Philippe Pasquier

Application de théories du langage naturel aux systèmes artificiels.

Directeur :

Prof. Brahim Chaib-draa

Examen de synthèse

Doctorat en informatique
Département d'informatique
Université Laval, Ste-Foy, PQ

Table des matières

Extended abstract

Since cognitive science are basicaly the multidisciplinary study of human intelligence from observations and their interpretations, the language takes a central place in it as the support and regulator of our symbolic and social layer. From that point, it is easy to understand why natural language motivates so many studies from philosophy of language, psycho and socio-linguistics and linguistics. But it may seem less clear why one should use the (obviously approximative) natural language formal theories for artificial systems.

However, many works of AI [Artificial Intelligence], DAI [Distributed Artificial Intelligence], MAS [Multi-Agents Systems], CSCW [Computer Supported Cooperative Work] and HCI [Human Computer Interface] use and produce natural language theories. From the communication view point, we can distinguish three types of artificial systems : machine-machine systems (MAS), human-machines-human (CSCW) and human-machine (HCI). We will illustrate the use of natural language theories in building those kind of systems.

We first present the COSMO MAS communication framework from Wong [62]. This framework is based on the SAT [Speech Act Theory] born in philosophy of language studies. We will show why this natural language theory bring usefull properties for communication in CPS [Cooperative Problem Solving] and CKBS [Cooperative Knowledge Based Systems].

Another use of SAT is the Formal Language for Business Communication from Moore and Kimbrough [40] which is an instance of communication framework for CSCW systems. We will argue the benefits of a general framework like SAT against more specifics ones like EDI [Electronic Data Interchange] or tagged languages.

With the standardized ACLs [Agent Communication Language], the speech act theory is nowadays the standard approach for communication in artificial systems. A number of limits were found with this approach, leading to the search of others (possibly complementary) natural language theories. In the Balkanski and Hurault-Plantet [7] phonedesk human-machine oral interface system another natural language theory based on planning is developed and implemented. We then discuss why this highly domain dependant theory and the empiric integration of the different parts of those systems leads to insatisfaction. As a complementary approach we will browse the common ground theory of Traum and Poesio [55]. This common ground framework is based on more linguistic tools like DRT [Discourse Representation Theory] and explicitly fills some of the previous theorical shortcomings.

Finally, we discuss the relevance and the feasibility of applying those more sophisticated theories to pure machine-machine communication. Then, we conclude this synthesis by remembering the conclusion of a French CNRS report [35] which states the need for a new theorical framework taking into account every aspects of communication, particularly semantics and pragmatics. Language should be considered as a whole expression meaning rather then empirically integrated parts.

Résumé

La pluridisciplinarité est au fondement des sciences cognitives. C'est particulièrement visible autour des méthodes formelles d'interactions. Les chercheurs en interface homme-machine vocale utilisent des notions de philosophie du langage, de philosophie analytique de l'esprit, de psycholinguistique et de linguistique apliquées à l'informatique et à l'intelligence artificielle. C'est le cas du travail de Traum et Poesio [55] sur la formalisation du fond commun (i.e les informations partagées) et de sa structuration dynamique : l'établissement[1]. Le formalisme proposé est une extension de la DRT (un formalisme classique de linguistique computationnelle). C'est le cas aussi pour le formalisme de Balkanski et Hurault-Plantet [7] qui étend les outils de représentation de plans partagés de Lochbaum avec la théorie du discours coopératif de Grosz et Sidner.

Parallèlement, le courant SMA[Systèmes Multi-Agents] cognitif utilise de manière moins approfondie (du fait des contraintes imposées par l'approche SMA) les mêmes éléments théoriques via les LCAs [Langage de Communication Agents]. Moore et Kimbrough [40] utilisent directement la théorie des actes de langage pour le traitement automatique des messages. Le travail de Wong [62] tire lui aussi partie de la théorie des actes de langage comme base pour l'élaboration des communications dans les systèmes de résolution coopérative de problèmes.

Dans ce rapport, on fait la synthèse de ces travaux tout en donnant des arguments justifiant l'utilisation de théories du langage naturel pour l'interaction dans les systèmes artificiels. On discute ensuite des limites des différentes théories présentées et des liens qui pourraient être établis entre ces deux types de travaux. Enfin, un troisième niveau de discours vise à mettre ces quatre travaux en perspective dans le cadre plus général des sciences cognitives.

[1] «fond commun» et «établissement» sont les traductions respective par la communauté francophone de «common ground» et «grounding» dont il est difficile de rendre excactement compte en français.

1 Introduction

On définit souvent les sciences cognitives comme l'étude de l'intelligence humaine à partir de ses manifestations observables et interprétables. De ce point de vue, il s'agit en grande partie d'analyser et de modéliser les activités et processus de perception, de mémorisation ou d'apprentissage qui dépendent du langage en tant que système support et surtout régulateur de nos expressions symboliques «supérieures» et socialisées.

Parallèlement aux recherches fondamentales de linguistique et de philosophie du langage, un certain nombre de travaux d'IA [Intelligence Artificielle], d'IAD [Intelligence Artificielle Distribuée], de SMA [Systèmes Multi-Agents] et d'IHM [Interfaces Homme Machine] utilisent et produisent des modèles formels du langage naturel. On peut se demander pourquoi les systèmes artificiels devraient dialoguer à la manière des humains ?

Ce rapport donne des éléments de réponse à cette question en présentant une synthèse des quatre articles suivants :

- Stephen T. C. Wong. COSMO : A Communication Scheme for Cooperative Knowledge-based Systems. *IEEE Transactions on Systems, Man and Cybernetic,* No.23, Juin 1993, Pages 809-824.
- Scott A. Moore et Steven O. Kimbrough. On Automated Message Processing in Electronic Commerce and Work Support Systems : Speech Act Theory and Expressive Felicity. *ACM Transactions on Information Systems,* Vol. 15, No.4, Octobre 1997, Pages 321-367.
- Cécile Balkanski et Martine Hurault-Plantet. Cooperative Requests and Replies in a Collaborative Dialogue Model. *Int. J. Human-Computer Studies.* No.53, Academic Press, 2000, Pages 915-968.
- David R. Traum et Massimo Poesio. Conversational Actions and Discourse Situations. *Computational Intelligence.* Vol. 13, No.3, 1997, Pages 1-53.

La communication dans les systèmes artificiels prend trois formes : machine-machine (SMA), homme-machine(s)-homme (Médiatisation), homme-machine (IHM). Dans ces systèmes, rien n'oblige à ce que la communication soit structurée comme dans les sociétés humaines. Pourtant, outre l'intérêt en terme de validation pour les sciences cognitives, un certain nombre d'arguments font penser que des interactions inspirées de celles des humains bénéficient au système. Nous verrons comment la théorie des actes de langage (section 2) peut être appliquée aux systèmes multi-agents avec le cadre de communication COSMOS de Wong [62](section 3) et aux systèmes de médiatisation avec les travaux sur le traitement automatique des messages de Moore et Kimbought [40](section 4). Après avoir discuté d'une manière plus générale les limites des langages de communication agents (section 5), on présentera deux autres théories plus complètes pour le traitement du langage naturel oral dans les IHM (section 6). D'une part le modèle du dialogue coopératif développé par Balkanski et Hurault-Plantet [7] puis, de manière complémentaire, la théorie du fond commun de Traum et Poesio [55]. Pour chacun de ces types d'application, des arguments justifiant l'utilisation de théories du langage naturel seront fournis.

Enfin, nous discuterons de l'application de ces théories plus raffinées pour le langage naturel oral aux systèmes multi-agents (section 7) avant de conclure par une mise en perspective de ces travaux dans le cadre plus général des sciences cognitives.

2 Théorie des actes de langage : «Quand dire c'est faire.»[2]

La théorie des actes de langage [4, 21, 45, 60] est issue de la philosophie du langage. Cette théorie rend compte du processus de communication entre agents cognitifs au sein d'une société : le langage par l'entremise de la langue. 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. 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* : l'interlocuteur (le ou les) a perçu l'énonciation. Il lui faut interpréter le sens de l'énoncé en terme de 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 interprété 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) ; 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 «inférentielles» [6] et d'autres plus «analytiques» [46].

2.1 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) consiste 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 texte. Par exemple, les

[2]Selon le titre de l'édition française [5] de l'ouvrage de J.L Austin [4].

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 (i.e. force illocutoire d'ordre et force illocutoire d'assertion pour le futur). D'après Searle et Vanderveken [48], 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 conditions de sincérité. Détaillons cela.

Le but illocutoire

Le but illocutoire relie la proposition énoncée au monde réel. Pour Searle, il existe cinq utilisations du langage possible, ce sont 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. *comissif / promissif / engageant* : le locuteur s'engage (vis à vis du destinataire) à accomplir l'action représentée par le contenu propositionnel. Exemples d'actes illocutoires comitifs : promesse, menace, renonciation, acceptation, voeux, serment...
4. *expressif* : le contenu propositionnel concerne l'humeur mentale et l'affect du locuteur. Exemples d'actes illocutoires expressifs : déclarations d'amour, félicitations, remerciements, insultes...
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 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 chose 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.

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 a satisfaire son contenu propositionnel. Les actes de langage ayant un but illocutoire directif ou comissif 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.
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.
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 toujours satisfait puisqu'il n'a pas affaire aux choses représentées mais à l'état mental du locuteur.

Au niveau linguistique, la théorie des actes de langages classique prévoit qu'il est possible de déterminer le type 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. Seule l'utilisation «d'expressions performatives explicites» de la forme «Je te <verbe performatif> que ...» donne le type de l'acte réalisé sans ambiguïté (et encore, on peut 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 langages indirects. Par exemple, l'énoncé S_1 s'interprète en situation 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, quand à l'étude de ce type d'actes, et ce même si les travaux de Grice [22] sur la notion d'implicature fournissent des indications sur le type d'indirection réalisée pour passer de la question littérale à la requête indirecte.

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 comissif 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 peuvent être associés un désir, un regret, une inquiétude...De tels états mentaux sont des attitudes propositionnelles $m(p)$ où :

- p est le contenu propositionnel
- 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 état mentaux qu'il exprime) des non-sincères. 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 comissifs 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.

2.1.1 Définition récursive de la force illocutoire

On distingue cinq forces illocutoires primitives qui sont les plus simples possibles :

1. *La force illocutoire primitive d'assertion* : cette force possède un but assertif, un mode d'accomplissement, un degré d'intensité et des conditions sur le contenu propositionnel neutres avec comme condition préparatoire que le locuteur ait des justifications ou des raisons de croire en la vérité du contenu propositionnel et comme condition de sincérité que le locuteur croit le contenu propositionnel.
2. *La force illocutoire primitive d'engagement* : cette force a un but comissif, un mode d'accomplissement et un degré d'intensité neutres avec comme condition sur le contenu propositionnel qu'il représente une action future du locuteur, la condition préparatoire que le locuteur soit capable d'accomplir cette tâche et la condition de sincérité qu'il ait l'intention de l'accomplir.
3. *La force illocutoire primitive directive* : cette force est définie par un but directif, un mode d'accomplissement et un degré d'intensité neutres avec la condition que le contenu propositionnel représente une action future de l'interlocuteur, la condition préparatoire que l'interlocuteur soit capable d'accomplir cette action et la condition de sincérité que le locuteur désire que l'interlocuteur accomplisse cette action.
4. *La force illocutoire primitive de déclaration* : cette force possède un but déclaratif, un mode d'accomplissement et un degré de puissance neutres avec la condition que le contenu propositionnel représente une action présente du locuteur, la condition préparatoire qu'il soit capable de l'accomplir (avec preuves) et la condition de sincérité qu'il en ait l'intention.
5. *La force illocutoire primitive expressive*[3] : cette force a tous ses attributs (but expressif, mode d'accomplissement, condition sur le contenu propositionnel, condition préparatoire, condition de sincérité et degré de puissance) neutres.

Les forces illocutoires non-primitives sont alors obtenues par application des opérations suivantes : ajout de conditions sur le contenu propositionnel, ajout de conditions préparatoires, ajout de conditions de sincérité, restriction du mode d'accomplissement et finalement modulation du degré d'intensité.

2.1.2 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, etc. Tous ces actes prennent la forme $F(F(p))$ ou plus généralement $F(...F(p)...)$ et il n'y a pas de limitations à ce type d'imbrication. En outre, certains actes illocutoires utilisant des connecteurs comme «et», «mais» ne sont pas de la forme $F(p)$. Il s'agit d'actes illocutoires dit complexe qui sont généralement de la forme $F_1(p_1) \wedge F_2(p_2)$. Par exemple A est à la fois une assertion et une question.

[3] Elle est réalisée syntaxiquement par des énoncés exclamatifs en français. Comme il n'existe pas de verbe performatif simple nommant cette force, il n'y a pas d'énoncé exclamatif exprimant seulement la force expressive primitive

A : Il est six heure, non ?

Un certain nombre de connecteurs illocutoires ont été proposés par Searle et Vanderveken dans le cadre de leur logique illocutoire [48].

2.2 Sémantique des actes de langage

2.2.1 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 la locuteur doit être prêt à s'engager à accomplir les actions représentées par le contenu propositionnel dans le futur.

2.2.2 Les conditions de satisfaction des actes illocutoires

La plupart des actes de langages 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 (trois composantes). 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[4].

Les conditions de succès et de satisfaction font partie de la sémantique des actes de langages. Pour analyser la forme logique des actes illocutoires, l'idée est de définir récursivement à la fois leurs conditions de succès et de satisfaction et les conditions de vérité de leur contenu propositionnel.

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

3 Les communications dans les SMAs (machine-machine)

3.1 Discussion : langage naturel et SMA.

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 ?

