Évolution des croyances au sein d'une théorie de l'intentionnalité : application au dialogue coopératif orienté tâches*

Dominique Longin

Groupe de Logique Appliquée
Institut de Recherche en Informatique de Toulouse
Université Paul Sabatier
118 Route de Narbonne, 31062 Toulouse Cedex 4
Email : Dominique.Longin@irit.fr
Site Web : http://www.irit.fr/ACTIVITES/EQ_ALG/

Résumé

Nous étudions la dynamique des croyances dans les dialogues coopératifs homme-machine orientés tâche. Nous supposons que, durant le dialogue, chaque participant peut mal comprendre une information, l'oublier ou tout simplement changer d'avis à son sujet.

Nous présentons tout d'abord quelques hypothèses relatives aux participants de façon à bien cadrer notre travail, puis quelques approches fondamentales dans ce domaine. Nous soulignons quelques faiblesses de ces dernières et introduisons une nouvelle logique de l'action, de la croyance et de l'intention, où cette dernière est définie dans une logique modale non normale pour éviter certaines propriétés contre-intuitives.

Nous nous concentrons sur l'interaction entre les différents opérateurs modaux d'une part, et les actes de langage accomplis *via* les énoncés d'un dialogue d'autre part. Notre notion de base pour décrire une telle interaction est celle de topique : nous supposons que nous pouvons associer un ensemble de topiques à tout agent, acte de langage et formule. Cela nous permet d'exprimer la compétence d'un agent, l'adoption de croyance et la préservation.

Mots clés

Logiques pour l'IA, mise à jour de connaissances, raisonnement sur l'action, représentation des connaissances, dialogue entre agents rationnels.

*Ce travail a été supporté par le centre de recherche et de développement de France Télécom, pôle *Interactions Intelligentes et Dialogue*, dans le cadre du marché 97 1B 046. Merci à Andreas Herzig pour l'aide précieuse qu'il a apportée à mon travail. Merci également à David Sadek et Philippe Bretier pour leurs remarques pertinentes.

1 Introduction

Les dialogues coopératifs homme-machine orientés tâche, qu'ils soient en langage naturel ou artificiel, sont un des challenges les plus importants des sciences informatiques. Les participants à de tels dialogues ont un but commun principal, qui est d'accomplir la tâche en question. Chaque participant a quelques informations permettant l'accomplissement de quelques sous-buts, mais aucun d'eux ne peut accomplir le but principal tout seul.

Le contexte de notre travail est un système effectif et générique de dialogues coopératifs temps réel. Ce système a été développé par France Télécom comme une instanciation de la technologie d'agent rationnel appelée ARTIMIS [39, 44]. Cette approche consiste en premier lieu à décrire le comportement d'un agent à l'intérieur d'une théorie logique de l'interaction rationnelle [40, 41, 42, 43], et dans un second temps à implémenter cette théorie par un système d'inférences [39, 5]. Ce dernier est le noyau d'ARTIMIS. Pour un ensemble fixe de domaines, ce système est capable d'accepter en entrée un langage spontané presque non contraint, et de réagir de façon coopérative. Les activités du système de dialogue sont doubles : prendre en compte les énoncés du locuteur, et générer des réactions appropriées. La partie réactive est complètement définie dans l'état courant de la théorie et de l'implémentation. À l'opposé, la consommation d'un énoncé est spécifiée seulement de façon partielle, en particulier la partie *changement de croyances* (cf. Sect. 2) du processus de consommation.

Nous basant en cela sur des travaux antérieurs [14, 24, 32, 25, 26], nous nous focalisons dans ce qui suit sur l'évolution des croyances[1] d'un agent participant à une conversation. Chaque énoncé accompli lors du déroulement de cette dernière conduit chacun des participants à modifier un sous-ensemble de ses croyances.

La *coopérativité* de chaque agent est une hypothèse non seulement utile mais fondamentale. Informellement, un agent est coopératif par rapport à un autre, si le premier aide le second à accomplir les buts que ce dernier s'est fixé (cf. les principes coopératifs de GRICE, ainsi que ses maximes conversationnelles [21]).

