

# Dialogue et dynamique des croyances

Dominique Longin\* et David Sadek†

\* Groupe de Logique Appliquée  
Institut de Recherche en Informatique de Toulouse  
118 Route de Narbonne, 31062 Toulouse Cedex 4  
Dominique.Longin@irit.fr  
[http://www.irit.fr/ACTIVITES/EQ\\_ALG/](http://www.irit.fr/ACTIVITES/EQ_ALG/)

† France Télécom R & D  
DIH/D2I  
Technopôle Anticipa – 2, avenue Pierre - Marzin  
22307 Lannion Cedex – France  
David.Sadek@rd.francetelecom.fr

## 1 INTRODUCTION

Parce que le monde est fondamentalement dynamique, nous sommes sans cesse confrontés à de nouvelles informations sur son état (passé, présent, ou futur). L'un des « canaux d'échanges » privilégié avec notre environnement est celui de la communication verbale, qui nous permet non seulement de percevoir des informations en provenance de ce dernier, mais également de le transformer. Au travers d'un type de dialogues particulier, nous proposons dans ce qui suit de faire un rapide tour d'horizon des approches dites « intentionnelles » (*i.e.* relatives à une théorie de l'Intentionnalité [43, 46, p. 83]) du dialogue, en nous focalisant sur la forte interaction qui existe entre l'évolution de l'*état intentionnel* (ou *état mental*) d'un agent et la dynamique du monde dans lequel cet agent évolue<sup>1</sup>.

Le dialogue coopératif personne-machine, qu'il soit en langage naturel ou artificiel, est l'un des challenges les plus importants des sciences informatiques. Dans le cadre de dialogues orientés-tâche, les participants ont un but commun principal, qui est celui d'accomplir la (ou les) tâche(s) en ques-

---

1. Nous appelons *monde*, tout ce qui est extérieur à l'état mental d'un agent donné (point de vue subjectif). En particulier, cela inclut l'état intentionnel de tous les autres participants au dialogue. Nous utilisons également comme synonyme le terme d'*environnement*.

tion. Chaque participant a des informations permettant l'accomplissement de sous-buts, mais aucun d'eux ne peut accomplir le but principal tout seul.

Les systèmes de dialogue développés aujourd'hui par France Télécom R&D comme des instanciations de sa technologie générique ARTIMIS d'agent rationnel [35, 41, 40], permettent d'ores et déjà de gérer en temps réel un dialogue coopératif en langage naturel. L'approche sous-jacente consiste d'abord à décrire le comportement d'un agent par une théorie logique de l'interaction rationnelle [36, 37], puis à mettre en œuvre cette théorie par un système d'inférence [6, 5, 41], qui constitue le noyau du composant central d'ARTIMIS, l'unité rationnelle. Pour rendre compte de l'interaction entre un agent et son environnement, les activités du système de dialogue sont doubles : observer et interpréter les interventions du locuteur, et calculer et générer les réactions appropriées. Dans son état actuel, la théorie décrit complètement le processus de réaction. Elle fournit également une spécification formelle générique du processus de *reconstruction des croyances* à l'issue de l'observation d'un événement (tel un acte communicatif). Cette spécification laisse toutefois ouverts différents aspects relatifs à la problématique du *changement de croyances*.

Nous nous focalisons ici sur l'évolution des croyances<sup>2</sup> d'un agent participant à un dialogue. Chaque intervention accomplie par l'un des participants conduit l'ensemble des participants à modifier une partie de leurs croyances.

Nous illustrons, à l'aide d'un exemple, les problèmes rencontrés<sup>3</sup> (Sect. 2). Nous mettons en évidence la part de ces problèmes qui est résolue par quelques unes des principales contributions du domaine, et celle qui ne l'est pas (Sect. 3). Après avoir brièvement discuté des théories formelles de la révision (Sect. 4), nous formulons quelques conclusions et perspectives (Sect. 5).

## 2 EXEMPLE DE DIALOGUE

Deux agents (le système *s* et l'utilisateur *u*) négocient la vente (resp. l'achat) d'un billet de train :

---

2. Le terme de *croyance(s)* peut être pris soit dans un sens restreint (littéral), soit dans un sens plus large, où tout agent ayant une attitude mentale pense qu'il a cette attitude mentale, et inversement. C'est ce dernier sens qui nous autorise à désigner l'ensemble des attitudes mentales (croyance, intention, ...) par le terme de *croyances*. Par la suite, nous employons le terme au singulier pour désigner le sens restreint et au pluriel pour le sens large.

3. De nombreux exemples similaires peuvent être trouvés dans [16].

*s*<sub>1</sub> : Bonjour. Que désirez-vous?  
*u*<sub>1</sub> : Un billet en première classe pour Marseille, s'il vous plaît.  
*s*<sub>2</sub> : 150 €.

*u*<sub>2</sub> : Hum... pardon, en seconde classe, s'il vous plaît.  
*s*<sub>3</sub> : 95 €.

*u*<sub>3</sub> : Puis-je payer les 85 € par carte bleue ?  
*s*<sub>4</sub> : Le prix n'est pas 85 € mais 95 €. Oui, vous pouvez payer par carte bleue.

*u*<sub>4</sub> : ...