Il y a évidemment plusieurs courants au sein des recherches concernant les SMAs [Systèmes Multi-Agents]. Nous nous limiterons ici aux SMAs cognitifs, basés sur le paradigme de représentation symbolique des connaissances et de formalisation du raisonnement issue du courant cognitiviste des sciences cognitives. L'interaction dans ce type de système requiert des techniques de communication plus sophistiquées que les solutions traditionnelles (passage de données ou appel de procédures). Même si on se limite, comme c'est souvent le cas, à la classe des systèmes coopératifs, il faut prendre en compte :

1. *l'hétérogénéité des agents* : les messages doivent être mutuellement compréhensibles alors que les points de vues des agents ne sont pas forcément mutuellement consistants ;
2. *l'échange de savoir* : un agent coopératif doit pouvoir manipuler des croyances sur les autres et en particulier sur leurs comportements, croyances et intentions. En effet, il doit être capable d'identifier et expliciter les conflits qui font obstacle à la résolution du problème. Pour ce faire, il doit pouvoir exprimer ces différents types de connaissances et non de simples données.
3. *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 prédéfinis. L'agent doit être capable de développer sa propre stratégie de communication dynamiquement. Notons que ce n'est pas le cas dans les systèmes actuels.
4. *la structure organisationnelle* : Pour éviter l'explosion combinatoire des 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. C'est ce que défend notamment le travail de Wong autour du projet COSMO [62].

3.2 Le cadre COSMO

COSMO [62] est un cadre de communication pour les systèmes de résolution coopérative de problème ou CPS [Cooperative Problem Solving] et les systèmes de bases de connaissances coopératives ou CKBS [Cooperative Knowledge Based Systems]. COSMO est issu des recherches pour le projet BDN [Building Design Network] du Research Center at Lehigh University.

3.2.1 Principes de base et composantes du modèle

Le premier principe est d'isoler la communication de la manière dont les agents représentent leurs connaissances (prise en compte du point 1 ci-dessus). Cela se fait grâce à l'utilisation d'un langage global. Dans COSMO, cela est pris en charge au sein de chaque agent par un module de gestion de connaissances appelé *knowledge handler* capable de coder et décoder les messages du langage de l'agent vers un langage global.

Le second principe est de baser la communication sur la théorie des actes de langages (voir section 2). Ainsi l'échange de savoir sera flexible (prise en compte du point 2 ci-dessus). Dans COSMO, l'idée est de typer les messages avec un nombre limité de forces illocutoires F et puis de se servir des types pour déclencher les réponses appropriées chez les agents. Le type de l'acte sert à sélectionner la procédure qui va le traiter.

3.2.1.1 Rôles organisationnels

Une des originalités de COSMO est de permettre la prise en compte des hiérarchies organisationnelles. Les rôles sont utilisés dans les systèmes de résolution coopérative de problème (CPS) pour : établir le problème à résoudre, segmenter le problème en sous-problèmes pouvant être résolus par de simples agents et coordonner les agents sur les activités et les tâches à accomplir.

Généralement, on distribue (à la conception) un certain nombre de rôles aux agents. Ces rôles rendent compte de la place de chaque agent dans la structure hiérarchique. Dans le cadre COSMO, une valeur est affectée à chaque rôle. Un agent a est un supérieur hiérarchique de b ssi $v(role_a) > v(role_b)$. Les différentes valeurs sont prédéterminées pour chaque agent par une table des rôles connue de tous. Ces rôles organisationnels sont utilisés pour déterminer le type de stratégie communicationnelle ou décisionnelle à adopter face à un autre agent. Par exemple, si un agent a possède un ensemble de stratégies $S_a = \{s_1, ..., s_n\}$. Face à b, l'agent a utilisera la stratégie s_i où i est fonction de la différence hiérarchique entre a et b : $i = v(role_a) - v(role_b)$. Ainsi, on peut imaginer un agent qui donne des ordres à ses subordonnés et des conseils à ses supérieurs hiérarchiques. Ce système permet une prise en compte simple des rôles hiérarchiques dans le choix de l'attitude communicationnelle.

3.2.1.2 Types de messages

Dans COSMO, les agents utilisent deux ensembles disjoints de types de messages. Il y a les messages d'actions et les messages de réponses associés. Pour chaque message d'action, les post-conditions indiquent quels types de messages de réponses peuvent y correspondre. Les types de messages sont donc organisés par paire d'adjacence. Plusieurs paires d'adjacences sont disponibles pour chaque but illocutoire afin de rendre compte des degrés d'intensité. C'est la position hiérarchique relative de l'interlocuteur qui sélectionne l'une ou l'autre [5].

Le modèle opérationnel pour les actes de communication dans COSMO est un modèle de pré/post-conditions classiques[6]. Les préconditions sont : l'honnêteté, le contenu propositionnel,

[5] Dès lors que la table des rôles est connue de tous dans son état le plus récent, le système de messages différents pour marquer les rôles semble inutile.

[6] C'est ce type de modèle qui sert pour la sémantique des langages formels, y compris des LCAs. Nous avons

les rôles et l'accessibilité. Les conditions d'honnêteté et les conditions sur le contenu propositionnel correspondent respectivement aux conditions de sincérité et aux conditions sur le contenu propositionnel de la théorie des actes de langage. Les conditions sur les rôles sélectionnent le type d'acte à utiliser comme nous l'avons vu dans la section précédente. Enfin, les conditions d'accessibilité indiquent que le locuteur doit s'assurer que l'interlocuteur est joignable.

Les postconditions sont que l'émetteur reçoive soit une réponse correspondant à son message soit un message système indiquant un problème dans la communication. Un acte est donc initié quand les préconditions en sont remplies. Lorsque plusieurs actes ont leurs préconditions remplies, ils sont évalués à l'aide d'une fonction d'utilité.

3.2.2 Les stratégies de communication des gestionnaires de connaissances

Chaque gestionnaire de connaissances (*knowledge handler*) enregistre chaque message avec son heure d'arrivée ou de départ et un statut (répondu ou en attente) dans une table de discours. Cette table permet de repérer les messages redondants ou circulaires (lorsqu'un agent reçoit une demande dont il est l'auteur initial).

Un des buts explicite de la communication est de pouvoir s'enquérir d'informations auprès des autres. Encore faut-il savoir qui ! Les agents à qui sont adressées les demandes sont obtenus grâce à une fonction d'utilité définie en fonction du rôle et d'informations statistiques sur les réponses acceptables des agents.

A un moment donné, un agent peut avoir un grand nombre de messages à traiter. Les messages de réponse que doit fournir l'agent sont prioritaires sur les messages de demande qu'il pourrait adresser. Puis les messages de contrôle (notifier un problème ou une circularité...) ont priorité sur les autres. En outre, plus le rôle de l'interlocuteur est élevé dans la hiérarchie, plus son message sera prioritaire. Finalement les messages les plus anciens sont examinés en premier.

Il peut arriver que le système soit tellement occupé que les messages de faible priorité ne soient jamais traités. L'agent avertit alors l'émetteur qui pourra se rabattre sur une autre solution.

3.2.3 Protocoles de communication

Un acte communicatif d'un agent peut en engendrer un certain nombre d'autres dans le système. Un processus de communication entre agents est une séquence d'étapes de communication et de calcul machine. Un processus de communication se termine lorsque l'initiateur de la première communication reçoit une réponse à son message initial. Les séquences ne se mélangent pas grâce à une référence au premier message dans tous les messages ultérieurs d'une séquence. Ces séquences sont des protocoles qui précisent les étapes qui permettent de communiquer des intentions plus complexes que les actes de langages simples. Les protocoles imposent un ordre pour les interactions qui les composent.

Les protocoles proposés sont des protocoles de résolution de conflits classiques : Contract Net, négociation et négociation multi-étapes (discuté dans la thèse de l'auteur [61]). Un protocole

déjà détaillé ces systèmes dans [42].

supplémentaire de négociation basé sur les préférences a été introduit pour gérer les conflits[7] qui paralysent l'action en accord avec la théorie du choix social [2]. Un agent gestionnaire (manager) lance un réseau à contrats (Contract Net) pour récupérer les préférences, les agents construisent alors une liste de préférences selon les critères discutés puis une agrégation est faite par le contractant. En cas d'insatisfaction, les agents peuvent négocier avec le coordinateur.

3.3 Conclusion

Le modèle de préconditions et postconditions et son extension dans COSMO permet de répondre aux questions :

Q_1 : Quand et comment initier un acte de langage A ?

Q_2 : Est-ce que l'acte A a été satisfait ?

Q_3 : Quelle est la relation entre les actes communicationnels et les rôles ?

Q_4 : Comment les actes sont liés les uns aux autres dans un dialogue ?

Q_5 : Est-ce que des intentions plus complexes peuvent être transmises avec les actes primitifs et si oui, comment ?

Pour ce faire, le système de communication COSMO utilise :

- un modèle opérationnel de pré/post-conditions, pour savoir quand un acte peut être employé et quand il a été satisfait (Q1 et Q2) ;
- les rôles organisationnels pour choisir quel type de message utiliser (Q3) ;
- les paires d'adjacences pour jouer sur le type de réponses attendues (Q4) ;
- des fonctions d'utilité pour sélectionner l'attitude communicationnelle préférée (Q1) ;
- des protocoles pour la négociation des conflits et les communications plus complexes (allocation de tâches...)(Q5) ;
- les gestionnaires de connaissances (*knowledge handler*) pour gérer les points ci-dessus, de sorte que des systèmes à base de connaissances différents puissent facilement communiquer ensemble, quels que soient leurs types de représentation et de fonctionnement internes.

Il existe un certain nombre de travaux comparables à COSMO : un projet de Huhns et Arni [28] qui proposent une architecture de systèmes experts distribués communiquant via des modules d'aide qui parlent tous un langage nommé SURF, les travaux de Shoham sur AOP [Agent Oriented Programming] [49] et ceux de Finin et KSI [Knownledge Shared Initiative] [41] qui débouchèrent sur KQML. Tous utilisent aussi des médiateurs qui ressemblent aux gestionnaires (*handler*) dans COSMO. Dans un précédent travail [42], on a présenté les LCAs [Langage de Communication Agents] standards : KQML et FIPA-ACL qui fonctionnent sur les mêmes principes (voir 3.2.1). L'application de théories du langage naturel pour modéliser les communications dans les systèmes multi-agents est donc une réalité motivée (voir 3.1) qui donne des résultats.

[7] Les conflits sont ici définis comme des inconsistances logique.

Comme nous le verrons plus loin, ces résultats sont à nuancer. Mais ce n'est pas tant l'idée d'appliquer une théorie du langage naturel aux SMAs que le choix de cette théorie qui fait l'objet de débats.

4 Langage naturel et système de médiatisation (homme-machine(s)-homme)

Dans les systèmes d'aide au travail ou collecticiels, la machine n'est plus un simple médiateur qui transmet les messages des utilisateurs. Son rôle est celui d'un assistant normatif qui doit permettre et même aider les utilisateurs à se conformer aux règles de l'interaction. Avec le FLBC [Formal Language for Business Communication] c'est exactement le but que se sont fixés Moore et Kimbrough [40]. Cette section présente leurs travaux avant de conclure sur la pertinence de l'utilisation d'outils développés pour le langage naturel à cette fin.

4.1 Un système de traitement automatique des messages pour le commerce électronique et l'aide au travail

Le FLBC est un langage formel basé sur la théorie des actes de langage. La définition de tous les actes possibles est donnée sous forme d'une grammaire formelle en deux parties. Une première partie contient les définitions de base pour la structure des messages et des contenus propositionnels tandis qu'une seconde introduit les buts illocutoires et les prédicats nécessaires pour le domaine d'application. Un acte de langage est représenté par les champs suivants : locuteur, interlocuteur, force illocutoire, contenu propositionnel, contexte et identificateur.

A titre de validation, le FLBC a été appliqué aux communications électroniques dans l'administration militaire américaine. L'étude d'un corpus et une série d'entretiens ont permis à Moore et Kimbrough d'isoler sept types de messages suffisants dans ce cadre. Ces sept types de messages qui servent de base pour le FLBC destiné à l'administration militaire, s'expriment de la façon suivante :

- *lire/évaluer/commenter* : ce type de message permet de demander à quelqu'un de lire, d'évaluer ou de commenter un texte inclus dans le message. Il s'agit d'un message directif, des paramètres optionnels concernant la nécessité d'une réponse et l'enveloppe temporelle sont également disponibles ;
- *rendez-vous* : c'est une demande de rendez-vous qui exige une réponse (dont la forme dépend des rôles respectifs des agents)
  $msg(From, To, request(N), appointement(From, To, Begin, End, Place), Ctxt, ID)$
  où $-5 < N < 5$;
- *dissémination d'information* : il s'agit d'un message pour la diffusion d'information. Pour l'envoyeur, ce type de message est du type «envoi et oubli». C'est au(x) récepteur(s) de décider si le message mérite réponse ou commentaire.
  $msg(From, To, assert, interesting(To, I), Ctxt, ID)$;

- *action de personnel* : c'est un type de message pour la demande d'action qui nécessite souvent une réponse (dans tous les cas l'action doit être effectuée) ;
- *demande d'information* : c'est un type de message pour la demande d'information, $msg(From, To, query, Ctnt, Ctxt, ID)$;
- *absence* : ce message permet de notifier une absence (planifiée ou autorisée), $msg(From, To, assert, absent(From, Begin, End), Ctxt, ID)$;
- *assertion* : ce type de message permet de transmettre une information crue (vraie), $msg(From, To, assert, Ctnt, Ctxt, ID)$;

Le FLBC étant basé sur la théorie des actes de langage, il est articulé syntaxiquement et admet une sémantique précise[8] ce qui facilite l'inférence. Dans le prototype Prolog développé pour l'administration militaire, quatre types d'inférences ont été implémentées :

- *l'inférence pour valider le message avant son envoi* : ce type d'inférence permet de valider le message avant son envoi. Cela permet d'éviter de donner un ordre à un supérieur hiérarchique ou de proposer un rendez-vous dans le passé...
- *l'inférence pour interpréter et traiter les messages à la réception* : l'interpréteur sémantique a accès au contenu sémantique des messages. L'application peut par exemple inscrire dans l'agenda de l'utilisateur les rendez-vous pris ... ;
- *l'inférence au niveau système* : ce type d'inférence permet de déterminer quels sont les messages en attente de réponse et quels sont les messages auxquels on a déjà répondu ;
- *l'inférence au niveau application* : ce type d'inférence permet le traitement des messages reçus et envoyés comme des faits dans une base de données ;

4.2 Comparaisons et justifications

En partant de l'exemple des systèmes d'inscription automatique aux listes de discussion et aux forums, les auteurs ont exploré les possibilités de traitement automatique des messages pour le commerce électronique et les systèmes d'aide au travail. Leur travaux ont été sous-tendus par les considérations suivantes :

- *le besoin de gestion automatique des messages* : le besoin existe déjà, d'ailleurs en 1996, 70000 entreprises utilisaient EDI [Electronic data Interchange] et rien que la SWIFT [Society for Worlwide Interbank Financial Telecommunication] gérait 2.8 millions de messages EDI par jour en 1996. Cela va de pair avec les nouvelles structurations horizontales du marché, en réseaux de partenaires plutôt qu'en grosses structures verticales.
- *le besoin d'en dire plus et de le dire à propos* : il est certain que le système des souscriptions (subscribe) pour les listes de diffusion (mailing lists) est insuffisant pour le commerce en général. Le problème était justement de déterminer ce qui est suffisant.

