Établissement de croyances mutuelles dans le dialogue homme-machine

Michael Brenner

DEA de Sciences Cognitives
1998 – 1999

Stage effectué sous la direction de
Cécile Balkanski et Martine Hurault-Plantet

du 20 mars au 20 septembre 1999
dans le Groupe Langage et Cognition
du LIMSI/CNRS, Orsay

Remerciements

J'ai passée une année formidable en France et au LIMSI. C'est en raison des réponses encourageantes de Gérard Ligozat à mes premiers méls (il y a déjà deux ans!) que ceci est devenu possible et c'est donc lui que je remercie en premier. Je remercie également mes professeurs Bernhard Nebel et Hans Spada à Freiburg qui ont soutenu ce projet et se sont montrés toujours généreux quand j'avais encore une fois besoin d'une lettre de recommandation.

L'atmosphère au LIMSI a été très agréable dès le (pot du) début de l'année et l'est toujours resté, pendant la période des cours ainsi que pendant mon stage. C'est surtout Mady Méheut qui dès le début a donné aux étudiants le sentiment d'être les bienvenus et qui a su résoudre tous les problèmes de salle. Je remercie tous les professeurs que j'ai eu pendant cette année. J'ai profité beaucoup de (presque) chacun de leurs cours. Quelques uns d'entre eux vont certainement retrouver dans ce travail des choses que j'ai apprises dans leurs cours. Je remercie mes camarades du DEA, non seulement pour m'avoir laissé photocopier tous ces cours qui allaient trop vite pour moi, mais surtout pour les discussions, les soirées, les pauses café – tout cela a créé une bonne ambiance entre nous.

Je remercie Anne Vilnat qui m'a soutenu beaucoup, surtout vers la fin de mon stage quand j'étais malade, ainsi qu'encore Gérard Ligozat, cette fois pour avoir lu et commenté le chapitre 5 à la dernière minute. En fait, je remercie tout le groupe Langage et Cognition qui m'a accueili si gentiment.

Enfin et surtout, je remercie Cécile et Martine. Elles n'ont pas seulement encadré ce stage, mais elles m'ont fourni une atmosphère de travail qui était plus qu'agréable. Elles m'ont proposé un sujet de stage passionnant qui a mené à de nombreuses discussions intéressantes et instructives pour moi.

Table des matières

Liste des tableaux

Table des figures

Chapitre 1

Introduction

1.1 Le processus de «grounding»

Pour qu'un dialogue puisse réussir, les participants ont besoin d'une «base commune», c'est-à-dire des connaissances connues et partagées par tous les participants et telles que chaque agent [1] sait que les autres les connaissent également. Un agent utilisera un mot, par exemple, seulement s'il croit que ce mot à une certaine signification dans sa langue, s'il croit que l'auditeur croit en la même signification, s'il croit que l'auditeur croit que le locuteur croit en la même signification que l'auditeur, etc. Si le locuteur n'était pas certain de cette «base commune», il ne pourrait pas être certain non plus que l'auditeur le comprenne. Puisque ce type de «base commune» est exprimée en termes de croyances et de croyances sur les croyances de l'autre, on parle de **croyances mutuelles**. La notion de croyance mutuelle (des expressions comme *connaissance commune*, *croyance partagée*, etc., sont également utilisées) est étudiée depuis longtemps et sa nécessité en pratique a été montrée, entre autres, tant dans des domaines linguistiques (signification d'un énoncé: Schiffer (1972), utilisation d'expressions anaphoriques: Clark et Marshall (1981)) que non linguistiques (comportement collaboratif: (Grosz et Sidner, 1990)).

Les croyances mutuelles représentent une condition nécessaire de bonne communication et en contrepartie la communication sert à établir de nouvelles croyances mutuelles. Le processus d'établissement de croyances mutuelles dans le dialogue a été nommé **grounding** par Clark et Schaefer (1989). Bien que les travaux en intelligence artificielle aient longtemps ignoré le problème ou aient supposé que l'établissement de croyances mutuelles est un effet automatique d'un énoncé, le processus de «grounding» [2] a été l'objet d'études en psycholinguistique (Clark et Schaefer (1987)) et a été décrit dans un modèle computationnel par Traum (1994). Le processus de «grounding» est nécessaire parce que la compréhension est incertaine: l'auditeur peut ne pas comprendre ou mal comprendre un énoncé et le locuteur peut donc ne pas être certain d'être compris quand il produit cet énoncé. Il a été démontré qu'en théorie une croyance mutuelle ne peut pas être atteinte si la communication est faillible (Halpern et Moses, 1990),

1. Quand nous utilisons le mot *agent*, nous désignons tout être capable de montrer un certain comportement. Un des buts de ce travail sera de montrer comment un agent *artificiel* (un programme ou robot) peut être conçu pour montrer un comportement langagier normalement reservé aux agents *humains*.

2. par manque d'expression appropriée nous allons utiliser le mot anglais comme terme technique, ainsi que sa version adjective «grounded» pour désigner un fait qui est devenu mutuellement connu.

(1)	Utilisateur:	Je voudrais parler à Corinne Savol.
(2)	Standard:	Corinne Savol?
(3)	Utilisateur:	Oui.
(4a)	Standard:	Je ne connais pas Corinne Savol.
(4b)		Quel est son poste?
(5)	Utilisateur:	Je ne sais pas.
(6a)	Standard:	Je n'ai pas le numéro de poste de Corinne Savol.
(6b)		Savez-vous avec qui travaille Corinne Savol?
(7)	Utilisateur:	Elle travaille avec Louise Abou.
(8)	Standard:	Je vous passe Louise Abou.

FIG. 1.1 – *Un dialogue avec le système de Balkanski et Hurault-Plantet (1999)*

mais selon Clark et Schaefer (1989) les participants à un dialogue semblent atteindre un *critère de grounding* qui signifie qu'ils croient mutuellement qu'ils ont chacun compris ce que le locuteur a voulu dire «à un degré suffisant pour l'objectif courant».

1.2 Buts et contributions de ce travail

La figure 1.1 montre la simulation d'un dialogue avec un standard téléphonique. Le standard a été modélisé par un système de dialogue développé par Balkanski et Hurault-Plantet (1999). Le système modélise les aspects *collaboratifs* d'un dialogue en ayant recours aux théories de Grosz et Sidner (1986, 1990), Grosz et Kraus (1996) et Lochbaum (1998). Bien que le système utilise une représentation élaborée d'actions et du plan sous-jacent au dialogue permettant ainsi d'interpréter et de générer des actions collaboratives, les croyances (notamment les croyances mutuelles) ne sont représentées qu'implicitement dans le plan du dialogue. Ceci provoque des énoncés comme (6a) qui est produit dans le but d'être coopératif, mais ne tient pas compte du fait que l'utilisateur connait déjà l'information qui est exprimée. En effet, avant d'énoncer (6a), le standard n'a pas le numéro cherché, l'utilisateur le sait, le standard sait que l'utilisateur le sait, etc. Il s'agit d'une croyance mutuelle qui ne nécessite plus d'être exprimée. Pourtant, au début du dialogue cette croyance ne semble pas encore être mutuelle. La motivation initiale de ce travail a donc été de déterminer à quel moment et comment cette croyance est établie dans le dialogue et comment éviter que le système produise de tels énoncés inutiles.

Les modèles existants ne se sont montrés que partiellement pertinents pour cette tâche, car certains phénomènes importants (comme la demande de confirmation dans l'énoncé (2)) ne peuvent pas, ou seulement difficilement, être expliqués avec ces modèles. Nous avons donc développé des notions pour une nouvelle théorie du processus du «grounding» qui, entre autres, tient compte de ces phénomènes. Cette théorie décrit comment un énoncé peut produire chez le locuteur un état d'*attente* d'un signal de compréhension et comment l'attente permet le «grounding» après une réponse adéquate de l'auditeur. Une notion clé de la théorie est l'état mental de *supposition* qui peut, entre autres, exprimer l'incertitude de l'auditeur sur son interprétation, une incertitude qui s'exprime, par exemple, dans une demande de confirmation. Pour pouvoir spécifier un agent artificiel se comportant selon les prédictions de la théorie, nous avons fourni une formalisation de la théorie en utilisant une logique multi-modale et nous avons démontré dans cette logique comment une séquence d'énoncés peut mener à l'établissement d'une croyance mutuelle. Les résultats de la formalisation ont ensuite été utilisé pour décrire les parties liées au «grounding» d'un modèle de gestion de dialogue et de la représentation de croyances. Finalement, ce modèle a été intégré dans le système de Balkanski et Hurault-Plantet (1999). Ce système comprend maintenant une représentation de croyances et des algorithmes permettant d'éviter la génération des énoncés dont le contenu est déjà connu de l'auditeur.

1.3 Plan du travail

Le chapitre suivant présente des travaux et notions fondamentaux pour notre recherche. Des travaux plus concrètement liés à l'établissement de croyances mutuelles seront présentés dans le chapitre 3. Nous proposons des éléments pour notre théorie du «grounding» dans le chapitre 4. Le chapitre 5 présente une formalisation de la théorie ainsi que des preuves de quelques propriétés d'un agent spécifié selon cette formalisation. Le chapitre 6 décrit une architecture pour un agent de dialogue qui intègre les résultats de la théorie. Cette architecture est à la base du système développé dans le cadre de ce travail et qui est présenté dans le chapitre 7. Finalement, le chapitre 8 revoit les contributions de notre travail et, surtout, présente des pistes à poursuivre.

Chapitre 2

Travaux fondamentaux

Ce chapitre présente un tour d'horizon de quelques travaux fondamentaux pour notre recherche. La section 2.1 donne un aperçu de quelques travaux importants en philosophie de la langue. Puisque ces travaux utilisent largement des notions cognitives comme «croyance» ou «intention», nous présentons en section 2.2 des études traitant de la représentation et de la formalisation de ces attitudes mentales. Ceci nous permet d'étudier les approches sur la conversation développées en intelligence artificielle en section 2.3. Dans le chapitre suivant nous regarderons de plus près les travaux qui se sont centrés sur le même thème que nous, l'établissement de croyances mutuelles dans le dialogue.

2.1 Philosophie de la langue

Quelques uns des choix que nous avons pris dans l'élaboration de notre théorie et quelques unes des critiques vis-à-vis d'autres modèles sont directement motivés par des travaux en philosophie de la langue que nous présentons donc dans cette section.

2.1.1 Les actes de langage

Austin (1962) remarque qu'il y a des phrases auxquelles ont ne peut pas naturellement attribuer une valeur de vérité. Une phrase comme «Je te couronne roi de France» ne décrit pas l'état du monde (et ainsi serait vraie ou fausse), mais est une tentative de *changer* ce monde (qui peut donc réussir ou échouer). La phrase peut donc avoir les caractéristiques d'une *action* et celui qui l'énonce réalise un **acte de langage**.

Austin distingue trois types d'actes qui sont réalisés simultanément lorsqu'on dit quelque chose:

1. **acte locutoire**: l'énonciation même, l'acte *de* dire quelque chose
2. **acte illocutoire**: l'action (communicative) réalisée *en* disant quelque chose: promettre, affirmer, commander, demander quelque chose, etc.
3. **acte perlocutoire**: l'acte de provoquer des effets chez les auditeurs *par* l'énoncé. Ces effets sont spécifiques aux circonstances de l'énonciation.

Un acte illocutoire a deux composantes: la *force illocutoire* spécifiant le type d'action et son *contenu propositionnel*. Searle (1969) remarque que cette décomposition devient apparente dans deux types de négation: la phrase «Je promets de venir» peut être niée ou bien par «Je promets de ne pas venir» (négation propositionnelle) ou bien par «Je ne promets pas de venir» (négation illocutoire).

Searle (1969) décrit des conditions nécessaires et suffisantes pour la réalisation d'actes illocutoires, les *conditions de félicité*. En distinguant entre certains types de conditions, Searle propose un classement des actes de langage. Il donne cinq catégories principales d'actes:

1. **assertifs** exprimant la croyance en la vérité d'une proposition énoncée
2. **directifs** essayant de faire faire quelque chose à l'auditeur
3. **promissifs** qui engagent le locuteur auprès du destinataire
4. **expressifs** informant sur un état psychologique du locuteur
5. **déclaratifs** qui agissent directement sur le monde en changeant un statut institutionnel

Cette catégorisation a été critiquée comme étant incomplète (manquent, par exemple, les injures, etc.) et arbitraire (Levinson, 1983). On peut se demander s'il est raisonnable d'essayer de trouver une catégorisation dans laquelle toute force illocutoire *possible* devrait se retrouver. Comme le remarque Levinson (1983, p. 241), il se peut que les illocutions soient en nombre fini, mais les perlocutions ne le sont sûrement pas et il n'y a pas de raison qu'un effet perlocutoire donné ne puisse pas être l'effet conventionnel et donc illocutoire dans une autre culture. Ceci indique que les forces illocutoires pourraient être aussi illimitées que les perlocutions.

Au lieu de trouver une catégorisation fixe des actes de langage, il nous semble donc plus important de regarder de plus près ce que Searle a appelé la *condition essentielle*: l'intention derrière l'acte de langage et qui déclenche sa réalisation. Cette *intention communicative* a été évoquée en premier par Grice (1957).

2.1.2 Grice: la communication intentionnelle

Partant de l'idée qu'on peut appeler *communication* seulement une transmission d'information qui a été *conçue* pour être comprise, Grice (1957) donne la définition suivante de **signification non-naturelle** (ou **communication intentionnelle**), citée ici dans la version française de Moeschler et Reboul (1994):

> Dire qu'un locuteur L a voulu signifier quelque chose par X, c'est dire que L a eu l'intention, en énonçant X, de produire un effet sur l'auditeur A grâce à la reconnaissance par A de cette intention.

L'acte communicatif consiste donc à *faire connaître* à l'autre une intention très particulière dont l'effet est atteint par simple reconnaissance. Nous parlons d'une *intention communicative*. Si un énoncé exprime une intention communicative nous allons dire qu'elle est sa *signification* (non-naturelle). Schiffer (1972) défend la thèse que la définition de l'intention communicative doit être complétée par d'autres intentions (sur la reconnaissance des premières intentions). Dans le chapitre 5, nous essayerons de fournir une telle définition.

Selon Grice, unne communication réussie nécessite la participation de l'auditeur A: A doit *reconnaître* l'intention du locuteur L. Un point important de cette approche est que le lien entre l'énoncé X et l'effet que L voulait produire peut être assez faible tant que le locuteur croit que l'auditeur peut quand-même reconnaître son intention. Comprendre un énoncé revient donc à reconnaître ce que le locuteur *voulait* dire plutôt qu'à ce qu'un énoncé exprime directement. Ceci évoque la notion de *contexte* d'un énoncé: dans différentes situations et avec différentes croyances sur l'autre, un auditeur va interpréter un même énoncé différemment et un locuteur va exprimer la même intention avec des énoncés différents.

2.2 Attitudes mentales

Nous avons vu que toute théorie de l'utilisation de la langue et donc toute théorie du dialogue doit s'appuyer sur les états mentaux des participants. Cette section présente des travaux philosophiques décrivant les différentes attitudes mentales ainsi que les approches formelles de description et de raisonnement sur ces attitudes.

2.2.1 Croyances

L'expression «je crois que φ» a au moins deux interprétations qui doivent être distinguées: l'expression peut signifier que le locuteur est *certain* que φ est vrai ou bien qu'il considère la possibilité d'avoir tort. En intelligence artificielle (non probabiliste), c'est la première définition qui est la plus utilisée et nous parlons alors des **croyances** d'un agent comme de ce dont un agent est *certain* d'être vrai. La deuxième interprétation du mot «croire» étant très importante pour ce travail, nous utilisons la notion de **supposition** pour décrire qu'un agent n'est pas entièrement certain si un fait est vrai. Informellement, «je suppose que φ» signifie «φ me semble plus probable que son contraire» [1]. Dans le chapitre 5 nous définissons formellement les opérateurs $\mathsf{Bel}_A\varphi$ et $\mathsf{Sup}_A\varphi$ qui signifient «A croit φ» et «A suppose φ», respectivement. Une troisième notion est celle de la **connaissance**. Le plus souvent, elle est définie comme «croyance + vérité», c'est-à-dire une croyance qui est en fait vraie. La connaissance est donc une notion qui est attribuée à un agent de l'extérieur. Dans ce travail, elle ne sera utilisé que quand un agent attribue à un autre une croyance que le premier croit vraie.

Contrairement à un fait propositionnel comme $\varphi =$ «les oranges sont bleues» qui doit être ou bien vrai ou bien faux, on doit distinguer trois cas différents pour les *croyances* d'un agent A sur φ:

1. A croit que φ est vrai
2. A croit que φ est faux
3. A ne sait pas si φ est vrai ou faux [2]

Pour pouvoir représenter ces différents faits, une logique des prédicats n'est donc pas suffisante. La solution la plus fréquemment utilisée et proposée initialement par Hintikka (1962) consiste en l'utilisation d'une **logique modale** de croyance (logique doxastique) ou de connaissance (logique épistémique). Une proposition peut être quantifiée avec l'opérateur modal Bel et ainsi

1. Cette description de l'incertitude provient de Sadek (1991)

2. C'est une propriété intéressante, au moins dans les langues que connait l'auteur, qu'une phrase comme *«A ne croit pas si φ» ne soit pas grammaticalement bonne.

les trois **états épistémiques** peuvent être distingués: 1. $\mathsf{Bel}_A\varphi$, 2. $\mathsf{Bel}_A\neg\varphi$ et 3. $\neg\mathsf{Bel}_A\varphi \wedge \neg\mathsf{Bel}_A\neg\varphi$. «Savoir-si φ» est souvent exprimé par $\mathsf{KIF}\varphi = \mathsf{Bel}\varphi \vee \mathsf{Bel}\neg\varphi$. Le cas de l'ignorance (3.) correspond donc à $\neg\mathsf{KIF}_A\varphi$.

Depuis Hintikka, la sémantique des logiques épistémiques et doxastiques a souvent été définie en termes de mondes possibles[3]. Cette sémantique étant très proche de celle d'une logique non modale, beaucoup de propriétés importantes en sont héritées – dont malheureusement une qui pose problème, celle de l'**omniscience logique**: la logique prédit que tout ce qui se déduit logiquement d'une croyance sera aussi cru. Mais ceci n'est pas toujours vrai pour la croyance humaine et peut ainsi mener à de nombreux problèmes. Dans l'interaction homme-machine, par exemple, de mauvaises prédictions sur les croyances de l'homme pourraient être faites. Fagin et al. (1995) proposent plusieurs solutions à ce problème. Dans la section 6.1.3 nous proposons comment en pratique les capacités de raisonnement peuvent être raisonnablement restreintes.

2.2.2 Croyances mutuelles

Certaines croyances sont dites *communes* ou *mutuelles* à un groupe d'agents. Ces expressions évoquent que non seulement tout agent dans un groupe croit qu'une proposition est vraie, mais aussi que chaque agent croit que les autres le croient également. Mais ceci ne suffit pas non plus: un agent doit aussi croire que les autres croient qu'il sait lui-même que la proposition est vraie. Il est facile de construire des exemples montrant que le sentiment intuitif que quelque chose est mutuellement connu implique des emboîtements profonds de croyances de chaque agent sur les croyances des autres et de leurs croyances sur les propres croyances de l'agent. Cette section présente des définitions formelles essayant de saisir les caractéristiques particulières de la notion de croyance mutuelle. Le phénomène a aussi été nommé connaissance commune, croyance commune, croyance partagée, etc., mais nous n'utiliserons que la notion de croyance mutuelle dans la suite. Nous utilisons la formule $\mathsf{MB}_{A,B}\varphi$ pour désigner le fait que les agents A et B croient mutuellement que φ est vrai.

La définition classique de la croyance mutuelle qui a souvent été utilisée en intelligence artificielle la décrit comme un conjonction infinie de croyances emboîtées. On parle alors de l'approche *itérée*.

$$\begin{aligned}\mathsf{MB}_{A,B}\varphi \quad =_{déf} \quad & \mathsf{Bel}_A\varphi \wedge \mathsf{Bel}_B\varphi \wedge \mathsf{Bel}_A\mathsf{Bel}_B\varphi \wedge \mathsf{Bel}_B\mathsf{Bel}_A\varphi \wedge \\ & \mathsf{Bel}_A\mathsf{Bel}_B\mathsf{Bel}_A\varphi \wedge \mathsf{Bel}_B\mathsf{Bel}_A\mathsf{Bel}_B\varphi \wedge \cdots\end{aligned}$$

Pourtant, le plus souvent les croyances mutuelles apparaissent sous la forme de croyances d'un agent particulier *sur* ce qui est mutuellement connu. On parle alors de **croyances mutuelles unilatérales**. Par exemple, Cohen et Levesque (1990c) n'utilisent qu'un opérateur BMB qui représente la conjonction suivante:

$$\mathsf{BMB}_{A,B}\varphi \quad =_{déf} \quad \mathsf{Bel}_A\varphi \wedge \mathsf{Bel}_A\mathsf{Bel}_B\varphi \wedge \mathsf{Bel}_A\mathsf{Bel}_B\mathsf{Bel}_A\varphi \wedge \cdots$$

À strictement parler, selon la définition de l'approche itérée, $\mathsf{MB}_{A,B}\varphi$ n'implique pas $\mathsf{BMB}_{A,B}\varphi$. Ceci est dû au fait qu'en principe, un agent pourrait avoir une infinité de croyances sans savoir

3. Pour une introduction cf., par exemple, (Genesereth et Nilsson, 1987)

que ce sont *toutes* les croyances nécessaires pour une croyance mutuelle. De la même manière, un groupe peut avoir une croyance mutuelle ($\mathsf{MB}_{A,B}\varphi$) sans que les membres du groupe sachent mutuellement qu'ils ont cette croyance mutuelle ($\mathsf{MB}_{A,B}\mathsf{MB}_{A,B}\varphi$). En fait, ce n'est donc qu'un observateur extérieur qui peut déterminer si un groupe a une croyance mutuelle ou pas.

Mais intuitivement, avoir une croyance mutuelle ne devrait pas seulement signifier avoir toutes les croyances individuelles, mais aussi être conscient d'avoir une croyance mutuelle. L'*approche du point fixe* décrit donc la croyance mutuelle ainsi

$$\mathsf{MB}_{A,B}\varphi \quad =_{déf} \quad \mathsf{Bel}_A(\varphi \wedge \mathsf{MB}_{A,B}\varphi) \wedge \mathsf{Bel}_B(\varphi \wedge \mathsf{MB}_{A,B}\varphi)$$

Dans cette définition, la conjonction infinie a donc été remplacée par une formule *autoréférentielle*. Cette définition peut être *développée*, c'est-à-dire la définition peut être appliquée à elle-même en donnant:

$$\begin{aligned} \mathsf{MB}_{A,B}\varphi &\rightarrow \mathsf{Bel}_A\mathsf{Bel}_B(\varphi \wedge \mathsf{MB}_{A,B}\varphi) \wedge \mathsf{Bel}_B\mathsf{Bel}_A(\varphi \wedge \mathsf{MB}_{A,B}\varphi) \\ &\rightarrow \mathsf{Bel}_A\mathsf{Bel}_B\mathsf{Bel}_A(\varphi \wedge \mathsf{MB}_{A,B}\varphi) \wedge \mathsf{Bel}_B\mathsf{Bel}_A\mathsf{Bel}_B(\varphi \wedge \mathsf{MB}_{A,B}\varphi) \end{aligned}$$

En utilisant le fait que $\mathsf{Bel}(p \wedge q) \leftrightarrow \mathsf{Bel}p \wedge \mathsf{Bel}q$, toutes les croyances de l'approche itérée sont produites, mais aussi les croyances *sur* les croyances mutuelles souhaitées. Airenti et al. (1993) utilisent une version unilatérale de l'approche du point fixe (attention: l'ordre des indices est important):

$$\mathsf{BMB}_{A,B}\varphi \quad =_{déf} \quad \mathsf{Bel}_A\varphi \wedge \mathsf{BMB}_{B,A}\varphi$$

Il en résulte que $\mathsf{BMB}_{A,B}\varphi$ implique la chaine infinie de la définition de BMB de Cohen et Levesque (1990c), mais aussi par exemple $\mathsf{Bel}_A\mathsf{Bel}_B\mathsf{MB}_{A,B}\varphi$.

L'approche itérée ainsi que l'approche du point fixe ont un inconvénient important: il n'est pas clair comment une croyance mutuelle peut se former. L'approche itérée semble nécessiter la vérification d'une infinité de croyances et l'approche du point fixe se définit en termes de croyances sur des croyances mutuelles – mais d'où viennent celles-ci?

La dernière possibilité de définition que nous présentons est donc de décrire les *situations* dans lesquelles naissent les croyances mutuelles. On parle de l'approche des *situations partagées*:

> Nous disons qu'il existe une croyance mutuelle, dans une population P, que X, si et seulement si la situation actuelle A est telle que:
>
> 1. Tout le monde dans P a des raisons de croire que A a lieu.
> 2. A indique à tout le monde dans P que tout le monde dans P a des raisons de croire que A a lieu.
> 3. A indique à tout le monde dans P que X.

Cette approche est à la base de l'heuristique de coprésence pour l'établissement de croyances mutuelles de Clark et Marshall (1981) qui sera présentée en section 3.1. Bien que l'approche des situations partagées décrit quand on peut parler de croyances mutuelles, le défaut de la définition est évident: elle décrit la situation, mais pas les croyances que cette situation entraîne et ainsi est peu utile pour décrire l'état mental d'un agent. À vrai dire, il ne s'agit pas d'une nouvelle définition, mais de la *description* d'une propriété de la croyance mutuelle décrite par l'approche du point fixe. Dans la formalisation que nous fournissons dans le chapitre 5, nous utiliserons donc l'approche des situations partagées comme *règle d'inférence* qui décrit comment une croyance mutuelle peut naître d'un ensemble de croyances «ordinaires».

L'établissement de croyances mutuelles étant le thème de ce travail, nous avons consacré le chapitre 3 entièrement aux travaux qui en traitent.

2.2.3 Intentions

L'intention est une attitude qui relie un agent aux actions qu'il a *décidé* de réaliser. Bratman (1990) propose une théorie détaillée de l'intention en mettant l'accent sur les propriétés suivantes: une intention ne peut pas avoir d'effets directs sur le futur, mais agit en guidant le raisonnement et le comportement d'un agent dans le présent. Ainsi, elle permet d'une part la coordination d'activités complexes et d'autre part de surmonter les limites de capacité du raisonnement humain au moment de l'action. Les intentions *contrôlent* le comportement de l'agent, contrairement aux *désirs* qui n'ont qu'une «influence potentielle» sur les actions de l'agent. Bratman précise trois rôles psychologiques des intentions: lorsqu'un agent a l'intention d'obtenir un certain effet, cette intention le force premièrement à un raisonnement sur la fin et les moyens (comment obtenir l'effet), deuxièmement à l'adaptation de ses autres intentions pour qu'elles soient cohérentes, et dernièrement, en exécutant l'action, l'agent va s'efforcer d'obtenir les effets qu'il a l'intention d'obtenir.

Il est raisonnable de distinguer entre deux types d'intentions: l'intention *de* réaliser une action dans le futur et l'intention *que* un certain état du monde soit réalisé. Grosz et Kraus (1996) les dénomment *Intend To* et *Intend That*, respectivement. Elles ont en commun de contraindre la planification de l'agent et de devoir être mutuellement cohérentes. Par contre, leur rôle dans le raisonnement est différent: Int.To force l'agent au raisonnement sur la fin et les moyens et à l'action (Bratman, 1990). Int.Th peut entraîner une Int.To mais est aussi la motivation pour la coordination de différents plans, aider un collaborateur, etc.

Bratman (1990) montre que l'attitude qu'on a vis-à-vis des *effets de bord* d'une action ne joue pas ces trois rôles. Un agent n'a donc pas nécessairement l'intention d'obtenir les effets impliqués par ses intentions. Par exemple, même si se faire arracher une dent provoque des douleurs, je peux avoir l'intention de me faire traiter sans l'intention d'avoir mal. Le problème de la *fermeture conséquentielle* se pose donc pour chaque formalisation de l'intention en logique modale. Comme notre formalisation actuelle (chapitre 5) ne fournit pas encore une définition de l'intention, nous présentons dans la suite un modèle de l'intention qui pourrait s'intégrer dans notre formalisme et qui évite le problème de la fermeture conséquentielle.

Konolige et Pollack (1993) représentent l'intention avec une logique modale non-normale. Chez eux, un monde représente un futur possible. Pour chaque intention d'un agent, les mondes possibles se divisent en mondes *désirés* (où l'objet de l'intention est vrai) et *non désirés* (où ce n'est pas le cas). L'idée principale est de dire qu'on a une intention sur un fait p seulement si les mondes où p est vrai sont désirés et si en plus les mondes où p est faux ne sont pas désirés. C'est cette deuxième condition qui distingue un fait désiré de sa conséquence logique. La figure 2.1 illustre cette idée. Supposons que p = 'dent arrachée' et q = 'douleurs'. Les mondes dans lesquels p est vrai sont représentés par la zone grise. On voit que dans tous ces mondes q est aussi vrai – c'est normal parce qu'il s'agit d'une conséquence logique de p. Mais bien que q soit vrai dans tous les mondes désirés, il n'est pas vrai que dans tous les mondes non désirées q soit faux. On n'a donc pas l'intention que q.

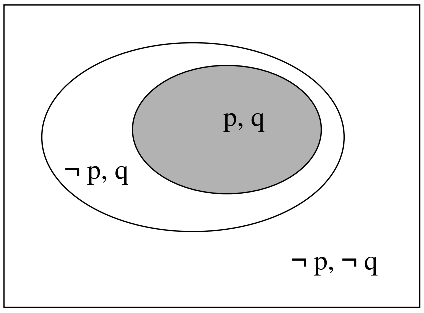

FIG. 2.1 – *Intentions et effets de bord*

Quelques philosophes (par exemple, Searle (1990)) proposent une notion d'*intention collective* qui doit correspondre à celle de la croyance mutuelle. Searle affirme qu'une intention collective (exprimée par une phrase comme «nous avons l'intention de faire x») ne se laisse pas exprimer comme la somme des intentions individuelles. Il affirme que ce n'est même pas le cas si on essaie d'exprimer une intention collective par les intentions individuelles *plus* des croyances individuelles et mutuelles. Ce qui manque est une expression de la volonté d'agir en *coopération*. Un groupe de gens courant vers une maison parce qu'il a commencé à pleuvoir va avoir la croyance mutuelle que chacun a l'intention d'y courir, mais cela ne donne pas l'impression d'une activité collective. Si, par contre, il s'agit d'un groupe de danseurs qui agissent ainsi en raison de leur danse commune, on peut parler d'une intention collective. Searle affirme donc que la *we-intention* est un phénomène primitif qui émerge d'un «sens de communauté» d'un agent. (Comme dans d'autres travaux, Searle semble sous-entendre que ce phénomène est réservé aux agents *biologiques*.)