Cet exemple illustre le fait que, dans un dialogue, les agents peuvent changer d'avis, se tromper (*i.e.* accomplir un énoncé différent de celui qu'ils souhaitaient accomplir), mal comprendre un énoncé (*i.e.* reconnaître un énoncé différent, d'un point de vue linguistique, de celui accompli), ou mal l'interpréter (*i.e.* lui donner un sens différent de celui souhaité initialement par le locuteur)<sup>4</sup>. Il est essentiel que de tels phénomènes soient pris en compte lors de la modélisation de l'évolution des états mentaux des agents.

Dans les dialogues considérés, nous étudions plus particulièrement l'évolution des croyances de l'auditeur<sup>5</sup>, qui doit être capable : (1) d'accepter des informations d'un certain type (par exemple, les informations sur la destination et la classe de transport dans *u*<sub>1</sub>) ; (2) de déduire des informations supplémentaires non directement contenues dans l'énoncé, en utilisant des connaissances relatives au domaine d'application (cf. la déduction du prix dans *s*<sub>2</sub>, étant données les informations contenues dans *u*<sub>1</sub>) ; (3) de continuer à croire une partie de ce qu'il croyait avant l'accomplissement du dernier énoncé (dans *u*<sub>2</sub> par exemple, la destination n'étant pas précisée de nouveau, l'agent *s* doit la croire inchangée à l'issue de cet énoncé : il est alors capable de calculer le nouveau prix – cf. *u*<sub>3</sub>) ; (4) d'accepter éventuellement des informations contradictoires avec ses propres croyances actuelles (pour pouvoir prendre en compte les changements d'avis et/ou les corrections de son interlocuteur, comme dans *u*<sub>2</sub>) ; (5) de refuser parfois de prendre en compte

4. Ceci nous conduit à préciser que, selon la dissociation de Kant entre les conditions objectives du savoir véritable, et les conditions purement subjectives du tenir-pour-vrai, il est d'usage de différencier respectivement le *savoir* (ou la *connaissance*) de la *croyance*. Faisant l'hypothèse que le canal de communication (perception, compréhension, ...) est imparfait, c'est ce dernier concept qui est choisi pour modéliser les participants à un dialogue.

5. En fait, comme nous l'avons montré dans [28], certains formalismes (dont les nôtres) rendent également compte de l'évolution des croyances de l'auteur de l'acte. Intuitivement, il suffit de considérer ce dernier comme un observateur particulier de cet acte.

certaines informations, en particulier si l'interlocuteur donne des informations à propos de faits sur lesquels il n'est pas compétent (cf.  $u_3$  et  $s_4$ ).

Partant de ces considérations, nous passons en revue, dans la section suivante, les principales approches au travers de l'éclairage de notre exemple.

### 3 LES CHANGEMENTS DE CROYANCES

Les travaux sur la *révision* [1] et la *mise à jour* [26, 25] de connaissances (ou de croyances) montrent que pour décrire de tels processus, on doit disposer d'au moins un ensemble d'informations et une nouvelle information à ajouter à cet ensemble. Bien que la révision et la mise à jour correspondent sémantiquement à des opérations différentes, nous les désignons dans ce qui suit par une expression neutre unique : un *changement de croyances*. Nous reviendrons sur la révision et la mise à jour à la fin de cette section.

Parallèlement, les travaux entrepris sur la notion d'agent ont conduit à des approches fondées sur une *théorie de l'Intentionnalité* [43, 4]. Dans cette théorie, un *agent rationnel* est représenté par son *état mental*, qui est l'ensemble des informations rendant compte de ses différentes *attitudes mentales* (*i.e.* croyances, buts, intentions, ...) <sup>6</sup>.

Une telle représentation s'inscrit au sein de théories formelles à deux volets : *l'équilibre rationnel* et *l'interaction rationnelle* [8, 9, 36, 37]. Le premier volet, au travers des propriétés des attitudes mentales et de l'action, décrit l'ensemble des relations qui doivent être maintenues, d'une part, entre les différentes attitudes mentales d'un agent et, d'autre part, entre ses attitudes mentales et plans et actions. Le second volet caractérise la composante « relation inter-agent » dans un environnement multi-agent (communication, coopération, ...).

Nous nous focalisons dans ce qui suit sur les capacités de certaines de ces théories à prendre en compte l'évolution du monde (au cours d'un dialogue) dans lequel les agents qu'elles formalisent évoluent. Les différents énoncés réalisent des actes de langage [3, 45, 44].

#### *Cohen et Levesque.*

COHEN et LEVESQUE [8, 9] ont jeté les bases formelles générales de telles théories. Selon eux, une théorie de la conversation devrait *expliquer la co-*

---

6. Cette théorie est à la base de ce qui est communément appelé les *architectures BDI* (pour *Belief-Desire-Intention*).

*hérence du dialogue en termes d'états mentaux des participants.* Ils souhaitent ainsi généraliser la théorie des actes de langage en une théorie de la communication, où les propriétés des actes de langage seraient dérivables de principes (plus généraux) de rationalité.

Dans leur approche, l'ensemble d'informations associé à un agent est un état mental et la nouvelle information correspond à une intention du locuteur de produire un certain effet sur l'auditeur (ces effets étant liés à l'énoncé accompli). La sincérité des agents n'est pas postulée et constitue le critère d'acceptation de cette nouvelle information.

Il est alors intéressant de remarquer qu'à ce niveau, l'agent doit avoir reconnu l'acte accompli par son interlocuteur, éventuellement de façon indirecte<sup>7</sup>. Cette approche présuppose donc tout un travail d'interprétation de l'énoncé effectivement accompli, travail non décrit dans la théorie.