[8] Voir [32] pour une présentation de la méthode par laquelle les messages FLBC sont transcris en logique du premier ordre.

– *le besoin de fondements théoriques* : des bases théoriques solides sont nécessaires pour la généralité et la robustesse du système. La théorie des actes de langage est acceptée en philosophie, en linguistique et pour les systèmes d'information en intelligence artificielle. Elle n'a dailleurs pas de compétiteur direct.
– *la praticabilité* : à titre de validation, le FLBC a été implémenté pour l'armée américaine.

Pour ce qui est des comparaisons, il existe quatre grandes approches (existantes ou faisant l'objet d'intenses recherches) pour la gestion automatique des messages pour le commerce électronique : la linguistique computationnelle, l'EDI [Electronic Data Interchange] , les systèmes de messages à balises et le FLBC. La linguistique computationnelle n'étant pas encore suffisamment exploitable pour ce type d'application, il reste à comparer le FLBC à l'EDI et aux langages étiquetés.

Les protocoles EDI ont été développés dans les années 80 pour remplacer les flux de documents papiers inter-firmes. Ils sont très utilisés. Il y a au moins cinq standards EDI en compétition mais une convergence vers le modèle X12 de l'ANSI est observable. Ce standard définit les structures de données correspondant aux documents à échanger. Les entreprises n'ont plus qu'à utiliser ces formats lors du développement des logiciels. Cette technique n'est pas très souple et correspond à une électronisation des systèmes papiers en place. Dès lors que le domaine des échanges varie et que les besoins d'expressivité vont grandissant, l'EDI n'est pas suffisant. L'EDI souffre de deux grands problèmes résolus par le FLBC :

– *un problème structurel* : l'EDI ne permet pas l'imbrication[9]. En effet, les différentes imbrications possibles font l'objet de messages distincts ce qui donne 25 types de message pour l'imbrication simple de 5 forces illocutoires ou 125 pour une double imbrication. Le FLBC propose donc une économie de représentation. Cette économie peut être utile pour l'inférence. En effet EDI (X12 ou SWIFT) a plusieurs centaines de types de message de sorte que lorsque l'on demande qu'elles sont les questions que l'on a posées en 1996, il faut qu'il y ait eu un recensement de tous les types de messages qui correspondent à des questions. Dans la réalité, c'est plus compliqué que cela dès lors qu'un type de message EDI peut servir à plusieurs actes illocutoires selon son contenu.
– *l'expression des dates et du temps* : toutes les dates de X12 vont de pair avec leur signification (codée) de sorte qu'une liste croissante de plus de 600 codes pour les dates est à gérer [40]. FLBC règle ce problème à l'aide de quelques prédicats et des opérateurs booléens (indisponibles dans EDI) ;

Enfin l'EDI n'a aucun fondement théorique comparable à la théorie des actes de langage. En ce sens, le FLBC est plus puissant que l'EDI (à condition que l'application soit conçue pour tirer partie de ses potentialités) car il est plus souple et a un plus grand pouvoir d'expression.

Les messages étiquetés sont composés de deux parties : une en-tête et un contenu. Le contenu peut être traité de manière rudimentaire (affiché, copié, transféré...) mais n'est pas signifiant (sauf définition ad-hoc). Seule l'en-tête, composée de balises, a une signification pour le système. Amplement utilisés, leur principal défaut est qu'à chaque nouveau sens doit correspondre une nouvelle étiquette. On aimerait un système qui, comme en linguistique, permet de composer une infinité de sens à partir de règles de formation et d'interprétation.

[9] Comme par exemples : X dit à Z que W lui a dit que P ou X demande à Z de lui promettre que P.

Le FLBC [Formal Language for Business Communication] peut être vu comme un message étiqueté où l'essentiel du contenu serait dans l'en-tête. Le message est une série de déclarations signifiantes ordonnées. Mais il tire partie de l'hypothèse d'universalité des actes de langage. L'infinité théorique des forces illocutoires est réduite grâce aux types primitifs et aux règles de composition (plus un certain nombre d'approximations comme la discrétisation du degré d'intensité...).

4.3 Conclusion

Les travaux sur le FLBC ont permis de mettre à jour l'intérêt de messages articulés syntaxiquement pour faciliter l'accès sémantique et donc l'inférence autant que pour garantir une grande richesse d'expression. Ces avantages sont directement hérités du fait de fonder les travaux sur des bases théoriques d'analyse du langage naturel.

En ce sens, le FLBC est un exemple d'application d'une théorie du langage naturel réussie. Évidemment, comme toute recherche appliquée tournée vers l'industrie, son adoption par les acteurs industriels reste très dépendante de facteurs extrinsèques. Un de ces facteurs est l'investissement déjà fourni dans l'utilisation d'autres méthodes comme les standards EDI ou les langages étiquetés.

Ainsi, dans les systèmes homme-machine(s)-homme, l'assistant logiciel ne participe pas forcément directement au dialogue. Mais pour autant, dès lors qu'il a la vocation d'être utilisé par des humains, il faut qu'il soit assimilable et donc fondé sur des principes naturels [37].

Pour conclure cette section, il convient de remarquer que le FLBC n'est qu'une reformulation de standards déjà existants basés sur la théorie des actes de langage. Ces standards sont les LCAs [Langage de Communication Agent]. Comme c'est aussi le cas pour le cadre COSMO[10] présenté dans la partie précédente, plutôt que d'analyser séparément les limites de ces deux approches, on va présenter celles des LCAs qui les subsume.

5 Problématiques autour des langages de communication agents

Les deux sections précédentes présentent des outils développés autour de la théorie des actes de langage qui pourraient être reformulés dans un cadre plus «standard». En effet, il peut sembler que pour les SMAs, les théories formelles du langage naturel telles que la théorie des actes de langage soient pertinentes (c'est ce que nous avons montré dans les sections précédentes) moyennant quelques extensions : mise en place de protocoles, prise en compte des rôles sociaux...L'idée d'un cadre standard a donné lieu aux LCAs.

Plutôt que de discuter des limites du cadre COSMO puis de celles du FLBC, on présente les limites du cadre plus standard (qui les subsume) des langages de communication agents. Et, comme le souligne Dignum [18], les problèmes soulevés par les LCAs tels qu'ils existent actuellement sont nombreux. Cela va des problèmes d'ontologie aux problèmes de complétude des LCAs. Par exemple, les LCAs ne sont pas adaptés aux sciences cognitives ou à ABSS [Agent

[10] Si on fait abstraction des mécanismes à base d'utilité intégrés à COSMO.

Based Social Simulation], mais plutôt aux domaines pratiques (commerce, industrie...). Ici, on se concentrera sur les problèmes concernant la sémantique, la pragmatique et les protocoles.

5.1 Problèmes concernant la sémantique des LCAs

- *Problème de la minimalité sémantique* : les systèmes de pré/post-conditions habituellement utilisés dans les LCAs 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 LCAs. D'un côté, on veut que la sémantique soit suffisamment générique pour rendre compte de toutes les situations d'utilisation des LCAs. 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 LCAs 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 déduction qu'il est typiquement formellement intractable [18]. Pourtant, l'usage d'un LCA avec une sémantique complète, facilement extensible, est un énorme atout pour des SMA hétérogènes et ouverts. 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 computationnel ne suit pas.
- *Problème de l'alignement des sémantiques* : l'alignement de la sémantique d'un LCA 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 le LCA. Ce type de problème peut advenir lorsque les contraintes (pourtant nécessaires) de la sémantique du LCA sont trop fortes pour l'agent. Cela arrive par exemple avec les conditions de sincérité. La plupart des sémantiques du LCA (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 LCAs 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 (mensonge...) si cela l'aide à atteindre ses buts[11]. 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) pour les LCAs peuvent poser des problèmes avec les modèles comportementaux des agents.
- *Problèmes liés à la distribution et à l'autonomie des agents* : les sémantiques des LCAs 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 LCAs une barrière entre l'effet espéré d'un acte de langage et son effet réel. Quand un agent A transmet l'information X à B, il a l'intention que B va au moins croire que A croit X.

[11] Cas fréquent en e-commerce et plus généralement en économie lors de stratégie de maximisation de gain.

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.

5.2 Problèmes concernant la pragmatique

- *Problème de prise en compte du contexte social* : l'interférence du contexte social sur l'utilisation des LCAs. 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 LCAs.
- *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. D'un certain point de vue, 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 LCAs.

5.3 Problèmes concernant les protocoles

Les protocoles palient à un problème majeur des LCAs 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 donne lieu à des discours et des conversations qui sont des ensembles d'énoncés inter-dépendants. 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 d'une unification (matching) de la sémantique du LCA 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 langages précédents). Cela pose le problème du lien entre la théorie sémantique d'une performative d'un LCA 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 performatives qui la constituent. De l'autre, certains chercheurs pensent que la sémantique de la conversation elle-même doit être vue comme primitive. Dans cette voie, les éléments de base de la sémantique des conversations doivent être sociaux plutôt qu'individuels. Les activités, actions et intentions communes sont alors les éléments de base de ces sémantiques, au même titre que les croyances, désirs et intentions personnels.

Actuellement, la prise en compte du contexte se fait de manière simplifiée par l'utilisation de conversations pré-planifiées, stéréotypées. Cela permet de réduire considérablement l'espace de recherche pour la réponse à donner tout en restant consistant avec la sémantique. Du fait de cet avantage computationnel, quasiment tous les SMA utilisant un LCA sont dotés d'une couche de «conversation», qu'elle soit explicite ou implicite. La spécification de ces conversations se fait à l'aide de polices de conversation qui sont en fait de petits protocoles. Même si les conversations peuvent être structurées, ce type d'approche semble trop rigide à la majorité des chercheurs.

En outre, malgré cette apparente simplicité, l'utilisation de polices de conversation soulève un certain nombre de questions. Tout d'abord, il existe plusieurs candidats non-équivalents pour la spécification de police de conversation : les réseaux de Pétri, les réseaux de transitions à états finis, les arbres de sous buts ainsi que d'autres approches fondées sur la logique. Ces formalismes offrent des degrés de souplesse des conversations envisageables très variables. Puis, comment implémenter les polices de conversation dans les systèmes multi-agents ? Doivent-elles faire partie de l'axiomatique de la communication ? Comment apprend on de nouvelles polices ? Comment intégrer les polices de conversation au sein du fonctionnement de l'agent ? etc...

5.4 Conclusion

On voit donc que les problèmes soulevés par les langages de communication agents sont nombreux. Notons que ceux qui concernent plutôt les aspects sociaux de la communication inter-agents sont neufs. Néanmoins, les réseaux d'obligation, de droit, d'influence, de pouvoir et de confiance des sociétés humaines sont de plus en plus étudiés et adaptés aux agents que ce soit pour les sciences cognitives (sociologie, linguistique...), l'ABSS, l'économie, l'e-commerce ou même l'industrie...

Ainsi, l'application des actes de langage a permis d'en montrer les limites à l'usage. Ces limites sont donc aussi celles de la théorie des actes de langage. Il s'agit d'une théorie de l'énoncé, unité de conversation incontournable mais qui ne rend pas compte des macro-structures langagières comme le discours ou le dialogue[12] ni d'un certain nombre d'aspects pragmatiques du langage naturel. A ce niveau, on a vu que les protocoles sont une alternative peu souple et sans fondement théorique. Malgré ces limites, la théorie des actes de langage reste une base solide pour traduire les énoncés, reste à voir si leur rôle est central ou pas.

Dans la section suivante, nous allons présenter deux théories du dialogue naturel oral. Ces théories utilisent les actes de langage uniquement pour la formalisation des énoncés. Par ailleurs, ce sont des théories du dialogue donc de l'échange d'énoncés liés entre eux. Dans une section ultérieure, nous discuterons de l'applicabilité de ces théories aux SMAs.

6 Systèmes pour le langage naturel oral (homme-machine)

Dans le cadre des systèmes d'interaction homme-machine, une compréhension du fonctionnement de la communication humaine est nécessaire. Il semble justifié, dans le souci de rendre les machines aptes à prendre part à des dialogues de plus en plus naturels de chercher à modéliser la manière dont les humains dialoguent entre eux. Toutefois cette démarche peut être critiquée car il est vrai que les utilisateurs modifient notablement leur manière de dialoguer en présence d'une machine : réduction du lexique, raccourcissement des phrases...Il reste difficile d'évaluer comment ces comportements vont évoluer au fur et à mesure que ces systèmes se perfectionneront

[12] Des solutions sont ébauchées par Chaib et Vongkasem dans [12].

(vraisemblable) et que leur utilisation se généralisera (également vraisemblable). On présente ici les travaux de Balkanski et Hurault-Plantet [7] puis de manière complémentaire ceux de Traum et Poesio [55].

6.1 Modèles du dialogue basés sur les plans

Cette famille de modèles est aussi fondée sur la théorie des actes de langage en ce que les énoncés y sont vus comme des actions communicatives plus que comme de simples chaînes de mots. Les modèles du dialogue basés sur les plans font l'hypothèse que les humains planifient leurs actions (y compris communicatives) pour accomplir leurs buts. Dès lors, le travail de l'interlocuteur est de reconnaître (à partir de l'acte de langage) le plan du locuteur et de répondre (coopérativement) à ce plan plutôt qu'à l'acte isolé.