La coopérativité n'entraîne pas toujours la *sincérité* (en particulier, on peut être à la fois coopératif et non sincère – cf. [32]). Dans ce qui suit, nous supposons que chaque participant est sincère. Cela signifie plus précisément que l'énoncé d'un locuteur reflète fidèlement ses croyances. Par exemple, si un participant dit « le ciel est bleu », alors c'est qu'à cet instant, forcément, il croit que le ciel est bleu. Une telle hypothèse a pour conséquence que les contradictions entre les présuppositions liées à l'accomplissement d'un énoncé, et les croyances de l'auditeur sur le locuteur juste avant cet énoncé, ne peuvent s'expliquer en terme de mensonge.

Nous illustrons les difficultés rencontrées à l'aide d'un exemple[2] qui nous servira pour toute la suite. Il y a deux agents, le système s et l'utilisateur u.

> $\boldsymbol{s_1}$ **:** *Bonjour. Que voulez-vous ?*
> $\boldsymbol{u_1}$ **:** *Un billet en première classe pour Lyon, s'il vous plaît.*
> $\boldsymbol{s_2}$ **:** *150 €.*
> $\boldsymbol{u_2}$ **:** *Hum... pardon, en seconde classe.*
> $\boldsymbol{s_3}$ **:** *100 €.*
> $\boldsymbol{u_3}$ **:** *Puis-je payer les 80 € par carte bleue ?*
> $\boldsymbol{s_4}$ **:** *Le prix n'est pas 80 € mais 100 €. Oui, vous pouvez payer par carte bleue.*
> $\boldsymbol{u_4}$ **:** *...*

Notre exemple illustre 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).

1. Le terme de *croyance(s)* peut être pris soit dans un sens restreint (c.-à-d. littéralement), soit dans un sens large (au sens où tout agent ayant une attitude mentale croit 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.* Afin de faire la distinction, nous emploierons le terme au singulier pour désigner le sens restreint (en accord avec le concept de *belief* de l'école d'Oxford), et au pluriel pour le sens large.

2. De nombreux exemples de dialogues de ce type peuvent être trouvés dans [17].

Ce dernier point nous oblige à spécifier que, selon la dissociation que Kant a faite 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 qu'un agent peut mal comprendre un énoncé, c'est ce dernier concept que nous choisissons pour modéliser nos agents.

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[3]. Celui-ci 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_1) ;
2. de déduire des informations supplémentaires non directement contenues dans l'énoncé, en utilisant des lois sur le monde (cf. la déduction du prix dans s_2, étant données les informations contenues dans u_1) ;
3. de continuer à croire une partie de ce qu'il croyait juste avant l'accomplissement du dernier énoncé (dans u_2 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_3) ;
4. d'accepter quelquefois des informations contradictoires avec ses propres croyances (en particulier quand l'utilisateur change d'avis, comme dans u_2) ;
5. de refuser parfois de prendre en compte certaines informations, en particulier si l'utilisateur donne des informations à propos de faits sur lesquels il n'est pas compétent (cf. u_3 et s_4).

Notre but est d'obtenir une sémantique ayant à la fois une axiomatisation complète et une procédure de preuve, et une implémentation effective. Ceci a motivé plusieurs choix, en particulier une sémantique des mondes possibles de type SAHLQVIST (pour laquelle il existe des résultats de complétude généraux), et une notion d'intention primitive (contrairement aux constructions complexes de la littérature) définie dans une logique modale non normale (mais qui peut malgré tout être réduite au cadre de travail de SAHLQVIST).

Après avoir présenté quelques approches existantes, nous introduisons, dans la prochaine section, notre cadre multimodal (Sect. 2). Celui-ci est basé sur une théorie métalinguistique de topiques (Sect. 3). Les topiques permettent d'intégrer des définitions appropriées de l'adoption et de la préservation de croyances dans notre théorie logique. Ces deux principes sont alors définis formellement (Sect. 4).

Dans ce cadre formel, des lois non logiques peuvent être formulées (en particulier des lois du domaine, des lois gouvernant les actes de langage, et des lois réactives). Pour ne pas alourdir inutilement cet exposé, nous ne présentons pas ici cette théorie non logique. Elle peut être trouvée dans [24], qui contient aussi une théorie logique moins élaborée. Nous raffinons ici cette dernière, en ajoutant une analyse sémantique et en introduisant une notion plus appropriée d'intention. Il en découle un changement de croyances plus fin, tenant compte en particulier des conditions d'abandon d'une intention selon [4, 9, 40].