En tout état de cause, l'approche de COHEN et LEVESQUE permet donc à un agent, soit de rejeter, soit d'accepter, la nouvelle information. En ce sens, elle prend en compte une évolution des états mentaux au cours d'un dialogue.

Néanmoins, leur théorie souffre du problème du décor (*frame problem*) [29] : l'acceptation d'une nouvelle information se fait au prix de l'abandon des croyances que l'agent possédait avant l'accomplissement de l'acte<sup>8</sup>. Dans notre exemple (Sect. 2), le système ainsi formalisé ne pourrait préserver la destination tout au long de la conversation, ce qui obligerait l'interlocuteur à fournir systématiquement, et en un seul énoncé, toutes les informations nécessaires (par exemple, l'énoncé  $u_2$  devrait, en plus de la nouvelle classe souhaitée, spécifier à nouveau la destination). Il est évident qu'un tel état de chose est peu souhaitable dans un système de dialogue (où les informations à fournir par l'utilisateur peuvent être quantitativement très importantes).

### ***Perrault.***

PERRAULT [30] essaie de résoudre le problème du décor dans un cadre fondé sur la logique des défauts de REITER [34]. Mais sa théorie souffre

---

7. L'énoncé « Passe-moi le sel » constitue une façon directe de demander le sel, au sens où le *sens de l'énoncé* correspond à ce que SEARLE appelle le *sens du locuteur*, *i.e.* au sens que le locuteur souhaite donner à son énoncé. En revanche, dans « Peux-tu me passer le sel? », le sens de l'énoncé et celui du locuteur ne coïncident pas : alors que le dernier correspond à une requête (identique à l'énoncé direct ci-dessus), le premier constitue littéralement une question où la réponse attendue est *oui* ou *non*. On dit alors que la requête a été accomplie de façon indirecte, ou encore que l'acte accompli est un acte indirect. Pour une ébauche du traitement des indirections, voir [5, 24, 14, 13].

8. Il s'agit de l'opération dite « triviale » de [1].

de plusieurs problèmes dont la plupart ont été mis en valeur par APPELT et KONOLIGE [2]. En particulier, les agents modélisés par la théorie de PERRAULT, ne remettent jamais en question leurs anciennes croyances et peuvent seulement faire des *expansions* (au sens de [1]) de leurs croyances. Ainsi, de tels agents ne pourraient, dans le cadre de notre exemple, prendre en compte l'énoncé  $u_2$ , où l'utilisateur abandonne l'idée de voyager en première classe, au profit d'un voyage en seconde. Un autre problème est que les agents peuvent s'auto-convaincre de la véracité d'un fait simplement en l'assertant.

### ***Appelt et Konolige.***

APPELT et KONOLIGE [2] recommandent l'utilisation d'une logique auto-épistémique hiérarchique (HAEL). En bref, l'avantage par rapport à une logique comme celle de PERRAULT est que, dans HAEEL, les « défauts » sont stratifiés et que l'on peut donc contrôler l'ordre de leur application. Cela peut être utilisé dans le but de supprimer certaines extensions non souhaitées.

Leur théorie peut représenter notre exemple correctement. Néanmoins, en laissant de côté l'outil logico-mathématique relativement complexe pour y arriver et une représentation des énoncés quelque peu contre-intuitive, il semble que leur processus de changement de croyances souffre d'un problème similaire à l'un de ceux rencontrés par PERRAULT.

Par exemple, supposons que l'agent  $j$  n'ait pas d'opinion à propos du temps qu'il fait, et que l'agent  $i$  l'informe qu'il fait beau. Alors sous certaines conditions, dans HAEEL,  $j$  adoptera la croyance selon laquelle il fait beau (*i.e.* « il fait beau » fera partie de ses croyances). Mais si l'on suppose maintenant que  $i$  informe  $j$  que «  $j$  pense qu'il fait beau », alors (toujours dans HAEEL),  $j$  devra aussi incorporer ce genre d'information dans son propre état mental. Ceci entre clairement en conflit avec notre intuition: au plus,  $j$  devrait-il penser que  $i$  considère que  $j$  pense qu'il fait beau.

Comme montré dans [24], la seule voie pour supprimer un tel effet indésirable semble être de déplacer l'ignorance de  $j$  à propos du temps qu'il fait au niveau 0 de la hiérarchie d'HAEL. Mais dans ce cas, ce niveau contenant les croyances de l'agent qui ne peuvent être changées, toute tentative d'accepter une assertion sur le temps qu'il fait sera bloquée, ce qui entre une fois de plus en conflit avec notre intuition (on retrouve le problème de PERRAULT selon lequel l'agent ne peut plus changer de croyance à propos du temps qu'il fait).