Le principal avantage de ces théories est de traiter le dialogue comme un cas particulier de comportement rationnel non-communicatif. Elles tirent partie des travaux en reconnaissance de plan et en planification. Elles utilisent donc des méthodes non-linguistiques. De plus, il est à noter que ces modèles ont généralement comme effet de bord de régler le problème des actes de langage indirects.

Balkanski et Hurault-Plantet [7] présentent un modèle computationnel du dialogue sous-tendu par une théorie de l'action. Tout ce qui est présenté correspond à l'implémentation qui en a été faite pour l'automatisation d'un standard téléphonique dans le cadre d'un projet du groupe «Traitement du Langage Parlé» du LIMSI [Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur]. Le travail se base sur l'étude d'un corpus de dialogue humain-humain du dit standard téléphonique.

6.1.1 Un cadre pour le dialogue oral coopératif

La figure 1 résume l'architecture du système. Les données sont encapsulées dans les modules *BASE de CONNAISSANCES* et *CONTEXTE du DIALOGUE*. Le module *CONTEXTE du DIALOGUE* contient les connaissances dynamiques du système qui sont le Rgraph et la Pile de Focus (détaillés plus loin). Les connaissances statiques sont une base de connaissances sur les actions, les recettes et les entités. Cette base regroupe les connaissances du domaine et les connaissances communicationnelles. Pour changer l'application, il suffit de changer les connaissances du domaine.

Le raisonnement est divisé en quatre modules :

- *le manager de dialogue* : il reçoit les énoncés en langage naturel de l'utilisateur et les transmet sous forme écrite au module d'interprétation ;
- *le module d'interprétation* : il met à jour le contexte du dialogue. Ce module est constitué de trois parties : un parseur/interpréteur sémantique, un module d'identification d'actes de langage et un module d'interprétation pragmatique ;
- *le module de génération* : il est lui aussi constitué de trois parties : planification/exécution, générateur d'actes de langages et un générateur de surface pour la formulation en langage

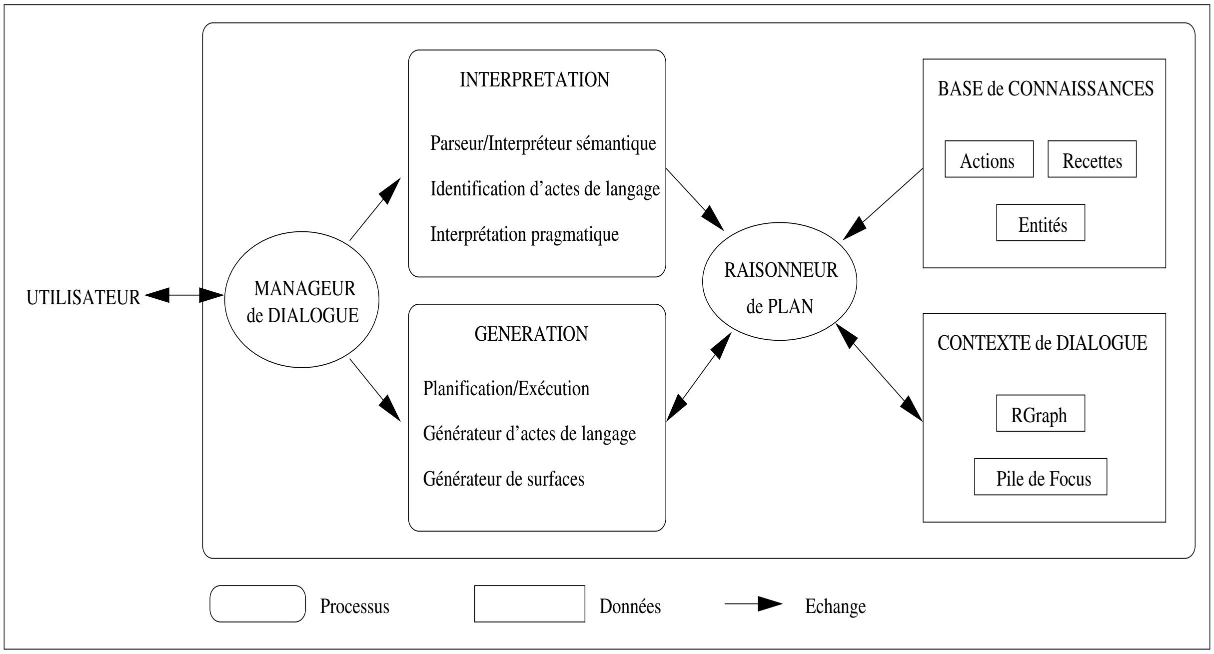

FIG. 1 – Architecture du système de dialogue en langage naturel [7].

écrit correct ;

– *le module de planification* : il gère le Rgraph et exécute l'algorithme d'avancement de tâches.

L'application a été implémentée en Smalltalk-80. L'analyse grammaticale et syntaxique repose sur la grammaire LFG [Lexical Functionnal grammar] [30] et l'interpréteur sémantique est basé sur les graphes conceptuels [51]. Dans les sous-sections suivantes, nous allons détailler le cadre théorique et les différentes composantes du système.

6.1.1.1 Cadre théorique

Le modèle calculatoire est basé sur la théorie du dialogue coopératif développée par Grosz, Kraus et Sidner [25, 24, 26] et étendue par Lochbaum [36]. L'idée est que le dialogue, comme beaucoup d'autres activités implique un comportement collaboratif. Cette théorie repose sur la notion de plans partagés proposée par Grosz et Kraus [27] qui permettent de prendre en compte les engagements des agents concernant le succès de leurs actes mais aussi de ceux d'autres agents. C'est cette dernière particularité qui mène les agents au comportement collaboratif, c'est-à-dire qui leur permet d'éviter les conflits. Un plan partagé est une formalisation de la notion de plan collaboratif. Un plan partagé pour une action inclut :

1. la croyance mutuelle d'avoir un moyen (une recette) pour réaliser l'action ;
2. les intentions individuelles que l'action soit accomplie ;
3. les intentions individuelles que les sous-actions prises en charge par les collaborateurs soient accomplies avec succès ;

4. les structures hiérarchiques de plans individuels et de plans partagés.

Les points 2 et 3 reflètent l'engagement que les agents collaboratifs doivent avoir concernant les actions du groupe entier. Ces engagements, desquels découle la coopération, sont formalisés grâce aux opérateurs *intention-que* (une proposition se réalise) et *intention-de* (faire quelque chose). L'opérateur *intention-de* suit les deux axiomes suivants :

*Axiome A*1

> Si un agent G croit que $prop$ n'est pas réalisé et G à l'intention que $prop$ et G croit qu'il peut faire A qui peut aider à établir $prop$, alors G va considérer faire A.

*Axiome A*2

> Si un agent G croit que $prop$ n'est pas réalisé et G à l'intention que $prop$ et G croit qu'il peut faire A qui peut permettre à un autre agent de faire quelque chose B qui aide à établir $prop$, alors G va considérer faire A.

Ce sont ces axiomes qui mènent les agents à l'adoption de comportements d'aide. Ces axiomes servent de cadre théorique pour le design de l'algorithme d'avancement de tâches de Balkanski et Hurault-Plantet.

6.1.1.2 Représentation du dialogue

La représentation des dialogues se fait à l'aide des Rgraph introduits par Lochbaum [36]. Les Rgraph sont constitués d'actions et de recettes. Dans la théorie de l'action qui les sous-tend, les actions peuvent être basiques ou complexes, communicatives ou pas. Les actions basiques peuvent être accomplies à volonté sous certaines conditions de faisabilité. Des recettes sont associées aux actions complexes et leur exécution nécessite l'exécution de toutes les actions de la recette. Une recette est donc un ensemble d'actions avec les contraintes qui leurs sont associées. On peut représenter une recette comme un arbre de profondeur deux dont la racine est l'action correspondant à la recette et les feuilles sont les actions ou recettes constituantes. A titre d'illustration, dans l'implémentation réalisée, il n'y a qu'une action communicative pour les connaissances communicatives indépendantes du domaine : *Obtenir_Valeur*. Celle-ci a trois recettes différentes. Les actions et recettes relatives au domaine (orienté tâche) ont été définies à partir de l'étude du corpus : une recette pour *Parler_Personne*, une pour *Transférer_Personnne* et deux recettes pour *Etablir_Communication* (voir figure 2).

Un Rgraph est une arborescence qui représente les croyances d'un agent à propos de la façon de réaliser une tâche aussi bien que l'état d'avancement de cette tâche. Les nœuds d'un Rgraph sont des actions (éventuellement communicatives). Un Rgraph n'est pas prédéfini mais est construit au cours du dialogue par agencement et instanciation de parties prédéfinies (les recettes). Les paramètres instanciés sont : les agents, le temps et certaines variables. Les actions peuvent être liées par la relation de contribution directe (*D-contribution*). En outre, un Rgraph peut inclure des actions qui ne font pas partie d'une recette mais qui sont nécessaires à l'exécution d'actions ultérieures. Elles sont alors liées aux autres actions avec une relation d'habilitation

(*Enablement*), comme c'est le cas par exemple pour *Chercher_Num_Tel* dans la figure 2. Pour chaque action, un statut est mémorisé. Les six statuts définis sont :

1. l'agent modélisé croit que l'action va faire partie du plan partagé ;
2. l'agent modélisé croit que les autres agents sont d'accord pour que l'action fasse partie du plan partagé ;
3. l'action est basique et l'agent modélisé croit que les autres seront d'accord pour la réaliser ;
4. l'action est complexe et l'agent modélisé croit que les autres seront d'accord avec une recette particulière pour cette action ;
5. l'action ne peut être accomplie (une réparation est nécessaire) ;
6. l'action a été accomplie (avec succès ou pas et dans ce dernier cas aucune réparation n'est plus possible)

Une action est ajoutée au Rgraph avec le statut 1, ce statut est mis à jour à mesure de l'avancement du dialogue. Cette notion de statut rend compte de l'intentionalité des agents et de sa dynamique. Pour Bratman [9] l'intention joue trois rôles : (1) poser un problème et entraîner un raisonnement sur les moyens, (2) contraindre les intentions des autres agents et, (3) guider l'agent dans sa conduite. En effet, ici un agent commence par avoir l'intention de réaliser une action (statut 1) qui devient une intention commune (statut 2). L'agent s'engage ensuite dans une phase de raisonnement sur les moyens (statut 3 ou 4 selon que l'action est basique ou complexe) puis enfin l'action est réalisée (statut 6) si possible (sinon statut 5). La figure 2 présente un Rgraph à un moment du dialogue où la communication entre l'utilisateur et le système a été établie avec succès (statut 6), le système s'apprête à chercher le numéro d'extension de la personne demandée par l'utilisateur (en effectuant *Chercher_Num_Tel*).

6.1.1.3 Interprétation des énoncés : l'algorithme d'interprétation

L'algorithme d'interprétation modélise le processus par lequel un agent détermine comment le but sous-tendu par un énoncé contribue aux buts ou sous-buts du plan collaboratif courant (i.e le Rgraph). On distingue deux types d'énoncés :

– *l'énoncé initial* : si l'énoncé contient explicitement un but (avec l'action correspondante *ActInit*[13]), un Rgraph est initialisé avec la recette correspondante sinon, c'est que l'énoncé fait implicitement référence à un but qu'il faut inférer. Beaucoup de recherches font l'hypothèse que le but est donné explicitement. Par exemple, dans le système TRAINS, Allen [1] ne fait pas cette hypothèse mais restreint l'espace de recherche en supposant que le but sera de prendre ou de rencontrer un train. L'algorithme proposé combine les deux approches. Dans tous les cas l'algorithme va initialiser un Rgraph avec une recette pour *ActInit* ou bien une recette qui contient *Actinit*. Un système de priorité permet de trancher si plusieurs recettes sont possibles.[14]
– *les autres énoncés* : Le Rgraph n'est pas vide. L'algorithme d'interprétation doit alors trouver un lien (contribution ou habilitation voir 6.1.1.2) entre l'acte A sous-tendu par

[13] *Actinit* est l'action sous tendue par l'énoncé initial.
[14] L'algorithme d'avancement de tâche permet de changer de recette plus tard si un problème survient.