2.3 Le dialogue en intelligence artificielle

Il s'est vite révélé très intéressant pour l'intelligence artificielle de suivre les travaux en philosophie de la langue présentés en section 2.1 et de considérer le langage en tant qu'action. Cela a permis aux chercheurs d'utiliser pour le traitement de la langue des méthodes développées pour raisonner sur des actions en général, notamment la planification.

2.3.1 Travaux basés sur la planification STRIPS

Traum (1994) mentionne Allen, Cohen et Perrault comme étant parmi les premiers à modéliser les actes de langage par des opérateurs de planification qui influencent les croyances du locuteur et de l'auditeur. Leur système de planification est une variante de STRIPS (Fikes et Nilsson, 1971) et utilise la même méthode pour résoudre le *frame problem* [4]: STRIPS suppose que seulement les changements mentionnés explicitement comme effets d'un opérateur vont

4. Comment peut-on prédire quels faits vont rester inchangés par l'exécution d'une action?

se produire. Cohen et Perrault utilisent deux types de préconditions: celles du type *cando* décrivent les faits que doivent être vrais pour pouvoir appliquer l'opérateur, les préconditions du type *want* décrivent les conditions de sincérité de Searle (1969) (cf. section 2.1.1). Partant d'un but du locuteur, le système de planification peut déterminer un acte de langage approprié.

Allen (1983) utilise le même formalisme pour un but complémentaire: il modélise la *reconnaissance* de plans à partir d'un énoncé. Un ensemble d'heuristiques permet de raisonner sur les intentions probables qui ont mené l'interlocuteur à produire un acte de langage. Ainsi, le système peut montrer un comportement collaboratif en donnant, par exemple, plus d'informations dans sa réponse que celles explicitement demandées. Allen montre aussi comment des actes de langage indirects peuvent être correctement interprétés.

2.3.2 Formalisations en logiques d'action

STRIPS s'est révélé limité dans sa capacité de décrire les actes de langage et de modéliser le raisonnement d'un agent. Des formalismes plus puissants ont été développés, souvent utilisant une logique générale d'action, la logique dynamique.

Cohen et Levesque (1990b) ont développé une logique dynamique multi-modale et l'ont utilisée pour montrer que les effets d'un acte illocutoire peuvent être décrits en termes d'une théorie générale d'interaction rationnelle et coopérative (Cohen et Levesque, 1990c). Ils définissent plusieurs caractéristiques (comme, par ex., la *sincérité*) en termes de cette logique. Ensuite, ils définissent ce qu'ils appellent la *disposition d'énonciation* (utterance mood), c'est-à-dire l'état mental du locuteur qui est communiqué à l'auditeur (si le locuteur est sincère). Un acte de langage est décrit comme une *tentative* (attempt) pour atteindre un certain effet. Cohen et Levesque affirment qu'il suffit à l'auditeur de reconnaître cette tentative du locuteur mais que la notion de *force illocutoire* peut être entièrement omise de leur modèle, et pour le locuteur et pour l'auditeur. Ils caractérisent plusieurs des catégories postulées par Searle (1969) en termes de leur logique. Dans Cohen et Levesque (1990a) ils décrivent le cas important des performatifs: les performatifs sont des actes où déjà la tentative fait réussir l'acte. Donc, la phrase «je t'effraie» n'est pas performative, mais «Je te demande de partir» est une demande dans tous les cas et donc performative. Cohen et Levesque prouvent que le verbe «mentir» ne peut pas être utilisée performativement.

Sadek (1991) poursuit une approche similaire, mais critique Cohen et Levesque (1990c) d'une part parce qu'ils n'indiquent pas comment leurs résultats pourraient être utile à un agent *pendant* le dialogue et d'autre part, parce que l'acte de langage est traité isolément et non en correspondance avec le plan poursuivi par son énonciation. Sadek définit une logique multimodale et décrit des actes de langage en termes de cette logique. Son idée principale, c'est que la logique est utilisée par les agents mêmes en planifiant leurs actions. Pour chaque acte, des *conditions de faisabilité* décrivent les conditions nécessaires et suffisantes à un agent pour énoncer l'acte, et les *effets perlocutoires* décrivent les effets qu'un agent veut atteindre en énoncant un tel acte. Ce sont ces effets qui guident la planification.

Perrault (1990) critique l'approche de Cohen et Levesque en argumentant que leur caractérisation des actes de langage ne tient pas compte du fait que les états mentaux du locuteur et de l'auditeur après un énoncé dépendent largement des états dans lesquelles ils étaient *avant*. Pourtant, il est difficile d'intégrer tous les états qui peuvent se révéler pertinents dans la définition des effets d'un acte de langage: Perrault donne des exemples pour les effets particuliers de mensonges, mensonges échoués ou ironie. Son approche ne consiste donc pas à intégrer toutes ces possibilités mais à décrire les effets «normaux» d'un acte en utilisant une logique de défauts. Perrault donne des règles de défauts disant que les anciennes croyances devraient être préservées, et que de nouvelles croyances communiquées par l'autre devraient être adoptées si elles ne contredisent pas les anciennes. Perrault admet que ce procédé permet d'adopter de *nouvelles* croyances, mais pas de *réviser* les anciennes. Il propose une variante de la règle pour la modification de croyances qui permet le changement d'une croyance si l'auditeur n'a pas de raisons de croire que le locuteur n'est pas sincère.

2.3.3 Approches basées sur les plans partagés

Pollack (1990) critique la notion de *plan* utilisée en intelligence artificielle comme inadéquate pour être utilisée dans une théorie de dialogue collaboratif. Un plan a longtemps été considéré comme une séquence d'actions pour atteindre un but, une *recette*. Pollack montre que les plans d'un agent sont plus que cela: il consistent en un ensemble de croyances sur la réalisabilité et les effets d'une recette et en des intentions associées. La reconnaissance du plan d'un agent doit donc prendre considération qu'il peut avoir d'autres croyances que l'observateur. Pollack définit un *plan simple* en termes de croyances et intentions. «Simple» signifie qu'une seule action est vraiment exécutée, les autres sont *générées* transitivement, c'est-à-dire *en* exécutant l'action a_1 l'agent réalise aussi l'action a_2 qui réalise l'action a_3, etc. Pour cette notion de plan bien définie (mais limitée) Pollack montre plusieurs types de «non-validité» d'un plan, selon les différentes croyances qui peuvent être fausses. Elle propose un modèle de reconnaissance de plans où l'observateur essaie de trouver des *explications* pour les croyances et intentions de l'autre agent en utilisant des heuristiques basées sur ses propres attitudes mentales.

Grosz et Sidner (1990) augmentent le modèle de Pollack (1990) pour décrire la *collaboration* entre plusieurs agents selon un **plan partagé**. Des croyances mutuelles remplacent donc les croyances de l'agent isolé. Grosz et Kraus (1996) ont précisé la définition initiale des plans partagés pour permettre, entre autres, de prendre en compte les critères pour un comportement intentionnel en groupe de Bratman (1992) et Searle (1990), mais ils montrent qu'une notion spécifique d'*intention collective* (cf. page 10) n'est pas nécessaire. La figure 2.2 (adaptée de (Grosz et Kraus, 1996)) présente les composantes essentielles de cette définition. Elles sont précisées et formalisées en utilisant les attitudes mentales Bel, MB, IntTo et IntTh introduites en section 2.2.

> Pour avoir un plan partagé pour une action, un groupe d'agents doit avoir
> **1.** la croyance mutuelle sur une recette à utiliser
> **2a.** des intentions individuelles qu'une action soit réalisée
> **2b.** des intentions individuelles que les collaborateurs réusissent en réalisant leurs sous-actions
> **3.** des plans individuelles ou collaboratifs pour les sous-actions

Fig. 2.2 – *Définition informelle d'un plan partagé*

Un plan partagé n'est pas réductible aux plans individuels des agents, mais il est décomposable hiérarchiquement et indique comment un agent peut *soutenir* des actions auxquelles il n'est pas directement lié. Grosz et Kraus introduisent la notion de *plan partagé partiel* (PSP) pour décrire le cas où un groupe a établi seulement un sous-ensemble des conditions pour un plan partagé complet (*full shared plan*, FSP). En particulier, un plan est partiel, si les agents ne se sont pas encore mis d'accord sur une recette complète pour réaliser l'action cible du plan. Grosz et Kraus (1999) décrivent comment un PSP peut être augmenté pour arriver à un FSP. Récemment, Hunsberger (1999) a simplifié les définitions des plans partagés et a pu ainsi démontrer des théorèmes importants sur les états mentaux des agents ayant ou établissant un plan partagé. Ces définitions permettront peut-être plus facilement d'intégrer une théorie formelle de l'établissement de croyances mutuelles, comme décrite dans le chapitre 5, dans la théorie des plans partagés.

L'utilisation des plans partagés dans le dialogue repose sur la théorie de la structure du discours de Grosz et Sidner (1986). Cette théorie distingue trois composantes différentes mais reliées: la structure linguistique (la séquence des énoncés), la structure intentionnelle (les buts des participants s'exprimant dans les segments du discours), et l'état attentionnel (le focus d'attention à un moment du discours). La structure linguistique décrit des aggrégations d'énoncés (appelées «segments du discours») selon des critères langagiers. La structure intentionnelle décrit le but poursuivi par un segment de discours (*discourse segment purpose*, DSP) et les relations entre les différents segments. Grosz et Sidner déterminent deux types de relation: la *domination* (DSP1 domine DSP2 si DSP1 sert à satisfaire une partie de DSP2) et la *précédence de satisfaction* (DSP1 s-précède DSP2 si DSP1 doit être satisfaite avant DSP2). L'état attentionnel est une structure dynamique qui comprend pour chaque segment du dialogue les objets, faits et relations pertinents. Ceci permet, par exemple, d'utiliser et de comprendre des expressions anaphoriques. Comme des agents peuvent s'engager dans des sous-dialogues et revenir plus tard à un thème de discours, l'état attentionnel est représenté par une *pile* où des *espaces de focus* peuvent être empilés et dépilés plus tard.

Grosz et Sidner (1986) argumentent que pour comprendre les énoncés d'un dialogue il faut reconnaître la structure intentionnelle de ce dialogue, c'est-à-dire les DSPs et leurs relations. En se basant sur les travaux de Grosz et Sidner (1990) et Grosz et Kraus (1996), Lochbaum (1998) identifie le DSP avec l'intention d'établir un *plan partagé complet*. Les relations entre les DSPs deviennent donc des relations entre les plans partagés correspondants. La décomposabilité des plans partagés correspond à la relation de dominance entre DSPs, alors que la précédence de satisfaction est représentée par un plan partagé étant la précondition pour la complétion d'un autre. L'évolution d'un plan partagé partiel (PSP) à un plan partagé complet (FSP) pendant le dialogue est modélisé par *l'algorithme d'augmentation* des plans partagés. Un auditeur va essayer d'interpréter un énoncé dans le contexte du PSP qu'il a déjà établi avec le locuteur. Un résumé de ce processus est donné en figure 2.3 (provenant de Lochbaum (1998, p.538)). Lochbaum utilise une structure particulière, le *graphe des recettes* ou Rgraph, pour modéliser les croyances d'un agent à un moment donné sur les relations entre les actions sous-jacentes au dialogue. Le Rgraph constitue un arbre qui comprend les actions déjà exécutées ainsi que les actions que l'agent modélisé considère comme étant mutuellement reconnues nécessaires pour atteindre le but commun aux agents. Le Rgraph évolue au cours du dialogue et indique à l'agent ce qui manque pour atteindre le but commun.

Un dialogue ne comprend pas seulement *l'établissement* d'un plan partagé, mais aussi la discussion des résultats des actions qui sont *exécutées* pendant la planification. Souvent, la planification est contrainte par la réussite ou l'échec d'une action, ce qui nécessite l'alternance des processus de planification et d'exécution. Balkanski et Hurault-Plantet (1999) fournissent un modèle capable non seulement de déterminer le moment approprié pour interrompre la

Soit donné:
$PSP(G_1, G_2, \alpha)$,
G_1 est l'agent modelisé, α l'action cible du plan partagé partiel
Soit $Prop$ la proposition communiqué par l'énocé U de G_2.

1. La communication a comme résultat que G_1 suppose que $MB(G_1, G_2, Bel(G_2, Prop))$
2. G_1 doit déterminer la relation de $Prop$ au contexte actuel du plan partagé:
   (a) Si G_1 croit que U ou $Prop$ indique l'initiation d'un plan partagé subsidiaire pour un acte β, alors G_1 va
      i. attribuer à G_2 $Int.Th(G_2, (FSP(G_1, G_2, \beta)))$
      ii. déterminer si lui aussi veut adopter une telle intention
   (b) Si G_1 croit que U ou $Prop$ indique le complément du plan partagé actuel, alors G_1 va
      i. attribuer à G_2 $Bel(G_2, (FSP(G_1, G_2, \alpha)))$
      ii. déterminer si lui aussi croit que le plan partagé actuel des agents est complet
   (c) Sinon G_1 va
      i. attribuer à G_2 une croyance que $Prop$ est pertinent pour le plan partagé actuel des agents
      ii. déterminer si lui aussi croit que c'est le cas
3. (a) Si (2) réussit, alors G_1 va signaler son accord (peut-être implicitement) et supposer MB sur la relation inférée en (2a), (2b) ou (2c), mettant à jour son regard du PSP
   (b) Sinon, G_1 va questionner G_2 ou communiquer son désaccord

FIG. 2.3 – *Augmentation d'un plan partagé selon Lochbaum (1998)*

planification par l'exécution d'une action ou la réalisation d'un acte de langage, mais aussi de replanifier en cas d'échec. Une implémentation du modèle de dialogue de Balkanski et Hurault-Plantet (1999) sera présenté dans le chapitre 7.

Chapitre 3

Établissement de croyances mutuelles: travaux existants

Si une croyance mutuelle comprend une conjonction infinie de croyances emboîtées, le problème (nommé **paradoxe des croyances mutuelles** par Clark et Marshall (1981)) se pose d'expliquer comment elle peut jamais être atteinte ou reconnue en un temps fini. Il y a peu de temps, la plupart des travaux en intelligence artificielle ignoraient simplement le problème. Dans ce chapitre, nous allons présenter les travaux qui traitent explicitement la question de l'établissement de croyances mutuelles.

3.1 Clark et Marshall: la coprésence

Clark et Marshall (1981) proposent deux heuristiques qui pourraient expliquer l'établissement d'une croyance mutuelle (MB) dans un temps fini. L'**heuristique de troncature** consiste à vérifier une partie seulement des croyances emboîtées et d'en inférer MB si elles sont trouvées vraies jusqu'à un niveau suffisamment profond. Mise à part la question où devrait se trouver ce niveau, les auteurs considèrent l'heuristique comme peu vraisemblable cognitivement, car des études empiriques ont montré que des humains (surtout des enfants) ont de grandes difficultés à traiter des croyances emboîtées, tout en étant capables d'utiliser des références anaphoriques et d'autres phénomènes considérés comme nécessitant des croyances mutuelles (cf. 2.2.2).

Clark et Marshall favorisent donc l'**heuristique de coprésence** qui repose sur l'approche des *situations partagées* discutée en section 2.2.2: certaines circonstances peuvent indiquer à tous les agents non seulement la véracité d'une proposition p mais aussi que tous les agents dans cette situation savent que p est vraie – et cela permet une sorte de preuve par induction qui vérifie toute la chaine simultanément. Les circonstances nécessaires sont décrites par Clark et Marshall comme la «triple coprésence» de l'agent A, l'agent B et l'objet de référence. Si, par exemple, A et B regardent une bougie et se voient mutuellement la regarder, cette situation va directement établir les croyances mutuelles qu'il y a une bougie, qu'ils la regardent, etc.

La coprésence n'est pas nécessairement physique, elle peut aussi être linguistique (référence à quelque chose dont on a déjà parlé). Une autre base pour une croyance mutuelle peut être l'appartenance à une même communauté où une certaine connaissance peut être présupposée. Reste cependant la question pour les agents de déterminer s'ils sont effectivement membres de

Base pour croyance mutuelle		Suppositions supplémentaires
1.	appartenance à une communauté	appartenance à la même communauté, connaissance universelle
2.	coprésence physique	
	a. immédiate	simultanéité, attention, rationnalité
	b. potentielle	simultanéité, attention, rationnalité, localisabilité
	c. antérieur	simultanéité, attention, rationnalité, possibilité d'être rappelé
3.	coprésence linguistique	
	b. potentielle	simultanéité, attention, rationnalité, localisabilité, compréhensibilité
	c. antérieur	simultanéité, attention, rationnalité, possibilité d'être rappelé, compréhensibilité
4.	coprésence indirecte	
	b. physique	simultanéité, attention, rationnalité (localisabilité ou possibilité d'être rappelé), associabilité
	c. linguistique	simultanéité, attention, rationnalité, (localisabilité ou possibilité d'être rappelé), associabilité, compréhensibilité

TAB. 3.1 – *Types de coprésence selon Clark et Marshall (1981)*

ce groupe (c'est-à-dire établir la croyance mutuelles qu'ils le sont). Clark et Marshall proposent que cela se passe en utilisant la coprésence physique ou linguistique ou la *coprésence indirecte* qui est un mélange de communauté commune avec coprésence physique ou linguistique.

Chacune de ces possibilités de coprésence nécessite des suppositions supplémentaires sur la situation et les agents (le tableau de Clark et Marshall (1981, p.43) qui les résume est reproduit dans le tableau 3.1). Par exemple, un agent peut croire qu'il y a une croyance mutuelle sur un fait, s'il suppose que tous les agents sont membres d'une communauté et s'il suppose que ce fait est connu universellement de tous les membres de cette communauté. Pour la coprésence physique ce sont la simultanéité, l'attention et la rationnalité qui doivent être supposés, c'est-à-dire que les agents doivent se trouver en même temps dans la même situation, ils doivent être attentifs et ils sont capables d'inférer rationnellement. On peut distinguer trois formes particulières de coprésence physique selon le moment où on se refère à la situation commune: si la référence se fait *dans* la situation même, les trois conditions de simultanéité, attention et rationnalité sont suffisantes. S'il s'agit d'un rappel, les agents doivent mutuellement se supposer capables d'un tel rappel. Et si c'est une situation qui n'a pas encore eu lieu, mais le pourrait facilement, on a besoin de «localisabilité», c'est-à-dire que la simple référence mènera à une situation de coprésence immédiate (par ex. A parle de «la bougie». B va regarder devant lui et ainsi établir la coprésence physique).

La coprésence linguistique (référence à un objet dont on a parlé auparavant ou qui n'a pas encore été mentionné) nécessite en plus une supposition nommée «compréhensibilité» de l'énoncé ce qui signifie que le destinataire doit pouvoir comprendre que l'énoncé se réfère à un objet. La coprésence indirecte combine les aspects de l'appartenance à une même communauté, de coprésence linguistique et de coprésence physique. Par exemple, la phrase «J'ai une nouvelle voiture, mais le moteur est déjà en panne» utilise non seulement la coprésence linguistique, mais aussi le savoir qu'une voiture a un moteur. Cette phrase nécessite donc la croyance mutuelle d'appartenir à une communauté où ce fait est mutuellement cru. La supposition nécessaire de l'«associabilité» signifie que l'autre doit pouvoir associer les faits avec ses connaissances en tant que membre de la communauté.

Traum (1994) critique l'approche de Clark et Marshall en montrant que les suppositions qu'ils donnent ne sont pas suffisantes pour la coprésence linguistique potentielle: la «localisabilité» va seulement mener à l'identification du bon objet, s'il n'y a pas déjà un référent commun établi et si donc l'auditeur ne voit pas la nécessité d'en localiser un autre. Clark et Marshall reconnaissent qu'une référence peut échouer, mais ils montrent aussi qu'elle peut être réparée. Ils constatent effectivement que cela se passe dans toutes les conversations: soit le locuteur se corrige lui-même pour préciser sa référence, soit l'interlocuteur demande un éclaircissement. Ils proposent deux possibilités pour réparer une référence: la réparation horizontale essaie de donner plus d'informations sur le référent mais utilise le même type de coprésence, alors que la réparation verticale essaie de compenser l'échec en changeant le type de coprésence (par exemple de pointerun objet pour changer d'une coprésence physique potentielle à une coprésence physique immédiate).

Traum remarque que, en raison de possibilités d'erreurs (et de réparations), les suppositions ne peuvent jamais être suffisantes pour établir une vraie croyance mutuelle: quand A fait une référence, il ne peut pas être sûr d'être compris par B. Et c'est pourquoi B, même s'il croit avoir compris la référence, ne peut pas être sûr que A croit que B a compris. Chaque croyance emboîtée supplémentaire va ajouter de l'incertitude, si bien qu'une vraie croyance mutuelle dont les agents sont entièrement convaincus ne peut pas être établie.

3.2 La théorie des contributions de Clark et Schaefer

Clark et Schaefer (1989) constatent que pour être effective, une conversation nécessite une base commune (**common ground**) des participants. En même temps, la conversation *augmente* cette base commune dans la mesure où le contenu des énoncés des participants devient mutuellement connu. Le processus d'augmenter cette base commune est nommé «**grounding**». Contrairement à l'ancien modèle de Clark et Marshall (1981), Clark et Schaefer (1989) ne supposent plus que le «grounding» se fait automatiquement au moyen de la coprésence linguistique, mais affirment qu'il s'agit d'un processus auquel participent le locuteur et les interlocuteurs. L'étude empirique d'un corpus de dialogues téléphoniques (Clark et Schaefer, 1987) leur montre que ce processus se produit dans toute conversation et qu'il nécessite la collaboration des agents pour augmenter la base commune.

Clark et Schaefer (1989) distinguent deux phases qui, ensemble, forment une «unité de conversation» qu'ils appellent une **contribution**. D'abord le locuteur («contributeur») spécifie le contenu de son énoncé et les auditeurs («partenaires») essaient de l'enregistrer. Ensuite, le contributeur et les partenaires essaient d'atteindre le **critère de grounding** (Clark et Schaefer, 1989, p. 262):

«Le contributeur et les partenaires croient mutuellement que les partenaires ont compris ce que le contributeur voulait exprimer à un degré suffisant pour l'objectif courant.»

Pour le cas de deux agents A et B les auteurs distinguent les deux phases suivantes d'une

1.	attention continue:	B montre qu'il continue d'être attentif et donc reste satisfait de la présentation de A
2.	initiation de la contribution suivante pertinente:	B commence la contribution suivante qui serait pertinente au même niveau que la contribution actuelle
3.	«acknowledgement»:	B fait un signe de tête (nodding) ou dit «mhm», «oui», etc.
4.	démonstration:	B montre tout ou des parties de ce qu'il a compris
5.	répétition:	B répète littéralement tout ou des parties de la présentation de A

TAB. 3.2 – *Types d'évidences de Clark et Schaefer (1989)*

contribution:

phase de présentation: A présente l'énoncé u à B en supposant que, si B fournit l'évidence e ou plus forte, ceci montre à A que B a compris ce que A a voulu dire.

phase d'acceptation: B accepte l'énoncé u en donnant l'évidence e' qu'il croit qu'il comprend ce que A veut dire par u. Il agit ainsi en supposant que A, lorsqu'il enregistrera l'évidence e, va aussi croire que B comprend.

Clark et Schaefer affirment: «Si ces deux phases sont exécutées correctement, A et B vont chacun croire qu'ils ont atteint la croyance mutuelle que B comprend ce que A veut dire par son acte» (Clark et Schaefer, 1989, p. 265f.). Il est très important de remarquer que cette notion n'exclut pas les malentendus: les deux agents peuvent tous les deux croire que B a compris tout en croyant des choses différentes sur ce qu'a dit A, si, par exemple, l'évidence e' de B est ambigüe.

La réaction de B, quelle qu'elle soit, est une nouvelle présentation et doit être à son tour acceptée. On se voit donc confronté au danger d'un processus qui entraine toujours de nouvelles acceptations. Clark et Schaefer proposent un **principe de force de l'évidence**: «Les participants supposent que, si l'évidence e_0 est nécessaire pour accepter la présentation u_0 et e_1 est nécessaire pour accepter e_0, alors e_1 va être plus faible que e_0.» L'introduction d'évidences plus ou moins faibles, nécessite une hiérarchie de «force» des évidences. La hiérarchie proposée par Clark et Schaefer (1989) est reproduite dans le tableau 3.2 (de l'ordre du plus faible au plus fort).

Cet ordre nous semble pourtant assez arbitraire: par exemple, une démonstration peut souvent être plus pertinente qu'une répétition littérale dans la mesure où elle montre mieux la compréhension du *contenu* de l'énoncé – pensons au perroquet ... Mais il ne suffit pas de changer la hiérarchie des forces: l'idée générale d'une telle hiérarchie nous semble redoutable: la «force» d'une évidence dépend moins de la manière dont elle est donnée, que de sa pertinence vis à vis du contexte. Par exemple, l'initiation d'une nouvelle contribution de B est une évidence de compréhension très forte lorsqu'elle n'est pas expliquable sans supposer que B a compris. Mais dans d'autres situations il se peut que la nouvelle contribution soit aussi pertinente sans la première contribution, et dans ce cas-là elle ne peut pas servir d'évidence de compréhension.

Clark et Schaefer affirment que chaque énoncé (y compris les acceptations) sont des présentations, mais ils ne disent pas clairement si cela implique que l'acceptation doit être acceptée elle-même pour compléter la contribution. Bien sûr, dans ce cas-là on doit se demander si cela n'entraine pas une récursion infinie qui fait que le «grounding» ne peut jamais être atteint. D'autre part, si ce n'était pas le cas, il faudrait deux types de présentation: celles qui doivent être acceptées et celles qui n'en ont pas besoin. Clark et Schaefer ne donnent pas d'explication concernant ces deux possibilités.

Un autre concept de la théorie a des conséquences importantes pour son applicabilité au dialogue homme-machine: selon Clark et Schaefer, les deux phases peuvent chacune comprendre de multiples énoncés de la part des deux parties. Réparations du locuteur lui-même et compléments apportés par l'auditeur sont considérées comme faisant partie de la phase de présentation, alors que demandes de confirmation, réparations faites par l'auditeur et «acknowledgements» («mhm») font partie de la phase d'acceptation. Ceci pose le problème de déterminer à quelle phase appartient un énoncé. Une réponse de B, est-elle un complément (présentation) ou une réparation (acceptation)? Très souvent, cette attribution n'est possible qu'en regardant de larges segments du dialogue avant et après l'énoncé. Il s'agit donc d'un modèle purement descriptif qui ne peut expliquer le dialogue qu'*a posteriori*. Traum (1994) a appelé cela la nature «off-line» des phases et a critiqué qu'il n'y ait pas de moyen facile de déterminer *l'état* de la contribution actuelle. Ce que nous cherchons est un modèle *cognitif* qui aide un agent à déterminer ses prochaines actions en fonction de ce qui s'est passé avant.

3.3 Traum: une théorie computationnelle de «grounding»

Le modèle de Traum (1994) est la première théorie qui ne néglige pas la nécessité du «grounding» et qui n'adopte pas de position extérieure pour l'étudier. Elle a été la référence principale pour notre travail et sera donc présentée en détail ici. Néanmoins, nous allons adopter une position profondément différente dans les chapitres qui suivent. Nos critiques du modèle de Traum seront présentées dans la section 3.4.

3.3.1 La grammaire du «grounding»

Reprochant à Clark et Schaefer (1989) de délimiter trop vaguement les phases de présentation et d'acceptation, Traum (1994) remplace la notion de *contribution* constituée de deux phases éventuellement récursives par celle de l'**unité de discours** (**discourse unit**, **DU**) qui a la même fonction que la contribution mais qui est composée d'actes individuels distincts, les **actes de grounding**. Ces actes constituent une de plusieures classes d'**actes de conversation** introduits par Traum et Hinkelman (1992). Selon cette théorie, des différents type d'actions sont réalisées à tout moment d'une conversation: les locuteurs changent (*actes de «prise de parole»*), ils essaient d'atteindre le «critère de grounding» (*actes de grounding*), réalisent des actes illocutoires (*actes de langage de base*) et agissent selon des structures supérieures conventionnelles (*actes d'argumentation*). Les actes des différents niveaux sont réalisés simultanément sans que les actes des niveaux plus hauts soient décomposables en actes de niveau inférieur.

Traum (1994, p.59) décrit les *actes de grounding* suivants:

Initiate: l'énoncé initial d'une unité de discours.

État	Acte d'entrée	Acte de sortie préféré
S	———	Initiate^{I}
1	Initiate^{I}	Ack^{R}
2	ReqRepair^{R}	Repair^{I}
3	Repair^{R}	Ack^{I}
4	ReqRepair^{I}	Repair^{R}
F	$\text{Ack}^{\{I,R\}}$	$\text{Initiate}^{\{I,R\}}$ (DU prochaine)
D	$\text{Cancel}^{\{I,R\}}$	$\text{Initiate}^{\{I,R\}}$ (DU prochaine)

TAB. 3.3 – *États d'une unité de discours (Traum, 1994)*

Continue: poursuite d'un acte précédent par le même locuteur.

Ack: «acknowledgement» [1] = signale la compréhension.

Repair: corrige les informations déjà énoncées ou ajoute de nouvelles informations.

ReqRepair: demande un acte de réparation.

ReqAck: essaie de faire que l'autre agent fournisse un «acknowledgement».

Cancel: ferme l'unité de discours (DU). Si l'intention communicative est encore actuelle, une nouvelle DU doit être ouverte.

Traum (1994) présente un automate à états finis dans lequel les *actes de grounding* représentent les transitions. Il définit ainsi une grammaire pour des séquences «bien formées» d'actes de grounding. Il affirme que toute séquence d'actes qui mène à un état final F va fournir le «grounding» de la DU. Les figures 3.1 à 3.3 présentent la fonction de transition de cet automate. Chaque figure montre pour un des états les transitions partant et arrivant à cet état. L'état S correspond à une DU qui n'a pas encore été commencée. L'état F représente une DU qui a été «grounded», alors que D correspond à une DU abandonnée qui ne peut plus être «grounded». Les états 1 à 4 représentent les étapes intermédiaires dans le processus: état 1 directement après la présentation, état 2 lorsque l'interlocuteur a demandé une réparation, état 3 lorsque l'interlocuteur lui-même a fourni un acte de réparation. L´état 4 correspond à une situation où le locuteur a eu des problèmes avec la réaction de l'autre et a demandé une réparation. Le tableau 3.3 montre les caractéristiques des états en donnant les actes d'entrées ainsi que les actes préférentiels de sortie, donc ceux qui vont dans la «direction du grounding».