**Sadek.**

SADEK introduit la notion de *reconstruction des croyances* [38, 36]. Le point de départ est ce qu'il appelle le *principe d'observation*, énoncé en trois postulats : (1) l'observation en tant qu'expérience perceptive est sûre ; (2) l'observation est possiblement non conforme à la réalité ; (3) lors d'une interaction, la cohérence des observations faites par un agent est une condition nécessaire et suffisante à la réussite (subjective) de la communication. Ce principe lui permet, en particulier, d'exprimer une distinction formelle entre une observation et l'« événement » qu'elle réalise et d'introduire notamment la notion d'*admissibilité* de l'action observée. SADEK formule sa théorie en quatre mécanismes : la mémoire et la persistance (concepts présents également chez PERRAULT, bien qu'exprimés différemment), ainsi que l'admission et la consommation : (1) la *mémoire* sert à l'agent pour se rappeler ce qu'il croyait juste avant l'accomplissement du dernier énoncé ; (2) l'*admission* permet de réunir les conditions favorables à l'acceptation de la nouvelle information (y compris au prix d'une révision des croyances) et, le cas échéant, de l'accepter ; (3) la *consommation* caractérise l'étape où les effets de l'acte (et ses qualifications persistantes) deviennent des croyances de l'agent ; (4) et la *persistance* consiste à conserver toutes les anciennes croyances consistantes avec le nouvel état engendré.

La théorie de Sadek permet de traiter notre exemple. Par ailleurs, elle présente un caractère générique, dans le sens où elle ne prescrit pas de stratégie particulière dans le cas où un acte ne peut être admis au vu des croyances courantes. En effet l'axiome d'admission laisse explicitement ouvert le choix à l'agent entre l'acceptation de l'acte qui vient d'être accompli et la remise en cause d'informations antérieures qui empêchent de l'accepter. Sur un tout autre plan, toutefois, l'interprétation autoépistémique qui est faite de la croyance dans l'axiome de persistance notamment, ne donne pas directement d'indication sur le processus de déduction sous-jacent.

La reconstruction des croyances s'appuie sur une formalisation originale des actes communicatifs [39] (ou actes de langage, dans l'esprit de SEARLE [45]). SADEK distingue les types d'effets suivants : l'*effet indirect* d'un acte est la persistance (à l'issue de l'accomplissement de celui-ci) de ses préconditions ; l'*effet intentionnel* correspond au point de vue gricéen de la communication [20], à savoir l'intention que le locuteur tente de rendre manifeste en accomplissant l'acte ; l'*effet rationnel* est l'effet *recherché* de l'acte par son auteur.

Par exemple, supposons que l'utilisateur  $u$  informe le système  $s$  que

sa destination est Marseille. Les préconditions de cet acte sont les suivantes :  $u$  pense que sa destination est effectivement Marseille (précondition de capacité);  $u$  ne pense pas que  $s$  sait si la destination de  $u$  est Marseille (précondition de pertinence au contexte). Les effets de cet acte sont les suivants : l'effet rationnel, *i.e.* l'effet attendu de l'acte par  $u$  sur  $s$ , est que  $s$  pense que la destination est Marseille; l'effet intentionnel traduit le fait que  $u$  a l'intention que  $s$  pense que  $u$  a l'intention de produire sur  $s$  un certain effet (l'effet rationnel); l'effet indirect correspond à la persistance des préconditions de capacité et de pertinence au contexte<sup>9</sup>.

### ***Herzig et Longin.***

Dans cette approche [12, 28, 21, 22, 23], un changement de croyances est réalisé par l'*adoption* et la *préservation* de croyances, toutes deux fondées sur le concept de topique. Ici, la nouvelle information correspond au fait qu'un acte de langage vient juste d'être accompli. L'acte accepté peut ou non être l'acte réellement accompli. Contrairement à SADEK ou GALLIERS [15] qui donnent à l'auditeur la possibilité de rejeter la nouvelle information, celle-ci est toujours acceptée<sup>10</sup>. En revanche, l'agent peut n'accepter qu'un sous-ensemble des conséquences de cette admission.

Dans ce but, HERZIG et LONGIN procèdent en deux étapes : l'auditeur consomme<sup>11</sup>, *via* des lois du domaine liant l'accomplissement d'un acte et ses effets, les effets indirect et intentionnel (au sens de SADEK), mais pas forcément toutes leurs conséquences : l'acceptation de ces dernières ne sera effective que si l'auditeur croit le locuteur compétent sur le ou les domaines sur le(s)quel(s) elles portent. Comme chez SADEK, l'effet rationnel n'est réellement consommé en tant que tel que s'il est dérivable de l'état mental de l'auditeur à l'issue du changement de croyances. Ainsi, la *compétence* du locuteur est ici le critère pour déterminer quelle part des conséquences de

9. Si, intuitivement, la persistance de la première semble toujours réalisée, il apparaît que préserver la seconde à l'issue de l'acte constitue un choix *a priori* de l'agent, selon lequel celui-ci considère que, en l'absence d'informations nouvelles, son acte n'a pas atteint son but. Dans [24], les auteurs procèdent différemment en considérant un état transitoire d'ignorance.

10. Bien qu'elle soit différente, cette phase correspond, dans son esprit, à la phase d'admission de SADEK [38]. Aussi, par abus de langage, parlons-nous ici parfois de l'*admission d'un acte* (ou de l'*admission d'une nouvelle information*).

11. Ce terme est employé par SADEK [36] pour décrire l'ajout, à l'état mental de l'agent, de la croyance correspondant à l'effet de l'acte. Là encore, bien que reprenant le terme, la phase de consommation de HERZIG et LONGIN est différente de celle décrite par l'axiome du même nom chez SADEK (cf. [28] pour plus de détails).



l'acceptation de la nouvelle information est consommée par l'auditeur.