2 Interaction rationnelle et changement de croyances

Les travaux sur la *révision* [1] et la *mise-à-jour* [29, 28] de connaissance (ou de croyance) illustrent que pour décrire de tels processus, nous devons disposer au moins d'un ensemble d'informations, et d'une nouvelle information à ajouter à cet ensemble. La révision et la mise-à-jour correspondant sémantiquement à des opérations différentes l'une de l'autre, nous éludons pour l'instant le problème du choix d'une approche plutôt que d'une autre, en les désignant toutes deux par une expression neutre unique : un *changement de croyances.*

Nous reviendrons sur les approches basées sur des opérateurs de révision et de mise-à-jour en fin de la présente section. Nous présentons en premier

3. En fait, comme nous l'avons montré dans [32], le formalisme rend 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.

lieu les travaux les plus importants basés sur une *théorie de l'intentionnalité* [46, 4], travaux dans la lignée directe desquels nous nous situons. Dans une telle théorie, les interactions entre agents rationnels sont analysées en termes d'états mentaux de ces agents. L'état mental d'un agent est un ensemble (clos ou non sous la conséquence logique) d'informations particulier, comprenant différentes attitudes mentales (croyance, but, intention, ...). Cette théorie est à la base de ce qui est communément appelé les *architectures BDI*[4].

Dans cet ordre d'idées, un changement de croyances doit se situer au sein d'une *théorie formelle de l'équilibre rationnel* et d'une *théorie formelle de l'interaction rationnelle* [9, 11, 42], de façon à structurer les états mentaux, et rendre les agents autonomes. La première de ces théories est en charge de décrire les propriétés de chaque attitude mentale, ainsi que les différentes interactions qu'il peut y avoir entre elles. La seconde de ces théorie caractérise, en termes d'attitudes mentales et d'actions, les interactions entre les agents et leur environnement.

Dans le présent travail, nous nous limitons à des actes purement linguistiques, et les énoncés sont représentés par des actes de langage [3, 48, 47], sur lesquels nous reviendrons dans ce qui suit.

Cohen et Levesque. COHEN et LEVESQUE [9, 11] ont jeté les bases formelles générales de telles théories. Selon eux, une théorie de la conversation devrait *expliquer la cohérence du dialogue en termes d'états mentaux des participants*. Ils tentent de 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 associées 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 (ou de rejet), 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, potentiellement de façon indirecte[5]. Cette approche présuppose donc tout un travail d'interprétation de l'énoncé effectivement accompli, travail non décrit dans la théorie. (Nous soulignons à ce titre que notre cadre de travail permet de traiter les indirections, ce que nous avons ébauché dans [27]. À titre de simplification, nous n'intégrons pas ici un tel traitement.)

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 permet la prise en compte d'une évolution des états mentaux au cours d'un dialogue.

Néanmoins, leur théorie souffre du problème (très connu en Intelligence Artificielle) du décor (*frame problem*) [33] : l'opération de changement de croyances ainsi formalisée se ramène à l'opération triviale de [1] : l'acceptation d'une nouvelle information se fait au prix de l'abandon de toutes les anciennes croyances (celles que possédait l'agent avant l'accomplissement de l'acte). Dans notre exemple, 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 [34] essaie de résoudre le problème du décor dans un cadre basé sur la lo-

4. Pour *Belief-Desire-Intention*, c.-à-d. les architectures d'agent basées sur la croyance, le désir (ou *but*) et l'intention.

5. 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 en question a été accomplie de façon indirecte, ou encore que l'énoncé accompli constitue un acte indirect.

gique des défauts de REITER [38]. Mais sa théorie souffre de sérieux problèmes dont la plupart ont été mis en valeur par APPELT et KONOLIGE [2]. En particulier, les agents (formalisés 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.