3.3.2 Modèle cognitif de la conversation

Afin qu'un agent puisse utiliser les connaissances pragmatiques exprimées dans l'automate, Traum a besoin d'un modèle *cognitif* correspondant, c'est-à-dire un modèle qui décrit comment changent les croyances de l'agent lorsqu'il produit ou reconnait des *actes de grounding*. La figure 3.4 montre ce modèle (repris de (Traum, 1994, p. 44)) du point de vue d'un agent X dans une conversation avec un autre agent Y.

1. Traum distingue précisément entre *acknowledgement* et *acceptation*. Il utilise le terme *acknowledgement* pour un signal de compréhension et *acceptation* pour un signal d'accord avec une proposition de l'autre. *Accept* est donc un *acte de langage de base* et implique la compréhension. *Ack* est un *acte de grounding* qui ne signifie rien pour l'acceptation. Exemple: la réponse «Non!» fournit un «acknowledgement», mais pas une acceptation. Nous adopterons cette distinction.

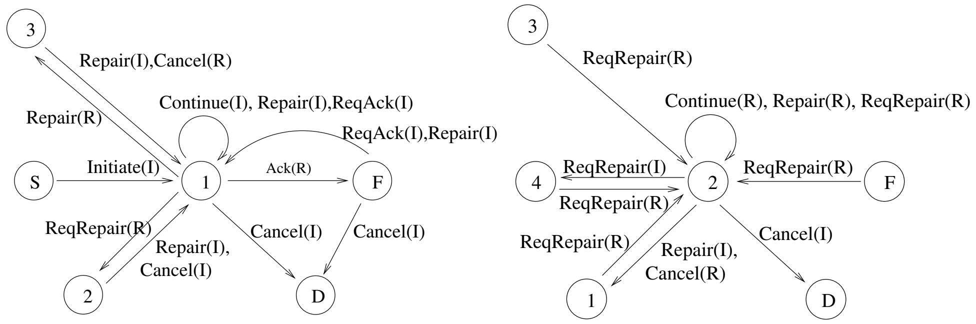

FIG. 3.1 – *Transitions partant de et arrivant à l'état 1 (gauche) et l'état 2 (droite)*

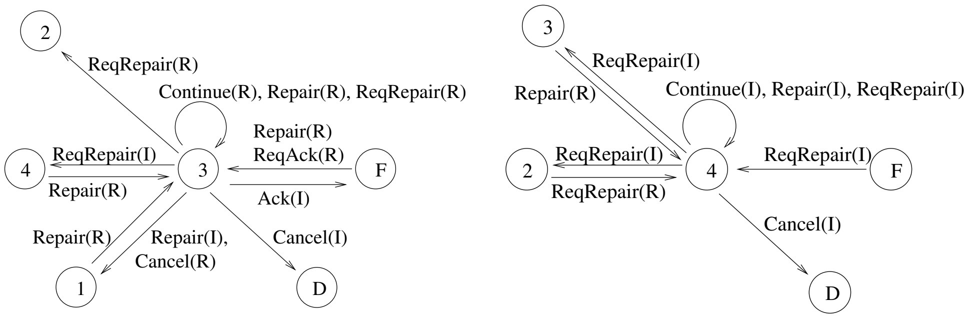

FIG. 3.2 – *Transitions partant de et arrivant à l'état 3 (gauche) et l'état 4 (droite)*

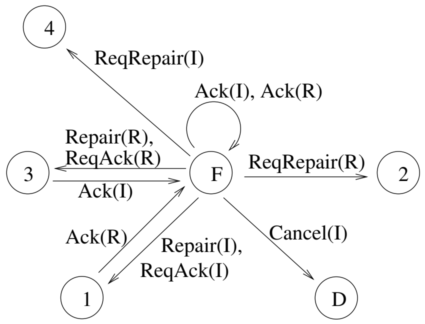

FIG. 3.3 – *Transitions partant de et arrivant à l'état F*

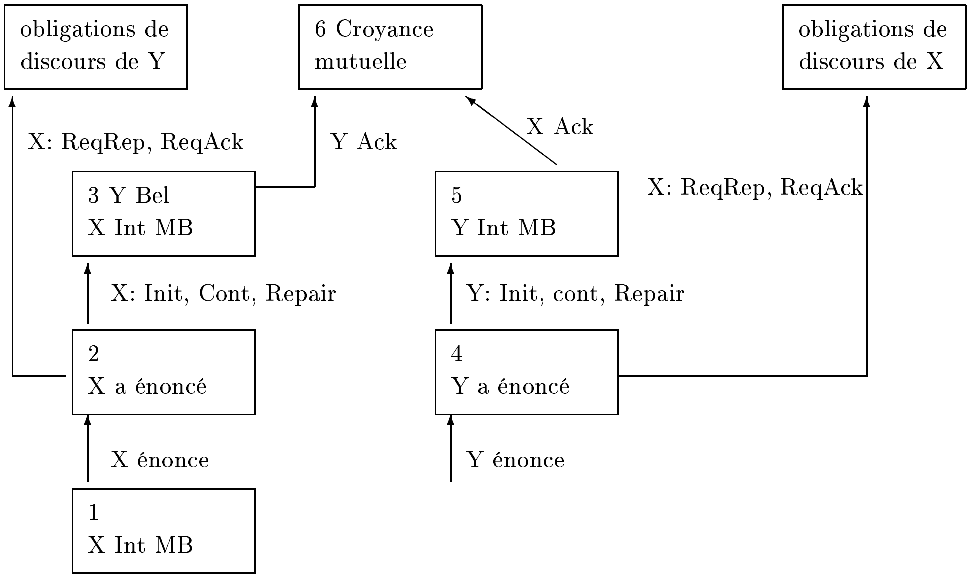

FIG. 3.4 – *Modèle cognitif de «grounding» du point de vue de X*

Examinons d'abord le cas de X entamant une nouvelle unité de discours: avant d'être énoncé, le contenu va se trouver dans la case 1. Lorsque X produit l'énoncé, le contenu de celui-ci va aller dans la case 2. Comme le module de génération est faillible, X va ensuite comparer le contenu des cases 1 et 2. S'ils sont différents, cela peut mener à un acte de réparation. X essaie d'inférer comment Y va interpréter son énoncé en utilisant (les croyances de X sur) les croyances, des méthodes de reconnaissance de plan, etc. Le résultat, ce que X croit que Y pense des intentions de X, est placé dans la case 3. De nouveau, si les contenus des cases 1 et 3 sont différents, X va produire d'autres énoncés pour clarifier la situation.

Un énoncé de la part de Y arrive toujours dans la case 4. S'il a un problème d'interprétation, X peut produire une demande de réparation où même une réparation. Si l'interprétation réussit, la suite dépend de l'acte que Y a réalisé: s'il s'agit d'un «acknowledgement» le contenu de la case sera transferé dans la case 6. Si l'énoncé est un acte d'Initiate, Continue ou Repair, alors X va essayer de reconnaître le plan et inférer les intentions de Y. Les résultats seront mis dans la case 5 d'où X peut fournir le «grounding» en donnant un «acknowledgement». Dans les cas de ReqRepair et ReqAck, X doit réconsidérer le contenu des cases 3 et 5 (en considérant les nouvelles informations données par l'énoncé de Y) et ajouter de nouvelles obligations de discours (fournir la réparation ou «l'acknowledgement» demandé, etc.).

3.3.3 États mentaux et conversationnels

Un agent capable de s'engager dans un dialogue a besoin des représentations internes du déroulement de la conversation ainsi que de ses interlocuteurs (*conversation model* et *user model*). Pour représenter les croyances de l'agent, Traum (1994) utilise plusieurs modalités distinctes mais liées qu'il appelle *contextes*. La structure de ces contextes est reproduite dans la figure 3.5 (l'agent est appelé *système*). Traum distingue entre les croyances actuelles sur le monde et les propositions faites par les agents. Il argumente qu'un fait peut être affirmé

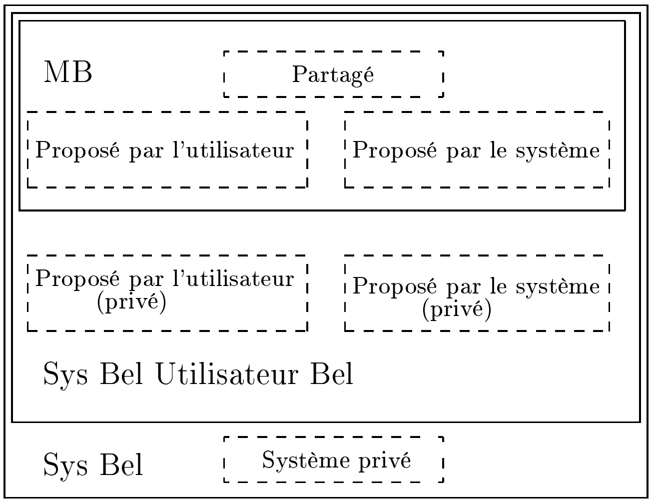

FIG. 3.5 – *Contextes de croyances et propositions de Traum (1994)*

(proposé) par un agent, mais que l'autre agent doit réaliser un acte (mental) supplémentaire d'acceptation pour en faire une croyance. Adoptant la perspective subjective du système, Traum (1994) représente toutes les croyances comme des emboîtements du contexte *Sys Bel*, donc des croyances du système. Quelques unes d'entre elles seront des croyances sur l'utilisateur et quelques unes de celles-ci des croyances mutuelles[2].

Ces contextes sont reliés au modèle cognitif de «grounding» (fig. 3.4). Considérant le système comme étant l'agent X, la modalité *système privé* correspond à la case 1, la modalité *Proposé par le système (privé)* à la case 3 et *Proposé par l'utilisateur (privé)* sera la case 5. La modalité *MB* correspond à la case 6, mais permet de distinguer encore plus finement: un énoncé peut être «grounded» sans que son contenu soit nécessairement accepté par les deux agents. La transition d'une proposition au niveau *MB* à la modalité *Partagé* peut se faire par des *actes de langage de base (core speech acts)*, par exemple par l'acte *Accept*.

3.4 Critiques du modèle de Traum

Pour illustrer nos critiques, nous allons considérer le dialogue représenté dans la figure 3.6 (adapté de Balkanski et Hurault-Plantet (1999)). Il semble clair que l'énoncé (1) donne une information qui doit être «grounded». (2) signifie que le standard n'est pas sûr d'avoir bien compris et pour cela demande une confirmation. (3a) fournit cette confirmation et l'utilisateur commence à donner une nouvelle information en (3b). Apparemment, le processus du «grounding» est fini après (3a). Les énoncés (1) à (3a) devraient donc pouvoir être expliqués comme une séquence d'*actes de grounding* partant de l'état S (figure 3.1, à gauche) et arrivant à l'état F (figure 3.3) qui signifie le «grounding». La seule possibilité de transition à partir de S est de produire un acte *Initiate(I)* qui mène à l'état 1. L'utilisateur joue alors le rôle

2. Il aurait été probablement plus clair d'appeler cette modalité *Sys Bel MB*. Pour Traum ce point n'est pas important parce qu'il ne traite pas des *fausses* croyances sur les croyances mutuelles d'un agent, mais cf. la section 6.1.1 dans le présent travail

(1)	Utilisateur:	Je voudrais parler à Corinne Savol.
(2)	Standard:	Corinne Savol?
(3a)	Utilisateur:	Oui.
(3b)		Elle travaille avec Louise Abou.
(4)	Standard:	Je vous passe Louise Abou.

FIG. 3.6 – *Exemple d'une demande de confirmation*

de l'initiateur (I). Pour que l'unité de discours puisse se trouver dans l'état F après (3a), cet énoncé doit être considéré comme un acte *Ack(I)*, parce que la figure 3.3 montre que le seul acte de l'initiateur qui mène à l'état F est un «acknowledgement». Cet «acknowledgement» doit être fait à partir de l'état 3[3]. Nous savons donc maintenant que l'énoncé (2) doit fournir une transition entre l'état 1 et l'état 3. La figure 3.1 montre que le seul acte qui fournit cette transition est *Repair(R)*. Ce résultat est peu convaincant: il n'est pas clair dans quelle mesure l'énoncé (2) «répare» l'énoncé (1). Mais si l'on attribue à (2) un acte plus plausible comme *ReqRepair(R)*, le «grounding» ne peut pas être atteint avec (3a).

Probablement, il faudrait introduire un nouvel *acte de grounding* pour expliquer l'énoncé (2) (et l'appeler *ReqConf(R)*, par exemple). Mais il n'est pas clair comment un tel acte devrait être intégré dans la classe des *actes de grounding* et dans l'automate. Ceci révèle un problème plus profond: les *actes de grounding* sont surtout caractérisé par leurs rôles dans l'automate, c'est-à-dire par leurs *fonctions* pour le processus de «grounding». Pourtant, un agent ne perçoit pas la fonction de l'acte, mais son contenu. L'automate donne quelques indications pour déterminer l'acte sous-jacent à un énoncé en contraignant les possibilités selon l'état actuel de l'automate. Mais dans chaque état (sauf S et D) *plusieurs* différents actes sont théoriquement possibles (sept ou plus pour chacun des états 1 à 4 et F). Comme nous avons vu en analysant l'exemple, il n'est même pas certain que l'ensemble des *actes de grounding* est complet. Il nous semble donc que l'automate ne facilite que peu l'interprétation d'un énoncé.

Traum (1994, p.74) donne quelques indications sur la manière dont un agent peut reconnaître un *acte de grounding*: par exemple, un acte de *ReqRepair* est dit se distinguer d'autres actes par «l'intonation de questionnement». Ceci pose un problème parce que ce type d'intonation pourrait aussi signaler un nouvel *acte de langage de base* comme *Request, Check, Suggest,* etc. Selon Traum ceci signalerait le début d'une nouvelle unité de discours et signifierait plutôt un «acknowledgement» implicite. Les autres indications pour la reconnaissance des *actes de grounding* présentent des ambiguïtés similaires qui ne peuvent être résolues qu'en tenant compte du *contexte* et des informations que communique un énoncé. Cela implique qu'un *acte de grounding* doit décrire le changement de croyances selon un certain contexte, mais pour déterminer cet acte il faut aussi avoir recours à ce contexte. Pourquoi ne pas simplement décrire les effets d'un énoncé sur les croyances d'un agent directement en fonction du contexte sans ayant recours aux *actes de grounding*? Cet argument devient encore plus pertinent quand nous considérons que la classe des *actes de grounding* décrite par Traum (1994) n'est apparamment pas complète. Nous croyons que le rôle d'un énoncé pour le «grounding»

3. ou à partir de l'état F, ce qui voudrait dire que le «grounding» serait déjà établi avec l'énoncé (2). Nous considérons qu'il est peut probable que l'énoncé (2) établit le «grounding», parce que le standard n'est pas certain d'avoir bien compris. Pendant qu'il considère encore la possibilté que l'utilisateur a dit d'autre chose, il ne va pas croire qu'une croyance mutuelle a été atteinte.

dépend largement du contexte (notamment des croyances d'un agent) et qu'il n'est possible de lui attribuer ce rôle qu'en fonction du contexte. Mais dans ce cas-là, il n'est pas nécessaire d'introduire toute une nouvelle classe d'actes de langage, mais il suffit de décrire les relations entre les informations communiquées par un énoncé et les croyances antérieures des agents.

En analysant le modèle de contributions de Clark et Schaefer (1989), Traum (1994, p.29) note qu'il n'était pas «clair si une contribution est jamais vraiment complète», parce que toute acceptation est une présentation qui doit elle-même être acceptée. Dans le modèle de Traum, un «acknowledgement» n'a pas besoin d'être accepté. Pourtant, Traum ne fournit pas de justification pour cette décision. Il n'explique pas pourquoi un «acknowledgement» peut entraîner une croyance mutuelle dans certains cas (états 1 et 3) et non dans d'autres (état 4). Nous considérons une telle justification comme indispensable pour un modèle formel du «grounding» et ce sera un des buts principaux des chapitres 4, 5 et 6 de décrire comment une séquence approprié d'énoncés peut établir une croyance mutuelle (ou au moins une croyance mutuelle unilatérale).

La théorie de «grounding» que nous développons dans les chapitres suivants se distingue fondamentalement de celle de Traum (1994) en ce qui concerne l'esprit philosophique sous-jacent: alors que Traum se base sur une *classification* d'actes de langage à plusieurs niveaux (dont un qui correspond aux actes de langage «classiques» de Searle (1969)), nous ne distinguons les énoncés qu'en fonction de l'intention communicative sous-jacente (Grice, 1957). Alors que Traum attribue à chaque énoncé un *acte de grounding* (et d'autres actes à d'autres niveaux), nous considérons qu'un énoncé n'a qu'une seul *signification*, mais qu'il peut produire des effets différents selon le contexte dans lequel il est énoncé.

Ces différences dans l'approche apparaissent dans chacun des chapitre suivants. Le chapitre 4 propose une théorie des changements d'états mentaux des participants à un dialogue. Ensuite, le chapitre 5 fournit une formalisation de cette théorie en utilisant une logique multi-modale. Le chapitre 6 explique comment la théorie peut être utilisée dans la construction d'un agent artificiel pour le dialogue homme-machine. Finalement, le chapitre 7 présente le système développé dans le cadre de ce travail.

Chapitre 4

Vers une nouvelle théorie de «grounding»

Puisque nous avons critiqué Traum (1994) pour ne pas avoir expliqué pourquoi certains énoncés (les «acknowledgements») peuvent mener à l'établissement d'une croyance mutuelle, nous devons essayer d'expliquer le processus de «grounding» nous-même. Ce chapitre présente donc nos idées concernant l'établissement de croyances mutuelles d'un point de vue psycholinguistique. Le problème que nous essayons d'élucider est le suivant: si un agent ne peut pas être certain que son énoncé sera compris (il n'y a donc pas de situation de coprésence comme chez Clark et Marshall (1981)), mais si, par contre, un processus de «grounding» est nécessaire dont la durée n'est pas déterminable à l'avance (comme dans la théorie de Clark et Schaefer (1989)) – comment un agent peut-il jamais être certain que le critère de «grounding» a été atteint? Une idée clé de notre approche est que tout participant d'un dialogue *sait* que la compréhension n'est pas un effet automatique et donc que le locuteur ne peut pas être certain d'être compris. À partir de cette *certitude de l'incertitude*, nous développons une notion d'*attente* qui décrira l'état du locuteur après avoir produit un énoncé. Les deux agents vont savoir mutuellement que le locuteur attend un signal de compréhension, un «acknowledgement» (qui peut être implicite) et c'est cette croyance mutuelle qui va permettre d'atteindre le «grounding», c'est-à-dire la croyance mutuelle que l'auditeur avait compris ce que le locuteur voulait dire.

4.1 Compréhension et critère de «grounding»

Suivant Grice (1957) nous considérons la signification (non-naturelle) d'un énoncé comme *ce que voulait exprimer* le locuteur, c'est-à-dire son *intention communicative* (cf. section 2.1.2). La signification d'un énoncé n'est donc pas toujours la même; elle dépend largement du contexte, notamment de ce que pense le locuteur *sur* le contexte et comment il pense pouvoir exprimer ses intentions dans ce contexte. Quand un locuteur produit un énoncé, l'auditeur va essayer d'attribuer à cet énoncé une intention communicative du locuteur; l'auditeur va essayer de *comprendre* l'énoncé. Un critère «absolu» de compréhension serait que l'auditeur attribue à un énoncé l'intention communicative qu'a voulu exprimer le locuteur. Mais pour détecter une telle «compréhension absolue» il faudrait un être supérieur connaissant l'état mental du locuteur ainsi que celui de l'auditeur. En réalité, la compréhension n'existe donc que dans le regard d'un des agents: l'auditeur peut *croire* qu'il a compris, le locuteur peut *croire* que l'auditeur a compris, l'auditeur peut *croire* que le locuteur croit qu'il a compris, etc. En fin de compte, le locuteur va «décider» si l'auditeur a compris son intention, mais comme il n'a pas accès à l'état mental de celui-ci, son jugement n'est pas objectif non plus.

Par souci de précision, nous introduisons les notations suivantes[1]: e_x est un énoncé par lequel l'agent X veut exprimer une certaine intention communicative $\mathsf{IntTh}_X\varphi$. La relation entre e_x et cette intention sera exprimé par $\textit{sig-nn}(X, e_x, \varphi)$. Quand le locuteur Y attribue l'intention $\mathsf{IntTh}_X\varphi'$ à e_x, ceci sera exprimé comme $\mathsf{Bel}_Y\textit{sig-nn}(X, e_x, \varphi')$. (Dans la suite, nous supposons que le locuteur connaît sa propre intention communicative et qu'alors $\textit{sig-nn}(X, e_x, \varphi) \leftrightarrow \mathsf{Bel}_X\textit{sig-nn}(X, e_x, \varphi)$.) Nous pouvons maintenant reformuler le critère de «grounding» de Clark et Schaefer (1989) (cf. page 17) de la manière suivante:

critère de «grounding»: un groupe d'agents a atteint le critère de «grounding», si chaque agent (y compris le locuteur) croit que la signification qu'il a attribué à l'énoncé est aussi mutuellement cru par tous les agents (y compris le locuteur).

Pour deux agents, le critère de «grounding» peut être représenté avec la formule suivante: $\mathsf{Bel}_X\mathsf{MB}_{X,Y}\textit{sig-nn}(X, e_x, \varphi) \wedge \mathsf{Bel}_Y\mathsf{MB}_{X,Y}\textit{sig-nn}(X, e_x, \varphi')$. Bien entendu, dans ce cas-là, chaque agent va croire que *son* interprétation de l'énoncé est partagée par les autres agents; la distinction entre φ et ψ n'existe donc que d'un point de vue *extérieur*.

En principe, cette définition est similaire à celle de Clark et Schaefer. Nous avons omis le «degré suffisant pour l'objectif courant», mais nous supposerons qu'une intention communicative est attribuée 'a un énoncé déjà en termes d'un «objectif courant». Le point important de cette réformulation, c'est que nous avons accentué l'aspect subjectif de la signification d'un énoncé et que nous avons décrit le critère de «grounding» uniquements en termes des croyances individuelles de chaque agent (y compris les croyances mutuelles unilatérales de l'agent). Nous pouvons donc commencer à décrire comment – pour chaque agent – son critère *individuel* de «grounding» peut être atteint.

4.2 La supposition communicative

Cette section décrit les états mentaux particuliers des agents engagés dans un dialogue après qu'un énoncé a été produit. La section suivante décrit comment à partir de ces états le critère de «grounding» peut être atteint.

Des personnes engagés dans un dialogue savent que la communication verbale est faillible. C'est pourquoi ils confirment qu'ils ont compris ou demandent de telles confirmations. Suivant la théorie de Clark et Schaefer (1989), un locuteur X ne peut pas être certain que l'auditeur Y a compris jusqu'à ce que Y fournisse un «acknowledgement» (explicitment ou implicitement). Mais si la faillibilité de la communication est connu de tous, l'auditeur Y qui a remarqué et compris un énoncé se trouve dans une situation intéressante: Y sait qu'il a compris (ou au moins, il le croit), mais il *sait* aussi que le locuteur X n'en peut pas encore être certain. Un auditeur qui croit qu'il a compris est donc *certain de l'incertitude* du locuteur.

1. Il ne s'agit pas encore d'une formalisation, mais seulement d'une notation pour préciser les formulations. Syntaxe, sémantique et axiomatisation précises des termes utilisés seront données dans le chapitre 5.

L'incertitude du locuteur n'est cependant pas telle qu'elle le «paralyse» et l'empêche de continuer jusqu'à ce qu'il reçoive un «acknowledgement». Le locuteur peut utiliser des expressions anaphoriques se référant aux parties précédantes de son énoncé, il peut continuer de parler et produire plusieurs phrases à la suite, etc. Le locuteur a donc quand-même une certaine confiance dans le fait d'avoir été compris. Bien qu'il sache que cela peut être faux, X *suppose* que Y a compris.

Cela peut expliquer l'utilisation d'anaphores, etc. Si X suppose que Y a compris, il peut aussi supposer que Y va comprendre l'expression anaphorique. Le point important c'est que l'état mental de supposition permet de raisonner, c'est-à-dire d'inférer d'autres suppositions, et d'agir sur leur base. Cependant la croyance mutuelle ne peut pas être inférée, parce qu'il y a toujours la possibilité que l'auditeur n'ait pas compris. Toutes les inférences restent des suppositions et ne sont valides que si Y a vraiment compris – si X arrive à croire que ce n'est pas le cas, il doit les abandonner.

Ces suppositions sont donc des *inférences prédictives* et jouent plusieurs rôles importants: elles permettent à X de produire un discours cohérent sans redondances (X ne va pas dire des choses que Y pourrait inférer s'il avait compris). De plus, elles aideront à déterminer si Y a compris l'énoncé: si, par exemple, X affirme qu'il croit un fait p, il va supposer après son énoncé que Y croit qu'il croit p. Si q est une conséquence logique de p, X peut inférer que Y croit maintenant aussi que X croit q. Si alors Y répond avec un énoncé qui nécessite pour être pertinent que Y croit que X croit p, X peut prendre cette réponse comme une évidence de compréhension. S'il n'avait pas raisonné sur ces suppositions, cela ne serait pas possible. Dans la section 4.4 nous essayons de mettre en rapport nos idées avec des théories psycholinguistiques sur les inférences prédictives.

Aprés avoir énoncé e_x, X va donc *supposer* que Y a compris. En utilisant l'opérateur Sup introduit en section 2.2.1 et la notion de compréhension de la section précédante nous pouvons représenter cet état mental comme $\mathsf{Sup}_X \mathsf{Bel}_Y \textit{sig-nn}(X, e_x, \varphi)$. Mais cette formule ne décrit que partiellement l'état mental de X. Si tous les agents savent qu'un locuteur ne peut pas être certain d'être compris et si Y a compris un énoncé de X, Y doit aussi savoir que X se trouve dans cet état d'incertitude maintenant. La situation suggère alors à X aussi $\mathsf{Sup}_X \mathsf{Bel}_Y$(X suppose Y a compris φ), c'est-à-dire $\mathsf{Sup}_X \mathsf{Bel}_Y \mathsf{Sup}_X \mathsf{Bel}_Y \textit{sig-nn}(X, e_x, s_x)$. Le même raisonnement entraîne que, s'il avait compris, Y devrait connaître aussi cette deuxième supposition, etc. Malgré son incertitude, X semble se trouver dans une situation partagée (cf. section 2.2.2, page 2.2.2), c'est-à-dire dans une situation particulière qui indique aux agents un fait ainsi que sa propre présence et qui peut ainsi entraîner des états mentaux particuliers comprenant une infinité d'attitudes mentales. En section 2.2.2 et dans la théorie de Clark et Marshall (1981), une telle situation a été utilisé pour décrire la *croyance mutuelle*, ici elle décrit un état mental de X que nous appellerons **supposition communicative**. La supposition communicative représentant un état mental de X, elle ressemble à une croyance mutuelle unilatérale. Et de même qu'on peut caractériser la croyance mutuelle unilatérale de X sur la compréhension de son énoncé par Y par la conjonction infinie

$$\mathsf{Bel}_X \mathsf{Bel}_Y \textit{sig-nn}(X, e_x, \varphi) \wedge \mathsf{Bel}_X \mathsf{Bel}_Y \mathsf{Bel}_X \mathsf{Bel}_Y \textit{sig-nn}(X, e_x, \varphi) \wedge \cdots$$

la supposition communicative est caractérisée par la formule

$$\mathsf{Sup}_X \mathsf{Bel}_Y \textit{sig-nn}(X, e_x, \varphi) \wedge \mathsf{Sup}_X \mathsf{Bel}_Y \mathsf{Sup}_X \mathsf{Bel}_Y \textit{sig-nn}(X, e_x, \varphi) \wedge \wedge \cdots$$

Cet état sera décrit par la suite avec par la formule $\mathsf{CSup}_X \mathsf{Bel}_Y \mathit{sig\text{-}nn}(X, e_x, \varphi)$. Dans le chapitre 5 une définition précise de l'opérateur CSup sera présentée. Dans la section 4.4 nous discutons la plausibilité de la supposition communicative comme état mental. Intuitivement, on a l'impression que cet état représente le «mi-chemin» envers le critère de «grounding» – le «grounding» serait atteint si toutes les suppositions devenaient des croyances.

Notons que Bunt (1989) utilise un opérateur *Suspect* d'une manière similaire. Mais pour lui, l'incertitude d'être compris remonte à l'effet suivant d'un acte de langage (Bunt (1989, p.60), présenté ici en utilisant notre terminologie): $\mathsf{Bel}_X \mathsf{MB}_{X,Y} \mathsf{Sup}_X \mathsf{Bel}_Y \mathit{sig\text{-}nn}(X, e_x, \varphi)$. Comme Bunt dit lui-même, cela signifie par exemple que X croit que Y croit que X suppose que Y croit que $\mathit{sig\text{-}nn}(X, e_x, s_x)$. Mais si X croit cela, il *croit* donc aussi que Y a compris, et dans ce cas-là $\mathsf{Sup}_X \mathsf{Bel}_Y \mathit{sig\text{-}nn}(X, e_x, \varphi)$ devrait être remplacé par $\mathsf{Bel}_X \mathsf{Bel}_Y \mathit{sig\text{-}nn}(X, e_x, \varphi)$. Nous croyons donc que l'alternance de Sup et Bel engendré par la supposition communicative est plus appropriée à représenter l'état du locuteur.

La supposition communicative peut aussi être utilisée pour décrire l'état mental de l'auditeur. Mais la nécessité du «grounding» naissant de la faillibité de la communication, la théorie doit considérer plusieurs cas de compréhension et d'incompréhension. Trois cas seront discutés ici: l'auditeur croit qu'il a compris, l'auditeur croit qu'il n'a pas compris et l'auditeur n'est pas certain s'il a bien compris ou pas. Nous avons déjà vu un exemple de ce troisième cas en figure 3.6 où une demande de confirmation de l'auditeur montre qu'il a en fait compris mais n'en est pas sûr.

Si l'auditeur Y est certain d'avoir compris, il est aussi certain de connaître la supposition communicative de X. Cet état peut donc être décrit comme $\mathsf{Bel}_Y \mathsf{CSup}_X \mathsf{Bel}_Y \mathit{sig\text{-}nn}(X, e_x, \varphi')$. Mais si Y ne fait que *supposer* que $\mathit{sig\text{-}nn}(X, e_x, \varphi')$, il peut aussi seulement supposer quel est l'état mental de X. L'état de Y correspond donc à $\mathsf{Sup}_Y \mathsf{CSup}_X \mathsf{Bel}_Y \mathit{sig\text{-}nn}(X, e_x, \varphi')$. Il est important de voir pourquoi à l'intérieur de CSup_X la croyance de Y ne devient pas aussi une supposition: la raison est qu'il s'agit de la supposition communicative de X qui ne peut pas encore savoir que Y n'est pas certain de son interprétation de l'énoncé. Si Y n'a pas du tout compris l'énoncé, il peut penser seulement qu'il doit y avoir une certaine signification non-naturelle et une certaine supposition communicative, mais il ne peut pas la préciser. Mais ce qui est important, c'est que dans tous les cas, l'auditeur Y sait que le locuteur n'est pas sûr que Y avait compris. Cela va fournir une motivation à Y pour donner un «acknowledgement» ou un signal d'incompréhension [2].