Dans l'exemple (Sect. 2),  $s$  accepte une des conséquences de l'accomplissement de  $u_2$  (l'information sur la nouvelle classe de transport), mais rejette une des conséquences de celui de  $u_3$  (la proposition de prix par l'utilisateur), car il considère que  $u$  est compétent sur les classes mais pas sur les prix. Formellement, dans cette approche, la compétence est une relation (métalinguistique) de dépendance entre un agent et une formule.

Que l'agent soit en mesure d'acquérir de nouvelles croyances à l'issue d'un acte ne suffit pas : encore doit-il pouvoir préserver certaines des croyances qu'il possédait juste avant (problème du décor). Le concept clé est ici une *relation d'influence* : s'il existe une telle relation entre l'acte accompli et une certaine croyance, alors celle-ci ne peut être préservée dans le nouvel état mental. Dans l'exemple, l'ancienne classe de transport ne peut être préservée après  $u_2$  car l'acte d'information sur la classe influence les croyances de l'auditeur portant sur les classes. D'un autre côté, la destination n'est pas influencée par  $u_2$  et peut ainsi être préservée.<sup>12</sup>

Tout cela présuppose la capacité de déterminer la (dépendance associée à la) compétence d'un agent et l'influence d'un acte de langage. Contrairement aux approches précédentes du raisonnement sur les actions (où le concept d'influence est primitif), le fondement de ces deux notions est ici fourni par le concept de topique : l'idée est qu'à tout agent, acte de langage, et formule, on peut associer des ensembles de topiques<sup>13</sup>.

Ainsi, un agent est compétent sur une formule  $A$  si et seulement si l'ensemble des topiques associés à  $A$  est un sous-ensemble de l'ensemble des topiques associés à cet agent (l'ensemble des topiques sur lesquels l'agent est compétent). De même, une formule  $A$  est préservée après l'accomplissement d'un acte de langage  $\alpha$  s'il n'y a aucun topique commun à l'ensemble des topiques associés à  $A$  et à l'ensemble des topiques associés à  $\alpha$ .

La notion de topique utilisée par HERZIG et LONGIN constitue une extension de celle d'EPSTEIN [11, p. 68] qui définit le *subject matter* d'une proposition  $A$ , et montre alors que deux propositions peuvent être définies comme étant en relation (de dépendance) si elles ont des *subject matters* en

---

12. Le concept d'influence (ou de dépendance) d'un acte est proche des notions qui ont été étudiées dans le domaine du raisonnement sur les actions, afin de résoudre le problème du décor (cf. par exemple l'occlusion de SANDEWALL [42], la relation d'influence de THIELSCHER [47], les opérateurs de changements possibles de GIUNCHIGLIA *et al.* [17], ou encore [7]).

13. Un topique correspond à ce dont parle quelque chose. Par exemple, un topique de « Il fait beau » peut être *le temps qu'il fait*.

commun<sup>14</sup>. Se fondant sur ce principe, HERZIG et LONGIN définissent un certain nombre d'axiomes permettant de calculer les topiques liés à un agent, un acte de langage, et une formule. *Via* l'utilisation des résultats d'EPSTEIN, les dépendances sont ainsi définies de façon non primitive.

### *Rao et Georgeff.*

RAO et GEORGEFF [31, 32] ont proposé des théories et des architectures pour les agents rationnels, ainsi qu'une procédure de preuve par la méthode des tableaux pour leurs logiques [33]. De façon similaire à STRIPS, les actions et les plans sont représentés par leurs préconditions, des « add-lists » et des « delete-lists ». Ces listes sont restreintes à des ensembles de formules.

Les problèmes liés à un tel cadre de travail sont bien connus : on ne peut *a priori* représenter des actions ayant des effets non déterministes ou indirects. Ainsi, on ne peut utiliser de connaissances sur le domaine pour dériver, par exemple, le prix du billet en fonction d'une destination et d'une classe de transport, comme c'est requis pour accomplir  $s_2$ . Plus important, les actions ne peuvent avoir que des effets factuels (*i.e.* représentés par des formules logiques sans modalité) : cela exclut la manipulation d'actes de langage dont les effets sont typiquement représentés par une imbrication d'opérateurs intensionnels (de croyance et intention, notamment).

## 4 LES THÉORIES FORMELLES AGM ET KM

Comme nous l'avons dit dans la section précédente, il existe d'importantes analyses formelles du changement de croyances. Les plus importantes sont la *révision* d'ALCHOURRÓN, GÄRDENFORS et MAKINSON (AGM) et la *mise à jour* de KATSUNO et MENDELZON (KM) [1, 26, 25]. La première décrit l'évolution des croyances dans un monde statique qui se précise peu à peu, au fur et à mesure que de nouvelles informations sont ajoutées. La seconde, au contraire, décrit les changements d'un monde dynamique au fur et à mesure que ceux-ci s'opèrent [26, 19]. Ainsi, la distinction entre ces deux opérations est fondée sur différentes hypothèses relatives à la nature de la nouvelle information. Cette distinction est reflétée par des postulats différents, et des sémantiques différentes sont données respectivement pour les opérateurs de révision et de mise à jour.

---

14. D'autres études de cette notion de topique existent dans la littérature logico-philosophique, en particulier [27, 18, 10].

En transposant cette distinction entre révision et mise à jour au contexte d'un dialogue, quand l'auditeur réalise que le locuteur a changé d'avis (le monde réel a évolué), alors l'auditeur doit *mettre à jour* ses croyances ; quand l'auditeur réalise qu'il a mal compris une information en provenance de son interlocuteur (le monde réel n'a pas évolué), alors l'auditeur doit les *réviser*.

