être pris en compte. La section suivante détaille certains de ces facteurs.

Contraintes personnelles, situationnelles et intentions individuelles

Les attitudes ne sont pas des objets isolés, elles sont entre autres liées aux autres attitudes et forment un système dans lequel plusieurs alternatives en compétition peuvent co-exister. Pour revenir à l'étude de La Piere, il est possible que l'attitude d'être poli et serviable dans une situation de face-à-face avec le client, ou encore que l'appât du gain propre au métier de commerçant l'ai emporté lors du comportement du réceptionniste sur ses éventuels pré-

jugés envers les asiatiques. Plus généralement, il est clair que certaines situations inhibent l'expression comportementale de certaines attitudes pour en favoriser d'autres.

C'est pourquoi de nombreuses études préfèrent utiliser les *intentions comportementales* plutôt que les comportements effectivement observés pour traiter du lien attitude-comportement. Les intentions comportementales sont effectivement moins susceptibles d'être distordues par la prise en compte de la situation et des contingences qui amènent au comportement observé. On peut très bien, par exemple, avoir l'intention d'acheter le super-savon bio après s'en être fait venter les mérites, mais le manque de moyens, le fait que l'on ne le trouve pas disponible en stock chez notre détaillant habituel ou le manque de temps peut bloquer le comportement sans invalider l'intention.

C'est pourquoi les psychologues ont introduit la notion d'intention comportementale comme tampon entre les attitudes et le comportement manifeste. Le lien entre attitude et comportement a donc été raffiné en : un lien entre les aspects cognitifs (et affectifs) des attitudes et les intentions comportementale et le lien entre le intentions et le comportement manifeste. Dans ce cadre plus riche, il a alors été établi que les attitudes sont de bonnes prédictrices des intentions mais que le réalisation de ces intentions est plus complexe à formaliser et dépend :

- *des contingences et de la situation* : certaines situation empêche l'expression comportementale d'une intention ;
- *spécificité de l'intention* : plus une intention est spécifique et précise, plus son accomplissement directe est probable. Ainsi, dès lors que la généralité et l'imprécision sont des obstacles à l'action, les intentions doivent être spécifiées de la manière la plus spécifique et précise possible ;
- *proximité temporelle de la réalisation* : plus court est le laps de temps entre le moment où l'intention est adoptée et l'occurrence de sa réalisation, plus cette réalisation est probable ;
- *type d'intention* : il est clair que les intentions qui dépendent ou qui concernent le comportement des autres sont plus complexes à accomplir puisque cette réalisation ne dépend pas simplement de l'agent possédant l'intention ;
- *origine de l'intention* : les intentions découlent des attitudes par délibération, les psychologues ont montré que les intentions issues d'attitudes acquises par perception directe seront plus probablement accomplies que celles acquises autrement (par cognition ou par communication) ;
- *l'importance subjective de l'intention* : l'importance subjective des attitudes et des intentions résultantes module l'intensité du lien avec le comportement ;

– *récompense matérielle ou punition associées* : plus la récompense matérielle associée à l'intention est élevée, plus sa réalisation est probable [Krishman et Smith, 1998].

Plus généralement, les liens entre attitudes et intentions ainsi que ceux entre les intentions et leur réalisation ont été approfondis par les psychologues. Le lecteur intéressé pourra consulter l'état de l'art de Petty et Cacioppo [1996] à cet égard.

C.3 Comportement anti-attitudinal et changement d'attitudes

Si, avec les réserves que nous avons vues, les attitudes dirigent les intentions comportementales et les comportements, il peut être intéressant d'étudier qu'elles peuvent être les conséquences d'un comportement anti-attitudinal sur les attitudes. Cela a été le sujet d'intérêt à l'origine d'une famille de théories regroupées sous le nom de théories de la cohérence cognitive. Ces approches considèrent que l'agent essaie de maintenir une balance, un équilibre entre les diverses composantes cognitives (cognitions composant les attitudes et comportements). L'inconfort et le stress créés par les déséquilibres sont source de changements. La théorie de la dissonance cognitive, notamment, a eu un impact majeur sur la compréhension de la manière dont les désaccords entre attitudes et comportements sont résolus. Concentrons-nous sur ce qui fut l'élément révélateur historique de l'importance de cette approche.

En 1959, Festinger et Carlsmith Festinger et Carlsmith [1959] réalisent la fameuse expérience des 20$/1$. Cette expérience, dissimulée comme un test de performance auprès des cobayes, se déroule comme suit : les participants devaient accomplir deux tâches pénibles d'une demi-heure chaque. Après quoi, une somme de 1$ ou 20$ leur était offerte pour les inciter à convaincre un futur participant que l'expérience est passionnante (sauf un 30% de participants, groupe de contrôle pour cette dernière tâche, qui n'ont rien reçu eu n'ont pas eu à convaincre qui que ce soit). Finalement, les participants sont interviewés (de manière apparemment indépendante et non-corrélée aux deux premières étapes de l'expérience) sur leur participation à cette expérience. Il s'agit donc d'étudier comment la récompense influence leur propre évaluation du comportement anti-attitudinal (dire du bien d'une expérience dont on pense du mal). Seuls les sujets ayant reçu 1$ ont effectué un changement d'attitude (pour justifier leur comportement incohérent) et trouvent finalement le moyen d'apprécier l'expérience. Les individus du groupe de contrôle et les « biens payés » ont gardé leur attitude négative envers la pénible tâche effectuée.

De manière générale, les attitudes tendent à expliquer les comportements, même si dans le monde réel de nombreux facteurs et la complexité de leurs intrications peuvent rendre cette relation moins visible et lisible. Parallèlement, et c'est le résultat levé par Festinger et Carlsmith, certains comportements anti-attitudinaux (notamment, s'ils ne sont pas justifiés autrement) peuvent déclencher des changements d'attitudes. Résultat lourd de conséquences.