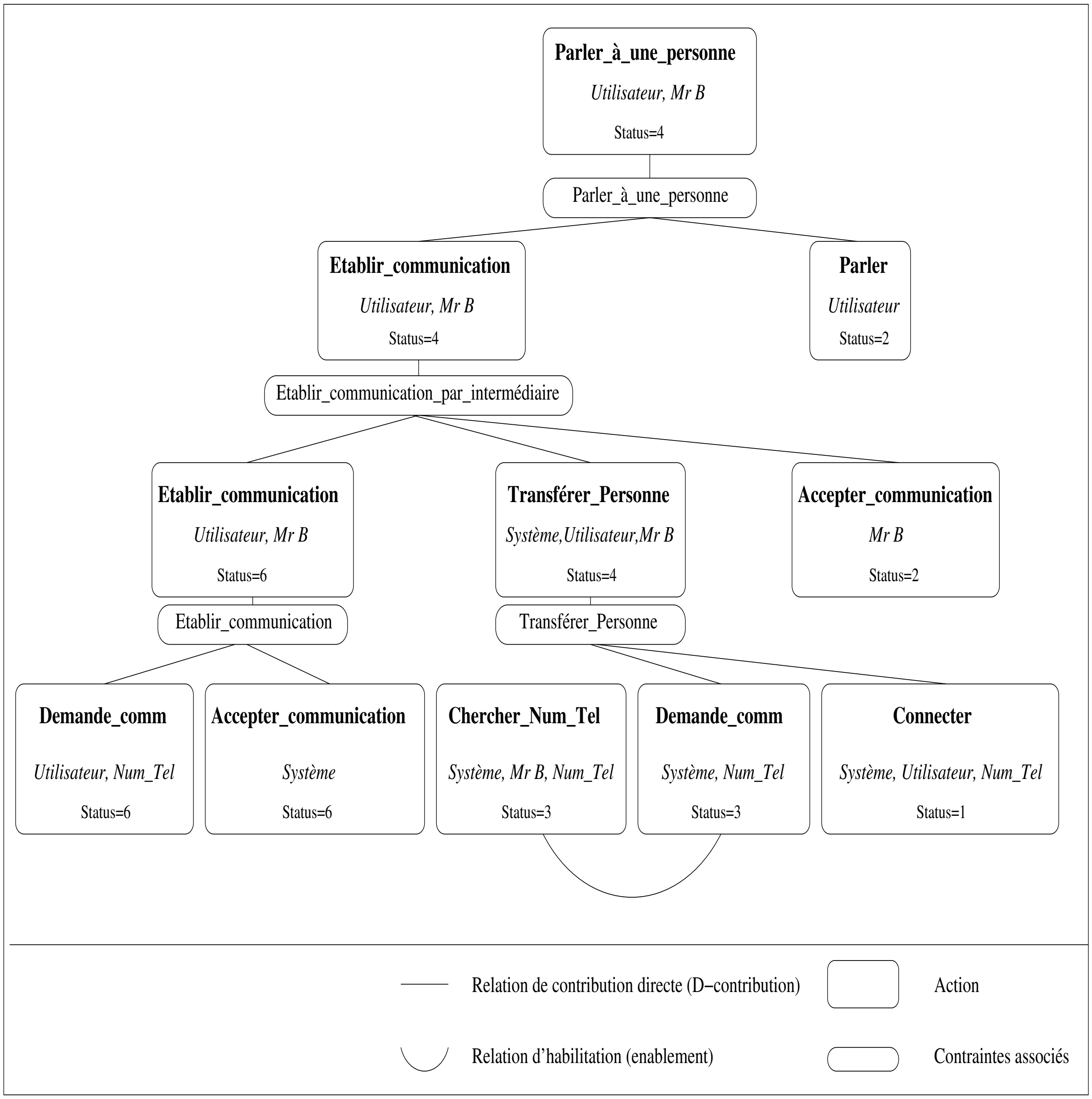

FIG. 2 – Exemple de Rgraph (adapté de [7]).

l'énoncé traité et l'acte B actuellement sur le dessus de la Pile de Focus (c'est l'action courante). Si aucun lien n'est trouvé l'algorithme remonte sur le chemin vers la racine et cherche un lien entre A et B' (auquel B contribue). Si aucun lien n'est trouvé une fois rendu à la racine, l'algorithme échoue. Si l'énoncé renvoie plutôt à une acceptation ou à un signe de compréhension, il est utilisé pour faire passer certains noeuds du Rgraph du statut 1 au statut 2.
Sauf échec, A est donc ajouté au Rgraph avec le statut 2, 4 ou 6 dépendamment du type d'ajout et du fait que l'acte soit à accomplir par le locuteur ou par son interlocuteur. Certaines hypothèses concernant cette affectation dépendent de la connaissance que le locuteur a à priori du domaine (ici, la tâche de standardiste).

6.1.1.4 Génération du discours : l'algorithme d'avancement de tâche

Cet algorithme génère un énoncé qui contribue à l'élaboration du plan collaboratif ou qui suggère une fin de dialogue si aucune élaboration supplémentaire n'est nécessaire. Le point crucial de cet algorithme est de permettre aux agents de progresser dans la réalisation du but commun (d'où son nom). Pour ce faire cet algorithme mélange planification et exécution. L'algorithme, guidé par les contraintes d'exécution, examine et met à jour le Rgraph. La boucle principale est constituée de trois parties :

- *initialisation* : A_i est initialisé avec l'action courante ;
- *conditions d'arrêt* : L'algorithme retourne un énoncé lorsqu'une des trois clauses suivantes est vérifiée : (1) A_i est une action communicative et c'est à l'agent modélisé de la réaliser (l'algorithme renvoie A_i après l'avoir mise au statut 6), (2) A_i est une action qui revient à un autre que l'agent modélisé (l'algorithme renvoie un acte communicatif indiquant le succès des actions qu'il avait à faire), (3) A_i est l'action racine du Rgraph (un acte communicatif est retourné indiquant le succès ou l'échec (motivé) de la tâche A_i) ;
- *corps de la boucle* : Le corps est constitué d'un ensemble de six règles. Une règle, déterminée par le statut de A_i est appliquée à chaque tour. Les règles renvoient A_i avec un nouveau statut ou bien une nouvelle action A_{i+1}.

6.2 Limites, comparaisons et approfondissements

La classe des modèles de dialogue coopératif basés sur les plans admet des limites théoriques et pratiques. Un certain nombre de ces limites traditionnelles [15] et certaines difficultés sont partiellement résolues par ce modèle :

- *les informations spontanément données par l'utilisateur* : dans les modèles basés sur les plans, il est difficile de prendre en compte les informations données par l'utilisateur sans que ce soit en réponse à une question. Dans ce modèle, toutes les interventions de l'utilisateur sont prises en charge par l'algorithme d'interprétation qui cherche alors un lien entre l'action courante (le dessus de la Pile de Focus) et l'énoncé de l'utilisateur. Le modèle permet donc ce que l'on appelle l'initiative mixte.
- *la génération d'explications* : la génération d'une explication lorsque le système échoue est généralement un point délicat des approches par planification. Dans le modèle de Balkanski

et ses collègues, lorsqu'une action ne peut être réalisée, les raisons (paramètre non instanciable, précondition ou contrainte non satisfaite...) sont ajoutées à la liste des messages à produire par le module de *GENERATION*.

– *la reconnaissance des actes illocutoires* : les algorithmes des théories basées sur les plans sont très liés à la reconnaissance des actes illocutoires. Il leur faut un acte illocutoire par énoncé, cela pose des problèmes lorsque l'acte est accompli par plusieurs énoncés ou lorsque l'énoncé accomplit plusieurs actes. Dans le modèle de Balkanski et Hurault-Plantet, en cas d'enchaînement de multiples énoncés par l'utilisateur, c'est le *DIALOGUE MANAGER* qui les gère.
– *plan du discours vs plan du domaine* : la distinction entre les actes de langage à propos de la tâche et ceux de gestion du dialogue est coûteuse. Le modèle de Balkanski et Hurault-Plantet permet les sous-dialogues de clarification qui appartiennent à la gestion du dialogue. Un sous-dialogue de clarification est généralement engagé pour déterminer ou confirmer la valeur de certains paramètres. L'action *Obtenir_Valeur* permet au système comme à l'utilisateur, grace à ses différentes recettes, d'entamer un tel sous-dialogue. Par exemple, lorsqu'il manque une information à l'algorithme d'avancement de tâche (par exemple un paramètre à instancier), celui-ci peut intégrer une recette de l'action *Obtenir_Valeur* au Rgraph. Réciproquement, si l'utilisateur demande la valeur d'un paramètre (connu du système ou pas), une autre recette d'*Obtenir_Valeur* est appliquée.
– *la replanification* : le modèle facilite la replanification. En effet, quand l'opérateur (le système) ne peut satisfaire la demande directement, il replanifie pour y parvenir d'une autre manière. La replanification est un point clé pour tous les systèmes de dialogue : la souplesse en dépend. Dans le système présenté ici, lorsqu'une action échoue, la recette échoue et si aucune réexécution n'est possible, il replanifie le niveau supérieur du Rgraph.

De l'aveu des auteurs, un certain nombre d'énoncés (on se limite aux questions de l'utilisateur) du domaine restent hors de portée de leur modèle. Pour certains d'entre eux, des réparations de plans y sont associées. Pour d'autres, il suffit d'ajouter des recettes et des connaissances («Je voudrais le numéro de fax de Mr.X»). Mais il en existe qui remettent tout le système en cause comme : «Bonjour, c'est encore moi.». Les énoncés autres que ceux qui donnent des informations sur une unique action sont laissés de coté. Les agents doivent être altruistes («helpfull») et vraiment collaboratifs et coopératifs ce qui va bien dans le cadre de tâches orientées comme celles d'un standard téléphonique.

Il reste aussi des limites théoriques inhérentes à cette classe de modèles et auxquelles le travail de Balkanski et ses collègues n'échappe pas, en particulier :

– *la complexité des inférences* : les algorithmes de reconnaissance de plan sur lesquels repose le modèle sont combinatoirement intractables dans le pire des cas, et indécidables dans certains cas [11].
– *le manque de bases théoriques* : même si ces modèles sont computationnels et ont permis de nombreuses avancées dans la compréhension des conversations, il leur manque des fondements théoriques. 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.

– *la dépendance au domaine* : le modèle est basé 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 modèle utilise la reconnaissance et l'élaboration de plan pour comprendre et accomplir des actes communicatifs dans un contexte orienté tâche bien délimité. Il existe des systèmes comparables. Certains modèles se basent aussi sur les plans : le système TRAINS [1], le raisonneur de plans de domaine de Ferguson [19]. D'autres sont plus axés sur le raisonnement : le Circuit Fixit System de Duke [50] et Atimis de France Télécom [44].

Les modèles du dialogue basés sur les plans comme celui développé par Balkanski et al. ne sont que l'intégration astucieuse d'une reconnaissance et d'une génération de plan dans un domaine bien clos. Cela laisse de côté la dimension sociale du dialogue et ne fournit aucune explication pour les phénomènes pragmatiques observables lors de conversations. Il est d'ailleurs difficile d'imaginer un tel système pour un domaine plus important ou pour un domaine non-circonscrit. Il convient cependant de noter, pour faire suite à l'introduction de cette partie, que l'utilisation de méthodes non-linguistiques comme celles basées sur les plans donnent des résultats comparables à l'utilisation des méthodes linguistiques les plus efficaces. Cela montre à quel point les méthodes linguistiques en sont à leurs balbutiements malgré les efforts qu'elles suscitent[15].

Un dernier défaut du modèle de Balkanski et Hurault-Plantet est que la gestion du fond commun y est implicite et incomplète. La section suivante détaille ce point tout en introduisant les sections ultérieures.

6.3 Arrière plan, fond commun et établissement[16]

La notion d'arrière plan est la clé de voûte de la philosophie Searlienne. 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 usages. Par exemple, le verbe ouvrir s'interprète différemment dans «Ouvrir les yeux», «Ouvrir un restaurant», «Ouvrir son coeur», «Ouvrir une porte», «Ouvrir le débat». Pour Searle [47], 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[17].

Le langage est essentiellement construction d'interprétations communes. En première approximation, on réduit l'arrière plan à ce qui est dit et communément accepté dans le cadre d'une conversation. La notion de croyance commune introduite par Clark et Schaeffer avec la notion d'établissement [14] semble acceptable. Une information est d'abord présentée, comprise (retour

[15] Cela signifie par ailleurs que les systèmes de planification spécifiques au traitement du langage donnent de «bons» résultats. On peut s'étonner que des recherches qui tourne autour d'un outil (la planification) aillent aussi loin que celles qui ont le langage pour objet (linguistique) et qui sont libres de développer des outils en conséquence.

[16] Ces termes sont les traductions «fragiles» de : background, common ground et grounding dont il est difficile de rendre excactement compte en francais.

[17] Cette idée est déjà en germe chez les philosophes de l'école Gesalt. 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.

de compréhension) puis acceptée, on dira établie. Elle devient alors commune. Cette façon de voir évite le problème de récursivité infini des connaissances mutuelles[18] : je sais qu'il sait que je sais qu'il sait que je...Ce qui rend néanmoins difficile l'implémentation de ces notions, c'est la diversité des types de présentation (directe ou co-construite...) et d'acceptation (implicite par continuation ou explicite comme «OK» ou conversation...). Ainsi, il est souvent difficile de dire si un énoncé fait partie de la présentation ou de l'acceptation [52]. Le modèle de Traum et Poesio permet de dire à un moment donné de la conversation ce qui est accepté (et donc partagé) ou ce qu'il faudrait pour que cela le soit. Le modèle du dialogue de Balkanski et Hurault-Plantet est compatible avec ce que dit Traum mais le traitement du fond commun y est implicite. Cela entraîne certaines limites dans des cas particuliers (voir section précédente). Par exemple si le système demande une information X à l'usager, le système ne devrait pas ensuite informer l'usager de son ignorance de X si sa recherche par demande à l'usager a échoué car l'usager le sait déjà...Un traitement plus complet de l'établissement permettrait au système de savoir quand il doit ou pas justifier un échec d'action. La section suivante présente le modèle d'informations communes de Traum et Poesio.

6.4 Un modèle du fond commun dans les conversations orales

La théorie du fond commun de Traum et Poesio est issue du travail sur le projet TRAINS [1] de l'université de Rochester. Cette théorie est complétée par d'autres travaux du projet TRAINS pour former un système de dialogue complet. De manière isolée comme elle est présentée dans [55], elle correspond à l'ensemble *CONTEXTE du DIALOGUE* de la figure 1. C'est à dire qu'elle présente un formalisme de représentation des conversations orales.

6.4.1 Une représentation minimale du contexte

On utilise le terme «contexte» en référence à l'information qu'un conversant utilise quand il interprète les énoncés d'une conversation. Personne n'est d'accord sur ce que sont ces informations mais tout le monde s'accorde pour penser qu'elles incluent au moins le fond commun, c.à.d les informations partagées. Clark et Marshall [13] ont montré que le succès d'utilisation de références dépend fortement de la qualité des hypothèses faites par le locuteur sur le fond commun.

Du point de vue de la résolution de références, l'information cruciale apportée par le contexte est de permettre de savoir quelles références sont disponibles. Une propriété fondamentale des énoncés est alors d'ajouter de nouveaux référents au fond commun. Dans les énoncés de l'exemple suivant, le pronom «il» de B_2 fait référence au groupe de musique expérimentale dont il est question dans A_1. La logique du premier ordre est inapte pour capturer ce phénomène.

A_1 *: Il y a un groupe de musique expérimentale à Nantes.*

B_2 *: Il s'appelle robonom*[19].