Ainsi, les approches intentionnelles précédentes réalisent différentes opérations des théories formelles AGM et KM : la révision triviale [1] pour COHEN et LEVESQUE (puisque si de nouvelles croyances sont adoptées, elles remplacent purement et simplement toutes les précédentes – rien n'est gardé) ; l'expansion pour PERRAULT (car la conjonction de ses axiomes de mémoire et de persistance fait que toute ancienne croyance est préservée dans le nouvel état, son critère d'admission étant la consistance entre la nouvelle information et ce nouvel état) ; la révision pour SADEK (pour qui une nouvelle information inconsistante avec l'ancien état passe par une remise en cause de cet état) ; enfin, la mise à jour pour APPELT et KONOLIGE et pour HERZIG et LONGIN (puisque'en cas de compétence du locuteur sur la nouvelle information, celle-ci remplace les anciennes croyances qui sont en contradiction avec elle).

Au moins trois difficultés surgissent lorsqu'il s'agit d'appliquer ces théories à celles dites intentionnelles. Premièrement, il apparaît que dans beaucoup d'exemples de dialogue, l'auditeur est dans l'incapacité de décider si la nouvelle information à laquelle il est soumis correspond à un changement du monde réel<sup>15</sup> ou non. Dans notre exemple, après  $u_2$ , le système peut être incapable de distinguer le cas où l'utilisateur a changé d'avis, de celui où il a mal compris l'énoncé  $u_1$  (cf. [25] pour une critique de KM).

Deuxièmement, certains postulats sont inappropriés ou inadaptés au cas du dialogue. En particulier, la nature sur-informative de certaines informations est négligée (en relation avec le postulat  $(K \circ A) \leftrightarrow K$  si  $K \rightarrow A$ ) ; la nouvelle information appartient toujours au nouvel état ( $(K \circ A) \rightarrow A$ ), ce qui va à l'encontre de certaines approches, ou n'est pas assez explicite pour celles où toutes les conséquences de la nouvelle information ne doivent pas forcément être ajoutées.

Troisièmement, une seule attitude mentale est traitée (la croyance), et aucune distinction n'est faite entre ses niveaux *factuel* (croyance portant sur un fait), *introspectif* (croyance portant sur ses propres croyances), et *alterné* (croyance portant sur les croyances des autres).

---

15. Cf. la note de bas de page n° 1.

## 5 CONCLUSIONS ET PERSPECTIVES

Au travers d'un exemple, nous avons mis en évidence les deux processus de base liés à l'évolution des croyances d'agents participant à un dialogue : la préservation de certaines croyances et l'adoption de certaines autres.

Comme nous l'avons aussi montré, puisque seulement un sous-ensemble des croyances mises en jeu est concerné par l'un ou l'autre de ces processus, il est nécessaire d'introduire un certain « critère » de discrimination entre les croyances. En ce qui concerne l'adoption de nouvelles croyances, ce rôle est parfois attribué à la sincérité (COHEN et LEVESQUE), la consistance (PERRAULT, et d'une certaine façon SADEK), ou la compétence (SADEK, HERZIG et LONGIN, et d'une certaine façon APPELT et KONOLIGE). En revanche, pour ce qui est de la préservation d'anciennes croyances, le critère en question est soit l'influence d'un acte sur une attitude mentale (HERZIG et LONGIN), soit la consistance (dans tous les autres cas, excepté COHEN et LEVESQUE où rien n'est dit à ce sujet).

En définitive, du point de vue du changement de croyances, les approches de COHEN et LEVESQUE et de PERRAULT sont incomplètement spécifiées. Celle de SADEK, dans le cas où l'information ne peut être acceptée au vu des croyances courantes, ouvre le choix à l'agent entre l'acceptation de la nouvelle information et la mise en cause des raisons (antérieures) qui bloquent son acceptabilité. Enfin, celles de APPELT et KONOLIGE et de HERZIG et LONGIN privilégient, dans leur explication du conflit de croyances, l'hypothèse d'un changement d'avis de la part du locuteur (la nouvelle information correspond à une évolution du monde). Une analyse formelle de ces différents systèmes peut être trouvée dans [28].

Nous avons enfin montré brièvement les difficultés liées à la définition d'un opérateur de changement de croyance (à partir des postulats AGM ou KM) qui, pour un état mental et une nouvelle information donnés, fournirait un nouvel état mental.

Pour terminer, nous mentionnons trois points qu'il nous semble intéressant de prendre en compte dans le changement de croyances. Le premier point concerne le fait (cf. Sect. 4) qu'aucune des approches intentionnelles présentées ne permet de gérer l'ensemble des opérations identifiées dans les théories formelles AGM et KM (expansion, révision, mise à jour, contraction), chacune de ces opérations étant adaptée à une situation particulière (par exemple, un changement d'avis au cours d'un dialogue correspond à une mise à jour). Il apparaît donc essentiel d'étendre ces théories par un mé-

canisme formel de gestion de *stratégies de changements de croyances*. Ceci permettrait alors à un agent de déterminer de façon contextuelle la « bonne » opération à appliquer face à une nouvelle information (voir, par exemple, le cas du choix laissé ouvert chez SADEK entre l'acceptation de la nouvelle information ou la mise en cause d'une information antérieure qui bloque son acceptabilité).