4.3 Atteindre le critère de «grounding»

La supposition communicative du locuteur X après avoir énoncé e_x joue un rôle important pour son interprétation de la réponse e_y de Y. Si ce n'était pas le cas, ce seraient seulement les rôles qui seraient changés, et après sa réponse Y se retrouverait dans l'état $\mathsf{CSup}_Y \mathsf{Bel}_X \mathit{sig\text{-}nn}(Y, e_y, \psi)$ où ψ serait, par exemple, $\mathsf{Bel}_Y \mathit{sig\text{-}nn}(X, e_x, \varphi)$. Intuitivement, la différence entre le premier énoncé de X et la réponse de Y est que les deux agents savent que X *s'attend* à un «acknowledgement». Le fait qu'il existe une information dont X attend un

2. Dans une ancienne version de ce chapitre deux propriétés de l'auditeur qui correspondaient à son état de «vigilance» et sa «coopérativité» précisaient quand l'auditeur remarque que le locuteur se trouve dans un état de supposition communicative et quand il va fournir un signal de compréhension. Ces notions ont été omises ici en raison de la clarté du travail, mais pourraient être utilisées pour préciser et élargir la théorie.

«acknowledgement» est même mutuellement connu, mais X n'est pas certain que Y sache de quelle information il s'agit.

Une telle situation d'**attente** d'un signal de compréhension a des propriétés particulières qui vont permettre le «grounding». Contrairement à l'interprétation d'un énoncé entièrement nouveau, l'agent en *attente* d'une certaine information a la possibilité de *chercher* activement cette information. Ceci réduit une partie du processus d'interprétation avec un résultat inconnu à la *décision* si l'information cherchée est fournie par l'énoncé ou pas[3]. L'interprétation guidée par le but (goal driven) peut être beaucoup plus efficace que l'interprétation uniquement guidée par les données (data driven), parce que l'agent «sait ce qu'il cherche» et va le reconnaître plus simplement et sans ambiguïté. Pour reconnaître des «acknowledgements» implicites, X peut profiter des inférences prédictives qu'il a tirées de la supposition que Y avait compris.

À part le traitement *data driven*, l'agent X peut profiter d'une autre circonstance: son état d'attente est mutuellement connu et les deux agents savent que X doit prendre une décision maintenant parce que cette décision détermine son prochain énoncé. S'il croit que Y n'avait pas compris, il va répéter ou reformuler son énoncé, s'il croit que Y avait compris, il va continuer le dialogue. Pour pouvoir continuer le dialogue le plus vite possible, Y va donc essayer de montrer à X sa compréhension ou son incompréhension de manière le moins ambigüe possible. Ceci est surtout facile dans le cas où Y sait qu'il n'avait pas compris l'énoncé de X; il est presque certain que Y peut exprimer cela sans ambiguïté. De même, si Y est certain d'avoir compris, il va aussi être certain de pouvoir donner une réponse cohérente (ou un «acknowledgement» explicite) qui montrera sa compréhension et continuera le dialogue. Le troisième cas, l'incertitude de Y sur ce qu'il avait compris résultera dans une demande de confirmation. La demande de confirmation répétant (cf. figure 3.6) ce qui a été compris, Y va au moins être certain que X va pouvoir décider si Y avait compris ou pas.

La situation après la réponse de Y se distingue donc essentiellement de celle après l'énoncé de X: Y est *certain* que X va correctement interpréter son énoncé – c'est-à-dire, bien sûr, que Y est certain que X va correctement interpréter son «acknowledgement», son signal d'incompréhension ou sa demande de confirmation; toutes les *nouvelles* informations dans l'énoncé doivent être «grounded» *séparément*. Y peut être certain parce qu'il sait que X s'attend à certaines informations. Mais ceci est mutuellement connu et il est donc aussi mutuellement connu que Y est certain que sa réponse va être comprise. Si alors cette réponse signifie que Y avait compris l'énoncé de X, ceci entraîne la croyance mutuelle que Y est certain d'avoir compris. Du point de vue de Y, ceci est équivalent au «grounding», parce qu'il convaincu d'avoir bien compris et que ce fait est maintenant mutuellement connu. Bien entendu, ceci ne garantit pas que Y avait vraiment compris l'énoncé, mais si X vient à la même décision il sait que pour Y le critère de «grounding» a déjà été atteint et aucune réaction de plus de X n'est nécessaire: X va aussi croire que son énoncé a été «grounded».

Le même raisonnement montre comment l'expression d'incompréhension par Y peut fournir le «grounding»: Y va être certain d'être compris et du fait que X sait que Y en est certain. Pour lui cette situation partagée (cf. section 2.2.2) engendre la croyance mutuelle que Y n'avait pas compris. Si X comprend la réponse, il va aussi croire que cette croyance mutuelle a été atteinte. X peut alors reessayer de se faire comprendre, probablement en considérant que le premier essai a échoué et en essayant d'éviter des raisons possibles de l'échec.

3. Cette décision est néanmoins difficile: un signal de compréhension peut être implicite, de plus de nouvelles informations peuvent être données avec l'«acknowledgement».

Souvent, l'incompréhension va être partielle plutôt que totale, le locuteur ne va pas être sûr d'avoir bien compris (par ex. parce que ce qu'on lui a dit ne lui semble pas pertinent). Un exemple d'une demande de confirmation qui exprime cette incertitude a déjà été présenté (fig. 3.6). Un deuxième exemple identique à une seule lettre près, montre une situation de mauvaise compréhension où les doutes de l'auditeur sont justifiés (fig. 4.1).

(1)	Utilisateur:	Je voudrais parler à Corinne Savol.
(2)	Standard:	Corinne Davol?
(3a)	Utilisateur:	Non,
(3b)		Corinne <u>S</u>avol!
(3c)		Elle travaille avec Louise Abou.

FIG. 4.1 – *Exemple d'une demande de confirmation*

La demande de confirmation présente une situation particulière: les deux agents croient que X peut déterminer selon cette demande si Y avait compris ou pas. Mais comme Y n'est pas certain d'avoir compris, il ne peut pas être certain de la décision de X non plus. X, par contre, va non seulement savoir si Y avait compris ou pas, mais aussi que Y est dans cette situation d'incertitude maintenant – c'est maintenant Y qui s'attend à un «acknowledgement» ou à une correction. Par rapport à la situation après le premier énoncé de X, les rôles ont changés, l'auditeur s'attend à un «acknowledgement» du locuteur. Mais le point important c'est qu'ici les incertitudes et ainsi le processus de «grounding» sont finis. Les deux agents savent que X va «prononcer son jugement» qui sera (quoi qu'il soit positif ou négatif) attendu par Y et certainement correctement interprété. Dans l'exemple en fig. 4.1, c'est l'énoncé (3a) qui a cette fonction. La certitude d'être compris va mener à une croyance mutuelle dans les deux cas: au «grounding» dans le cas positif, à la croyance mutuelle en l'échec dans le cas négatif. Dans ce cas, il est très probable que le locuteur X va tout de suite réessayer de se faire comprendre, peut être en mettant l'accent sur les parties non comprises de son premier énoncé (indiqué dans l'exemple par un soulignement).

Il y a de grandes subtilités langagières et dans les «acknowledgements» et les demandes de confirmations. Par exemple, la répétition du nom dans l'énoncé (2) de l'exemple en figure 4.1 ne représente pas seulement une demande de confirmation de ce nom, mais aussi un «acknowledgement» du reste de la phrase (1) qui n'est pas répétée. Le standard à bien compris que l'utilisateur veut parler à quelqu'un et cette partie est «grounded» avec (2). C'est pourquoi dans (3b), l'utilisateur n'a plus besoin de le répéter. Notre théorie peut traiter ces subtilités sans changement: le point essentiel, c'est que le locuteur *sait* ce qu'il veut savoir confirmé par l'auditeur et qu'il va pouvoir utiliser une interprétation guidée par ce but pour déterminer si la réponse fournit cette évidence. Pour avoir cette granularité plus fine, on doit alors voir un énoncé comme composé de nombreuses parties qui peuvent être «grounded» individuellement.

Notre théorie prédit qu'après au plus tard trois énoncés (si une demande de confirmation a été posée) une croyance mutuelle unilatérale sera établie pour chacun des agents. Ceci peut être la croyance mutuelle que l'auditeur a compris l'énoncé ou bien la croyance mutuelle qu'il ne l'a pas compris. Dans ce cas-là, la contribution (dans le sens de Clark et Schaefer (1989)) peut continuer (et peut-être que le «grounding» ne sera jamais atteint), mais en principe à ce moment-là le processus commence de nouveau. Nous croyons donc qu'au lieu de considérér une contribution comme un tout, il est plus utile de regarder ces *unités de grounding* de deux ou trois énoncés (des «coups» de la théorie de jeux) que nous avons présentées dans ce chapitre et où chaque énoncé joue un rôle précis. Contrairement aux *unités de discours* de Traum (cf. section 3.3.1) nos unités de «grounding» sont d'une longueur fixe de ou bien deux (présentation → «acknowledgement» ou présentation → signal d'incompréhension), ou bien trois «coups» (présentation → demande de confirmation → «acknowledgement» positif ou négatif). Une contribution (Clark et Schaefer, 1989) ou une unité de discours (Traum, 1994) peuvent être construites[4] à partir d'unités de «grounding» en les combinant de telle manière que les «résultats intermédiaires» sont toujours des croyances mutuelles sur l'échec du «grounding» et seul le dernier échange établit la croyance mutuelle que l'auditeur a (finalement) compris.

4.4 Validité psychologique de la théorie

À plusieurs points importants de notre théorie nous avons eu recours à des théories psychologiques. Ce sont aussi les points où nos hypothèses sont certainement critiquables. Revoyons ces points:

Les processus de raisonnement que nous attribuons aux agents sont parfois très complexes. On pourrait donc reprocher à la théorie de supposer une rationnalité parfaite des agents que les êtres humains n'ont sûrement pas. Il faut pourtant distinguer entre les cas où nous supposons un véritable raisonnement déductif et les cas où nous supposons la naissance d'une croyance mutuelle ou supposition communicative à partir d'une situation partagée. Quant aux déductions décrites ici, nous croyons que le processus plutôt «standardisé» du «grounding» permet qu'un agent puisse obtenir ses résultats sans avoir besoin de les déduire chaque fois. Il s'agirait donc d'une sorte de *Chunking*, un résumé intensionalisé (avec un 's'!) et généralisé d'un processus mental (Laird et al., 1986). Les règles et résultats du processus du «grounding» sont appris en interagissant socialement, probablement sans en être conscient. Quant aux croyances mutuelles et suppositions communicatives dont nous considérons qu'elles naissent directement d'une situation particulière, nous nous appuyons sur Clark et Marshall (1981) qui citent plusieurs études psychologiques favorables à l'hypothèse que des croyances mutuelles unilatérales sont des états mentaux primitifs. Quant à la notion de supposition communicative introduite dans notre travail, nous ne connaissons pas de travaux psychologiques traitant ce type de sentiment d'incertitude dans un groupe. Pourtant, il nous semble probable que d'autres états mentaux existent qui pourraient décrire des croyances dans un groupe moins «certaines» que la croyance mutuelle.

L'introduction de l'attitude Sup est la distinction la plus importante par rapport aux autres modèles de «grounding». Elle est particulièrement justifiée par l'étude sur les inférences prédictives pendant la compréhension de textes par Keefe et McDaniel (1993). Ces auteurs montrent que les inférences prédictives existent, mais sont d'une nature bien différente que d'autres croyances: elles disparaissent rapidement si elles ne sont pas confirmées ou si elles se montrent non pertinentes par la suite. C'est exactement dans ce sens-là que nous utilisons ce que nous

4. Les modèles de Clark et Schaefer (1989) et Traum (1994) comprennent des éléments qui ne sont pas constructibles avec notre modèle; il s'agit notamment du concept de «réparation». Nous croyons que la réparation d'un énoncé est une action qui nécessite que l'auditeur *a compris* l'énoncé, mais n'est pas d'accord sur son contenu. À notre avis, une réparation ainsi qu'une demande de réparation représente plutôt un «acknowledgement» implicite.

avons appelé les «suppositions». Elles disparaissent si la réponse de l'autre ne les confirme pas. Mais tant qu'elles existent, elles se trouvent dans une sorte de «symbiose» avec l'énoncé de l'auditeur: elles aident à l'interprétation de cet énoncé et la réponse aide à les justifier. Notre idée d'une supposition est donc la suivante: si elle a contribué à interpréter l'énoncé, cela montre qu'elle est vraie et elle va pouvoir se transformer en une croyance. Si, par contre, une interprétation de l'énoncé est trouvée qui contredit la supposition, elle est abandonnée.

Une troisième hypothèse contestable est la suivante: la certitude que le locuteur va trouver une décision *immédiate* sur la compréhension de son énoncé par l'auditeur. Il serait aussi possible qu'un agent *garde* les suppositions au lieu de les transformer en croyances ou de les abandonner. Mais ceci pose un problème concernant les capacités mentales d'un agent: la réponse de l'auditeur Y va sûrement entraîner de nouvelles inférences et il est évident que celles-ci doivent être limitées quelque part. Si toute ancienne supposition pouvait mener à des nouvelles suppositions en fonction de l'énonce de Y, le nombre de suppositions exploserait. Ces chaines d'inférences deviendraient de plus en plus nombreuses et de plus en plus improbables (parce qu'il s'agit de suppositions sur suppositions) – ceci ne serait ni cognitivement gérable ni utile. La théorie *construction-intégration* de Kintsch (1988) propose donc le mécanisme de *pruning*, la «coupure» des inférences devenues inutiles. Alors que Keefe et McDaniel (1993) pensent plutôt à une dégradation des inférences prédictives, Kintsch propose le processus *actif* du «pruning». Sans vouloir entrer dans une discussion détaillée, nous considérons l'abandon de la supposition de compréhension comme processus *actif*, et même conscient, qui sert à rendre toute autre inférence dans cette direction inutile.

Bien que nous ayons mentionné des travaux en accord avec nos idées, il reste la nécessité d'une étude empirique. La théorie repose sur des observations de l'auteur et des résultats des autres théories présentées plutôt que sur une étude soigneuse expérimentale. Un corpus de dialogues téléphoniques (Castaing, 1993) a été étudié et semble ne pas s'opposer à nos idées. Pourtant, une étude statistique de ce corpus n'a pas été faite; nous ne croyons pas qu'une telle étude pourrait valider ou invalider la théorie, parce que les processus cognitifs impliqués ne sont pas mesurable dans un corpus de dialogue. Une vraie validation ne peut se faire que de manière expérimentale. Ceci n'a pas (encore) été fait, et nous n'avons pas dévéloppé un paradigme expérimental approprié. Néanmoins, nous croyons que cela *pourrait* être fait. Di Eugenio et al. (1999) présentent une étude empirique sur le processus de trouver un accord (agreement process) qui pourrait inspirer une telle expérience. D'autres possibilités naissent des expériences citées dans cette section.

En sciences cognitives, une deuxième méthodologie est utilisée pour démontrer la validité *pratique* d'une théorie: la modélisation en un système artificiel. Si un agent artificiel montre un comportement plus réaliste qu'un autre, cela ne suffit pas pour prouver que la théorie sous-jacente est «correcte», mais donne probablement quand-même une indication sur le chemin qu'on pourrait suivre.

C'est cette approche que nous allons poursuivre dans la suite du mémoire. Dans le prochain chapitre nous formalisons la théorie et prouvons quelques propriétés d'un agent modélisé suivant notre approche.

Chapitre 5

Éléments d'une formalisation

Ce chapitre présente une logique multi-modale que nous avons développée pour formaliser la théorie du «grounding» décrite dans le chapitre précédent. Ce qui distingue notre logique des approches similaires est l'introduction de l'opérateur de supposition Sup. Il va nous permettre de tenir compte de l'incertitude sur les effets d'un acte de langage en décrivant cette incertitude *comme un effet* de l'acte. Nous décrivons syntaxe et sémantique de la logique et prouvons quelques unes de leur propriétés en section 5.1. Cette section présuppose une familiarité avec les concepts de la logique modale, parce que pour des raisons de brièveté (et de pertinence pour le cadre du travail) certains aspects ne sont abordés que superficiellement. Pourtant, il n'est pas nécessaire de comprendre la définition formelle de la logique pour pouvoir suivre les sections suivantes. En section 5.2 nous décrivons des notions importantes de la théorie en termes de la logique. Enfin, la section 5.3 présente plusieurs théorèmes qui montrent entre autres l'établissement du «grounding» pour plusieurs types de situation.

À quoi une telle formalisation sert-elle? Nous croyons que la spécification est le meilleur moyen pour détecter les inconvénients d'une théorie, permettre une unification avec d'autres théories et, surtout, projeter et vérifier des systèmes qui se basent sur la théorie. Les résultats prouvés dans ce chapitre vont nous permettre de développer un modèle de dialogue dans le chapitre 6 et de l'implémenter dans un système de dialogue réel (chapitre 7).

5.1 Une logique multi-modale

Cette section décrit syntaxe, sémantique et axiomatisation d'une logique multi-modale temporelle de premier ordre. Nous allons nous concentrer sur l'aspect particulier de cette logique: l'opérateur de supposition Sup et son interaction avec l'opérateur de croyance Bel. D'autres aspects de la logique qui ne sont pas pertinent pour le cadre général du travail ne sont pas spécifié. En particulier, aucune notion formelle d'action n'est intégrée dans le formalisme. À part des raisons d'espace, ceci est dû au fait que nous ne décrivons ni la réalisation d'un acte de langage ni son interprétation, mais seulement les résultats de ces processus dans ce chapitre. Ainsi, nous pouvons limiter le traitement des actions aux croyances sur un prédicat $\mathit{done}(,)$ qui «simule» le résultat de la perception et de l'exécution d'une action.

Nous ne donnons pas non plus une sémantique formelle pour l'opérateur modal d'intention-que IntTh. Nous favorisons une approche similaire à celle de Konolige et Pollack (1993), mais la complexité de leur modèle est disproportionnée par rapport au rôle que joue l'intention dans le présent travail.

5.1.1 Motivation

La nécessité du «grounding» naissant de la faillibilité de la compréhension, nous avons besoin d'une logique capable de traiter l'incertitude. Il existe des propositions pour l'intégration d'un raisonnement probabiliste dans une logique doxastique. Mais cela ne correspondrait pas à nos besoins: nous croyons qu'un agent peut avoir un sentiment d'incertitude *qualitatif*, c'est-à-dire qu'il va être *conscient* de la possibilité que ce qu'il pense peut être faux sans pouvoir exprimer cette incertitude en termes de probabilités concrètes.

Mais une supposition est plus que l'incertitude. «Je suppose que x» inclut la possibilité que x est faux, néanmoins cela n'exprime pas l'ignorance totale («Je ne sais pas si x»). Nous caractérisons la supposition ainsi: Un agent suppose que x, si pour lui x est plus probable que son contraire. En utilisant Bel et Sup, nous aurons quatre degrés de certitude à exprimer et nous croyons que dans la plupart des situations, les hommes se contentent de ces quatre possibilités: «Je suis certain que x est vrai», «je suppose que x est vrai», «je suppose que x n'est pas vrai» et «je suis certain que x n'est pas vrai» [1].

En utilisant une sémantique de mondes possibles, la croyance est normalement exprimée ainsi: un agent A croit x, si dans tous les mondes qu'il considère possibles, x est vrai. Comment intégrer la notion de supposition dans ce modèle? Nous allons poursuivre une approche directe et simple: au lieu d'exiger que x doit être vrai dans tous les mondes considérés *possibles*, il nous suffit que x soit vrai dans tous les mondes considérés *probables* par A pour dire que A suppose que x. $\mathsf{Sup}_A x$ sera donc un état plus facilement atteignable: Je peux supposer que Jaques Chirac est président de la France, car dans tous les mondes probables que je peux imaginer ceci est le cas. Mais puis-j'en être certain? Il pourrait bien avoir démissioné ce matin ... Le lien entre les mondes «possibles» et «probables» est donc défini ainsi: tout monde «probable» est aussi «possible», mais un monde «possible» n'est pas nécessairement «probable».

Il y a un lien étroit entre l'opérateur $\mathsf{Sup}_A x$ et l'opérateur de *possibilité* $\neg\mathsf{Bel}_A\neg x$, mais il est crucial de voir leurs différences:

- $\mathsf{Sup}_A x$ implique $\neg\mathsf{Bel}_A\neg x$, mais l'inverse n'est pas nécessairement vrai
- $\neg\mathsf{Bel}_A\neg x$ permet que $\neg\mathsf{Bel}_A x$ soit vrai en même temps – en fait, ce cas définit *l'ignorance*, souvent exprimée comme $\neg\mathsf{KIF}_A x$ (ne pas savoir si x). Pourtant, $\mathsf{Sup}_A x$ exclut $\mathsf{Sup}_A\neg x$: ou bien x est considéré plus probable ou bien $\neg x$.

Formaliser ces propriétés sera la tentative du reste de cette section.

5.1.2 Syntaxe

Nous formalisons notre théorie dans une logique multi-modale du premier ordre avec égalité. Par souci de brièveté et de pertinence pour ce qui suit, nous décrirons les parties concernant la logique de prédicats très superficiellement.

1. L'ignorance représente donc le cas où l'agent n'aucune de ces croyances et suppositions.

Pour un alphabet S comprenant des variables, des constantes, des symboles de fonctions, des symboles de relations et les symbole suivants:

$$\neg, \wedge, \vee, \rightarrow, \leftrightarrow, \equiv, \forall, \exists,), ($$

nous définissons le langage de la logique des prédicats du premier ordre $\mathcal{L}_{Pr}$ comme usuellement.

Pour pouvoir parler du temps, nous ajoutons des règles suivantes qui définissent un langage $\mathcal{L}_T$

SYN-1. Toutes les règles qui définissent la syntaxe de $\mathcal{L}_{Pr}$ sont aussi des règles pour définir $\mathcal{L}_T$ [2]

SYN-2. Si $\varphi \in \mathcal{L}_T$, alors $\mathsf{Next}\ \varphi \in \mathcal{L}_T$ et $\mathsf{Last}\ \varphi \in \mathcal{L}_T$

Nous élargissons le langage en ajoutant les opérateur modaux Bel, Sup et IntTh qui représentent des attitudes mentales. De plus nous utiliserons les opérateurs GBel, MB et CSup pour des attitudes concernant plusieurs agents. Chaque opérateur sera défini pour tous les membres d'un groupe G (qui est un sous-ensemble de l'ensemble de constantes dans S). Le langage $\mathcal{L}_M$ sera donc défini avec les règles suivantes:

SYN-3. Toutes les règles qui définissent la syntaxe de $\mathcal{L}_T$ sont aussi des règles pour définir $\mathcal{L}_M$

SYN-4. Si $\varphi \in \mathcal{L}_M$ et $A \in G$, alors $\mathsf{IntTh}_A\varphi \in \mathcal{L}_M$, $\mathsf{Bel}_A\varphi \in \mathcal{L}_M$ et $\mathsf{Sup}_A\varphi \in \mathcal{L}_M$

SYN-5. Si $\varphi \in \mathcal{L}_M$ et $A, B \in G$, alors $\mathsf{GBel}_{A,B}\varphi \in \mathcal{L}_M$, $\mathsf{MB}_{A,B}\varphi \in \mathcal{L}_M$ et $\mathsf{CSup}_A\varphi \in \mathcal{L}_M$

5.1.3 Sémantique formelle

Nous définissons une *structure de Kripke M* comme le tuple $(W, T, Agt, B, S, I, \pi)$. W est un ensemble de mondes. Nous parlons d'un monde comme un point dans le temps. T est une relation qui associe à chaque monde le monde au moment suivant [3]. Agt est un ensemble d'agents, B, I et S sont des familles de relations binaires sur W, c'est-à-dire des pairs $(w, w') \in W \times W$. Les familles B, I et S sont indexées avec les éléments de Agt, c'est-à-dire que pour chaque agent A dans Agt, il existe une relation $W \times W$ correspondante dans les ensembles B, I et S. Pour un agent nous nous référons aux relations correspondantes par $\mathcal{B}_A$ et $\mathcal{S}_A$ (comme mentionné ci-dessus, nous ne donnerons pas de sémantique pour IntTh dans ce travail). Pour chaque monde, π associe aux constantes, symboles de fonctions et symboles de relation des valeurs, fonctions et symboles du monde.

Nous définissons la *relation de modèle* ($\models$) pour assigner une valeur de vérité à une expression du langage $\mathcal{L}_M$. Cette valeur va être différente pour chaque monde; la relation de modèle sera donc exprimée en fonction de la structure M, un monde $w \in W$ et une expression du langage $\mathcal{L}_M$. Pour définir la sémantique des opérateurs modaux, nous allons utiliser les notations $\mathcal{B}_A(w) = \{w' | (w, w') \in \mathcal{B}_A\}$ et $\mathcal{S}_A(w) = \{w' | (w, w') \in \mathcal{S}_A\}$, respectivement.

2. cette formulation un peu compliquée doit exprimer que les opérateurs de $\mathcal{L}_{Pr}$ sont maintenant aussi applicable aux opérateurs modaux: non seulement $\mathsf{Next}\ (\varphi \wedge \psi)$, mais aussi $\mathsf{Next}\ \varphi \wedge \mathsf{Next}\ \psi$

3. Pour simplicité nous donnons au temps une structure isomorphe aux nombres entiers: un ordre total discret sans premier ni dernier élément

Nous définissons par induction sur la structure des expressions (mais omettons ici les parties concernant la logique du premier ordre):

SEM-1. $M \models_w \mathsf{Next}\ \varphi$ *ssi* $M \models_{w'} \varphi$ où $(w, w') \in T$
SEM-2. $M \models_w \mathsf{Last}\ \varphi$ *ssi* $M \models_{w'} \varphi$ où $(w', w) \in T$
SEM-3. $M \models_w \mathsf{Bel}_A \varphi$ *ssi* $M \models_{w'} \varphi$ pour tout $w' \in \mathcal{B}_A(w)$
SEM-4. $M \models_w \mathsf{Sup}_A \varphi$ *ssi* $M \models_{w'} \varphi$ pour tout $w' \in \mathcal{S}_A(w)$
SEM-5. $M \models_w \mathsf{GBel}_{A,B} \varphi$ *ssi* $M \models_w \mathsf{Bel}_A \varphi$ et $M \models_w \mathsf{Bel}_B \varphi$
SEM-6. $M \models_w \mathsf{MB}_{A,B} \varphi$ *ssi* $M \models_w \mathsf{GBel}^k_{A,B} \varphi$ pour tout $k > 0$
SEM-7. $M \models_w \mathsf{CSup}_{A,B} \varphi$ *ssi* $M \models_w \mathsf{Sup}_A (\mathsf{Bel}_B \mathsf{Sup}_A)^k \varphi$ pour tout $k \geq 0$

5.1.4 Théorie de preuve

Nous poserons maintenant des contraintes sur les ensembles de relations d'accessibilité $\mathcal{B}$ et $\mathcal{S}$. Ces contraintes vont donner aux opérateurs Bel et Sup des propriétés sémantiques qui nous semblent souhaitables.

Pour tout agent $A \in Agt$, la relation $\mathcal{B}_A$ est:

- *sérielle*: pour tout monde $w \in W$, il existe $w' \in \mathcal{B}_A(w)$.
- *transitive*: si $w' \in \mathcal{B}_A(w)$ et $w'' \in \mathcal{B}_A(w')$ alors $w'' \in \mathcal{B}_A(w)$.
- *euclidiénne*: si $w' \in \mathcal{B}_A(w)$ et $w'' \in \mathcal{B}_A(w)$ alors $w'' \in \mathcal{B}_A(w')$

Nous allons choisir $\mathcal{S}_A$ tel que ces propriétés seront vraies aussi, mais plutôt comme conséquence de sa relation avec $\mathcal{B}_A$. Pour cela, nous définissons deux propriétés nouvelles au lieu de la transitivité et l'euclidienne. Nous exigeons d'abord:

- *inclusion*: Pour tout agent $A \in Agt$ et pour tout monde $w \in W$, $\mathcal{B}_A(w) \supseteq \mathcal{S}_A(w)$.

Ensuite, nous donnons à $\mathcal{S}_A$ les propriétés d'être

- *sérielle*: pour tout monde $w \in W$, il existe $w' \in \mathcal{S}_A(w)$.
- $\mathcal{B}_A$*-transitive*: si $w' \in \mathcal{B}_A(w)$ et $w'' \in \mathcal{S}_A(w')$ alors $w'' \in \mathcal{S}_A(w)$.
- $\mathcal{B}_A$*-euclidiénne*: si $w' \in \mathcal{B}_A(w)$ et $w'' \in \mathcal{S}_A(w)$ alors $w'' \in \mathcal{S}_A(w')$

Nous construisons maintenant une *théorie de preuve*, c'est-à-dire un ensemble de formules de $\mathcal{L}_M$ (les *axiomes*) et de *règles d'inférence* qui génèrent de nouvelles formules à partir des anciennes. Ces formules sont appelés *théorèmes* et nous désignons par $\vdash \varphi$ que φ est un théoreme. Une théorie de preuve est *valide* si les axiomes et les formules générées par les règles d'inférences sont valides (vraies dans tous les mondes). La théorie de preuve est *complète* si toute formule valide peut être générée.

Il est bien connu que la sérialité, la transitivité et la propriété euclidiéenne correspondent aux axiomes suivants que nous utilisons donc pour l'opérateur Bel.
Axiome 1 ($\mathbf{D_B}$). $\mathsf{Bel}_A \varphi \to \neg \mathsf{Bel}_A \neg \varphi$ *(cohérence)*
Axiome 2 ($\mathbf{4_B}$). $\mathsf{Bel}_A \varphi \to \mathsf{Bel}_A \mathsf{Bel}_A \varphi$ *(introspection positive)*
Axiome 3 ($\mathbf{5_B}$). $\neg \mathsf{Bel}_A \varphi \to \mathsf{Bel}_A \neg \mathsf{Bel}_A \varphi$ *(introspection négative)*

Les axiomes suivants décrivent Sup et sa relation avec Bel:
Axiome 4 (BS). $\mathsf{Bel}_A \varphi \to \mathsf{Sup}_A \varphi$

Axiome 5 ($\mathbf{D}_S$). $\mathsf{Sup}_A\varphi \rightarrow \neg\mathsf{Sup}_A\neg\varphi$ *(cohérence)*
Axiome 6 ($\mathbf{4}_{BS}$). $\mathsf{Sup}_A\varphi \rightarrow \mathsf{Bel}_A\mathsf{Sup}_A\varphi$ *(introspection positive)*
Axiome 7 ($\mathbf{5}_{BS}$). $\neg\mathsf{Sup}_A\varphi \rightarrow \mathsf{Bel}_A\neg\mathsf{Sup}_A\varphi$ *(introspection négative)*

Nous allons voir que les propriétés 4 et 5 normales sont valides pour Sup.