L_1 : $(\exists x, \exists w,\ groupeXP(x) \wedge Nantes(w) \wedge situer(x, w))$

[18] En fait, le problème ressurgit dès que l'on pense à distribuer cet objet commun.

[19] On utilisera cet exemple dans toute la section.

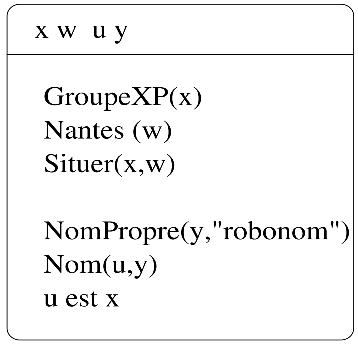

FIG. 3 – DRS pour le discours A_2.

Dans L_1, la variable x n'est plus disponible pour des références ultérieures dès lors que la portée de la quantification existentielle est limitée à cet énoncé. Les formalismes de l'IA quand à eux ne posent pas ce genre de problème et sont donc inadéquats pour le traiter. Ce problème trouve sa solution avec les formalismes de linguistique computationnelle, en particulier la DRT [Discourse Representation Theory] [29] qui est une théorie sémantique développée expressément pour représenter les liens de cohésion (anaphoriques ou autres) entre énoncés d'un discours. Traum et Poesio étendent ce formalisme aux conversations orales pour les besoins de leur théorie du fond commun.

6.4.2 La DRT [Discourse Representation Theory]

Avec la DRT [29], un discours est modélisé dans une DRS [Discourse Representation Structure] qui est une paire d'ensembles. Il y a l'ensemble des référents [Discourse Referents] et l'ensemble des faits à propos de ces référents [Conditions]. Une DRS est typiquement représentée par une boîte comme le montre la figure 3. Les DRSs sont des expressions logiques qui capturent les intuitions des modèles de discours de philosophes du langage comme Karttunen [31] et Webber [59] tout en admettant une sémantique précise. Ici, la DRS est vrai dans un modèle M et une situation (ou monde) s s'il y a une façon d'assigner les objets de s aux référents du discours x et w de telle façon que les conditions sur les objets de s soient vraies. Cette sémantique rend la DRS logiquement équivalente à la formulation existentiellement quantifiée L_1. En général, la caractérisation d'un contexte par la DRT consiste en plusieurs DRSs emboîtées dans une DRS racine représentant le fond commun. Cette DRS racine est construite incrémentalement par l'application de l'algorithme de construction de DRS [DRS Construction Algorithm] sur chaque nouvel énoncé. Par exemple l'effet du discours A_2 sur la DRS-racine est représenté par la figure 3 qui spécifie une interprétation pour une conversation type $<A_1, B_2>$.

A_2 : *Il y a un [groupe de musique expérimentale]$_i$ à Nantes. Il$_i$ s'appelle robonom.*

On note l'introduction de deux référents supplémentaires : y et u pour «robonom» et le pronom «il» respectivement. u est x indique que la dénotation pour le référent de discours u et la dénotation pour le référent de discours x sont identiques.

La DRT comme d'autres modèles du discours capture l'anaphore[20] pronominale, les descriptions définies[21] et l'ellipse[22]. Ainsi en basant le modèle du contexte sur ce formalisme on peut espérer un bon fonctionnement sémantique. La DRT version «vanilla» [29] à trois caractéristiques problématiques. Premièrement, la DRT met l'emphase sur la représentation des aspects sémantiques du contexte, laissant de côté tous les aspects pragmatiques y compris les informations nécessaires à la résolution de référence. Deuxièmement, la mise à jour du fond commun est simpliste alors qu'une grande partie des conversations visent à synchroniser les participants sur leur vision du fond commun. Finalement, l'algorithme de construction de [29] n'assigne pas un sens à chaque énoncé et laisse de côté les énoncés tronqués pourtant très présents dans les conversations (ce problème est connu et nommé : «manque de compositionnalité»)[23].

6.4.3 Actes conversationnels et situation de discours

En modélisant le fond commun par une DRS-racine on capture deux aspects du savoir commun des participants : les antécédences et les contenus propositionnels des assertions. La DRT capture les conditions de vérité du discours. Mais d'abord, tous les énoncés ne sont pas des assertions et puis beaucoup d'énoncés n'ont pas de conditions de vérité. Il en est ainsi des marqueurs comme «ok», «oui», «d'accord» et des indexes comme «je» ou «tu» dont le sens dépend de facteurs pragmatiques. On trouve encore des expressions qui concernent le discours lui-même (tout ou partie) comme «d'abord», «ensuite» ou «vice-versa» et «la liste précédente», «ci-dessous». Il est donc nécessaire d'intégrer les informations pragmatiques dans le fond commun. Cela se fait via les actes conversationnels.

6.4.3.1 Actes conversationnels

On trouve dans la littérature un grand nombre de théories des structures discursives. Le modèle de Traum et Poesio est basé sur la théorie des actes de langage à la façon de Bunt [10] dans son modèle DIT [Dialogue Interpretation Theory]. C'est ce qui différencie cette approche de la STDRT de Asher [3] qui s'intéresse aux discours textuels plutôt qu'aux conversations.

Venons en maintenant à la théorie des actes de conversation multi-niveaux de Traum et Hinkelman [54] telle qu'utilisée dans ce modèle. Cette théorie inclut des actes de langage noyaux qui sont les actes de langages classiques mais qui ne prennent leur plein effet qu'après établissement (confirmation). Plutôt que d'être des actions effectuées par le locuteur, ils sont considérés comme des actions communes qui sont ajoutées au fond commun. Cette théorie des actes conversationnels suppose en outre trois autres types d'actes de langages : les actes de tour de parole, les actes

[20]Anaphore : n. f. figure de style caractérisée par la répétition d'un terme en tête d'un groupe de mots ou d'une phrase. Ex : Marcher à jeun, marcher vaincu, marcher malade (Hugo). On note que cet énoncé comporte aussi un parallélisme de construction.

Anaphore pronominale : Figure de style de type anaphore ou la répétition ce fait via un pronom. Ex : Pedro a un âne, il le bat.

[21]Description désignant une entité déterminée sans la nommer.

[22]Ellipse : n. f. procédé syntaxique ou stylistique consistant à omettre un ou plusieurs mots à l'intérieur d'une phrase, leur absence ne nuisant ni à la compréhension ni à la syntaxe. Il y a ellipse du verbe dans la deuxième partie de la phrase "Pierre mange des cerises, Paul des fraises".

[23]C'est un problème que l'on retrouve dans la théorie des actes de langage.

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, réparrer, demande-réparrer, demande-valider, accepter, rejeter
Unité de discours (DU)	actes-noyau	informer, demander, s'engager
Multiple DUs	argumentation	élaborrer, résumer, clarifier, question-réponse, convaincre, trouver un plan

TAB. 1 – Actes conversationnels [55].

d'établissement et les actes d'argumentation plus complexes car ils mettent en jeu plus d'un acte noyau. Les actes conversationnels sont récapitulés dans le tableau 1.

6.4.3.2 Situation du discours et situation décrite

Cette tentative d'unification de la DRT avec la théorie des actes de langage est inspirée par les idées sur le fond commun de la sémantique des situations [17]. La sémantique des situations est basée sur la théorie de l'information selon laquelle ce que l'on sait est organisé en situations. En particulier, le fond commun conversationnel contient des informations partagées sur la situation de discours qui est la situation dans laquelle se trouvent les participants de la conversation. La théorie de Traum est une théorie des effets des actes conversationnels sur les situations du discours. Traum note lui-même que la situation du discours n'est qu'une partie du fond commun et laisse à d'autres le soin d'étudier les autres parties (dans quelle société, organisation, lieu ou postures sont les conversants...).

La situation du discours inclut les actions (communicatives) des agents et des informations sur leurs états mentaux (y compris leurs situations visuelles, leurs postures...). Mais la situation du discours inclut aussi des informations sur les situations décrites qui sont les situations dont les conversants parlent. En général, la situation du discours ne coïncide pas avec la ou les situations décrites.

6.4.3.3 Dynamique des actes de langage

La DRT doit être adaptée aux conversations. Avec la DRT traditionnelle, dans la mini conversation $<A_1, B_2>$ le référent «groupe de musique expérimentale» n'est plus accessible pour le second énoncé. De plus, la représentation traditionnelle ne nous indique pas que les deux énoncés se réferrent à la même situation.

En outre, la représentation traditionnelle ne propose pas de référent pour la situation du discours. Par exemple, l'énoncé A_3 se réfère à l'énoncé de B, B_2. Un évènement conversationnel doit pouvoir servir de référence pour les évènements futurs.

A_3 : *C'est une question ?*

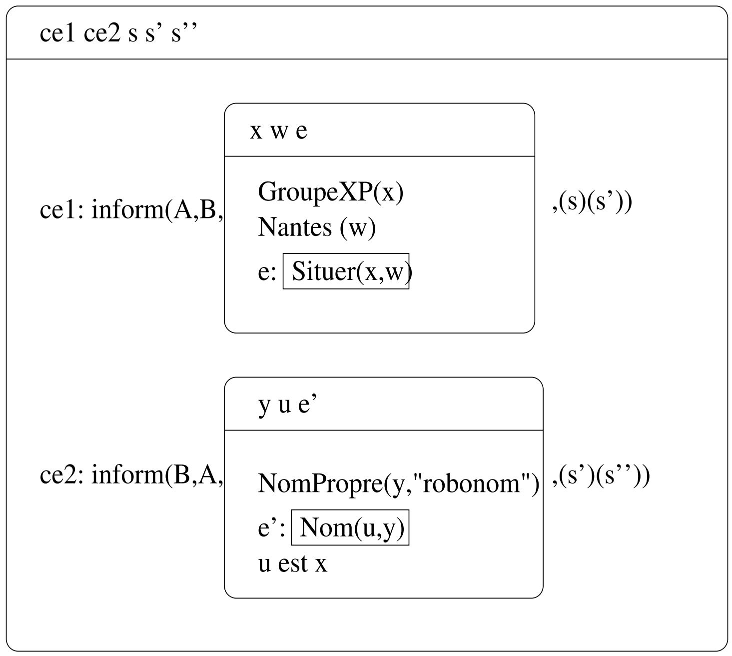

FIG. 4 – DRS pour le dialogue.$<A_1, B_2>$.

Traum et Poesio modifient la notion de DRS de sorte que les aspects ci-dessus soient capturés. Le dialogue $<A_1, B_2>$ y prend la forme représentée en figure 4. On y trouve les référents $ce1$ et $ce2$ pour les évènements conversationnels et s, s', s'' pour les situations décrites. La situation s' décrite par $ce1$ étend la situation s et s'' étend s' de sorte que tous les constituants de s' restent accessibles. Ainsi un DRS indique une relation entre deux situations.

6.4.3.4 États mentaux

Les états mentaux des agents jouent un rôle important dans la reconnaissance des actes conversationnels, la résolution de référence et le management du dialogue. L'information sur les intentions, les croyances, les obligations et les perceptions communes fait partie de la situation de discours. Dans le modèle, tous les faits, qu'ils concernent les états mentaux ou pas, sont représentés par des conditions dans la DRS racine. La connaissance du contenu de la DRS racine est commune.

6.4.4 Structure du discours et structure des évènements

Traum et Poesio incorporent à leur théorie des situations de discours, la théorie de la structure du discours de Grosz et Sidner [26]. Le modèle du discours de Grosz et Sidner inclut trois composantes inter-reliées : (1) la structure intentionnelle qui décrit les buts des interlocuteurs et leurs liens, (2) la structure attentionnelle qui indique les entités les plus saillantes à un moment du discours et, (3) la structure linguistique qui rend compte de la structuration syntaxique des segments du discours.

6.4.4.1 Structure intentionnelle et structure hiérarchique des évènements

Pour Grosz et Sidner, la reconnaissance d'actes de langage ne se limite pas à la reconnaissance de l'acte illocutoire associé à un énoncé mais est aussi liée aux intentions exprimées par les actes précédents. Dans leur théorie, un segment de discours est associé au sujet dont il traite via un DSP [Discour Segment Purpose]. Les DSP sont liés entre eux par des liens de dominance (quand un DSP participe à un autre noté $\alpha \uparrow \beta$) ou de précédence-satisfaction (quand un DSP est un pré-requis pour un autre, noté $\alpha \prec \beta$). Dans la théorie des actes de conversation de Traum, un énoncé génère un ou plusieurs actes illocutoires : actes de tour de parole, d'établissement ou actes noyaux. Pour caractériser les actes conversationnels de plus haut niveau (englobant plusieurs actes noyaux), Traum introduit la notion de brin conversationnel [Conversationnal Thread]. Les actes d'argumentation sont des cas particuliers de brins conversationnels. L'idée de brin conversationnel est à rapprocher de celle de script de conversation[24] et des micro et macro-structures de conversation de l'analyse conversationnelle. Ces idées sont explorées en IA via les jeux de dialogues [38] et les scripts de discours. Si $ce1 \uparrow ce2$ alors la situation décrite de $ce1$ doit être incluse dans la situation décrite de $ce2$. Si $ce1 \prec ce2$ (ce qui est le cas dans notre exemple) alors la situation décrite $ce2$ étend la situation décrite $ce1$.

6.4.4.2 État attentionnel

Des conditions de la forme $sujet(t)$ *is* s', où t est une date et s' une situation indiquent à tout moment le sujet de la discussion. Par défaut, les évènements conversationnels étendent le sujet courant plutôt que de le changer. Les situations sont structurées selon une relation d'inclusion hiérarchique.