Conditions du changement d'attitude

Festinger a utilisé sa théorie pour faire un certain nombre de prédictions non intuitives concernant le changement d'attitude. En outre, la théorie a été raffinée et reformulée au cours du temps [Brehm et Cohen, 1962; Aronson, 1968; Wickland et Brehm, 1976]. Comme les autres théories motivationnelles, le mode de représentation des éléments cognitifs et de leurs relations (qui peuvent être différentes des relations logiques ou physiques « réelles » entre ces éléments) permet à la théorie de la dissonance cognitive de déterminer les attentes individuelles subjectives. Ainsi, elle permet d'expliquer comment certains comportements peuvent être rationalisés . Les résistances au changement des cognitions sont particulièrement déterminantes à cet égard. Festinger a montré que l'effet d'éléments externes dissonants avec des croyances internes particulièrement fortes (associées à une forte résistance au changement) peut être de ne pas en tenir compte et de trouver d'autres éléments d'explication pour au contraire renforcer la croyance en question. Ces phénomènes sont désormais bien compris et acceptés dans les théories de la communication et leur impact sur les théories de la persuasion est considérable.

En fait, pour que le changement d'attitude survienne il faut : (a) que l'attitude soit bien établie (b) qu'il y ai des engagements sociaux qui soient associés à ce changement (c) qu'il y ai disconfirmation et (d) que le support social pour le changement soit suffisant. Dans une étude célèbre, réalisée par Festinger sur les membres d'une secte, c'est ce dernier point qui n'est pas présent chez les sujets impliqués dans une secte. En présence d'informations anti-attitudinales, c'est l'isolement « collectif » des membres de la secte et l'absence d'un support social pour le changement d'attitude qui les amène à former un réseau de justifications alternatives pour justifier leurs croyances et rétablir la consonance (ma croyance doit être vraie puisque les autres y croient aussi). Si la disconfirmation de croyances fortes est assez rare, la prise de décisions (de diverses importances) est commune.

Les recherches ont également montré que la dissonance serait plus forte après un choix entre deux alternatives de désirabilités comparables (et importantes) qu'après un choix entre alternatives de désirabilités très différentes. En effet, dans ce dernier cas, la décision peut être justifiée par l'abondance d'éléments positifs en faveur de la solution retenue par rapport à celle rejetée. Dans le premier cas, les possibilités étant comparables en désirabilité, la

dissonance induite par le choix sera plus grande. C'est un cas particulier du phénomène de justification insuffisante (ce phénomène sera présenté section 8.2.3).

Le fait que les cas de justification insuffisante entraînent un plus grand changement d'attitude que les autres découle de ce que la dissonance provoquée est plus intense. Ce résultat, confirmé par l'expérience du 1$/20$, est tout à fait contre intuitif dans la mesure où il indique que les gens qui acceptent d'avoir un comportement anti-attitudinal vont produire un changement d'attitude les alignant à ce comportement dans la mesure ou les justifications sont insuffisantes (sinon les justifications seules peuvent justifier le comportement et il n'y a pas dissonance) tandis que les théories classiques, « hédonistes », indiquent que le changement d'attitude aura lieu si les récompenses (les justifications) sont suffisantes. En fait, les études ultérieures ont montré la nécessité de différencier deux types de changements d'attitudes (introduits section 8.2) : le changement d'attitude partiel, justifié externalement et le changement d'attitude complet. Finalement, les aspects temporels et physiologiques de la réduction de dissonance ont été raffinés comme l'indique les sections suivantes.

Aspects temporels de la réduction de dissonance

Goethals et Cooper [1975] ont montré que la réduction n'intervient généralement pas avant que la dissonance apparaisse comme inévitable (ce qui ne signifie pas qu'elle est imminente). Zanna [1975], en montrant que la distraction d'une personne en cours de réduction interrompt et parfois stoppe la réduction en cours, ont confirmé le point de vue de Festinger considérant la dissonance comme un processus de traitement de l'information actif. En effet, dans la lignée des travaux de Riess et Schlenker [1977], Higgins et al. [1979] ont montré que l'individu réfléchit aux modalités de la réduction dès que toutes les cognitions impliquées dans la dissonance lui sont explicitement connues.

En résumé, quand les éléments impliqués sont explicitement connus, la dissonance est détectée, l'individu cherche alors une cause ou une solution externe, qui réclame peu ou pas de réorganisation cognitive de sa part [Zanna et Cooper, 1976]. Si aucune condition de neutralisation ne peut être trouvée alors on s'oriente vers un changement d'attitude [Riess et Schlenker, 1977].

Nature physiologique de la dissonance cognitive

De nombreuses études ont caractérisé les corrélats et effets physiologiques de la dissonance cognitive. Même si les mécanismes physiologiques sous-jacents restent mal compris, les réponses nerveuses correspondant aux tensions attendues ont bien été observées.

C.4 Conclusion

La notion d'attitude est une notion centrale en psychologie sociale. La grande diversité des théories motivationnelles des attitudes, du changement d'attitude ainsi que des théories de la persuasion qui en résultent rendent compte de l'importance de cet apport pour les sciences cognitives.

Pour ce qui est de l'intelligence artificielle distribuée et des systèmes multi-agents, nous défendons la thèse selon laquelle la psychologie sociale offre une source d'inspiration basée sur un savoir empirique et expérimental solide qui vient compléter les analyses hypothétiques de la philosophie analytique généralement prisent comme références (voir section 3.5 à ce sujet).