Comme déjà mentionné en sections 2.2.1 et 2.2.3, la propriété K ne dépend pas des relations d'accessibilité mais est une propriété naissant de l'utilisation d'une logique modale normale avec sémantique de mondes possibles. Nous avons donc:
Axiome 8 ($\mathbf{K}_B$). $(\mathsf{Bel}_A\varphi \wedge \mathsf{Bel}_A(\varphi \rightarrow \psi)) \rightarrow \mathsf{Bel}_A\psi$
Axiome 9 ($\mathbf{K}_S$). $(\mathsf{Sup}_A\varphi \wedge \mathsf{Sup}_A(\varphi \rightarrow \psi)) \rightarrow \mathsf{Sup}_A\psi$

Les derniers trois axiomes décrivent les opérateurs de groupe GBel, MB et CSup.
Axiome 10. $\mathsf{GBel}_{A,B}\varphi \rightarrow \mathsf{Bel}_A\varphi \wedge \mathsf{Bel}_B\varphi$
Axiome 11 ($\mathbf{FP}_{MB}$). $\mathsf{MB}_{A,B}\varphi \rightarrow \mathsf{GBel}_{A,B}(\varphi \wedge \mathsf{MB}_{A,B}\varphi)$
Axiome 12 ($\mathbf{FP}_{CSup}$). $\mathsf{CSup}_{A,B}\varphi \rightarrow \mathsf{Sup}_A(\varphi \wedge \mathsf{Bel}_B\mathsf{CSup}_{A,B}\varphi)$

Quant aux règles d'inférence, nous donnons des règles usuelles pour la logique du premier ordre et des règles usuelles de nécessitation pour Bel, Sup et Next . Deux autres règles d'inférence décrivent comment inférer MB et CSup:

($\mathbf{R}_{MB}$) De $\vdash \varphi \rightarrow \mathsf{GBel}_{A,B}(\varphi \wedge \psi)$ infère $\vdash \varphi \rightarrow \mathsf{MB}_{A,B}\psi$
($\mathbf{R}_{CSup}$) De $\vdash \varphi \rightarrow \mathsf{Sup}_A(\psi \wedge \mathsf{Bel}_B\varphi)$ infère $\vdash \varphi \rightarrow \mathsf{CSup}_{A,B}\psi$

Ces règles d'inférences sont appelées *règles d'induction* parce qu'elles décrivent des situations où la croyance mutuelle (ou la supposition communicative) est le résultat d'une sorte de preuve par induction, comme utilisée pour l'heuristique de coprésence de Clark et Marshall (1981) (cf. section 3.1).

Halpern et Moses (1992) prouvent la validité et la complétude d'une logique similaire à la nôtre, mais il s'agit d'une logique propositionnelle sans les opérateurs Sup et CSup. Nous ne pouvons pas fournir la preuve ici, mais au moins la *validité* de Sup et CSup nous semble claire.

5.1.5 Propriétés de l'opérateur Sup

Nous présenterons maintenant quelques propriétés de l'opérateur Sup, surtout concernant l'interaction avec Bel.
Proposition 1. *Les formules suivantes sont des théorèmes:*

$$\mathsf{Sup}_A\varphi \rightarrow \mathsf{Sup}_A\mathsf{Sup}_A\varphi$$
$$\neg\mathsf{Sup}_A\varphi \rightarrow \mathsf{Sup}_A\neg\mathsf{Sup}_A\varphi$$

Preuve: Par application de l'axiome 6 (7, respectivement) et l'axiome 4. □

Sup a donc les propriétés d'une logique KD45.
Proposition 2. *Les formules suivantes sont des théorèmes:*

$$\mathsf{Bel}_A\mathsf{Bel}_A\varphi \rightarrow \mathsf{Bel}_A\varphi$$
$$\mathsf{Sup}_A\mathsf{Sup}_A\varphi \rightarrow \mathsf{Sup}_A\varphi$$
$$\mathsf{Bel}_A\neg\mathsf{Bel}_A\varphi \rightarrow \neg\mathsf{Bel}_A\varphi$$
$$\mathsf{Sup}_A\neg\mathsf{Sup}_A\varphi \rightarrow \neg\mathsf{Sup}_A\varphi$$

Preuve: Par application de D et 5 pour les deux premières formules, respectivement D et 4 pour les deux autres. □

Proposition 3. *Les formules suivantes sont des théorèmes:*

$$\begin{aligned} \mathsf{Bel}_A\varphi &\rightarrow \mathsf{Sup}_A\mathsf{Bel}_A\varphi \\ \neg\mathsf{Bel}_A\varphi &\rightarrow \mathsf{Sup}_A\neg\mathsf{Bel}_A\varphi \\ \mathsf{Bel}_A\mathsf{Sup}_A\varphi &\rightarrow \mathsf{Sup}_A\varphi \\ \mathsf{Bel}_A\neg\mathsf{Sup}_A\varphi &\rightarrow \neg\mathsf{Sup}_A\varphi \end{aligned}$$

Preuve: Les deux premières formules sont des conséquences directes des axiomes d'introspection pour Bel et de l'axiome 4. Les deux autres sont des conséquences de l'axiome 4 et de proposition 2. □

Proposition 4. *Les formules suivantes sont des théorèmes:*

$$\begin{aligned} \mathsf{Sup}_A\mathsf{Bel}_A\varphi &\rightarrow \mathsf{Bel}_A\varphi \\ \mathsf{Sup}_A\neg\mathsf{Bel}_A\varphi &\rightarrow \neg\mathsf{Bel}_A\varphi \end{aligned}$$

Preuve: Par application de D_S, BS et 5_B (ou 4_B, respectivement). □

5.2 Notions et règles conversationnelles

5.2.1 Notions mentalistes basées sur des situations partagées

La règle d'inférence (R_{MB}) représente le moyen pour inférer une croyance mutuelle dans notre logique. Cette règle décrit comment une croyance mutuelle «naît» d'une situation particulière, une *situation partagée* comme nous avons dit en section 2.2.2. Le point important de cette règle est donc que pour atteindre une croyance mutuelle un agent n'a pas besoin de vérifier un nombre de croyances infini, mais de la sensation de se trouver dans une situation partagée.

Une idée cruciale du chapitre 4 était qu'il peut y avoir encore d'autres types de situation partagée. Une situation comprenant l'incertitude sur la réussite de la communication a été décrite par la formule

$$\mathsf{Sup}_A compris(B,e,\varphi) \wedge \mathsf{Sup}_A\mathsf{Bel}_B\mathsf{Sup}_A compris(B,e,\varphi) \wedge \\ \mathsf{Sup}_A\mathsf{Bel}_B\mathsf{Sup}_A\mathsf{Sup}_A\mathsf{Bel}_B\mathsf{Sup}_A compris(B,e,\varphi) \wedge \cdots$$

Pour pouvoir décrire de telles formules, nous avons défini l'opérateur CSup qui a la propriété suivante (axiome FP_{CSup}):

$$\mathsf{CSup}_{A,B}\varphi \quad = \quad \mathsf{Sup}_A(\varphi \wedge \mathsf{Bel}_B\mathsf{CSup}_{A,B}\varphi)$$

Intuitivement, on peut expliquer cette formule de la manière suivante: A a une **supposition communicative** s'il *suppose* qu'il se trouve dans une situation X qui lui indiquerait φ et indiquerait à B que A se trouve dans X. Appliqué au cas de la compréhension d'un énoncé cela nous donne: X = B a compris l'énoncé de A. A *suppose* donc que X et que B sache que A suppose que X.

En développant $\mathsf{CSup}_{A,B}\varphi$ nous avons

$$\begin{aligned}\mathsf{CSup}_{A,B}\varphi &= \mathsf{Sup}_A\varphi \wedge \mathsf{Sup}_A\mathsf{Bel}_B\mathsf{Sup}_A\varphi \wedge \\ &\quad \mathsf{Sup}_A\mathsf{Bel}_B\mathsf{Sup}_A\mathsf{Bel}_B\mathsf{Sup}_A\varphi \wedge \cdots\end{aligned}$$

Mais, ici aussi, CSup ne nécessite pas la vérification de toutes les suppositions individuelles. La règle d'inférence (R_{CSup}) décrit le type de situation partagée où un agent a une supposition communicative.

Nous introduisons une troisième notion basée sur l'approche de point fixe. Il s'agit de l'**intention communicative** telle qu'elle est utilisée par Grice (1957) dans la description de la signification non-naturelle (cf. section 2.1.2). Selon Grice, un énoncé exprime non naturellement deux intentions du locuteur: l'intention d'obtenir un certain effet et l'intention que la première intention soit reconnu par l'auditeur. Schiffer (1972) a argumenté que pour décrire l'état mental d'avoir une intention communicative, il faut ajouter une intention que la deuxième intention soit reconnue et une intention que la troisième intention soit reconnue, etc. De plus, l'intention communicative va aussi comprendre l'intention du locuteur de savoir que son intention a été reconnue, et que l'auditeur sait que le locuteur sait que l'auditeur a reconnu l'intention, etc. Une définition de l'intention communicative doit alors comprendre à la fois un emboîtement d'intentions et un emboîtements de croyances. Dans l'esprit des deux autres attitudes en termes de l'approche du point fixe, nous proposons la relation suivante entre l'intention communicative, l'intention-que et la croyance mutuelle:

$$\mathsf{CIntTh}_{A,B}\varphi = \mathsf{IntTh}_A\mathsf{MB}_{A,B}(\varphi \wedge \mathsf{CIntTh}_{A,B}\varphi)$$

Notons que l'intention communicative (comme la supposition communicative) présente une attitude mentale *unilatérale*, c'est-à-dire qu'il s'agit de l'intention d'un seul agent, bien que cette intention comprenne l'établissement d'une croyance mutuelle entre deux agents. Les différents niveaux d'intentions gricéennes deviennent visibles dans la formule

$$\begin{aligned}\mathsf{CIntTh}_{A,B}\varphi &= \mathsf{IntTh}_A\mathsf{MB}_{A,B}\varphi \wedge \\ &\quad \mathsf{IntTh}_A\mathsf{MB}_{A,B}\mathsf{IntTh}_A\mathsf{MB}_{A,B}\mathsf{Bel}_B\varphi \wedge \\ &\quad \mathsf{IntTh}_A\mathsf{MB}_{A,B}\mathsf{IntTh}_A\mathsf{MB}_{A,B}\mathsf{IntTh}_A\mathsf{MB}_{A,B}\mathsf{Bel}_B\varphi \wedge \cdots\end{aligned}$$

Le premier niveau décrit l'*effet* que veut obtenir le locuteur, les niveaux plus emboîtés décrivent l'intention de faire reconnaître l'intention communicative. Nous allons voir que le *grounding* peut être naturellement décrit comme la satisfaction de l'intention communicative à partir du «deuxième» niveau d'emboîtement. La satisfaction en plus du premier niveau correspond à l'*adoption* de l'intention et sera, par exemple, l'acceptation d'une affirmation, la réponse à une demande, etc. [4]

4. Comme nous n'avons pas donné de sémantique à l'intention-que, celle de CIntTh n'est pas définie non plus. Mais au moins une contrainte problématique est posée par la dernière formule: l'extensionalisation présentée nécessite la propriété $\mathsf{IntTh}(\varphi \wedge \psi) \rightarrow \mathsf{IntTh}\varphi \wedge \mathsf{IntTh}\psi$, ce qui n'est probablement pas souhaitable pour l'intention. Les intentions n'apparaissant presque pas directement dans ce chapitre, nous négligeons ce problème ici.

5.2.2 Notions importantes de la théorie

Nous exprimons maintenant quelques notions informelles de la théorie du «grounding» dans la logique. Pour la plupart d'entre elles, des abbréviations sont introduites. Il faut soigneusement distinguer entre les symboles du langage $\mathcal{L}_M$ et les abbréviations qui ne sont là que pour condenser des formules complexes.

done() sera un symbole de relation de $\mathcal{L}_M$. Il remplace un traitement plus sophistiqué d'action dans notre logique. $\mathit{done}\,(A, e)$ représente le fait que entre le moment précédent et le moment actuel l'agent A a réalisé une action e (ici uniquement: l'acte de langage e). Nous pouvons donc exprimer le fait que A croit que B a réalisé un acte de langage e_B par la formule $\mathsf{Bel}_A \mathit{done}\,(B, e_B)$.

Nous avons déjà défini ce que veut dire avoir une *intention communicative*, mais le lien entre un acte de langage et cette intention n'est pas encore spécifié. Nous définissons donc la **signification non-naturelle** (cf. section 2.1.2):

Définition 5.1 (signification non-naturelle).

$$\mathit{sig\text{-}nn}(A, e_A, \varphi) \quad =_{abbr} \quad \mathit{done}\,(A, e_A) \rightarrow \mathsf{CIntTh}_{A,B}\varphi$$

$\mathit{sig\text{-}nn}(A, e_A, \varphi)$ n'est qu'une abbréviation, c'est-à-dire que dans toute formule, elle est remplacée par l'implication $\mathit{done}\,(A, e_A) \rightarrow \mathsf{CIntTh}_A\varphi$. De toute façon, *sig-nn* ne peut pas être un symbole de relation de $\mathcal{L}_M$, car un de ses arguments n'est pas un objet, mais la formule φ. Notons aussi que φ n'est pas nécessairement le *seul* fait que voulait exprimer le locuteur avec cet énoncé. La définition est telle que $\mathit{done}\,(A, e_A)$ peut impliquer plusieures intentions qu'avait le locuteur.

La relation entre un énoncé et sa signification non-naturelle reste (exprès) très vague selon cette définition. Intuitivement, elle exprime qu'un locuteur a toujours une *raison* pour produire un énoncé, son *intention communicative* sous-jacente. Ce que *signifie* donc cet énoncé c'est ce que *voulait exprimer* le locuteur. Et ceci peut être différent selon la situation. La signification non naturelle décrit donc la notion de *contexte* en termes du lien entre un énoncé et l'intention qu'il exprime.

Cette définition ne décrit pas *comment* le locuteur peut trouver un énoncé adéquat à son intention (génération) ou comment l'auditeur peut reconnaître une telle intention (interprétation), mais décrit seulement le *résultat* de ces deux processus. La définition reste donc purement *descriptive*. Mais ce qui sera important dans ce chapitre, c'est que c'est la signification non-naturelle que le locuteur *veut faire reconnaître*. Quand il dit «oui», par exemple, il n'a pas seulement l'intention de faire reconnaître la valeur sémantique de l'énoncé, mais une intention communicative plus large, comme «oui, je vais faire la vaisselle». Et c'est sur la signification non-naturelle qu'il veut établir le «grounding»; il veut être certain que l'autre sait ce qu'il *a voulu exprimer* avec un tel énoncé.

Nous ne traitons pas le processus d'interprétation même dans ce chapitre, mais commençons au point ou l'auditeur a attribué une signification à un énoncé. Il peut en être certain ou seulement supposer qu'il a trouvé la bonne signification. Il peut aussi échouer à donner une signification à l'énoncé. Nous désignons ce dernier cas par le symbole de relation $\mathit{incompris}()$. Le résultat de l'interprétation d'un énoncé est donc de la forme $\mathsf{Bel}_B \mathit{sig\text{-}nn}(A, e_A, \varphi)$ ou $\mathsf{Sup}_B \mathit{sig\text{-}nn}(A, e_A, \varphi)$ ou alors $\mathsf{Bel}_B \mathit{incompris}(B, e_A)$.

Contrairement à l'incompréhension, nous pouvons décrire la **compréhension** directement en termes de la signification d'un énoncé. D'un point de vue absolu (qui n'existe jamais en réalité) un agent a *compris* un énoncé s'il lui a attribué une signification et si celle-ci est correcte (c'est-à-dire que locuteur voulait en fait exprimer l'intention communicative qu'a trouvé l'auditeur). Cette définition est donc similaire à la définition de la *connaissance* comme croyance correcte ($\mathsf{Bel}_X \varphi \wedge \varphi$).

Définition 5.2 (compréhension).

$$compris(B, e_A, \varphi) =_{abbr} sig\text{-}nn(A, e_A, \varphi) \ \wedge \ \mathsf{Sup}_B sig\text{-}nn(A, e_A, \varphi)$$

Attention: *compris*() est un méta-symbole, une abbréviation qui remplace la conjonction décrite ici. *incompris*() cependant est un symbole de relation qui ne peut pas être réduit à d'autres formules.

Cette définition de la compréhension peut sembler trop faible en première vue, parce qu'elle éxige seulement que l'auditeur ait une *supposition* correcte. Mais le point est que la compréhension dépend de l'interprétation qu'a trouvée l'agent, mais non de sa certitude de cette interprétation. Le résultat de l'interprétation va être la signification *la plus probable* pour lui, celle sur laquelle il a au moins une supposition. Si cette supposition est correcte, il a compris même sans le savoir. Ce qui, par contre, dépend *largement* de la certitude de son interprétation, c'est s'il *croit* lui-même d'avoir compris. En d'autres mots: un agent qui n'est pas certain de son interprétation, n'est pas certain qu'il a compris non plus. Le lemme suivant résulte directement du développement de *compris*() et des propriétés de Bel et Sup.

Lemme 5.

$$\begin{gathered}
\mathsf{Bel}_B compris(B, e_A, \varphi) \ \leftrightarrow \ \mathsf{Bel}_B sig\text{-}nn(A, e_A, \varphi) \\
\mathsf{Sup}_B sig\text{-}nn(A, e_A, \varphi) \ \rightarrow \ \mathsf{Sup}_B compris(B, e_A, \varphi) \\
\mathsf{Sup}_B sig\text{-}nn(A, e_A, \varphi) \wedge \mathsf{Bel}_B compris(B, e_A, \varphi) \ \rightarrow \ \mathsf{Bel}_B sig\text{-}nn(A, e_A, \varphi) \\
\mathsf{Bel}_A compris(B, e_A, \varphi) \ \rightarrow \ \mathsf{Bel}_A \mathsf{Sup}_B sig\text{-}nn(A, e_A, \varphi)
\end{gathered}$$

Un résultat intéressant de ce lemme est que d'autres agents peuvent croire qu'un agent a compris un énoncé avant qu'il ne le croit lui-même. Ce résultat va nous aider à expliquer le «grounding» dans le cas d'une demande de confirmation.

Un acte de langage est **grounded** s'il est mutuellement connu que l'auditeur a compris la signification de cet acte[5].

Définition 5.3 (grounded).

$$grounded(A, B, e_A, \varphi) \quad =_{abbr} \quad \mathsf{MB}_{A,B} compris(B, e_A, \varphi)$$

grounded décrit donc une vraie croyance mutuelle sur la signification d'un énoncé. Il faut bien voir que le **critère de grounding** est une notion plus faible:

Définition 5.4 (critère de grounding). *Pour deux agents A et B, deux formules φ et ψ,*

5. Dans la défintion de Clark et Schaefer (1989), la compréhension n'est exigée que jusqu'à un «degré suffisant pour l'objectif courant». Notre définition permettrait de tenir compte de cet aspect, parce que nous n'exigeons nulle part que le locuteur n'attribue qu'une seule intention communicative à un acte. On pourrait donc définir le «grounding» par rapport à celle des intentions communicatives qui est «suffisante pour l'objectif courant» du locuteur. Nous négligeons ce point ici.

et un énoncé e_A de A nous disons que le critère de grounding *est atteint si*

$$\mathsf{Bel}_A grounded(A, B, e_A, \varphi) \ \wedge$$
$$\mathsf{Bel}_B grounded(A, B, e_A, \psi)$$

Si $\varphi = \psi$, une vraie croyance mutuelle est atteinte. Mais si $\varphi \neq \psi$, les agents vont *croire* qu'ils ont attribué la même signification à l'acte et qu'ils ont atteint la croyance mutuelle sur cette signification, mais ce n'est pas le cas. Pourtant, comme pour tous les deux le «grounding» est atteint, ils vont continuer leur conversation et ne se rendre compte du malentendu que plus tard (ou peut-être jamais).

Enfin, nous définissons la notion d'**attente**. Un agent A s'attend à un énoncé de B concernant φ, si A croit que cet énoncé va soit signifier φ soit $\neg\varphi$.

Définition 5.5.

$$attente(A, B, \varphi) \quad =_{abbr} \quad \mathsf{Bel}_A \mathsf{Next}\ \exists e_B[done(B, e_B) \wedge (sig\text{-}nn(B, e_B, \varphi) \vee sig\text{-}nn(B, e_B, \neg\varphi))]$$

5.2.3 Connaissances conversationnelles

Ici, nous donnons quelques règles de conversation censées d'être connues de tout participant à un dialogue. Ce sont des règles élémentaires qui expriment des connaissances générales sur le déroulement d'un dialogue, notamment un dialogue où les participants sont collaboratifs et essayent de s'assurer mutuellement de leur compréhension. Ce que les règles ne décrivent donc pas, ce sont les cas des agents non coopératifs ou inattentifs. Ces règles décrivant des faits et des conventions connus des utilisateurs de la langue, chaque règle est décrite comme une croyance mutuelle, c'est-à-dire qu'elle peut être utilisée *par les agents eux-mêmes* en raisonnant sur la situation de dialogue.

La première règle décrit que c'est le locuteur qui détermine la signification non-naturelle d'un énoncé.

Règle 1.

$$\mathsf{MB}[sig\text{-}nn(A, e_A, \varphi) \ \leftrightarrow \ \mathsf{Bel}_A sig\text{-}nn(A, e_A, \varphi)]$$

La relation entre *compris*() et *incompris*() est telle que

Règle 2.

$$\mathsf{MB}[sig\text{-}nn(A, e_A, \varphi) \ \rightarrow \ (compris(B, e_A, \varphi) \leftrightarrow \neg incompris(B, e_A))]$$

Les règles suivantes décrivent (d'une manière très simplifiée) quelques aspects de la perception et de l'énonciation d'un acte de langage.

Règle 3.

$$\mathsf{MB}[done(A, e_A) \leftrightarrow \mathsf{Bel}_A done(A, e_A) \ \wedge \ \mathsf{Bel}_B done(A, e_A)]$$

Règle 4.

$$\mathsf{MB}[done(A, e_A) \wedge sig\text{-}nn(A, e_A, \varphi) \rightarrow \mathsf{Sup}_A \mathsf{Bel}_B sig\text{-}nn(A, e_A, \varphi)]$$

Ces dernières règles expriment l'hypothèse que les participants (y compris le locuteur même) *remarquent* toujours l'énonciation d'un acte de langage, mais que le locuteur ne peut que *supposer* que la signification de son énoncé a été reconnu.

Les règles suivantes décrivent la coopérativité de l'auditeur qui exprimera avec sa réponse s'il croit avoir compris ou pas. Nous avons déjà mentionné que notre définition de la signification non-naturelle d'un énoncé n'exige pas qu'un locuteur n'ait qu'une seule intention communicative qu'il veut exprimer avec un énoncé. Ces règles n'exigent donc pas non plus que l'auditeur ne donne pas de nouvelles informations dans sa réponse. Mais *au moins* sa réponse va signaler son état de compréhension.

Règle 5.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last}\ \mathit{done}(A,e_A)\wedge \mathit{incompris}(B,e_A)\\&\quad\rightarrow\ \mathsf{Next}\ \exists e_B[\mathit{done}(B,e_B)\wedge \mathit{sig\text{-}nn}(B,e_B,\mathit{incompris}(B,e_A))]]\end{aligned}$$

Règle 6.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last}\ \mathit{done}(A,e_A)\wedge \mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)\\&\quad\rightarrow\ \mathsf{Next}\ \exists e_B[\mathit{done}(B,e_B)\wedge \mathit{sig\text{-}nn}(B,e_B,\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi))]]\end{aligned}$$

Règle 7.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last}\ \mathit{done}(A,e_A)\wedge \mathsf{Sup}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)\wedge\neg\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)\\&\quad\rightarrow\ \mathsf{Next}\ \exists e_B[\mathit{done}(B,e_B)\wedge \mathit{sig\text{-}nn}[B,e_B,(\mathsf{Sup}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)\wedge\neg\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi))]]]\end{aligned}$$

Les règles suivantes décrivent la naissance de l'attente et sa propriété particulière: quand un agent s'attend à un énoncé concernant φ, c'est certain qu'il comprend cet énoncé.

Règle 8.

$$\mathsf{MB}[\mathsf{CSup}_{A,B}\varphi\rightarrow \mathit{attente}(A,B,\varphi)]$$

Règle 9.

$$\mathsf{MB}[\mathsf{Last}\ \mathit{attente}(A,B,\varphi)\wedge \mathit{done}(B,e_B)\wedge \mathit{sig\text{-}nn}(B,e_B,\varphi)\rightarrow \mathsf{Bel}_A\mathit{sig\text{-}nn}(B,e_B,\varphi)]$$

Deux cas particuliers sont décrits par les deux règles suivantes: quand un agent s'attend à une réponse à son énoncé et que l'autre exprime qu'il n'a pas compris l'énoncé, ceci est censé d'être toujours reconnu. Et si un agent A attend que B exprime une croyance sur φ, A va aussi comprendre si B n'exprime qu'une supposition sur φ.

Règle 10.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last}\ \mathit{done}(A,e_A)\wedge \mathit{done}(B,e_B)\wedge\ \mathit{sig\text{-}nn}(B,e_B,\mathit{incompris}(B,e_A))\\&\quad\rightarrow\ \mathsf{Bel}_A\mathit{incompris}(B,e_A)]\end{aligned}$$

Règle 11.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last}\ \mathit{attente}(A,B,\mathsf{Bel}_B\varphi)\wedge \mathit{done}(B,e_B)\wedge\ \mathit{sig\text{-}nn}(B,e_B,\mathsf{Sup}_B\varphi\wedge\neg\mathsf{Bel}_B\varphi)\\&\quad\rightarrow\ \mathsf{Bel}_A(\mathsf{Sup}_B\varphi\wedge\neg\mathsf{Bel}_B\varphi))]\end{aligned}$$

Une dernière règle concerne le traitement d'un autre type de réponse. Un agent A peut exprimer que *l'autre* (B) n'avait pas compris le dernier énoncé de A. (Ceci correspond à ce qui est appelée *allo-feedback négatif* par Bunt (1995).) Nous considérons qu'un tel acte est certain d'être compris si B s'attend encore à une information de A à propos du premier énoncé de celui-ci.

Règle 12.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last\ Last}\ \mathit{done}\,(A,e_A) \wedge \mathit{sig\text{-}nn}(A,e_A,\psi) \wedge \mathit{attente}(B,A,\mathit{sig\text{-}nn}(A,e_A,\varphi))\\ &\quad \wedge \mathit{done}\,(A,e'_A) \wedge\ \mathit{sig\text{-}nn}(A,e'_A,\neg \mathit{compris}(B,e_A,\psi))\\ &\quad \rightarrow\ \mathsf{Bel}_B \mathit{incompris}(B,e_A)]\end{aligned}$$

5.3 Établir des croyances mutuelles

Après ces longs travaux préparatoires, nous pouvons maintenant prouver assez facilement les effets d'un acte de langage et de séquences d'actes de langage. Nous allons montrer comment le critère de «grounding» individuel de chaque agent peut être atteint dans plusieurs cas, y compris des demandes de confirmation. Pour des raisons d'espace, nous ne donnons que des esquisses de preuves, mais essayons quand-même de mentionner tous les éléments contribuant à chaque preuve.

5.3.1 Effets d'un acte de langage

Les particularités de la situation communicative entraînent des effets immédiats de la réalisation d'un acte de langage.

Lemme 6.

$$\mathsf{MB}[\mathit{done}\,(A,e_A)\ \rightarrow\ \mathsf{MB}_{A,B}\mathit{done}\,(A,e_A)]$$

Preuve: Par application de la règle d'inférence (R_{MB}) à la règle conversationnelle 3 (avec $\varphi = \psi = \mathit{done}\,(A,e_A)$) □

Lemme 7.

$$\mathsf{MB}[\mathit{done}\,(A,e_A) \wedge \mathit{sig\text{-}nn}(A,e_A,\varphi) \rightarrow \mathsf{CSup}_A \mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)]$$

Preuve: Soit $\varphi = \mathit{done}\,(A,e_A) \wedge \mathit{sig\text{-}nn}(A,e_A,\varphi)$. À cause des règles 3 et 4, nous avons $\varphi \rightarrow \mathsf{Bel}_A \mathsf{Bel}_B\, \mathit{done}\,(A,e_A)$ et $\varphi \rightarrow \mathsf{Sup}_A \mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A)$. Soit $\psi = \mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A)$. À cause de la règle 4, nous avons $\varphi \rightarrow \mathsf{Sup}_A(\psi \wedge \mathsf{Bel}_B \varphi)$. Le lemme résulte ensuite de l'application de la règle d'inférence (R_{CSup}). □

Comme conséquence immédiate, nous pouvons maintenant décrire l'état mental du locuteur après l'acte de langage, celui de l'auditeur étant certain d'avoir compris ou supposant seulement avoir compris.

Théorème 8. *Les formules suivantes sont valides*

$$\begin{aligned}\mathsf{Bel}_A \mathit{done}\,(A,e_A)\ \wedge\ \mathsf{Bel}_A \mathit{sig\text{-}nn}(A,e_A,\varphi)\ &\rightarrow\ \mathsf{CSup}_{A,B}\mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)\\ \mathsf{Bel}_B \mathit{done}\,(A,e_A)\ \wedge\ \mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)\ &\rightarrow\ \mathsf{Bel}_B \mathsf{CSup}_{A,B}\mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)\\ \mathsf{Bel}_B \mathit{done}\,(A,e_A)\ \wedge\ \mathsf{Sup}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)\ &\rightarrow\ \mathsf{Sup}_B \mathsf{CSup}_{A,B}\mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)\end{aligned}$$

A: Je voudrais parler à Corinne Savol.
B: Je suis désolé, Corinne Davol ne travaille pas ici.

FIG. 5.1 – *Dialogue résultant dans une «fausse» croyance mutuelle unilatérale*

5.3.2 «Grounding» et certitude sur la compréhension

Nous pouvons maintenant prouver un premier théorème important: si l'auditeur B est certain d'avoir bien compris, il est aussi certain qu'il peut exprimer cela sans ambiguïté et ainsi établir le «grounding».