6.4.5 Évènements et interprétations micro-conversationnels

Le discours oral spontané consiste la pluspart du temps en brides d'énoncés plutôt qu'en énoncés complets. Ces fragments sont mélangés aux pauses, hésitations, répétitions, corrections. . .Le fait que des actes de tour de parole et d'établissement surviennent bien souvent avant que l'acte noyaux n'ait été terminé montre bien que le fond commun est mis à jour avant que l'acte noyau soit complètement accompli[25]. Pour prendre en compte tous les sons qui sont émis par les interlocuteurs, Traum et Poesio récupèrent la notion d'évènement micro-conversationnel (prédicat *utter* pour tous les actes locutoires) introduite dans la théorie des actes locutoires de Poesio. Les informations inférées à partir d'un évenement micro-conversationnel, MCE [Micro-Conversationnal Event], sont : la catégorie syntaxique, le sens suggéré par le locuteur et en utilisant le reste du fond commun les références, les liens anaphoriques, le sens des actes illocutoires et perlocutoires dans lesquels le MCE s'inscrit. *cat* est la fonction qui donne la catégorie syntaxique d'un MCE et $\sim>$ celle qui en donne le sens. Les MCE se combinent (la plupart du temps) pour former un des actes conversationnel classique. Quand un acte illocutoire est reconnu, il est alors ajouté au fond commun. Les actes illocutoires sont donc générés par les actes locutoires au sens de

[24] Qui sont applicables aux situations routinières : réservation de train, restaurant. . .

[25] Traum et Poesio, comme Stalnaker considèrent que les actes locutoires même s'ils ne correspondent pas à un acte illocutoire complet sont intégrés au fond commun et qu'ils déclenchent l'interprétation de la même manière que les actes illocutoires complets.

Acte	Description
initiate	Début d'un nouveau DU
continue	Le locuteur apporte d'autres éléments
acknowledge	Preuve de bonne compréhension des éléments antérieurs par les autres conversants
repair	Corrige la mécompréhension (potentielle) du contenu du DU
request repair	Signale un problème de compréhension
request ack	Signale aux autres d'aquiesser
cancel	Retire le DU du common ground, le rendant ingroundable

TAB. 2 – Actes d'établissement.

Goldman [20]. De la même façon ici, c'est le prédicat *generate* qui lie l'acte illocutoire aux actes locutoires qui l'ont généré.

Dès lors que les actes locutoires et illocutoires font partie du fond commun la question est : quel acte génère les références ? La réponse n'est pas claire mais les auteurs choisissent les actes illocutoires, auxquels la dynamique est associée.

Dans le modèle, l'interprétation est vue comme un remplissage des «trous» dans l'information disponible aux interlocuteurs. Ceci est habituellement caractérisé par la notion de sous-spécification. De nombreuses théories de l'interprétation des énoncés partent de l'hypothèse que toutes les interprétations d'un énoncé ne sont produites qu'après filtrage par le contexte. Ici, les règles d'interprétation sont formulées comme des règles d'inférence de la logique des défauts de Reiter [43]. Les interprétations désambiguïsées sont obtenues en calculant l'extension de la théorie des défauts $< D, W >$ où D est l'ensemble des règles d'inférence et W l'interprétation sous-spécifiée initiale. Ce calcul d'extension ce fait incrémentalement après chaque micro-mise à jour du fond commun et est suivi d'une révision des hypothèses.

Dans la reconnaissance de l'acte illocutoire, le contexte joue un très grand rôle. Comme l'a indiqué Austin [4], le même énoncé peut correspondre à des actes illocutoires très différents selon le contexte. Par exemple, l'énoncé A_1 peut être : un $inform$ de l'existence du groupe, un *check* pour vérifier la connaissance de l'autre ou encore une *suggestion* pour l'organisation d'une rencontre ou d'une tournée...L'algorithme de reconnaissance d'acte illocutoire utilisé est celui développé pour le système TRAINS et est basé sur des travaux antérieurs.

6.4.6 Établissement (grounding)

Jusqu'ici, on considérait que tout ce qui a été énoncé est incorporé au fond commun mais c'est une idéalisation. En fait, pour que cela le soit réellement il faut que tous les participants aient contribué. Ce sont ces contributions que l'on appelle établissement. Ces contributions peuvent introduire de nouveaux éléments, compléter, corriger ou confirmer les éléments existants. Lorsqu'un locuteur accomplit un acte conversationnel noyau, il accomplit en parallèle un ou plusieurs actes d'établissement. Les actes d'établissement sont repris dans le tableau 2.

Pour un même segment de discours (DU), on a donc souvent deux brins conversationnels

[Conversationnal Thread] à savoir l'établissement et le but du locuteur dans l'énoncé en rapport aux autres énoncés. Mais l'établissement peut aussi se faire de manière explicite : «Ok», «Ouai», «D'accord» sans être associé à un acte noyau. On appelle script de discours un enchaînement de brins de conversations routiniers.

Il est important de pouvoir distinguer les parties du discours qui sont établies de celles qui ne le sont pas. Cela détermine à un moment donné ce qui est communément admis de ce qui ne l'est pas encore. Néanmoins, il faut avoir accès aux parties non établies qui peuvent servir de référence autant que les parties établies.

A_4 : *Il y a un groupe de musique expérimentale à Nantes.*

Le suite étant un des énoncés suivants :

A_a : *Invitons le.*

B_b : *Invitons le.*

B_c : *Hum.*

B_d : *Un quoi ?*

A_e : *Je veux dire un collectif de vidéastes.*

A_f : *Est-ce que tu m'entends ?*

Les suites A_a et B_b montrent que les deux interlocuteurs peuvent utiliser A_4 comme référence. Dans B_b, la référence peut d'ailleurs être interprétée comme un acte d'établissement. B_c est juste un acte d'établissement. Mais B_d montre que A_4 peut ne pas être compris et donc absent du fond commun. A_e montre qu'il faut pouvoir revenir en arrière. Finalement A_f montre que le locuteur ne suppose pas toujours avoir été entendu et compris avant d'avoir eu de retour (feedback). Pour rendre compte des effets de ces styles d'énoncés sur le contexte il nous faut un modèle du discours qui puisse inclure à la fois les informations établies et les non-établies. Le fond commun consiste donc en deux parties, toutes deux accessibles pour les références : la DRS-racine établie et la DRS-racine non-établie. La seconde est une extension de la première qui inclut tous les DUs non établis. Les actes d'établissement peuvent être vus comme faisant passer l'information d'un coté à l'autre.

6.4.7 Conclusion, critique, discussion

Si tout le monde s'accorde sur le slogan «le langage est action commune», par contre les théories du fond commun ou de l'action commune diffèrent toutes. Traum et Poesio [55] veulent unifier ces théories incompatibles et issues de domaines comme : la résolution de références, la reconnaissance d'intention ou le management du dialogue. Par exemple les théories linguistiques du contexte développées par Kamp et Reyle [29] pour asseoir la sémantique par anaphore sont incompatibles avec les théories du contexte développées pour capturer les effets des actes de langage par Cohen et Levesque [16]. De plus, elles sont incompatibles sans que l'on puisse choisir l'une ou l'autre puisqu'elles adressent des points différents. De même, les théories du fond commun de la tradition des actes de langages ne sont pas compatibles avec celles développées par l'analyse conversationnelle [34]. Ces dernières étant plus concentrées sur les aspects de coordinations comme les tours de parole, complètement délaissés par la théorie des actes de langages.

Cette théorie des informations partagées lors d'une situation de discours est adaptée aux conversations orales (micro-évènements conversationnels). Elle étend la DRT qui est un outil des sémanticiens formalistes. C'est en cela une approche mixte à différents niveaux : cohérence-informationnelle (DRT) / intentionnelle (Actes de langage-états mentaux), syntaxe (situation de discours) / sémantique (situation décrite). Malgré la complexité de ce modèle, il reste un certain nombre d'éléments qui ne sont pas pris en compte par le formalisme, notamment les ressources cognitives limitées, discutées par Walker [58].

7 Discussion

Les recherches sur le dialogue en langage naturel tentent de répondre à deux questions majeures : Quelles sont les informations contenues dans une séquence d'énoncés qui ne sont contenues dans aucun des énoncés qui la composent ? Comment est-ce que le contexte influe sur le sens d'un énoncé ? L'IA distribuée, les systèmes multi-agents et la linguistique computationnelle essaient de capturer ces notions avec des systèmes formels. Il y a deux approches : l'approche informationnelle et l'approche intentionnelle [23].

Pour l'approche informationnelle, la cohérence du discours découle des relations sémantiques qui existent entre les informations exprimées par les différents énoncés. Les outils employés sont l'abduction et l'inférence sur les contenus propositionnels des énoncés. Avec l'approche intentionnelle, la cohérence du discours découle des intentions des interlocuteurs et la compréhension dépend de la reconnaissance de celles-ci. Les travaux présentés ici s'inscrivent tous dans cette seconde famille d'approches qui prennent leurs racines dans les travaux de philosophes du langage tels que Grice [22] ou Searle [45] ainsi qu'en psycholinguistique et sociolinguistique. Le modèle de dialogue coopératif par plans partagés de Balkanski et Hurault-Plantet basé sur la théorie du discours de Grosz et Kraus en est l'exemple type.

Une des originalités de la théorie de l'arrière plan conversationnel développée par Traum et Poesio est de partir d'un outil de linguistique computationnelle classique, la DRT qui fait partie de l'approche informationnelle pour l'enrichir de sorte qu'il soit compatible avec une approche intentionnelle. Voyons maintenant si il est pertinent d'imaginer utiliser ces cadres formels pour l'interaction d'agents logiciels.

7.1 Application des approches du langage naturel aux SMAs

Comme on l'a vu, la théorie des actes de langage est classiquement utilisée pour les systèmes multi-agents. Mais est-il également utile d'utiliser des théories comme celles présentées dans la section précédente pour les SMAs ?

Les méthodes basées sur les plans sont comparables à des protocoles modulaires qui se construiraient dynamiquement par emboîtement et pré-séquencement de leurs parties. Le modèle de Balkanski et Hurault-Plantet peut être appliqué aux SMAs d'autant plus simplement qu'il est issu de techniques d'IA telle que la planification. En outre, la grande difficulté de cette classe de modèles qui est la reconnaissance des actes illocutoires sous-jacents aux énoncés en langage naturel (voir la section 6.2) disparait lorsque ce sont des systèmes artificiels qui communiquent entres eux. En effet, les actes de langage échangés pas les agents artificiels sont généralement codés de façon à ce que les actes illocutoires y soient explicites[26].

Dans le modèle de Balkanski et Hurault-Plantet, c'est le système - unique agent artificiel de l'application - qui maintient le plan partagé. Dans un modèle où plusieurs agents artificiels communiquent ensemble, si le plan est un objet partagé (physiquement) par les agents, on retombe sur un modèle centralisé (au moins au niveau de cette donnée). On suppose donc que chaque agent dispose de sa version du plan partagé. Une des contraintes pour la distribution de ce modèle (et donc son application aux SMAs) est qu'alors, pour que la tâche soit menée à bien, il faut que les recettes des agents coïncident, au moins au niveau des actions communicatives. Lorsque deux agents communiquent pour accomplir ensemble une tâche, deux possibilités sont donc envisageables :

– les deux agents disposent des mêmes recettes : ce n'est pas très intéressant car cela signifie que les agents connaissent leurs plans réciproques pour la tâche. Néanmoins, cela permet de concevoir les recettes de manière générique ce qui évoque la notion de compétence. Les agents seront alors homogènes.
– les agents disposent de recettes différentes, asymétriques : dans cette autre approche un agent peut avoir une recette pour la tâche et l'autre une autre recette pour la même tâche (avec les actes communicatifs qui se correspondent). De sorte que chacun exécute la recette parallèlement et que la tâche est accomplie de manière commune sans pour autant que les agents sachent tous les détails du plan de l'autre. Les agents peuvent alors être hétérogènes.

Dans les deux cas, une des qualités de ce modèle est d'être complet dans le sens où tout le fonctionnement de l'agent est spécifié et les recettes sont mixtes (actions et actes communicatifs). Le modèle est donc prêt à l'usage, le développeur n'a plus qu'à mettre au point les recettes correspondantes aux tâches du domaine que les agents auront à effectuer. Les problèmes rencontrés avec les utilisateurs humains disparaissent dès lors qu'un agent utilise forcément une recette connue. Une des limites de ce modèle est la construction ad-hoc d'un grand nombre de recettes. Il faudrait une étude poussée de la question pour situer une telle approche en rapport aux méthodes à base de protocoles.

Comme nous l'avons indiqué dans un précédent rapport [42], le langage est co-construction d'interprétations communes. Le modèle du fond commun de Traum et Poesio (plus orienté linguistique) pose la réalisation des actes illocutoires comme collective grâce au phénomène de l'établissement. C'est une évolution en ce sens que d'opposer aux actes de langages traditionnels (cadre $F(p)$), les actes multi-niveaux. Les niveaux sont liés entre eux (au moins les trois premiers) de sorte qu'un acte noyau (type $inform$) est réalisé avec succès seulement si une certaine structure d'actes d'établissement est remplie. Néanmoins, il n'y a pas de relation entre le type

[26]Notons que d'autres problèmes subsistent, comme indiqué à la section 5.1 .

de séquence d'actes d'établissement et l'acte noyau réalisé. L'avantage de cette façon d'envisager les actes de dialogue est d'établir précisément le fond commun. Concrètement, dans le langage naturel, les actes d'établissement ne sont pas nécessairement distincts des actes noyaux. Pour les SMAs cela reste une question de savoir s'ils doivent l'être ou non. Les actes de tour de parole peuvent aussi être profitables dans les SMAs pour établir de nouveaux protocoles où les interventions des agents seraient plus libres. On peut par exemple imaginer des conversations entres les agents d'un groupe où chacun participe selon ses possibilités. Et dynamiquement, par établissement, un point de vue du groupe sur une question donnée s'établit.

La théorie du fond commun de Traum et Poesio est dédiée aux conversations orales et les micro-évènements conversationnels y jouent un rôle important, un certains nombre d'aménagements seraient nécessaires pour l'adapter efficacement au mode écrit utilisé par les agents. Un des problèmes à surmonter est la quantité de messages nécessaires. De manière simpliste : lorsque cinq agents conversent, chaque message doit être distribué puis établi par les quatre autres agents, soit $4 + 4 * 4 = 20$ messages pour la réalisation d'un acte de langage quelconque sans compter les actes de tour de parole[27]. Un autre élément fondamental de la théorie de Traum et Poesio est la distinction entre situation du discours et situation décrite. Le système de référence proposé permettrait une bien plus grande flexibilité dans l'organisation du dialogue multi-agents. Là encore, une étude qui dépasse le cadre de cette synthèse serait nécessaire pour évaluer l'intérêt de cette approche dans un contexte SMA[28].