Le deuxième point peut se résumer ainsi : lorsqu'un agent vient d'accomplir un acte, que devient l'intention<sup>16</sup> à l'origine de cet acte ? La réponse dépend du comportement adopté par l'agent : soit il considère (par défaut) que son acte a été convenablement interprété, et il doit alors abandonner cette intention (puisque'elle a été satisfaite) ; soit il considère que, jusqu'à preuve explicite du contraire, son acte n'a pas atteint son objectif (et l'intention de l'agent doit donc persister) ; soit il est dans un état transitoire d'ignorance où il ne sait rien à ce sujet (et là aussi son intention doit persister). Il semble donc nécessaire de doter un agent de *stratégies* lui permettant, là encore, de choisir un comportement de façon dynamique au cours du dialogue, stratégies qui influenceraient donc le processus de changement de croyances.

Enfin, le dernier point est relatif à l'inférence des effets indirects d'un acte par un observateur, c'est-à-dire les effets supplémentaires par rapport à ceux de l'acte littéral (*i.e.* interprété au premier degré). Ces effets peuvent s'avérer inconsistants avec les croyances courantes sans que les effets directs ne le soient. Par conséquent, l'inférence des effets indirects d'un acte doit également être associée au processus de changement de croyances.

## RÉFÉRENCES

- [1] Carlos E. Alchourrón, Peter Gärdenfors et David Makinson. On the logic of theory change. *The Journal of Symbolic Logic*, 50(2):510–530, Juin 1985.
- [2] Douglas Appelt et Kurt Konolige. A nonmonotonic logic for reasoning about speech acts and belief revision. In M. Reinfrank, J. de Kleer, M.L. Ginsberg et E. Sandewall, éditeurs, *Proc. of Int. Work. on Non-Monotonic Reasoning*, volume 346 of *LNAI*. Spr.-Verlag, 1989.
- [3] John L. Austin. *How To Do Things With Words*. Oxford Univ. Press, 1962.
- [4] Michael E. Bratman. *Intention, Plans, and Practical Reason*. Harvard Univ. Press, 1987.
- [5] Philippe Bretier. *La communication orale coopérative: contribution à la modélisation logique et à la mise en œuvre d'un agent rationnel dialoguant*. Thèse, Paris XIII, 1995.

---

16. Une hypothèse de base est que si un agent accomplit une action (intentionnelle) alors il a l'intention d'accomplir cette action (réciproque fausse), cette intention ayant elle-même pour origine l'intention de produire sur son interlocuteur un certain effet. C'est de cette dernière intention dont il s'agit ici.

- [6] Philippe Bretier et David Sadek. A rational agent as the kernel of a cooperative spoken dialogue system: Implementing a theory of interaction. In J.P. Müller, M.J. Wooldridge et N.R. Jennings, éditeurs, *Intelligent Agents III - Proc. of ATAL'96*. Spr.-Verlag, 1997.
- [7] Marcos Alexandre Castilho, Olivier Gasquet et Andreas Herzig. Formalizing action and change in modal logic I: The frame problem. *Journal of Logic and Computation*, 1999.
- [8] Philip R. Cohen et Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence Journal*, 42(2-3):213-261, 1990.
- [9] Philip R. Cohen et Hector J. Levesque. Rational interaction as the basis for communication. In Philip R. Cohen, Jerry Morgan et Martha E. Pollack, éditeurs, *Intentions in Communication*. MIT Press, 1990.
- [10] Robert Demolombe et Andrew J.I. Jones. On sentences of the kind «sentence 'p' is about topic t»: some steps towards a formal-logical analysis. In Hans Jürgen Ohlbach et Uwe Reyle, éditeurs, *Essays in Honor of Dov Gabbay*. Kluwer, 1998.
- [11] R. L. Epstein. *The Semantic Foundations of Logic*. Kluwer Academic Pub., 1990.
- [12] Luis Fariñas del Cerro, Andreas Herzig, Dominique Longin et Omar Rifi. Belief reconstruction in cooperative dialogues. In Fausto Giunchiglia, éditeur, *Proc. AIMSA '98*, volume 1480 of *LNAI*. Spr.-Verlag, 1998.
- [13] Andreas Herzig Faure, Rémi et Dominique Longin. Une formalisation de la détection des actes de langage indirects. Rapport technique. À paraître, IRIT, 2000. cf. [www.irit.fr/ACTIVITES/EQ\\_ALG](http://www.irit.fr/ACTIVITES/EQ_ALG).
- [14] Rémi Faure. « Peux-tu me passer le sel ? » : formalisation de la détection des actes indirects directifs dans des dialogues entre agents coopératifs autonomes. Rapport de DEA, Université Paul Sabatier - IRIT, 2000.
- [15] Julia Rose Galliers. Autonomous belief revision and communication. In Peter Gärdenfors, éditeur, *Belief Revision*. Cambridge Univ. Press, 1992.
- [16] S. Gitton. *Mise en oeuvre d'un système expérimental de dialogue oral et modèle formel de traitements d'erreurs*. Thèse, Univ. de Rennes I, 1995.
- [17] E. Giunchiglia, G. N. Kartha et V. Lifschitz. Representing action: indeterminacy and ramifications. *Artificial Intelligence Journal*, 95:409-438, 1997.
- [18] N. Goodman. *About*. *Mind*, LXX(277), 1961.
- [19] Gösta Grahne. Updates and counterfactuals. In J. A. Allen, R. Fikes et E. Sandewall, éditeurs, *Proc. KR'91*. Morgan Kaufmann Pub., 1991.
- [20] H. Paul Grice. *Studies in the way of words*. Harvard Univ. Press, USA, 3<sup>e</sup> édition, 1989.
- [21] Andreas Herzig et Dominique Longin. Belief dynamics in cooperative dialogues. *Journal of Semantics*, 2000. À paraître.
- [22] Andreas Herzig et Dominique Longin. Beliefs, intentions, speech acts and topics. Rapport Technique 00-08-R, IRIT, Mars 2000. cf. [www.irit.fr/ACTIVITES/EQ\\_ALG](http://www.irit.fr/ACTIVITES/EQ_ALG).
- [23] Andreas Herzig et Dominique Longin. A topic-based framework for rational interaction. In *Proc. TALN-2000*, 2000. Poster Session.
- [24] Andreas Herzig, Dominique Longin et Jacques Virbel. Towards an analysis of dialogue acts and indirect speech acts in a BDI framework. In Massimo Poesio et David Traum, éditeurs, *Proc. GötaLog-2000*, Sweden, Juin 2000.
- [25] Andreas Herzig et Omar Rifi. Propositional belief base update and minimal change. *Artificial Intelligence Journal*, 115(1):107-138, Décembre 1999.