Théorème 9.

$$\mathsf{MB}[\mathsf{Last}\ \mathit{done}\,(A,e_A) \wedge \mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi) \wedge \mathit{done}\,(B,e_B) \\ \rightarrow\ \mathsf{Bel}_B\mathit{grounded}\,(A,B,e_A,\varphi)]$$

Preuve: Supposons $\mathsf{Last}\ \mathit{done}\,(A,e_A)\ \wedge\ \mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)\ \wedge\ \mathit{done}\,(B,e_B)$. En raison du théoreme 8, B croit $\mathsf{Bel}_B\mathsf{CSup}_{A,B}\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$. La règle 8 entraîne $\mathsf{Bel}_B\mathit{attente}(A,B,\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi))$. Étant coopératif (règle 6), B va exprimer que $\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$ et en raison de l'attente de A, B va être certain d'être compris. B croit alors $\mathsf{Bel}_B\mathsf{Bel}_A\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$. Selon B, les agents se trouvent donc dans une situation qui indique à tous les deux que $\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$ et qu'ils se trouvent dans cette situation. Par application de (R_{MB}), on atteint $\mathsf{Bel}_B\mathsf{MB}_{A,B}\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$. Comme $\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$ entraîne aussi $\mathsf{Bel}_B\mathsf{Bel}_A\mathit{sig\text{-}nn}(A,e_A,\varphi)$, cela se réduit à $\mathsf{Bel}_B\mathsf{MB}_{A,B}\mathit{sig\text{-}nn}(A,e_A,\varphi)$ □

Bien entendu, cela ne décrit ni que B a vraiment compris ni qu'une vraie croyance mutuelle a été atteinte. Le théorème décrit seulement que B étant certain d'avoir compris, va aussi croire qu'il va atteindre le critère de «grounding» par sa réponse. Par exemple, le théorème 9 prédit aussi qu'après un dialogue comme celui en figure 5.1, B va croire que le «grounding» sur sa compréhension de l'énoncé de A est atteint. Il s'agit donc d'une croyance mutuelle unilatérale qui est «fausse» en ce que les croyances attribuées à A par B ne sont pas vraies. Nous traiterons ces cas de mauvaises compréhensions en section 5.3.4.

En ce qui concerne le locuteur A, nous pouvons d'abord constater qu'après la réponse de B, il va ou bien croire que celui avait compris ou bien qu'il n'avait pas compris.

Théorème 10.

$$\mathsf{MB}[\mathsf{Last}\ \mathit{done}\,(A,e_A) \wedge \mathit{sig\text{-}nn}(A,e_A,\varphi) \wedge \mathit{done}\,(B,e_B) \rightarrow \mathsf{KIF}_A\mathit{compris}(B,e_A,\varphi)]$$

Preuve: Comme montré ci-dessus, $\mathsf{Last}\ \mathit{done}\,(A,e_A)$ implique $\mathsf{Last}\ \mathit{attente}(A,B,\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi))$. Si B n'a pas compris, en raison des règles 5 et 10, $\mathit{done}\,(B,e_B) \rightarrow \mathsf{Bel}_A\mathit{incompris}(B,e_A)$. La règle 2 entraîne alors $\mathsf{Bel}_A\neg\mathit{compris}(B,e_A,\varphi)$. Si B a compris (en étant certain ou pas) il va au moins exprimer $\mathsf{Sup}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$ (règle 6 ou 7). Comme A s'attend à une telle réponse, la règle 11 entraîne $\mathsf{Bel}_A\mathsf{Sup}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$, ce qui correspond à $\mathsf{Bel}_A\mathit{compris}(B,e_A,\varphi)$. Si B n'avait pas compris, il y a deux possibilités: A n'arrive pas à interpréter e_B ou il y réussit. Si $\mathit{incompris}(A,e_B)$, à cause des règles 2 et 11 (contreposition) il n'est pas possible que $\mathit{compris}(B,e_A,\varphi)$. Et si A arrive à interpréter l'énoncé et conclut que B n'avait pas attribué la signification $\mathit{sig\text{-}nn}(A,e_A,\varphi)$ à l'énoncé, il sait aussi que B n'a pas compris. Dans tous les cas, nous avons donc $\mathsf{Bel}_A\neg\mathit{compris}(B,e_A,\varphi)$ ou $\mathsf{Bel}_A\mathit{compris}(B,e_A,\varphi)$. □

Le théorème suivant décrit que A atteint son critère de «grounding» individuel si B exprime qu'il est certain d'avoir compris et si A croit que ceci est correct.
Théorème 11.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last}\ \mathit{done}\,(A,e_A) \wedge \mathit{sig\text{-}nn}(A,e_A,\varphi) \wedge \mathit{done}\,(B,e_B) \wedge \mathit{sig\text{-}nn}[B,e_B,\mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)]]\\ &\quad \mathsf{Bel}_A \mathit{grounded}\,(A,B,e_A,\varphi)\end{aligned}$$

Preuve: En raison de la règle 9 A va savoir que $\mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)$ et comme A sait en fait que $\mathit{sig\text{-}nn}(A,e_A,\varphi)$, nous avons $\mathsf{Bel}_A \mathit{compris}(B,e_A,\varphi)$. Le théorème 9 entraîne alors $\mathsf{Bel}_A\mathsf{Bel}_B\mathsf{MB}_{A,B}\mathit{grounded}\,(A,B,e_A,\varphi)$. Une croyance mutuelle peut être «partitionnée» ainsi: $\mathsf{MB}_{A,B}\alpha = \mathsf{Bel}_A\alpha \wedge \mathsf{Bel}_B\mathsf{MB}_{A,B}\varphi \wedge \mathsf{Bel}_A\mathsf{Bel}_B\mathsf{MB}_{A,B}\varphi$. Pour $\alpha = \mathit{sig\text{-}nn}(A,e_A,\varphi)$, ceci nous donne $\mathsf{Bel}_A\mathsf{MB}_{A,B}\mathit{sig\text{-}nn}(A,e_A,\varphi)$. □

5.3.3 «Grounding» et incertitude de la compréhension

Si l'auditeur n'est pas sûr d'avoir compris, mais a une supposition sur la signification de l'énoncé, il va demander une confirmation du locuteur. Lorsque B exprime ainsi son incertitude sur sa propre interprétation, le «grounding» immédiat est devenu impossible: A doit au moins informer B qu'il avait en fait compris. Si B n'avait pas compris, A doit répéter ou reformuler son premier énoncé. Le théorème suivant décrit l'état après que B ait exprimé son incertitude.
Théorème 12.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last}\ \mathit{done}\,(A,e_A) \wedge \mathsf{Sup}_B \mathit{sig\text{-}nn}(A,e_A,\varphi) \wedge \neg\mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi) \wedge \mathit{done}\,(B,e_B)\\ &\quad \rightarrow\ \mathsf{CSup}_{B,A}\mathit{sig\text{-}nn}(A,e_A,\varphi)]\end{aligned}$$

Preuve: Si $\mathit{sig\text{-}nn}(A,e_A,\varphi)$, A se trouvait dans l'état d'attente *attente*(A,B,$\mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)$) et B pourrait être certain que A comprend son énoncé (règle 11). Mais comme B n'en est pas sûr, il se trouve dans une situation de supposition communicative. En raison de sa coopérativité (règle 7), sa réponse exprime $\alpha = \mathsf{Sup}_B \mathit{sig\text{-}nn}(A,e_A,\varphi) \wedge \neg\mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)$. En raison du théorème 8, ceci entraîne $\mathsf{CSup}_B\mathsf{Bel}_A\alpha$. Mais cela comprend $\mathsf{CSup}_B\mathsf{Bel}_A\mathsf{Sup}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$. Comme $\mathsf{Sup}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$ est vrai aussi, ces deux formules forment une «partition» de $\mathsf{Sup}_B\mathit{sig\text{-}nn}(A,e_A,\varphi)$ (cf. l'axiome (FP$_{CSup}$). □

Les rôles ont donc changés après que B ait exprimé son incertitude: A sait si B a compris (théorème 10) et B ne le sait pas. Mais nous pouvons prouver que le prochain énoncé de A va établir le «grounding», si B avait vraiment compris.
Théorème 13.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last}\ \mathsf{Last}\ \mathit{done}\,(A,e_A) \wedge \mathsf{Last}\ \mathit{done}\,(B,e_B) \wedge\\ &\quad \mathit{compris}(B,e_A,\varphi) \wedge \neg\mathsf{Bel}_B\mathit{sig\text{-}nn}(A,e_A,\varphi) \wedge \mathit{done}\,(A,e'_A)\\ &\quad \rightarrow\ \mathit{grounded}\,(A,B,e_A,\varphi)]\end{aligned}$$

Preuve: En raison de la règle de coopérativité 7, B a exprimé son incertitude. Comme prouvé pour le théorème 11, A sait maintenant que B a compris, mais il sait aussi (à cause du théorème 12) $\mathsf{Bel}_A\mathsf{CSup}_{B,A}\mathit{sig\text{-}nn}(A,e_A,\varphi)$, donc que B n'est pas certain d'avoir compris. La règle 8 entraîne alors $\mathsf{Bel}_A\mathit{attente}(B,A,\mathit{sig\text{-}nn}(A,e_A,\varphi))$. A peut donc être certain que B va comprendre sa réponse, un «acknowledgement» que B avait correctement interprété e_A. Le reste de la preuve est donc équivalente à celle du théorème 11. □

A:	Je voudrais parler à Corinne Savol.
B:	Je suis désolé, elle ne travaille pas ici.

FIG. 5.2 – *Une mauvaise compréhension presque indétectable*

5.3.4 Incompréhension et mauvaise compréhension

Il nous reste à décrire deux cas d'échec de la communication: celui de l'auditeur ne comprenant pas un énoncé et celui de l'auditeur croyant faussement avoir compris.

Théorème 14.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last}\ \mathit{done}(A,e_A)\wedge \mathit{incompris}(B,e_A)\wedge \mathit{done}(B,e_B)\\&\quad\rightarrow\ \mathit{grounded}(A,B,\mathit{incompris}(B,e_A))]\end{aligned}$$

Preuve: En raison de la règle 10, A va croire $\mathsf{Bel}_A \mathit{incompris}(B,e_A)$. Le reste de la preuve correspond à celles des théorèmes 9 et 11. Les deux agents vont savoir mutuellement que A va comprendre. □

Un cas plus intéressant est celui d'une demande de confirmation, c'est-à-dire B n'est pas sûr d'avoir compris et, de fait, il n'a vraiment pas compris.

Théorème 15.

$$\begin{aligned}&\mathsf{MB}[\mathsf{Last}\ \mathsf{Last}\ \mathit{done}(A,e_A)\wedge \mathit{sig\text{-}nn}(A,e_A,\psi)\wedge \mathsf{Last}\ \mathit{done}(B,e_B)\wedge\\&\quad \mathsf{Sup}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)\wedge\neg\mathsf{Bel}_B \mathit{sig\text{-}nn}(A,e_A,\varphi)\wedge \mathit{done}(A,e'_A)\\&\quad\rightarrow\ \mathit{grounded}(A,B,\mathit{incompris}(B,e_A))]\end{aligned}$$

Preuve: En raison de la règle de coopérativité 7, B a exprimé son incertitude. Comme $\mathit{attente}(A,B,\mathit{sig\text{-}nn}(A,e_A,\psi))$, la règle 11 entraine $\mathsf{Bel}_A\neg \mathit{sig\text{-}nn}(A,e_A,\psi)$ et donc $\mathsf{Bel}_A\neg \mathit{compris}(B,e_A,\psi)$. Par la règle 2, cela entraîne $\mathsf{Bel}_A \mathit{incompris}(B,e_A)$. A va donc fournir un «acknowledgement» négatif (*allo-feedback négatif* chez Bunt (1995)). Mais comme il est mutuellement connu que B se trouve dans un état d'attente (parce qu'il n'était pas certain d'avoir compris) A peut être certain (règle 12) que B va comprendre et abandonner sa mauvaise interprétation. La règle 12 étant mutuellement connu, l'effet de l'acknowledgement négatif devient une croyance mutuelle aussi. □

En figure 5.1 nous avons donné l'exemple d'une mauvaise compréhension. Il s'agit d'un cas plutôt heureux, car l'auditeur répète ce qu'il a compris et A peut éventuellement détecter l'erreur. La figure 5.2 montre un exemple où une détection est moins évidente. Les différences sont tellement fines qu'il est très difficile de distinguer entre ces deux exemples. Nous n'abordons donc pas le thème de la détection de mauvaise compréhension à part les demandes de confirmation. Il est cependant important de voir que nos règles de conversation *permettent* de traiter de telles mauvaises compréhensions. Par contre, une simplification que nous avons faite c'est que nous ignorons un auditeur «maladroit» qui *a* compris, mais répond de telle manière que le locuteur ne le croit pas.

Ce chapitre à présenté quelques éléments d'une nouvelle logique modale et son utilisation pour formaliser la théorie du «grounding» développée dans le chapitre 4. Une telle logique est un outil très puissant pour la description des processus mentaux d'un agent. Pourtant, c'est aussi un outil lourd; décider de la validité d'une formule (par exemple, par un démonstrateur de théorème) est computationnellement inefficace (cf. Halpern et Moses (1992) pour quelques preuves de complexité de telles logiques). Notre approche sera donc de «compiler» les résultats de ce chapitre dans un modèle de gestion de dialogue. Ce modèle est présenté dans le chapitre suivant.

Chapitre 6

Gestion de dialogue

Nous avons argumenté à plusieurs endroits dans ce travail que des processus de raisonnement peuvent être résumés dans des règles «conventionnelles», c'est-à-dire que les résultats du raisonnement sont tellement stables qu'ils sont devenus des conventions entre les utilisateurs d'une langue. Dans le chapitre précédent nous avons montré comment de telles «règles conversationnelles» peuvent entrainer d'autres connaissances conventionnelles, sur les effets d'un acte de langage et même sur des séquences d'actes de langage. Il est donc inutile pour un système de dialogue d'utiliser des procédures de raisonnement compliqués afin d'obtenir des résultats qui suivront toujours le même schéma. Les «schémas» que nous avons présentés dans le chapitre précédent vont former la base d'un modèle de gestion de dialogue qui est décrit dans ce chapitre. La section 6.1 décrit comment croyances, croyances mutuelles et suppositions communicatives peuvent être représentées d'une manière efficace et souple. La section 6.2 présente comment des actes de langage peuvent agir sur les structures de représentation, comment l'interprétation et la génération d'énoncés sont influencées par le contexte, notamment par l'état du processus de «grounding».

Une simplification importante sera faite dans ce chapitre: nous supposerons que les agents sont *sincères*. Jusqu'à maintenant, nous avons décrit l'état mental du locuteur d'un énoncé en termes de son *intention communicative*. Cette notion était définie afin de représenter seulement qu'un agent *veut communiquer* un fait à un autre. Il n'était pas exigé que le locuteur croit ce qu'il affirme. En demandant la sincérité nous enlevons la possibilité de mensonges, mais aussi d'ironie et d'autres nuances d'une langue (ce qui nous expose aux mêmes critiques que Perrault (1990) a faites au modèle de Cohen et Levesque (1990c)). Mais nous gagnons un formalisme de représentation beaucoup plus simple pour les croyances et les actes de langage. Si nous supposons qu'un agent sincère croit ce qu'il énonce[1], nous pouvons distinguer entre la *compréhension* et l'*acceptation* de son contenu d'une manière très simple. La compréhension, c'est-à-dire la reconnaissance de l'intention communicative, remonte à la reconnaissance des croyances du locuteur exprimées avec l'énoncé, alors que l'acceptation revient à l'adoption des mêmes croyances (ou bien des intentions et désirs qui font l'objet de la croyance). Quand le locuteur A dit «Il fait beau» nous allons dire que l'auditeur B (croit qu'il) a *compris* au moment

1. Ceci n'est pas limité aux affirmations de croyances. Un agent qui dit «J'aimerais tellement voir ta collection de timbres» exprime un désir. Il est sincère seulement s'il a ce désir *et* s'il croit qu'il l'a. Un agent qui croit mentir en exprimant des attitudes mentales qu'il croit ne pas avoir, mais ne se rend seulement pas compte qu'il les a en fait, est en dehors des capacités de description de notre modèle.

où il atteint la croyance $\mathsf{Bel}_B \mathsf{Bel}_A$(il fait beau). B a *accepté* cette affirmation quand il y a aussi Bel_B(il fait beau). Supposer la sincérité ne nous semble pas être une limitation importante pour un système de dialogue *homme-machine*, alors que pour une théorie du «grounding» générale nous avons préféré utiliser la notion plus fine d'intention communicative.

6.1 Représentation d'états mentaux et conversationnels

Dans cette section, nous décrirons un modèle de représentation de croyances pour le dialogue. Nous distinguons deux types de croyances: les croyances de l'agent sur le monde et les croyances liées à l'état actuel du dialogue. Ces dernières peuvent être vues comme une sorte de *mémoire de travail* dont le contenu change toujours. C'est dans cet **état conversationnel** que se trouveront les suppositions communicatives qui vont ou bien disparaître après un certain temps ou bien être transformées en croyances plus persistantes. De telles croyances sont stockées dans ce que nous appellerons une **mémoire propositionnelle**. En principe, il y persistent toujours, même s'il se peut qu'un agent n'a pas toujours accès. Les croyances de la mémoire sont rassemblées en soi-disant **espaces de croyance**, des structures de données dont les propriétés sont décrites maintenant.

6.1.1 Espaces de croyance

Nous ne représentons pas «littéralement» les formules multi-modales décrivant les croyances, mais attribuons à chaque formule propositionnelle plusieurs *espaces de croyance*. Un espace de croyance est un rassemblement de plusieurs croyances sur une proposition. Si une proposition se trouve dans un espace de croyance, *toutes* les croyances rassemblées dans l'espace de croyance sur la proposition doivent être vraies.

Ce procédé nécessite la certitude que les croyances regroupées dans un espace de croyance sont toujours vraies simultanément. Vu les résultats des chapitres précédents, nous devons donc au moins avoir des espaces de croyances différents pour les croyances individuelles et les croyances mutuelles. Mais nous devons aussi distinguer les propres croyances de l'agent représenté et ses croyances concernant les croyances de son partenaire. Un modèle qui permet de combiner les quatre espaces «croyance individuelle», «croyance mutuelle», «croyance propre» et «croyance sur l'autre» est représenté en figure 6.1. En combinant les quatre espaces, on peut exprimer des croyances complexes de l'agent. Si, par exemple, pour l'agent X la proposition $p =$«*il pleut*» se trouve dans les trois espaces marqués gris en figure 6.1, cela signifie que X croit qu'il pleut, que X croit qu'il est mutuellement connu qu'il croit qu'il pleut et qu'il croit aussi que l'autre agent (Y) croit qu'il pleut. Cette dernière croyance n'est pas mutuellement connue (probablement, Y vient d'énoncer sa croyance et X n'a pas encore fourni un «acknowledgement»).

En principe, le modèle devrait permettre de combiner librement les quatre espaces, mais il y a un inconvénient: une croyance mutuelle comprenant la croyance individuelle correspondante, les deux dimensions «horizontales» du modèle ne sont pas indépendantes. Cela n'est pas souhaitable quand une croyance individuelle change *après* qu'une croyance mutuelle ait été établie. Si, dans le même exemple, X remarque qu'il a arrêté de pleuvoir, $p =$«*il pleut*» ne se trouve plus dans l'espace de la croyance individuelle, mais non plus dans l'espace pour la croyance mutuelle, parce que sans la croyance individuelle il n'y a pas de croyance mutuelle. Mais ce que nous aimerions exprimer, c'est que X (alors qu'il n'a pas encore informé Y de

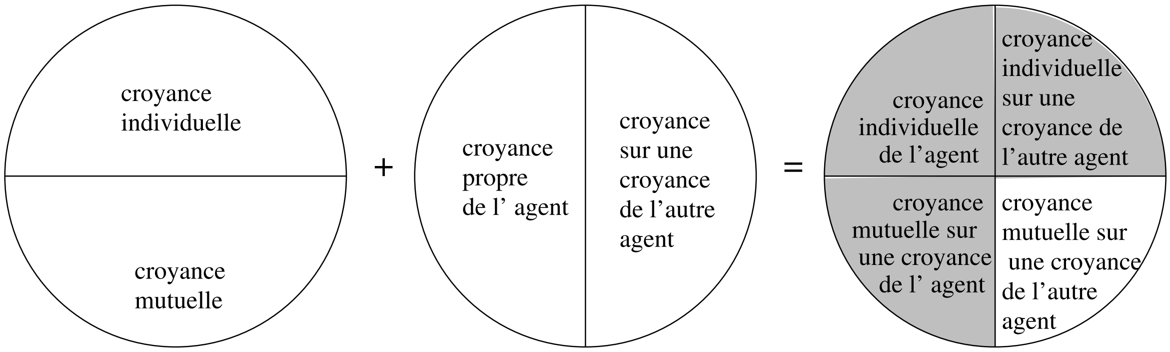

FIG. 6.1 – *Modèle de base pour quatre espaces de croyance*

l'arrêt de la pluie) croit que Y croit toujours en la croyance mutuelle. Mais une telle croyance n'est pas encore exprimable dans le modèle de base. Le modèle final pour les quatre espaces de croyance est représenté en figure 6.2. Cette fois, les croyances exactes exprimées par chaque espace sont indiquées.

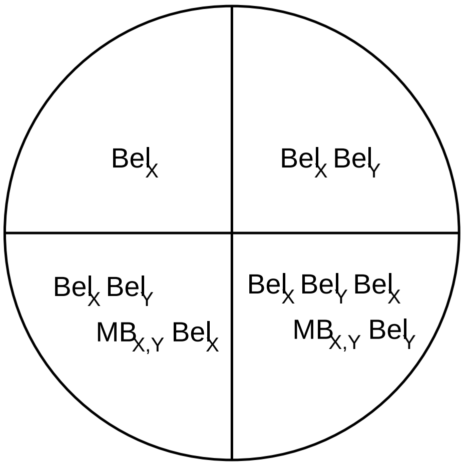

FIG. 6.2 – *Modèle d'espaces de croyance indépendants*

La figure 6.3 donne deux exemples importants de combinaison entre les espaces. L'exemple à gauche montre comment représenter la croyance mutuelle sur une croyance de l'agent représenté (X); nous allons nous référer à cet état comme *croyance grounded* de X. Le deuxième exemple en figure 6.3 montre ce que nous allons appeler une *croyance partagée* des deux agents. Les deux combinaisons expriment les formules $\mathsf{Bel}_X \mathsf{MB}_{X,Y} \mathsf{Bel}_X \varphi = \mathsf{Bel}_X \varphi \wedge \mathsf{Bel}_X \mathsf{Bel}_Y \mathsf{MB}_{X,Y} \mathsf{Bel}_X \varphi$ et $\mathsf{Bel}_X \mathsf{MB}_{X,Y} \varphi = \mathsf{Bel}_X \mathsf{MB}_{X,Y} \mathsf{Bel}_X \varphi \wedge \mathsf{Bel}_X \mathsf{MB}_{X,Y} \mathsf{Bel}_Y \varphi$, respectivement.

Chaque proposition dans chaque espace de croyance se trouve dans un **état**. Cet état exprime si l'agent a les croyances représentées par l'espace ou s'il ne les a pas. Comme nous avons vu en section 2.2.1, il y a *trois états épistémiques* que doit pouvoir traiter le modèle: la croyance en la vérité d'une proposition, en la vérité de sa négation ou l'ignorance de la valeur de vérité. Pour les espaces représentant des croyances emboîtées, les possibilités se multiplient. Le modèle

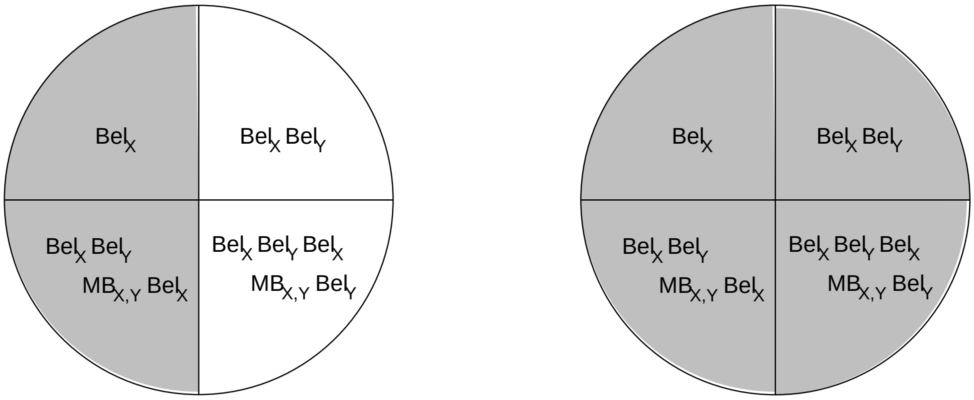

FIG. 6.3 – *«Grounding» d'une croyance de X* $\mathsf{Bel}_X\mathsf{MB}_{X,Y}\mathsf{Bel}_X$ *(gauche) et croyance partagée* $\mathsf{Bel}_X\mathsf{MB}_{X,Y}$*(droite)*

doit, par exemple, pouvoir représenter des croyances comme «X croit que Y ne sait pas si φ» ou «il est mutuellement cru par X et Y que X ne sait pas si φ».

Il est important de bien distinguer entre les espaces de croyance et leurs états. Les espaces représentent les différents *types* de croyance qu'un agent peut avoir sur une même proposition, alors que les états représentent les valeurs épistémiques pour chaque espace. Les espaces sont mutuellement indépendants, mais les états s'excluent mutuellement (on ne peut pas croire φ et en même temps l'ignorer). Tout espace se trouve dans *un* état, mais celui-ci peut être différent pour chaque espace (X peut croire φ et croire que Y ignore φ).

Le modèle traite une proposition φ et sa négation $\neg\varphi$ comme des propositions différentes, mais liées. Les différents états épistémiques seront décrits comme une combinaison des croyances sur φ et de celles sur les propositions qui ne peuvent être vrai seulement si φ est fausse. Il ne s'agit pas nécessairement des *négations* de φ, mais plutôt des propositions qui sont *mutuellement exclusives* de φ. Ce procédé a plusieurs avantages: la forme verbale d'une proposition et celle de sa négation sont souvent différentes (comme dans «il est marié» et «il est célibataire») et peuvent être utilisées différemment selon le contexte. De plus, il peut y avoir plusieures formes logiquement équivalentes de la même proposition («il est célibataire», «il n'est pas marié», «il n'a pas d'épouse») et notre modèle ne pose pas le problème de décider quelle est la «vraie» négation et quelle est une autre proposition avec la même la valeur de vérité. La section 6.1.3 décrira les relations logiques entre croyances et le raisonnement de l'agent.

Plus une croyance est emboîtée, plus il existe de possibilités d'ignorance. Les tableaux 6.1 et 6.2 présentent des exemples pour les espaces de croyance Bel_X et $\mathsf{Bel}_X\mathsf{Bel}_Y$. Chaque état se décompose en une croyance sur la proposition même et une croyance sur une proposition mutuellement exclusive de la première. Ensemble, les deux formules décrivent exactement les états épistémiques souhaités. Le choix des formules représentant les états ne sera pas discuté ici, mais il a été fait de telle manière que le nombre de formules différentes est le plus petit possible tout en pouvant exprimer tous les états épistémiques. L'opérateur $\mathsf{KIF}_X\varphi$ (savoir-si) est une abréviation pour $\mathsf{Bel}_X\varphi \vee \mathsf{KIF}_X\neg\varphi$[2]. Pour l'espace Bel_X il nous suffit de deux états

2. Bien que pour l'espace Bel_X on pourrait exprimer les trois états sans avoir recours à KIF, ceci n'est plus possible pour les autres espaces. Sans entrer dans le détail on peut dire que cela est dû au fait qu'une croyance comme $\mathsf{Bel}_X(\varphi \vee \psi)$ n'est pas décomposable en deux croyances sur des formules atomiques et nous avons donc

Croyance à exprimer	état de *marié(A)*	état de *célibataire(A)*
X croit que A est marié	KIF_X	$\neg\mathsf{Bel}_X$
X croit que A n'est pas marié	$\neg\mathsf{Bel}_X$	KIF_X
X ne sait pas si A est marié	$\neg\mathsf{Bel}_X$	$\neg\mathsf{Bel}_X$

TAB. 6.1 – *États épistémiques pour l'espace de croyance* Bel_X

Croyance à exprimer	état de *marié(A)*	état de *célibataire(A)*
X croit que Y croit que A est marié	$\mathsf{Bel}_X\mathsf{KIF}_Y$	$\mathsf{Bel}_X\neg\mathsf{Bel}_Y$
X croit que Y croit que A n'est pas marié	$\mathsf{Bel}_X\neg\mathsf{Bel}_Y$	$\mathsf{Bel}_X\mathsf{KIF}_Y$
X croit que Y sait si A est marié	$\mathsf{Bel}_X\mathsf{KIF}_Y$	$\mathsf{Bel}_X\mathsf{KIF}_Y$
X croit que Y ne sait pas si A est marié	$\mathsf{Bel}_X\neg\mathsf{Bel}_Y$	$\mathsf{Bel}_X\neg\mathsf{Bel}_Y$
X ne sait pas si Y sait si A est marié	$\neg\mathsf{Bel}_X\mathsf{Bel}_Y$	$\neg\mathsf{Bel}_X\mathsf{Bel}_Y$

TAB. 6.2 – *États épistémiques pour l'espace de croyance* $\mathsf{Bel}_X\mathsf{Bel}_Y$

(KIF_X, $\neg\mathsf{Bel}_X$) pour exprimer les trois cas épistémiques. Pour les cinq cas épistémiques de l'espace $\mathsf{Bel}_X\mathsf{Bel}_Y$, nous avons besoin de trois états ($\mathsf{Bel}_X\mathsf{KIF}_Y$, $\mathsf{Bel}_X\neg\mathsf{Bel}_Y$ et $\neg\mathsf{Bel}_X\mathsf{Bel}_Y$).

Les espaces comprenant des croyances mutuelles sont plus compliqués à représenter. Nous ne discutons que brièvement l'approche que nous avons choisie; l'exemple dans le tableau 6.3 montre plus clairement ce qu'on peut et ce qu'on ne peut pas représenter avec ce modèle. Nous simplifions la représentation en supposant ce qui suit: un agent croit ou bien qu'aucune croyance mutuelle sur une proposition n'est établie ou bien que l'autre agent a une croyance mutuelle unilatérale qui est connue de l'agent. C'est-à-dire que nous ne représentons pas des propositions comme $\neg\mathsf{KIF}_X\mathsf{Bel}_Y\mathsf{MB}_{X,Y}\mathsf{Bel}_X\varphi$. Il suffit donc de décrire si une croyance mutuelle unilatérale existe et si oui, sur lequel des états décrits pour les croyances individuelles.