En outre, appliquer les théories dédiées aux systèmes homme-machine oraux aux SMAs pourrait permettre de structurer l'interaction entre agents logiciels et entre agents logiciels et humains de manière uniforme. Malheureusement, comme on l'a vu, les théories formelles du langage naturel sont encore approximatives et la portée des systèmes développés est limitée à de petits domaines bien circonscrits. Devant la difficulté de construire des théories plus générales, on peut questionner l'épistémologie.

7.2 Questions épistémologiques

Pour mettre en perspective les travaux sur la formalisation du langage naturel, on peut se référer aux nombreux débats de sciences cognitives ou d'épistémologie à ce sujet. Il apparaît qu'une grande part de la linguistique-informatique et des travaux concernant la communication inter-agents se basent sur les systèmes formels logiques (la version analytique searlienne de la théorie des actes de langage, la DRT ...). On connaît un certain nombre de limites de ces systèmes formels [33]. Pour ce qui est de leurs applications concernant le langage et la langue, l'épistémologie fournit des contre-indications [8]. Au sein-même des sciences cognitives, la possibilité de rendre compte de phénomènes complexes comme la communication avec les concepts du cognitivisme (auxquels souscrivent toutes les approches logiques) est remise en cause [56, 57]. Même si ces limites sont théoriques et discutées, il est bon de garder en tête des questions comme : Quel est le sens de ces tentatives ? Quels en sont les objectifs ? Quels en sont les espoirs explicites ? Et plus généralement encore : les propriétés «scientifiques» du langage lui sont-elles intrinsèques ou relèvent-elles simplement des méthodes d'analyse qu'on lui applique ?

[27] Même si ces actes de niveaux différents peuvent être fait de manière simultanée.

[28] Traum [53] donne des éléments dans ce sens.

8 Conclusion

L'étude du langage est classiquement décomposée en cinq niveaux : phonologie / morphologie, lexique, syntaxe, sémantique et pragmatique. Chacun de ces niveaux peut être vu de manière synchronique[29] ou diachronique[30]. Dans tous les cas, les deux derniers posent encore de gros problèmes fondamentaux. L'intégration de tous ces niveaux est effectuée de manière empirique et le besoin d'une véritable théorie du langage vu comme moyen d'expression se fait sentir [35]. Devant ce besoin, on comprend la perplexité des informaticiens - et non d'eux seuls - devant la complexité des productions des linguistes. Ce besoin est celui de devoir composer les problèmes syntaxiques, relativement bien résolus, aux problèmes sémantiques, longtemps délaissés.

En outre, l'échec de l'entreprise chomskienne a des conséquences méthodologiques que l'on ne doit pas oublier. Il s'agit de ré-intégrer l'empirique dans l'étude et la formalisation du langage et de garder à l'esprit que toute formalisation linguistique ne signifie rien d'autre qu'un souci de rigueur et d'objectivité dans l'explicitation de cet empirique [39]. Cependant, dans cette synthèse, on a montré l'utilité des théorisations du langage naturel pour les systèmes artificiels, que ce soit pour : les SMA avec le cadre COSMO [62], les systèmes de médiatisation avec le FLBC [40] ou bien les interfaces homme-machine avec le système TRAINS [55] ou l'automatisation d'un standard téléphonique [7]. En outre, on a discuté de l'intérêt de théories raffinées comme celles du langage naturel oral utilisée en IHM pour les SMAs et les systèmes de médiatisation.

Plus généralement, l'intérêt des modèles du langage naturel pour les systèmes artificiels est multiple. Pratiquement, cela donne un cadre unificateur à des recherches souvent difficiles à comparer. De plus, l'observation des conversations humaines est une base pour l'élaboration de protocoles ou de mécanismes d'interaction sophistiqués. En outre, les humains ont développé des techniques d'interactions très perfectionnées (le langage naturel a «le plus grand» pouvoir d'expression) qui s'accommodent de leur rationalité limitée. Dès lors que les agents artificiels sont eux aussi limités, on peut s'en inspirer.

Tout cela signifie bien qu'au lieu d'abandonner les efforts de formalisation du langage naturel, il s'agit de valider les théories résultantes par l'adéquation empirique qu'elles établissent. C'est l'intérêt réciproque que les systèmes artificiels décrits ici ont vis-à-vis des théories formelles du langage naturel. Cet aller-retour participe du projet de l'IA cognitive et plus largement de celui des sciences cognitives.

Références

[1] J. F. Allen, L. K. Schubert, G. Ferguson, P. Heeman, C. H. Hwang, T. Kato, M. Light, N. Martin, B. Miller, M. Poesio, and D. R. Traum. The TRAINS project : A case study in building a conversationnal planning agent. *Journal of Experimental and Theorical AI, MA*, (7) :7–48, 1995.

[2] K. J. Arrow. *Social Choice and Individual Values*. Wiley, 1951.

[3] N. Asher. *Reference to Abstract Objects in English*. Dordrecht, 1993.

[29] étude et description d'un phénomène linguistique à un moment donné.

[30] étude et description d'un phénomène linguistique selon sa construction et son évolution temporelle.

[4] J. L. Austin. *How to Do Things With Words.* Oxford University Press : Oxford, England, 1962.

[5] J. L. Austin. *Quand dire, c'est faire.* Seuil, 1962.

[6] K. Bach and R. M. Harnish. *Linguistic Communication and Speech Acts.* MIT Press, Cambridge, Mass., 1979.

[7] C. Balkanski and M. Hurault-Plantet. Cooperative requests and replies in a collaborative dialogue model. In Academic Press, editor, *Int. J. Human-Computer Studies*, number 53, pages 915–968, 2000.

[8] H. Barreau. *L'épistémologie.* Presses Universitaires de France, 1995.

[9] M. E. Bratman. *Intentions, Plans, and Practical Reason.* Harvard University Press : Cambridge, MA, 1987.

[10] H. Bunt. *The Structure of Multimodal Dialogue*, chapter Dynamic Interpretation and Dialogue Theory. John Benjamin, Amsterdam, 1996.

[11] E. Bylander. Complexity results for planning. In *Prooceedings of the 12th International Joint Conference on Artificial Intelligence*, pages 274–279, Sydney, Autralia.

[12] B. Chaib-draa and L. Vongkasem. ACL as a joint project between participants : A preliminary report. In F. Dignum and M. Greaves, editors, *Issues in Agent Communication*, number 1916 in Lecture Notes in Artificial Intelligence, pages 235–248. Springer, 2000.

[13] H. H. Clark and C. R. Marshall. *Elements of Discourse Understanding*, chapter Definite Reference and Mutual Knowledge. Cambridge University Press, New York, 1981.

[14] H. H. Clark and E. F. Schaeffer. Collaborating on contributions to conversations. *Language and Cognitive Processes*, (2) :19–41, 1987.

[15] P. R. Cohen. *Survey of the State of Art in Natural Language Technology*, chapter Discourse and Dialogue, pages p.234–241. Universitat des Saarlandes, 1996.

[16] P. R. Cohen and H. J. Levesque. *Intentions in communication*, chapter Persistence, Intention and Commitment. Number 12. Morgan Kaufmann, 1990.

[17] K. Devlin. *Logic and Information.* Cambridge University Press : Cambridge, England, 1991.

[18] F. Dignum. Issues in agent communication : An introduction. In F Dignum and M. Greaves, editors, *Issues in Agent Communication*, Lecture Notes in Artificial Intelligence, pages 1–16, 2000.

[19] G. Ferguson. *Knowledge Representation and Reasoning for Mixed-Initiative Planning.* Tr-562, University of Rochester, Department of Computer Sciences, 1995.

[20] A. Goldman. *A Theory of Human Action.* Princeton University Press, Princeton, NJ, 1970.

[21] H. P. Grice. Meaning. *Philosophical review*, 1957.

[22] H.P Grice. Utterer's meaning and intentions. *Philosophical Review*, 1969.

[23] B. J. Grosz. *Survey of the State of the Art in Human Language Technology*, chapter Discourse and Dialogue. National Science Foundation, European Commission, 1996.

[24] B. J. Grosz and S. Kraus. Collaborative plans for group activities. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)*, pages 367–373, Chambéry, France, 1993.

[25] B. J. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, pages 269–357, 1996.

[26] B. J. Grosz and C. L. Sidner. Attention, intentions and the structure of discourse. *Computationnal Linguistics*, 12 :175–204, 1986.

[27] B. J. Grosz and C. L. Sidner. Plans for discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 417–444. The MIT Press : Cambridge, MA, 1990.

[28] M. H. Huhns, D. M. Bridgeland, and N. V. Arni. A DAI communication aide. In *Procedings of the 10th International Workshop on Distributed Artificial Intelligence*, number 29, 1990.

[29] H. Kamp and U. Reyle. *From Discourse to Logic : Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, volume 42 of *Studies in Linguistics and Philosophy*. Kluwer Academic Publishers : Boston, MA, Institute for Computational Linguistics, University of Stuttgart, 1993.

[30] R. Kaplan and J. Bresnan. *The Mental Representation of Grammatical Relations*, chapter Lexical Functional Grammar : A Formal System for Grammatical Representation, pages 173–281. The MIT Press : Cambridge, MA.

[31] L. Karttunen. The syntax and semantic of questions. *Linguistique and Philosophy*, pages 1–44, 1977.

[32] S. O. Kimbrough. On representation schemes for promising electronically. *Decision Support Systems*, 2(6) :99–122, May 1990.

[33] J. Ladrière. Les limites de la formalisation. *Logique*, pages 313–333, 1996.

[34] S. C. Levinson. *Pragmatics*. Cambridge University Press : Cambridge, England, 1983.

[35] J. S. Lienard. Communication homme machine ; rapport sur la définition, l'état de l'art et les perspectives scientifiques. Technical report, Rapport du Comité d'Objectif Scientifique et Technique (COST), LIMSI, SPI-CNRS, avril 1991.

[36] K. E. Lochbaum. *Using Collaborative Plans to Model the Intentionnal Structure of Discourse.* PhD thesis, Harvard University, Cambridge, MA., 1994.

[37] A. Lodder. On structure and naturalness in dialogical models of argumentation. In J. Hage, T. Bench-Capon, and A. Koers, editors, *The 11th Conference JURIX*, 1998.

[38] N. Maudet. *Modéliser les Conventions des Interactions Langagières : la Contribution des Jeux de Dialogue.* PhD thesis, ENSEIHT-UPS, 2001.

[39] J. C. Milner. *Introduction à une Science du Langage.* Seuil, Paris, 1989.

[40] S. A. Moore and S. O. Kimbrough. On automated message processing in electronique commerce and work support systems : Speech act theory and expressive felicity. *ACM Transactions on Information Systems*, 15(4) :321–367, October 1997.

[41] R. Neches, R. Fikes, T. Finin, T. Gruber, R. Patil, and T. Senator. Enabling technology for knowledge sharing. *AI Magazine*, Fall :36–56, 1991.

[42] P. Pasquier. Communication entre agents. Rapport de synthèse, Laboratoire DAMAS, Université Laval, Québec, Canada, juin 2001.

[43] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 1-2(13) :81–132, April 1980.

[44] D. Sadek, P. Bretier, and F. Panaget. ARTIMIS : Natural dialogue meets rational agency. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97)*, Yokohama, Japan, 1997.

[45] J. R. Searle. *Speech Acts : An Essay in the Philosophy of Language*. Cambridge University Press : Cambridge, England, 1969.

[46] J. R. Searle. *Expression and Meaning*. Cambridge University Press : Cambridge, England, 1979.

[47] J. R. Searle. *La Redécouverte de l'Esprit*. NRF Essais. Gallimard, 1992.

[48] J. R. Searle and D. Vanderveken. *Foundations of Illocutionary Logic*. Cambridge University Press, NY, 1985.

[49] Y. Shoham. Agent-oriented programming. Technical Report STAN–CS–1335–90, Computer Science Department, Stanford University, Stanford, CA 94305, 1990.

[50] R. W. Smith, D. R. Hipp, and A. W. Biermann. An architecture for voice dialogue systems based on prolog-style theorem proving. *Computationnal Linguistics*, (21) :281–320, 1995.

[51] J. Sowa. *Conceptual Structures : Information processing in Man and Machine*. Reading. Addison-Wesley, 1984.

[52] D. R. Traum. *A Computationnal Theory of Grounding in Natural Language Conversation*. PhD thesis, Departement of Computer Sciences, University of Rochester, 1994.

[53] D. R. Traum. *Foundations and Theories of Rational Agents*, chapter Speech Acts for Dialogue Agents, pages 169–201. Kluwer Academic Publishers, 1999.

[54] D. R. Traum and E. A. Hinkelman. Conversation acts in task-oriented spoken dialogue. *Computationnal Intelligence*, 1992. Special Issue on Non-literal Language.

[55] D. R. Traum and M. Poesio. Conversational actions and discourse situations. *Computationnal Intelligence*, 13(3) :309–347, 1997.

[56] F. Varela. *Invitation aux Sciences Cognitives*. Seuil, 1996.

[57] G. Vignaux. *Les Sciences Cognitives : une Introduction*. le Livre de Poche : biblio essais, Paris, la découverte edition, 1991.

[58] M. A. Walker. *Informational Redundancy and Ressource Bounds in Dialogue*. PhD thesis, University of Pennsylvania, 1993.

[59] B. L. Webber. *A Formal Approach to Discourse Anaphor*. Garland, New York, 1979.

[60] L. Wittgenstein. *Philosophical Investigations*. MacMillan, 1953.

[61] S. T. C. Wong. *A framework for Description and Design of Cooperative Knowledge based systems*. PhD thesis, Lehigh University, Oct. 1991.

[62] S. T. C. Wong. COSMO : A communication scheme for cooperative knowledge-based systems. *IEEE Transactions on Systems, Man and Cybernetics*, 23 :809–824, May/June 1993.