- [26] Hirofumi Katsuno et Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. In Peter Gärdenfors, éditeur, *Belief Revision*. Cambridge Univ. Press, 1992.
- [27] D.K. Lewis. General semantics. In D. Davidson et G. Harman, éditeurs, *Semantics of natural language*. D. Reidel Publishing Company, 1972.
- [28] Dominique Longin. *Interaction rationnelle et évolution des croyances dans le dialogue : une logique basée sur la notion de topique*. Thèse, Univ. Paul Sabatier, Novembre 1999. [www.irit.fr/ACTIVITES/EQ\\_ALG/](http://www.irit.fr/ACTIVITES/EQ_ALG/).
- [29] J. McCarthy et P. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer et D. Michie, éditeurs, *Machine Intelligence*, volume 4, pages 463–502. Edinburgh University Press, 1969.
- [30] C. Raymond Perrault. An application of default logic to speech act theory. In Philip R. Cohen, Jerry Morgan et Martha E. Pollack, éditeurs, *Intentions in Communication*. MIT Press, 1990.
- [31] Anand S. Rao et Michael P. Georgeff. Modeling rational agents within a BDI-architecture. In J. A. Allen, R. Fikes et E. Sandewall, éditeurs, *Proc. KR'91*. Morgan Kauf. Pub., 1991.
- [32] Anand S. Rao et Michael P. Georgeff. An abstract architecture for rational agents. In Bernhard Nebel, Charles Rich et William Swartout, éditeurs, *Proc. KR'92*. Morgan Kaufmann Pub., 1992.
- [33] Anand S. Rao et Michael P. Georgeff. Decision procedures for BDI logics. *Journal of Logic and Computation*, 8(3), Juin 1998.
- [34] Ray Reiter. A logic for default reasoning. *Artificial Intelligence Journal*, 13:81–132, 1980.
- [35] David Sadek. Design considerations on dialogue systems: From theory to technology - the case of Artemis -. In *Proc. of the ESCA TR Workshop on Interactive Dialogue Systems (IDS'99)*, Klöster Irsee, Germany, 1999.
- [36] M. D. Sadek. *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication*. Thèse, Univ. de Rennes I, 1991.
- [37] M. D. Sadek. A study in the logic of intention. In Bernhard Nebel, Charles Rich et William Swartout, éditeurs, *Proc. KR'92*. Morgan Kaufmann Pub., 1992.
- [38] M. D. Sadek. Towards a theory of belief reconstruction: Application to communication. *Speech Communication Journal'94, special issue on Spoken Dialogue*, 15(3–4), 1994.
- [39] M. D. Sadek. Dialogue acts are rational plans. In M.M. Taylor, F. Néel et D.G. Bouwhuis, éditeurs, *The structure of multimodal dialogue*, pages 167–188, Philadelphia/Amsterdam, 2000. John Benjamins publishing company. From Proc. ESCA/ETRW (Venaco II), 1991.
- [40] M. D. Sadek, A. Ferrieux, A. Cozannet, P. Bretier, F. Panaget et J. Simonin. Effective human-computer cooperative spoken dialogue: The AGS demonstrator. In *Proc. ICSLP'96 (Int. Conf. on Spoken Language Processing)*, 1996.
- [41] M.D. Sadek, P. Bretier et F. Panaget. ARTIMIS: Natural dialogue meets rational agency. In Martha E. Pollack, éditeur, *Proc. IJCAI'97*. Morgan Kauf. Pub., 1997.
- [42] Erik Sandewall. *Features and Fluents*. Oxford Univ. Press, 1994.
- [43] J. R. Searle. *Intentionality*. Cambridge Univ. Press, 1983.
- [44] J. R. Searle et D. Vanderveken. *Foundation of illocutionary logic*. Camb. Un. Press, 1985.
- [45] John R. Searle. *Speech acts*. Cambridge Univ. Press, 1969.
- [46] John R. Searle. *Du cerveau au savoir*. Hermann, 1985.
- [47] Michael Thielscher. Computing ramifications by postprocessing. In *Proc. IJCAI'95*, 1995.