Croyance à exprimer	état de *marié(A)*	état de *célibataire(A)*
$\mathsf{Bel}_X\mathsf{MB}$ que X croit que A est marié	$\mathsf{Bel}_X\mathsf{MB}_{X,Y}\mathsf{KIF}_X$	$\mathsf{Bel}_X\mathsf{MB}_{X,Y}\neg\mathsf{Bel}_X$
$\mathsf{Bel}_X\mathsf{MB}$ que X croit que A n'est pas marié	$\mathsf{Bel}_X\mathsf{MB}_{X,Y}\neg\mathsf{Bel}_X$	$\mathsf{Bel}_X\mathsf{MB}_{X,Y}\mathsf{KIF}_X$
$\mathsf{Bel}_X\mathsf{MB}$ que X sait si A est marié	$\mathsf{Bel}_X\mathsf{MB}_{X,Y}\mathsf{KIF}_X$	$\mathsf{Bel}_X\mathsf{MB}_{X,Y}\mathsf{KIF}_X$
$\mathsf{Bel}_X\mathsf{MB}$ que X ne sait pas si A est marié	$\mathsf{Bel}_X\mathsf{MB}_{X,Y}\neg\mathsf{Bel}_X$	$\mathsf{Bel}_X\mathsf{MB}_{X,Y}\neg\mathsf{Bel}_X$
$\mathsf{Bel}_X\neg\mathsf{MB}_{X,Y}$ si X croit que A est marié	$\mathsf{Bel}_X\neg\mathsf{MB}_{X,Y}\mathsf{KIF}_X$	$\mathsf{Bel}_X\neg\mathsf{MB}_{X,Y}\mathsf{KIF}_X$

TAB. 6.3 – *États épistémiques d'une croyance mutuelle (espaces* Bel_X+$\mathsf{Bel}_X\mathsf{Bel}_Y\mathsf{MB}_{X,Y}\mathsf{Bel}_X$*)*

Nous ne présentons pas ici les représentations d'intentions et d'autres modalités. Comme la notion d'*intention commune* fait objet de discussions philosophiques (cf. section 2.2.3), cela devrait être justifié en détail. Notons seulement que nous avons fait une tentative pour représenter les intentions et les croyances sur les intentions avec des espaces de croyances et que cette approche semble fournir exactement les distinctions nécessaires pour représenter des *plans partagés*. Une telle représentation d'intentions pourrait donc être utilisée afin qu'un

besoin d'au moins un opérateur qui comprend la croyance sur une disjonction.

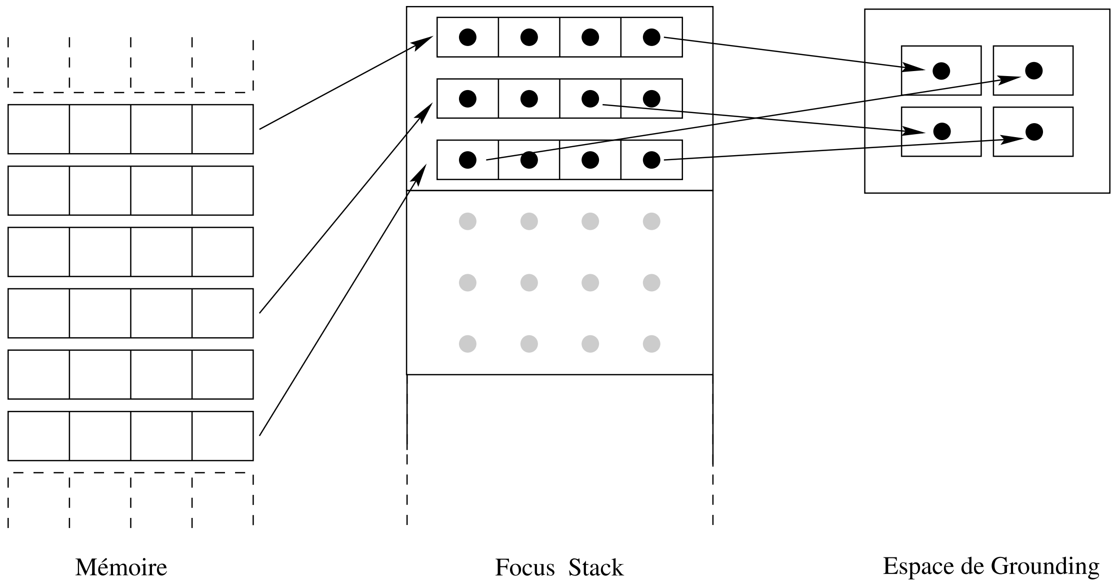

FIG. 6.4 – *Relations entre mémoire, focus stack et espace de grounding*

agent puisse déterminer facilement quelles croyances et intentions manquent encore à un *plan partagé complet* (FSP).

6.1.2 État conversationnel

Un modèle compututationnel de dialogue concret (celui de Balkanski et Hurault-Plantet (1999)) sera présenté dans le chapitre suivant. Ici, nous ne décrivons que les composantes liées au processus de «grounding». Déterminer le contenu d'un acte de langage dépend largement du contexte de son énonciation, des informations présupposées par les agents dans la situation actuelle. Le «grounding» se passe donc dans ce que Grosz et Sidner (1986) ont appelé le **focus d'attention** (cf. section 2.3.3). Le focus d'attention comprend les objets, croyances et relations entre elles pertinentes à un moment donné du dialogue. C'est ici que nous situons les attitudes mentales éphémères comme les suppositions communicatives.

Au cours d'un dialogue, le focus d'attention va changer quand des sous-dialogues spécifient les détails d'un thème ou quand de nouveaux thèmes sont abordés. Grosz et Sidner (1986) proposent de représenter les changements de l'attention par une *pile d'espaces de focus* (focus space stack). Lorsqu'un sous-dialogue est commencé, un nouvel espaces de focus est posé sur la pile. Ceci permet aux participants de «revenir» sur un thème qui n'était pas clos.

Dans notre modèle, la mémoire représente toutes les croyances d'un agent. Quand le dialogue actuel concerne une proposition, les espaces de croyances de l'agent sur cette proposition vont se trouver dans le focus d'attention. Si de plus, une proposition est l'objet d'un acte de langage, celles des croyances sur la proposition qui sont explicitement exprimées dans l'énoncé vont aller dans un **espace de grounding** qui contient seulement des croyances qui doivent encore être «grounded». La figure 6.4 montre une représentation graphique des relations entre ces trois modules.

Dernier évènement	supposition communic.	cas de base
A vient de dire $Esp(A,\varphi)$	$\mathsf{CSup}_{A,B} Esp(A,\varphi)$	$\mathsf{Sup}_A \mathsf{Bel}_B Esp(A,\varphi)$
A croit que B vient de dire $Esp(B\varphi)$	$\mathsf{Bel}_A \mathsf{CSup}_{B,A} Esp(B\varphi)$	$\mathsf{Bel}_A \mathsf{Sup}_B \mathsf{Bel}_A Esp(B,\varphi)$
A suppose que B a dit $Esp(B,\varphi)$	$\mathsf{Sup}_A \mathsf{CSup}_{B,A} Esp(B,\varphi)$	$\mathsf{Sup}_A \mathsf{Sup}_B \mathsf{Bel}_A Esp(B,\varphi)$
A a demandé confirm. sur $Esp(B,\varphi)$	$\mathsf{CSup}_A Esp(B,\varphi)$	$\mathsf{Sup}_A Esp(B,\varphi)$
B a demandé confirm. sur $Esp(A,\varphi)$	$\mathsf{Bel}_A \mathsf{CSup}_B Esp(A,\varphi)$	$\mathsf{Bel}_A \mathsf{Sup}_B Esp(A,\varphi)$

TAB. 6.4 – *Variantes de suppositions communicatives dans l'*espace de grounding

L'*espace de grounding* représente une partie du focus d'attention particulière. Il ne représente que des suppositions communicatives, c'est-à-dire tout ce qui n'est pas encore «grounded». Les suppositions communicatives sont considérées comme des croyances très éphémères: ou bien elles sont confirmées par le prochain énoncé et atteignent ainsi un état plus stable dans la mémoire ou bien elles se révèlent fausses et disparaissent. Ainsi, seul le focus d'attention même et non les niveaux plus bas dans la pile peuvent contenir des suppositions communicatives.

Une supposition communicative dans l'*espace de grounding* d'un agent A est, en principe, d'une des deux formes principales suivantes: $\mathsf{CSup}_{A,B} Esp(A,\varphi)$ ou $\mathsf{Bel}_A \mathsf{CSup}_{B,A} Esp(B,\varphi)$, ce qui correspond aux deux possibilités élémentaires d'un énoncé fait par l'agent lui-même ou, respectivement, par l'autre. La notation $Esp(X,\psi)$ représente un espace (ou même une combination d'espaces) de croyance de l'agent X sur une proposition φ dont parle l'énoncé de X. D'autres formes résultent des différentes formes d'incertitude qui peuvent être exprimées par (ou reconnues dans) un énoncé. Le tableau 6.4 présente des exemples pour chacun des cas. Bien qu'une supposition communicative exprime un sentiment de situation partagée qui entraîne une infinité de suppositions emboîtées avec des croyances, on peut résumer chacune par le *cas de base*, la supposition la moins emboîtée. Le tableau 6.4 montre les cas de base pour chacun des exemples. À partir des cas de base des suppositions communicatives dans l'*espace de grounding*, un agent peut tirer des *inférences prédictives*. Par exemple, un agent A qui vient d'énoncer «il fait beau» (qu'on peut exprimer comme Bel_A(beau)) va supposer $\mathsf{Sup}_A \mathsf{Bel}_B \mathsf{Bel}_A$(beau). À partir de cette supposition, A peut maintenant inférer d'autres croyances de B qui seraient vraies si B avait compris. C'est aussi cette supposition qui permet à A de continuer par une phrase comme «ça me plaît»: seulement parce que A suppose que B a compris la première phrase, il peut aussi supposer que l'anaphore de la deuxième phrase peut être résolue par B.

6.1.3 Raisonnement

Deux propositions peuvent être logiquement interdépendantes. L'interdépendance entre une proposition et une autre qui exprime son contraire a déjà été utilisée dans la section 6.1.1. Nous appelons ce type de relation entre propositions *l'exclusion mutuelle*. D'autres types de relations incluent *l'équivalence* et *l'implication*. Quand une croyance change pendant le dialogue, le module de raisonnement doit adapter de façon appropriée les croyances sur les propositions reliées à la proposition qui est l'objet de la croyance changée.

Les relations entre les propositions ne sont pas fixes, mais sont aussi objets de croyances. Ainsi, un dialogue peut aussi servir à établir de nouvelles relations entre propositions. Quand l'agent X énonce «Je suis toujours malheureux quand il pleut», l'agent Y peut «apprendre» que la relation d'implication existe entre les propositions *il pleut* et *X est malheureux*, et cette relation peut être «grounded» par un «acknowledgement» de X. Si plus tard dans le dialogue, X dit qu'il pleut, Y peut conclure du fait que l'implication fait maintenant partie de la base commune des agents, que X voulait peut-être exprimer qu'il est malheureux.

Alors que les relations établies entre les propositions peuvent changer, les règles de raisonnement sont fixes et correspondent à une théorie de preuve équivalente ou similaire à celle du chapitre 5. L'agent spécifié est donc un agent rationnel, ce qui nous confronte au problème de l'omniscience logique (cf. section 2.2.1). Il y a deux possibilités de traiter ce problème sans influencer la validité des déductions effectuées par l'agent: on peut limiter le domaine ou la profondeur du raisonnement. Nous proposons une approche utilisant les deux moyens: nous ne posons pas de limites de raisonnement sur les croyances dans le focus d'attention, mais n'admettons qu'une profondeur limitée quand l'objet du raisonnment ne se trouve plus dans le focus.

6.2 Changer d'état: parler et écouter

Cette section décrira la mise à jour de croyances au cours d'un dialogue. La section 6.2.1 décrira notre représentation d'actes de langage, les sections 6.2.2 et 6.2.3 décrivent la modification d'un contexte par un énoncé et comment l'ancien contexte peut être utilisé pour interpréter un énoncé (6.2.2) ou choisir un acte à produire (6.2.3).

6.2.1 Actes de langages

Notre approche de représentation et d'utilisation d'actes de langage dans le dialogue est largement inspirée par les travaux de Bunt (1989, 1995). Bunt décrit un acte de langage en termes de ses *effets de changement de contexte*. La faillibilité de la communication étant au sein de nos idées, nous pouvons décrire ces changements de contexte *avec certitude* seulement pour le locuteur même. Le chapitre 5 nous a fourni la notion de supposition communicative pour décrire les modifications des croyances (ou bien: du *contexte*) du locuteur quand il produit un acte de langage. Quant à l'auditeur, le chapitre 5 nous donne au moins les effets précis d'un acte de langage lorsque l'interprétation a résulté en l'une des trois possibilités (incompréhension, certitude et incertitude de compréhension). Cela nous donne une description générale des effets (illocutoires) d'un acte de langage.

Quant au *contenu* d'un acte de langage, c'est-à-dire les informations qu'il communique, nous suivons une autre approche de Bunt. Pour un agent sincère, il y a toujours des conditions qui doivent être satisfaites avant qu'il puisse produire un certain acte de langage: pour produire une affirmation sur φ, l'agent doit croire φ; pour poser une question sur φ, l'agent doit avoir l'intention de savoir φ et il ne doit pas déjà savoir φ, etc; pour fournir un «acknowledgement» l'agent doit savoir que l'autre se trouve dans un état de supposition communicative, etc. Ces conditions sont appelées *conditions de pertinence* (correspondant aux *conditions de félicité* de Searle ou bien aux *préconditions* d'un opérateur de planification) et décrivent donc quand un agent peut envisager de produire un tel acte. Bien entendu, ces conditions joueraient un rôle différent pour un agent malhonnête: elles n'empêcheraient pas de produire un certain acte, mais décriveraient plutôt ce qu'il doit faire quand il veut obtenir un certain effet.

Nous croyons que les conditions de pertinence (CPs) ne sont ni absolues ni inaltérables. Dans certaines situations, les conditions pour un acte peuvent même s'opposer. La question «Tu sais ce que font treize fois quatorze?» a d'autres conditions de pertinence quand posée par le professeur que quand elle est chuchotée à son voisin par l'élève pendant l'examen. Le professeur est censé connaître la réponse, alors que l'élève exprime en fait qu'il ne la connaît pas. Nous croyons donc qu'il s'agit ici aussi d'une forme de *convention* entre les utilisateurs d'une langue, largement dépendante du contexte de l'énonciation et de la personne du locuteur. Une telle convention peut être représentée par $\mathsf{MB}(\text{acte} \rightarrow \text{CP})$, c'est-à-dire que les agents croient mutuellement qu'un acte présuppose une certaine CP. [3]

Ce qui est crucial dans la notion de *condition de pertinence*, c'est qu'elle décrit aussi quelle information est communiquée par l'acte de langage correspondant. Lorsqu'un auditeur a compris un acte de langage, il va connaître les conditions dans lesquelles il a été énoncé. Contrairement à un opérateur de planification nous pouvons donc caractériser un acte de langage en décrivant uniquement ces *préconditions*; ses *effets* en dérivent.

6.2.2 Interprétation

Un agent peut se servir de l'*espace de grounding* afin de vérifier *activement* si un énoncé fournit le grounding. La figure 6.5 montre une partie d'un algorithme d'interprétation qui utilise l'*espace de grounding* pour déterminer la signification d'un énoncé. L'*espace de grounding* sert d'entrée supplémentaire et permet en plus d'un traitement uniquement conduit par les données (data-driven) de *chercher* des éléments de l'énoncé que l'auditeur attend (goal-driven). Ainsi, il peut servir à la désambiguïsation de l'énoncé. Les inférences prédictives (des suppositions déduites du cas de base des suppositions communicatives dans l'*espace de grounding*) indiquent ce que *pourrait* dire le locuteur. Ainsi elles peuvent transformer une tâche d' *interprétation* avec de nombreuses possibilités en une tâche de *reconnaissance* ayant pour but concret de vérifier une information attendue.

L'algorithme va chercher pour toute information pas encore «grounded» et qui est donc objet d'une supposition communicative dans l'*espace de grounding* de l'auditeur A, si l'énoncé du locuteur B concerne cette information. Selon le type de supposition communicative, l'algorithme va chercher d'autres signaux: Si A vient de faire une demande de confirmation, l'algorithme va chercher un signal (positif ou négatif) que A avait bien compris. Si A s'attend à un acknowledgement (explicite ou implicite), l'algorithme va chercher des signes qui montrent que le nouvel énoncé de B continue le dialogue d'une manière cohérente. Si l'énoncé de B indique que B avait mal compris, n'est pas certain d'avoir bien compris ou n'avait pas du tout compris, l'algorithme va ajouter de nouvelles croyances à l'*espace du grounding* qui vont servir à fournir une réponse appropriée.

L'algorithme en figure 6.5 ne décrit l'utilisation de croyances dans l'interprétation que pour des informations qui se trouvent *déjà* dans l'*espace de grounding*. Pourtant, normalement la plupart des informations d'un énoncé vont être nouvelles. Pour les déterminer, le focus d'attention joue un rôle crucial. Bien que l'information qui doit être fournie ne soit souvent pas aussi précisément prédeterminée que pour l'*espace de grounding*, le thème et aussi largement le contenu

3. Bien sûr, il est possible que les agents aient en fait de différentes croyances sur les CP d'un acte: un enfant pourrait croire que le professeur ne connait pas la réponse à sa propre question. Nous considérons ce cas comme une *hypothèse fausse* sur les conventions d'une communauté langagière.

Input: A auditeur (agent modélisé), B locuteur, e_B énoncé à interpréter, M mémoire de A, G espace de grounding
Output: G' nouveau espace de grounding
$G' \leftarrow \emptyset$
for all suppositions communicatives S dans G **do**
  if S est de la forme $\mathsf{CSup}_A \mathsf{Bel}_B Esp(A,\varphi)$ **then** {A attend acknowledgement de son dernier énoncé e_A}
    if B continue le dialogue comme s'il avait compris **then**
      if existe une inférence prédictive $\mathsf{Sup}_A \alpha$ tirée à partir de S que e exprime ou présuppose **then**
        $M \leftarrow Update(M, \mathsf{Bel}_A \mathsf{MB}_{A,B} \mathsf{Bel}_B Esp(A,\varphi))$
        $M \leftarrow Update(M, \mathsf{Bel}_A \alpha)$
      else if existe φ' similaire à φ et e_B semble exprimer $\mathsf{Bel}_B Esp(A,\varphi')$ **then** {A a détecté un malentendu}
        $M \leftarrow Update(M, \mathsf{Bel}_A \mathsf{Bel}_B \mathsf{MB}_{A,B} Esp(A,\varphi'))$
        $G' \leftarrow Update(G', \neg compris(B, Esp(A,\varphi)))$ {corriger le malentendu dans le prochain énoncé}
    else if B exprime incertidude sur dernier énoncé e_A de A **then**
      if e_B indique à A que B avait compris e_A **then**
        $G' \leftarrow Update(G', \mathsf{Bel}_A \mathsf{CSup}_{B,A} Esp(A,\varphi))$ {A doit fournir un acknowledgement}
      else
        $G' \leftarrow Update(G', \neg compris(B, Esp(A,\varphi)))$
    else if B exprime incompréhension **then**
      $G' \leftarrow Update(G', \neg compris(B, Esp(A,\varphi)))$
  else if S est de la forme $\mathsf{CSup}_A Esp(B,\varphi)$ **then** {A s'attend à un signal de confirmation}
    if signal de confirmation pour $Esp(B,\varphi)$ trouvé **then**
      $M \leftarrow Update(M, \mathsf{Bel}_A \mathsf{MB}_{A,B} Esp(B,\varphi))$
    else
      réinterprète ancien énoncé (en utilisant nouvelles informations dans e)
$G' \leftarrow$ ajoute interprétations des nouvelles informations dans e_B en tenant compte du focus d'attention
if interprétation est considérée comme certain **then**
  $G' \leftarrow \mathsf{Bel}_A \mathsf{CSup}_B \mathsf{Bel}_A Esp(B,\varphi)$
else if une interprétation a été trouvée, mais elle est incertaine **then**
  $G' \leftarrow \mathsf{Sup}_A \mathsf{CSup}_B \mathsf{Bel}_A Esp(B,\varphi)$
else {e_B n'a pas pu être interprété}
  $G' \leftarrow incompris(A, e_B)$

FIG. 6.5 – *Intégration du «grounding» dans un algorithme d'interprétation*

du prochain énoncé sont prévisibles à partir des informations en focus. Particulièrement, comment le signalent Grosz et Sidner (1986), le focus comprend les intentions poursuivies par le segment actuel du discours. L'algorithme en figure 6.5 n'aborde pas cet aspect important de l'interprétation, mais le chapitre 7 montre comment dans un système pratique ces deux composantes interagissent. C'est donc une deuxième possibilité pour une interprétation guidée par ce que l'agent peut prévoir. En fin de compte, l'interprétation reste quand-même toujours un processus concernant les structures de l'énoncé même et ne peut pas être entièrement prévue; l'agent va donc souvent ne pas être certain de son interprétation. Les nouvelles informations sont ajoutées à l'*espace de grounding* selon les trois possibilités suivantes: A est certain d'avoir compris ($\mathsf{Bel}_A \mathsf{CSup}_B \mathsf{Bel}_A Esp(B, \varphi)$), A suppose avoir compris ($\mathsf{Sup}_A \mathsf{CSup}_B \mathsf{Bel}_A Esp(B, \varphi)$) et A n'a pas compris ($incompris(A, e_B)$).

6.2.3 Génération

Dans le processus de génération de son prochain énoncé, un agent doit fournir de nouvelles informations à son partenaire, mais aussi agir afin d'établir la croyance mutuelle sur le dernier énoncé de l'autre. Ceci implique une certaine hiérarchie entre les informations à exprimer dans le nouvel acte de langage. La figure 6.6 présente un algorithme pour déterminer quelles informations sont les plus importantes à exprimer dans le prochain énoncé et décrit la mise à jour des croyances lors de la génération d'un énoncé. L'algorithme a deux «sources» d'informations pour le prochain énoncé: une liste de *nouvelles* informations L qui est fournie par des algorithmes gérant le dialogue en général, et l'*espace de grounding* G qui après l'interprétation (cf. figure 6.5) contient les suppositions communicatives qui ne sont pas encore «grounded».

L'information la plus pertinente à exprimer pour un locuteur A sera qu'il n'a pas compris le dernier énoncé de son partenaire B. Si ce n'est pas le cas, le plus important sera de répondre à un signal d'incompréhension ou à une demande de confirmation de B sur le dernier énoncé de A afin de fournir au plus vite le «grounding» de cet énoncé. Seulement si tout cela n'est pas nécessaire, A peut tenter de répondre à l'énoncé de B: soit par une demande de confirmation (s'il n'est pas certain d'avoir compris), soit par une réponse qui montre sa compréhension en ce qu'elle est liée aux inférences qu'il a tirées de l'énoncé de B. Le reste des informations sont énoncées en utilisant le focus d'attention et particulièrement les informations déjà exprimées. Ainsi les suppositions sur la compréhension des premières parties d'un énoncé sont, par exemple, utilisées pour déterminer si des expressions anaphoriques peuvent être utilisées.

Cet algorithme serait le lieu où les *demandes d'acknowledgement* que nous avons négligées jusqu'ici, pourraient être expliquées et intégrées dans le modèle. Quand un agent enchaîne plusieurs phrases, il essaie de se référer aux informations et objets dans le focus pour produire un discours cohérent. Ce comportement ne correspond pas seulement aux fameuses maximes de Grice (1975), mais il facilite énormément le «grounding». Quand A produit une suite de phrases (ou parties de phrases) $p_1, p_2, \ldots, p_n$ où la compréhension de p_i présuppose la compréhension de p_{i-1}, le «grounding» de p_n fournit aussi le «grounding» de $p_{n-1}, p_{n-1}, \ldots, p_1$. Mais si A énonce une phrase q qui peut être comprise sans avoir recours à $p_1, \ldots, p_n$ énoncées avant, un signal de compréhension de q donné par B ne va pas assurer A que B a aussi compris $p_1, \ldots, p_n$. Il pourrait alors énoncer une *demande d'acknowledgement* avant d'énoncer q. Cette idée n'a pas encore été suffisamment explorée pour pouvoir être intégrée dans le modèle. Une étude de corpus pourrait montrer à quels moments d'une conversation des demandes d'acknowledgement sont énoncées.

Input: A locuteur, B auditeur, M mémoire de A, G espace de grounding, L liste des nouvelles informations à donner dans le prochain énoncé
Output: G' nouveau espace de grounding

$G' \leftarrow \emptyset$
if existe énoncé e_B où $incompris(A, e_B)$ dans G **then**
  exprime incompréhesion de e_B
else if existe énoncé e_A où $incompris(B, e_A)$ dans G **then**
  répète ou laisse tomber e_A
else if existe supposition S dans G de la forme $\mathsf{Bel}_A \mathsf{CSup}_B Esp(A, \varphi)$ **then** {demande de confirmation de B}
  if $Esp(A, \varphi) \in M$ **then**
    fournit confirmation
  else
    il doit y avoir $Esp(A, \psi) \in M$ où $\neg compris(B, Esp(A, \psi)) \in G$
    refuse $Esp(A, \varphi)$ et exprime $Esp(A, \psi)$
else {seulement si B semble avoir compris le dernier énoncé de A}
  if existe $\mathsf{Sup}_A \mathsf{CSup}_B \mathsf{Bel}_A Esp(B, \varphi)$ dans G **then** {A n'est pas certain d'avoir bien compris le dernier énoncé de B}
    exprime incertiude sur $Esp(B, \varphi)$
    $G' \leftarrow Update(G, \mathsf{CSup}_A Esp(B, \varphi))$
  else {A est certain du dernier énoncé de B}
    for all suppositions communicatives S dans G **do**
      for all inférences I de la forme $\mathsf{Bel}_A \mathsf{Sup}_B \alpha$ tirées à partir de S **do**
        if existe une information l dans L qui est logiquement liée à I et cette relation est MB **then**
          exprime l tel que B va savoir $\mathsf{Bel}_B I$
          $M \leftarrow Update(M, \mathsf{Bel}_A \mathsf{MB}_{A,B} \mathsf{Bel}_A Esp(B, \varphi))$
          $M \leftarrow Update(M, \mathsf{Bel}_A \mathsf{MB}_{A,B} \alpha)$
for all $Esp(A, \varphi) \in L$ pas encore exprimés **do**
  if pas encore $\mathsf{Bel}_A \mathsf{MB}_{A,B} Esp(A, \varphi)$ **then** {l'information donnée n'est pas encore mutuellement connue}
    exprime $Esp(A, \varphi)$ (en tenant compte du focus et des suppositions prédictives)
    $G' \leftarrow Update(G', \mathsf{CSup}_A \mathsf{Bel}_B Esp(A, \varphi)$
$G \leftarrow G'$ {Abandonne ancien espace de grounding}

FIG. 6.6 – *Fournir le «grounding»: un algorithme de gestion de croyances lors du processus de génération*

Chapitre 7

Implémentation

Nous avons intégré le modèle développé dans le chapitre précedent dans un système de dialogue homme-machine réel qui a été développé par Balkanski et Hurault-Plantet (1999). Nos extensions incluent une représentation des croyances capable de représenter l'évolution des croyances d'un agent au cours du dialogue, une représentation des actes de langage en termes de leurs conditions de pertinence, ainsi que plusieurs algorithmes qui gèrent la mise à jour des croyances pendant le dialogue et suppriment la génération d'énoncés inutiles. La section 7.1 présente le système de Balkanski et Hurault-Plantet (1999); l'intégration des concepts développés dans ce travail est décrite en section 7.2.

7.1 Le système existant

Le système de Balkanski et Hurault-Plantet (1999) modélise un standard téléphonique qui entame un dialogue avec un appelant dans le but de lui passer une personne. Le comportement du système est modélisé selon un corpus de dialogues transcrits d'un enregistrement d'un standard téléphonique du CNRS (Castaing, 1993). Ce système implémente le modèle computationnel de dialogue développé et décrit par Balkanski et Hurault-Plantet (1999). Le modèle repose sur la théorie de discours de Grosz et Sidner (1986), la notion du plan partagé de Grosz et Sidner (1990) et Grosz et Kraus (1996) et son utilisation pour la description et reconnaissance de la structure intentionnelle du discours par Lochbaum (1998). Ces travaux ont été présentés en section 2.3.3. Alors que Lochbaum est surtout concernée par l'utilisation et l'augmentation du plan partagé pendant l'*interprétation* d'un énoncé (cf. figure 2.3), Balkanski et Hurault-Plantet (1999) traitent l'exécution d'actions et la génération d'énoncés en fonction de l'état du plan partagé. Un *algorithme d'avancement dans la tâche* détermine quand un agent va exécuter une action du plan, générer un énoncé ou poursuivre la planification de la tâche sous-jacente au dialogue. L'algorithme permet ainsi l'alternance de ces trois processus d'une manière naturelle.

L'implémentation du système de Balkanski et Hurault-Plantet a été effectuée en Smalltalk. Un modèle de l'architecture du système est représenté en figure 7.1. Les quatre composantes principales du système sont un module de gestion de dialogue, un module d'interprétation, un module de génération et un raisonneur de plan. Actuellement, seulement le raisonneur sur

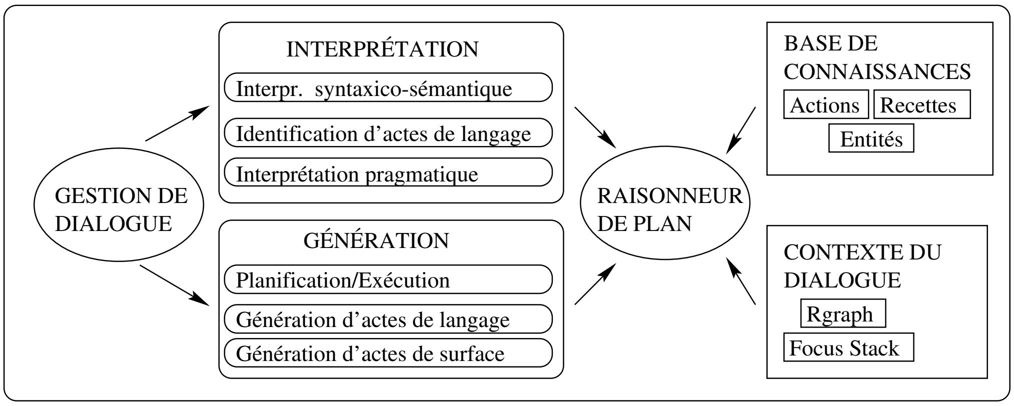

FIG. 7.1 – *Architecture du système de Balkanski et Hurault-Plantet (1999)*

le plan et les algorithmes qui lui sont associés dans les sous-modules d'interprétation pragmatique et de planification/exécution sont complètement implémentés. Les autres modules ont également été implémentés, mais en utilisant des suppositions simplificactrices. Par conséquent, certaines composantes du modèle de dialogue développé dans le chapitre précedent sont également implémentées d'une manière simplifiée ou ne sont que simulées.

La structure du dialogue est représentée par le Rgraph. Le Rgraph (cf. section 2.3.3) représente les croyances de l'agent concernant les actions de la tâche sous-jacente au dialogue ainsi que la relation entre ces actions. Quand il exécute l'algorithme d'avancement dans la tâche, le raisonneur sur le plan essaie d'augmenter le Rgraph en planifiant les actions qui pourraient contribuer au but commun des agents et en déclenchant l'exécution des actions si nécessaire (c'est-à-dire si la future planification dépend des résultats d'une action). L'algorithme d'avancement dans la tâche décrit quand le raisonneur de plan va interrompre la planification ou l'exécution d'actions et où il va proposer de générer un énoncé. C'est à ce moment-là que nos extensions (la représentation des croyances et leur mise à jour en fonction des actes de langage) deviennent importantes: en fonction des croyances à exprimer et des croyances déjà «grounded» (le *contexte*), des actes de langages différents peuvent être appropriés – dans certains cas il n'est même pas nécessaire de produire un énoncé proposé par le raisonneur sur le plan, parce que les informations qu'il exprime sont déjà mutuellement connues.

7.2 Intégration du processus de grounding

L'architecture du système résultant de l'intégration des modules et processus décrits dans le chapitre 6 est représenté en figure 7.2. Les zones grises représentent les modules concernés par le processus du grounding. Trois nouvelles structures ont été ajoutées. La représentation du contexte a été augmentée par une *mémoire propositionnelle* qui représente les croyances d'un agent évoluant pendant le dialogue en utilisant des espaces de croyance. De plus, un *espace de grounding* comme décrit en section 6.1.2 contient des suppositions communicatives sur les informations qui ne sont pas encore «grounded». Dans la base de connaissances, une représentation des actes de langage en termes de leurs conditions de pertinence a été intégrée. Le reste de la section décrit les composantes ajoutées et les algorithmes qui leur sont associés.

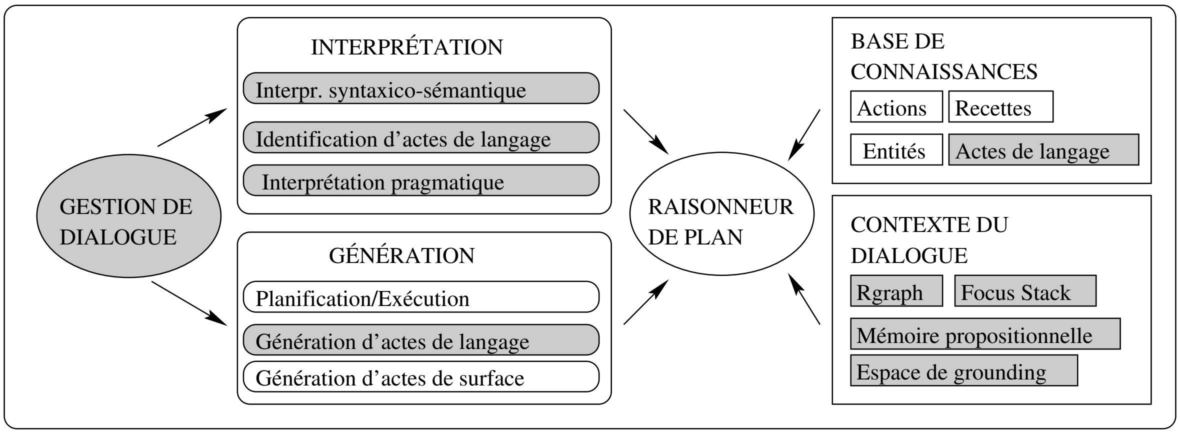

FIG. 7.2 – *Extensions du système de Balkanski et Hurault-Plantet (1999)*

La *mémoire propositionnelle* est une mise en pratique directe du modèle de la mémoire présenté en section 6.1.1. La mémoire d'un agent consiste d'un ensemble d'*éléments*, chaque élément comprenant toutes les croyances d'un agent sur une proprosition. Cela inclut les espaces de croyances de l'agent dans lesquelles se trouve la proposition, son état dans chacun de ces espaces, les relations logiques avec d'autres propositions et les liens avec le Rgraph. Si une partie du Rgraph est liée à une certaine proposition, cela engendre que les croyances sur cette proposition vont aller dans le focus d'attention quand l'action est exécutée ou quand le sous-dialogue actuel concerne cette action.

Pour chaque proposition, un *état de grounding* décrit si elle est objet d'une supposition communicative. Les élements de la mémoire dont la proposition est objet d'une supposition communicative font partie de l'*espace de grounding*. Les états des suppositions communicatives correspondent aux différentes possibilités représentées dans le tableau 6.4 en page 56. Actuellement, la représentation ne permet pas de raisonnement sur les suppositions. Ce type de raisonnement est surtout important pour la recherche active d'informations pendant l'interprétation et la génération d'énoncés cohérents. Les modules qui devraient intégrer ce type de raisonnement (interprétation sémantique, génération d'actes de surface) ne sont actuellement encore partiellement simulés. Une meilleure représentation des suppositions communicativives devrait faire partie de l'extension de ces modules.

Un acte de langage n'est pas représenté dans la forme «classique» comprenant le contenu propositionnel et la force illocutoire, mais (suivant Bunt (1989)) en termes du contenu propositionnel et de la condition de pertinence (cf. section 6.2.1). La condition de pertinence comprend des contraintes sur les croyances qu'un agent doit avoir sur le contenu propositionnel. Par exemple, un acte du type *Inform* nécessite que pour le locuteur X le contenu propositionnel φ se trouve dans l'espace de croyance Bel_X, c'est-à-dire le locuteur ne va produire l'acte *Inform* seulement s'il *croit* le contenu propositionnel[1]. L'acte *Question fermée* nécessite également que le contenu sémantique se trouve dans l'espace de croyance Bel_X, mais seulement dans l'état $\neg\mathsf{KIF}_X$: le système va donc poser une question seulement s'il ne connaît

1. En principe, cela n'est pas obligatoire. Pour un agent malhonnête, un acte *Inform* aurait la condition de pertinence $\mathsf{CIntTh}_X\mathsf{Bel}_Y\varphi$. Mais l'implémentation actuelle ne comprend pas encore la représentation des intentions.

pas encore la réponse (cf. la discussion en page 57). Les conditions de pertinence peuvent donc contraindre l'agent plus ou moins fortement: elles peuvent simplement exiger qu'une proposition se trouve dans un certain espace de croyances ou elles peuvent en plus déterminer l'*état* de la proposition dans cet espace (cf. page 52). Si les conditions de pertinence d'un acte comprennent des contraintes sur l état de grounding du contenu sémantique (c'est-à-dire qu'elles exigent que la proposition se trouve dans l'espace de grounding comme objet d'une supposition communicative), l'acte de langage concerne le processus de grounding. De tels actes sont, par exemple, *Acknowledge* ou *DemandeConfirmation*.

Deux algorithmes, correspondant aux algorithmes de gestion de croyances lors de l'interprétation et de la génération qui sont représentés dans les figures 6.2.2 et 6.2.3, gèrent la mise à jour des croyances et suppositions communicatives après la perception et la génération d'un acte de langage. Des éléments qui ont été «grounded» sont enlevés de l'espace de grounding, les nouvelles informations exprimées par l'énoncé y sont ajoutées.

Actuellement, le *module d'identification d'actes de langage* n'utilise pas encore les croyances existantes de l'agent pour déterminer l'acte sous-jacent à un énoncé. La «recherche» des informations attendues, comme décrite dans le chapitre 4 et dans l'algorithme d'interprétation (figure 6.2.2), devrait être intégrée dans ce module. Quant à la génération, c'est le raisonneur sur le plan qui détermine le type d'acte de langage à produire. À moyen terme, ce choix sera précisé en fonction des croyances de l'agent dans le *module de génération d'actes de langage*. L'exemple suivant illustre ce processus: si l'information à exprimer est «Je ne connais pas Corinne Savol», le raisonneur sur le plan va déterminer que l'acte de langage *Inform* est approprié. Un algorithme dans le *module de génération d'actes de langage* va ensuite essayer de préciser ce choix en vérifiant quelles croyances et suppositions liées à la proposition ¬connaître(Système,CS) sont disponibles. Si, par exemple, le dernier énoncé de l'utilisateur avait été «Vous connaissez Corinne Savol?», l'algorithme va détecter une supposition communicative ainsi que la croyance du système que $\neg\mathsf{KIF}_{Uti}$(connaître(Système,CS)). Ces croyances et suppositions permettent donc qu'un acte de langage plus «spécifique» soit produit: le module de génération va constater que les conditions de pertinence pour l'acte *ConfirmationNégative* sont satisfaites et va énoncer «Non» au lieu de «Je ne connais pas Corinne Savol».

Deux cas particulièrement intéressants et complètement implémentés seront décrits en utilisant l'exemple présenté dans l'introduction du mémoire, et qui est reproduit dans la figure 7.3: d'une part la gestion des «acknowledgements» implicites et d'autre part la suppression de la génération d'énoncés inutiles (parce que toutes les informations qu'ils pourraient communiquer sont déjà mutuellement connues).

Dans l'énoncé (4a), le système informe l'utilisateur qu'il ne connait pas Corinne Savol (CS). Ce fait implique que le système ne connait pas le numéro de CS non plus[2]. Quand le système énonce (4a), l'information ¬AvoirNuméro(Sys,CS) devient objet d'une supposition communicative et est intégrée dans l'espace de grounding: le système attend un «acknowledgement». Comme décrit dans le chapitre 4, ceci n'empêche pas le système de produire un deuxième énoncé (4b) sur la base de la supposition que l'utilisateur a compris le premier. Après que l'utilisateur ait produit son énoncé (5), le système va essayer de déterminer si cet énoncé fournit des «acknowledgements» des informations exprimées par (4a) et (4b). La réponse (5) continue le dialogue de manière cohérente et représente donc un «acknowledgement» impli-

2. Dans l'implémentation nous simplifions la tâche en supposant que, pour un standard téléphonique, «ne pas connaître» et «ne pas avoir le numéro» sont équivalents.

(1)	Utilisateur:	Je voudrais parler à Corinne Savol.
(2)	Standard:	Corinne Savol?
(3)	Utilisateur:	Oui.
(4a)	Standard:	Je ne connais pas Corinne Savol.
(4b)		Quel est son poste?
(5)	Utilisateur:	Je ne sais pas.
(6a)	Standard:	(Je n'ai pas le numéro de poste de Corinne Savol.)
(6b)		Savez-vous avec qui travaille Corinne Savol?
(7)	Utilisateur:	Elle travaille avec Louise Abou.
(8)	Standard:	Je vous passe Louise Abou.

FIG. 7.3 – *«Acknowledgements» implicites et suppression d'énoncés inutiles*

cite[3]. Alors que l'algorithme d'interprétation du chapitre 6 prédit que cet «acknowledgement» sera reconnu parce que les inférences prédictives du système permettent de trouver le lien avec le dernier énoncé, l'implémentation traite les «acknowledgements» implicites de manière encore simplifiée: si un énoncé ne fournit ni un «acknowledgement» explicite ni une demande de confirmation ni un signal d'incompréhension, l'auditeur est considéré comme ayant compris le dernier énoncé et donc comme ayant fourni un «acknowledgement» implicite. Cette heuristique est robuste pour des domaines où il n'y a pas de malentendus. Si par contre l'auditeur a mal compris, ne s'en rend pas compte et continue le dialogue, notre implémentation va faussement considérer son dernier énoncé comme «grounded». Ce défaut va être éliminé lors de l'intégration des processus décrits par l'algorithme de gestion d'interprétation du chapitre 6.

Dans l'exemple, c'est donc l'énoncé (5) qui fournit un «acknowledgement» implicite de $\neg$AvoirNuméro(Sys,CS). L'algorithme de mise à jour des croyances après la perception d'un acte de langage va déterminer que ce fait est maintenant grounded. Ensuite, le raisonneur sur le plan va déterminer l'échec du sous-plan qui consistait à trouver le numéro de poste de Corinne Savol et va donc essayer un nouveau plan: il va essayer de trouver un collaborateur de Corinne Savol. Mais l'ancien plan étant un plan partagé des deux agents, le raisonneur sur le plan va d'abord décider d'informer l'utilisateur de l'échec et envoyer au module de génération un message qui pourrait mener à la génération de (6a). Mais le *module de génération d'actes de langage* a été augmenté par un algorithme qui détermine pour un énoncé quelles croyances il peut produire chez l'auditeur. Cet algorithme renvoie que l'information communiquée par (6a) est déjà mutuellement connue et va alors supprimer cet énoncé.

3. Un «acknowledgement» *explicite* d'une question n'est que difficilement imaginable. Il devrait s'agir d'une *démonstration* dans le sens de Clark et Schaefer (1989) (cf. tableau 3.2), comme «Ah, vous voulez savoir si je connais Corinne Savol» – mais un tel exemple reste purement académique ...

Chapitre 8

Conclusion et perspectives

8.1 Résumé du travail

Ce mémoire a traité le phénomène de l'établissement de croyances mutuelles dans le dialogue. Partant de l'exemple d'un dialogue qui nous a montré la nécessité des croyances mutuelles pour la conversation, nous avons abordé le thème sous différents angles. Dans le chapitre 2, des travaux fondamentaux en philosophie, logique et intelligence artificielle ont été étudiés. Dans le chapitre 3, des travaux existants sur l'établissement des croyances mutuelles ont été présentés et critiqués. Partant de ces critiques, nous avons essayé dans le chapitre 4 d'expliquer le phénomène d'un point de vue psycholinguistique. Nous avons proposé des idées pour une théorie décrivant les changements dans les états mentaux des participants à un dialogue. Cette théorie a recours à des attitudes mentales (notamment la *supposition*) qui ne sont pas courantes en intelligence artificielle. Nous avons donc essayé dans le chapitre 5 de les caractériser formellement en définissant une logique multi-modale. Cette logique a été utilisée pour formaliser la théorie du chapitre 4. Nous avons montré comment les prédictions de la théorie peuvent être démontrées dans la logique à partir d'un ensemble de règles conversationnelles. Les résultats de ces preuves ont été à la base d'un modèle de gestion du dialogue qui a été décrit dans le chapitre 6. Nous y avons proposé une représentation des croyances et croyances mutuelles ainsi que des algorithmes permettant la mise à jour des croyances pendant le dialogue. Ce modèle de gestion de dialogue a été à la base d'une implémentation dans un système de dialogue qui nous a montré que, au moins dans les limites de cette application, notre modèle augmente les capacités d'un agent artificiel dans un dialogue.

Ce travail est un mémoire de DEA de *Sciences Cognitives* et présente en fait (presque) toutes les étapes du développement d'un *modèle cognitif*: une théorie psychologique est proposée, formalisée, transformée en un modèle cognitif et finalement implémentée dans un système où les prédictions de la théorie peuvent être testées en pratique. Dans le cadre d'un DEA, une telle approche «globale» entraine un défaut majeur: chaque partie reste nécessairement incomplète. La section suivante discute quelques éléments indispensables mais encore manquants, la section 8.3 présente des perspectives pour l'extension du travail et son application dans d'autres domaines.

8.2 Travaux restant à effectuer

La manque d'une validation empirique de la théorie a déjà été discutée en section 4.4, mais même une étude plus soigneuse des travaux pertinents en psycholinguistique devrait se révéler utile pour la justification de la théorie. Si une telle justification pouvait être faite, elle pourrait combler une lacune, notamment celle qui exitse entre les travaux de Clark et Marshall (1981) et de Clark et Schaefer (1989). Pendant que Clark et Marshall (1981) offrent une explication pour l'établissement de croyances mutuelles qui apparemment est trop simplifiée, Clark et Schaefer (1989) décrivent le processus plus complexe (et apparemment plus réaliste) de «grounding», mais n'offrent pas d'explication de la raison pour laquelle celui-ci devrait établir une croyance mutuelle. Le chapitre 4 a essayé de décrire l'établissement de croyances en tenant compte des évidences empiriques pour un processus complexe de «grounding», mais nos explications nécessitent un approfondissement et un ancrage dans d'autres théories psycholinguistiques.

Quant à la formalisation fournie dans le chapitre 5, la définition de la sémantique pour l'opérateur d'intention-que (IntTh) et son axiomatisation est la plus grande urgence. À moyen terme, un traitement explicite des actions doit être ajouté, c'est-à-dire la logique doit être enrichie d'une logique *dynamique*. Ayant intégré les intentions-que et les actions dans le formalisme, l'opérateur IntTo de Grosz et Kraus (1996) pourrait également être exprimé. Ce point atteint, l'intégration de notre théorie du «grounding» dans la théorie des plans partagés serait possible (probablement le plus facilement dans la version de Hunsberger (1999)). Un exemple pour une logique dynamique similaire à celle que nous avons en vue est présenté dans Singh et al. (1999).

Une structure correspondant aux *espaces de croyance* présentés dans le chapitre 6 doit être introduite pour représenter les *intentions* d'un agent. Des esquisses pour un tel modèle ont été faites et ont montré des propriétés intéressantes d'une représentation avec des *espaces d'intention*. Il semble que ces espaces pourraient représenter exactement les différentes conditions sur les intentions des agents qui sont nécessaires pour décrire que ces agents ont un plan partagé. Une telle approche permettrait donc une mise en œuvre assez directe du concept de plan partagé dans un modéle d'agent.

Le lien à la théorie de structuration du discours de Grosz et Sidner (1986) n'a été que vaguement indiqué dans le chapitre 6. La relation entre le *focus stack* et ce que nous avons appelé *espace de grounding* reste à éclaircir. En particulier, nous devrions spécifier comment le «grounding» influence l'enlèvement et le placement d'un *espace de focus* sur la pile et vice versa. Le changement du *focus* est un point «instable» dans le dialogue, parce que le «thème» du dialogue change aussi ce qui complique l'interprétation et donc le «grounding». Il serait important de déterminer comment les informations dans le *focus stack* peuvent être utilisées dans l'algorithme de mise à jour de croyance lors de l'interprétation (figure 6.5). D'autre part, l'état du processus de grounding pourrait aussi contraindre le changement du focus: il sera possible, par exemple, d'interdire d'enlever un élément du *focus stack* si une demande de confirmation restée sans réponse se trouve dans *l'espace de grounding*.

En ce qui concerne l'implémentation, en plus de la suppression des simplifications déjà mentionnées dans le chapitre 7, les améliorations les plus urgentes concernent d'une part la relation entre les croyances et le plan partagé et d'autre part l'utilisation explicite de croyances par le raisonneur sur le plan. Les représentations des croyances propositionnelles et des différentes autres croyances (sur des recettes, des entités et des actions) étant actuellement encore très hétérogènes, le raisonneur sur le plan ne peut pas vraiment se servir des croyances propositionnelles pour déterminer les prochaines actions.

8.3 Perspectives à long terme

Nous avons considéré dans ce travail le «grounding» comme un processus qui naît de la conscience des participants à un dialogue qu'ils ne peuvent pas toujours être certain d'être compris. Nous croyons que pour des raisons similaires, d'autres formes de «grounding» existent dans l'interaction entre agents et que donc la notion de «grounding» pourrait être généralisée.

Il est bien connu que certaines actions non langagières peuvent avoir des fonctions communicatives. Dans le système de Balkanski et Hurault-Plantet (1999), il s'agit, par exemple, des actions d'appeler le standard ou de décrocher. Il nous semble justifié de dire que le décrochement représente une sorte d'acknowledgement de l'appel: quand l'auditeur décroche, il montre qu'il a «compris» que quelqu'un l'appelle. D'autres actions n'ont pas de telles fonctions communicatives, notamment les actions qu'un agent effectue et que l'autre agent ne peut pas *percevoir*, comme, par exemple, la recherche d'un numéro de téléphone dans nos exemples. Nous avons vu que, si ces actions font partie du plan partagé, l'agent va essayer d'informer son partenaire sur les résultats (positifs ou négatifs) de telles actions. Nous croyons qu'il fait cela pour atteindre un autre *critère de grounding*: la croyance mutuelle sur le résultat de son action.

Une différence intéressante peut être remarquée ici: un agent va informer son partenaire sur le résultat de sa recherche pour un numéro de téléphone, mais il ne va pas dire «Je viens de vous appeler» – ceci est inutile parce que l'action elle-même exerce déjà cette fonction. Nous croyons que la différence consiste en la *perceptibilité* des deux actions. Pour une action qui n'est pas *perceptible*, les agents ne peuvent atteindre la croyance mutuelle sur son résultat que par l'intermédiaire d'un autre type d'action qui est (par définition) perceptible: les actes de langage. Un acte de langage peut donc être «grounded» par une réaction adéquate du partenaire et ce «grounding» peut entraîner le «grounding» du résultat de l'action non perceptible. Nous croyons que ce point de vue pourrait se révéler intéressant pour expliquer pourquoi et quand des énoncés sont produits pendant un dialogue.

En gros, ce problème revient à la question sur la réalisation d'un plan partagé: suffit-il que chaque agent exécute sa partie des actions du plan commun? Certainement pas. La collaboration nécessite la *coordination*; tout agent doit savoir jusqu'à quel point le plan commun a été réalisé et si les autres agents le savent aussi. Ce processus de coordination nécessite la *planification individuelle* de chaque agent: il doit planifier non seulement quelles actions du plan partagé il doit exécuter à un moment donné, mais aussi quelles informations il doit donner à l'autre agent, à quels évènements il doit s'attendre avant de réaliser une action, etc. Ce type de planification ayant nécessairement besoin d'intégrer les *perceptions* de l'agent, une planification du type STRIPS ne serait pas appropriée pour cette tâche. Levesque (1996) spécifie comment la perception peut être intégrée dans un modèle de planification. Lespérance et al. (1999) présentent un modèle d'agent qui intègre ces spécifications et décrivent comment des croyances peuvent être «produites» par les perceptions d'un agent. Il nous semble très intéressant d'étudier comment un tel modèle pourrait être utilisé pour décrire l'étalissement de croyances mutuelles dans l'évolution et la réalisation d'un plan partagé.

Bibliographie

AIRENTI, G., BARA, B. G., et COLOMBETTI, M. (1993). « Conversation and Behavior Games in the Pragmatics of Dialogue ». *Cognitive Science*, 17:197–256.

ALLEN, J. F. (1983). Recognizing Intentions From Natural Language Utterances. Dans BRADY, M. et BERWICK, R. C., éditeurs, *Computational Models of Discourse*, pages 107–166. MIT Press, Cambridge, MA.

AUSTIN, J. L. (1962). *How to Do Things With Words*. Harvard University Press.

BALKANSKI, C. T. et HURAULT-PLANTET, M. (1999). « Cooperative Requests and Replies in a Collaborative Dialogue Model ». *International Journal of Human-Computer Studies*, à paraître. Special Issue on???

BRATMAN, M. E. (1990). What Is Intention? Dans Cohen et al. (1990), Chapitre 2, pages 15–31.

BRATMAN, M. E. (1992). « Shared Cooperative Activity ». *The Philosophical Review*, 101:327–341.

BUNT, H. C. (1989). Information Dialogues as Communicative Action in Relation to Partner Modelling and Information Processing. Dans TAYLOR, M. M., NÉEL, F., et BOUWHUIS, D. G., éditeurs, *The Structure of Multimodal Dialogue*, volume 1, pages 47–71. Elsevier Science Publishers B.V.

BUNT, H. C. (1995). Dynamic Interpretation and Dialogue Theory. Dans TAYLOR, M. M., NÉEL, F., et BOUWHUIS, D. G., éditeurs, *The Structure of Multimodal Dialogue*, volume 2, pages 47–71. John Benjamins, Amsterdam.

CASTAING, M. (1993). « *Corpus de dialogues enregistrés dans un standard téléphonique* ». Notes et documents LIMSI 93-9, LIMSI/CNRS, Orsay, France.

CLARK, H. H. et MARSHALL, C. R. (1981). Definite Reference and Mutual Knowledge. Dans JOSHI, A. K., WEBBER, B. L., et SAG, I. A., éditeurs, *Elements of Discourse Understanding*. Cambridge University Press.

CLARK, H. H. et SCHAEFER, E. F. (1987). « Collaborating on Contributions to Conversation ». *Language and Cognitive Processes*, 2:1–23.

CLARK, H. H. et SCHAEFER, E. F. (1989). « Contributing to Discourse ». *Cognitive Science*, 13:259–294.

COHEN, P. R. et LEVESQUE, H. J. (1990a). « Performatives in a Rationnally Based Speech Act Theory ». Dans *Proceedings ACL-90*, pages 79–88.

COHEN, P. R. et LEVESQUE, H. J. (1990b). Persistence, Intention, and Commitment. Dans Cohen et al. (1990), Chapitre 3, pages 33–69.

COHEN, P. R. et LEVESQUE, H. J. (1990c). Rational Interaction as the Basis for Communication. Dans Cohen et al. (1990), Chapitre 12, pages 221–255.

COHEN, P. R., MORGAN, J., et POLLACK, M. E., éditeurs (1990). *Intentions in Communication*. MIT Press, Cambridge, MA.

DI EUGENIO, B., JORDAN, P. W., THOMASON, R. H., et MOORE, J. D. (1999). « The Agreement Process: An Empirical Investigation of Humand-Human Computer-Mediated Collaborative Dialogues ». *International Journal of Human-Computer Studies*, à paraître. Special Issue on???

FAGIN, R., HALPERN, J. Y., MOSES, Y., et VARDI, M. Y. (1995). *Reasoning About Knowledge*. MIT Press, Cambridge, MA.

FIKES, R. E. et NILSSON, N. J. (1971). « STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving ». *Artificial Intelligence*, 2:189–208.

GENESERETH, M. R. et NILSSON, N. J. (1987). *Logical Foundations of Artificial Intelligence*. Morgan Kaufman.

GRICE, H. P. (1957). « Meaning ». *Philosophical Review*, 66:377–88.

GRICE, H. P. (1975). Logic and Conversation. Dans COLE, P. et MORGAN, J. L., éditeurs, *Syntax and Semantics 3: Speech Acts*, pages 41–58. Academic Press, New York.

GROSZ, B. J. et KRAUS, S. (1996). « Collaborative Plans for Complex Group Action ». *Artificial Intelligence*, 86(2):269–357.

GROSZ, B. J. et KRAUS, S. (1999). The Evolution of SharedPlans. Dans Rao et Wooldridge (1999).

GROSZ, B. J. et SIDNER, C. L. (1986). « Attention, intention, and the structure of discourse ». *Computational Linguistics*, 12(3):175–204.

GROSZ, B. J. et SIDNER, C. L. (1990). Plans for Discourse. Dans Cohen et al. (1990), Chapitre 20, pages 417–444.

HALPERN, J. Y. et MOSES, Y. (1990). « Knowledge and Common Knowledge in a Distributed Environment ». *Journal of the ACM*, 37(3):549–587.

HALPERN, J. Y. et MOSES, Y. (1992). « A Guide To Completeness and Complexity for Modal Logics of Knowledge and Belief ». *Artificial Intelligence*, 54.

HINTIKKA, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press.

HUNSBERGER, L. (1999). « Making SharedPlans More Concise and Easier to Reason About ». Dans MULLER, J., SINGH, M., et RAO, A., éditeurs, *Intelligent Agents V – Agent Theories, Architectures and Languages (ATAL-98)*.

KEEFE, D. E. et MCDANIEL, M. A. (1993). « The Time Course And Durability of Predictive Inferences ». *Journal of Memory and Language*, 32:446–463.

KINTSCH, W. (1988). « The Role of Knowledge in Discourse Comprehension: A Construction-Integration model ». *Psychological Review*, 95:163–182.

KONOLIGE, K. et POLLACK, M. E. (1993). « A Representationalist Theory of Intention ». Dans *Proceedings of the 1993 International Joint Conference on Artificial Intelligence (IJCAI-93)*, pages 390–395.

LAIRD, J. E., ROSENSBLOOM, P. S., et NEWELL, A. (1986). « Chunking in SOAR: the Anatomy of a General Learning Mechanism ». *Machine Learning*, 1:11–46.

LESPÉRANCE, Y., LEVESQUE, H. J., et REITER, R. (1999). A Situation Calculus Approach to Modelling And Programming Agents. Dans Rao et Wooldridge (1999).

LEVESQUE, H. J. (1996). « What Is Planning in the Presence of Sensing? ». Dans *Proceedings of the 13th Nat. Conf. on AI (AAAI-96)*, Portland, Oregon.

LEVINSON, S. C. (1983). *Pragmatics.* Cambridge University Press.

LOCHBAUM, K. E. (1998). « A Collaborative Planning Model of Intentional Structure ». *Computational Linguistics*, 24(4):525–572.

MOESCHLER, J. et REBOUL, A. (1994). *Dictionnaire Encyclopédique de Pragmatique.* Éditions du Seuil.

PERRAULT, C. R. (1990). An Application of Default Logic to Speech Act Theory. Dans Cohen et al. (1990), Chapitre 9, pages 161–185.

POLLACK, M. E. (1990). Plans as Complex Mental Attitudes. Dans Cohen et al. (1990), Chapitre 5, pages 77–103.

RAO, A. S. et WOOLDRIDGE, M., éditeurs (1999). *Foundations and Theories of Rational Agents.* Kluwer.

SADEK, M. D. (1991). « Dialogue Acts Are Rational Plans ». Dans *Proceedings of the ESCA/ETR Workshop on Multi-Modal Dialogue*, Maratea, Italy.

SCHIFFER, S. R. (1972). *Meaning.* Clarendon Press, Oxford.

SEARLE, J. R. (1969). *Speech Acts.* Cambridge University Press, New York.

SEARLE, J. R. (1990). Collective Intentions and Actions. Dans Cohen et al. (1990), Chapitre 19, pages 401–415.

SINGH, M. P., RAO, A. S., et GEORGEFF, M. P. (1999). Formal Methods in DAI: Logic-Based Representation and Reasoning. Dans WEISS, G., éditeur, *Multiagent Systems: A Modern Approach*, Chapitre 8. MIT Press, Cambridge, MA.

TRAUM, D. R. (1994). « *A Computational Model of Grounding in Natural Language Conversation* ». PhD thesis, Department of Computer Science, University of Rochester.

TRAUM, D. R. et HINKELMAN, E. A. (1992). « Conversation Acts in Task-Oriented Spoken Dialogue ». *Computational Intelligence*, 8(3):575–599. Special Issue on Non-literal Language.