Appelt and Konolige. APPELT et KONOLIGE [2] recommandent l'utilisation d'une logique autoépistémique hiérarchique (HAEL). Globalement, l'avantage par rapport à une logique comme celle de PERRAULT, est que, dans HAEL, 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 la technologie 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 un de ceux rencontrés par PERRAULT : supposons que l'auditeur n'a pas d'opinion à propos du temps qu'il fait. Si le locuteur informe l'auditeur qu'il fait beau, alors, sous certaines conditions, l'auditeur adoptera la croyance selon laquelle il fait beau (*i.e.* « il fait beau » fera partie de ses croyances). Mais si on suppose maintenant que le locuteur informe l'auditeur que « l'auditeur croit qu'il fait beau » alors l'idée selon laquelle l'auditeur devrait incorporer ce genre d'information dans son propre état mental, entre clairement en conflit avec notre intuition : au plus, devrait-il croire que le locuteur croit qu'il croit qu'il fait beau.

Comme montré dans [27], la seule voie pour supprimer un tel effet indésirable, semble être de déplacer l'ignorance de l'auditeur à propos de p au niveau 0 de la hiérarchie d'HAEL. Mais dans ce cas, ce niveau contenant les croyances de l'agent ne pouvant pas être changées, toute tentative d'accepter une assertion de p serait bloquée, ce qui entre clairement en conflit, une fois de plus, avec notre intuition (puisqu'on retrouve le problème selon lequel l'agent ne pourrait plus changer de croyance à propos de p).

Sadek. SADEK, quant à lui, introduit la notion de *reconstruction des croyances* [40, 42, 43]. Celle-ci est définie par quatre axiomes : la mémoire et la persistance (qui étaient déjà présentes chez PERRAULT), ainsi que l'admission et la consommation :

- la mémoire sert à l'agent pour se rappeler ce qu'il croyait juste avant l'accomplissement du dernier énoncé ;
- l'admission permet de réunir les conditions favorables à l'acceptation de la nouvelle information (y compris au prix d'une révision des croyances) ;
- la consommation est l'étape pendant laquelle les effets de l'acte sont ajoutés aux croyances de l'agent ;
- et la persistance consiste à conserver toutes les croyances issues de la mémoire, consistantes avec le nouvel état engendré.

Bien que les axiomes de SADEK permettent de traiter notre exemple, en raison de leur accent autoépistémique, ils ne donnent aucune définition formelle constructive de la déduction.

Quoiqu'il en soit, la reconstruction des croyances s'appuie sur une formalisation originale intéressante des actes de langage [41], dans l'esprit de SEARLE [48]. SADEK distingue plusieurs effets, que nous reprenons dans notre travail en simplifiant leur définition formelle : l'*effet indirect* d'un acte est la persistance (à l'issus de l'accomplissement de celui-ci) de ses préconditions ; l'*effet intentionnel* correspond au point de vue gricéen de la communication [22] ; l'*effet rationnel* est l'effet attendu de l'acte par son auteur.

Par exemple, supposons que l'utilisateur u informe le système s que sa destination est Lyon. Alors :

- la précondition de sincérité est que u croit que sa destination est effectivement Lyon;
- la précondition de pertinence au contexte est que u ne croit pas que s sait si la destination de u est Lyon.

Dans ces conditions, les effets de cet acte d'information sont les suivants :

- l'effet indirect stipule que juste après l'accomplissement de l'acte, les préconditions de sincérité et de pertinence au contexte sont toujours vraies [6] ;
- l'effet intentionnel traduit le fait que u a l'intention que s croit que u a l'intention de produire l'effet rationnel ci-dessous ;
- enfin, l'effet rationnel (*i.e.* l'effet de l'acte sur s attendu par u) est que s croit que la destination est Lyon.

Rao et Georgeff. RAO et GEORGEFF [35, 36] ont proposé des théories et des architectures pour les agents rationnels. Récemment, ils ont défini une procédure de preuve par la méthode des tableaux pour leurs logiques [37].

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* ni représenter des actions non déterministes, ni des actions ayant des effets indirects (obtenus *via* des contraintes d'intégrité). Ainsi, on ne peut utiliser de lois sur le monde 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.* des effets 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 tels que des intentions et des croyances.

Approches AGM et KM. Comme nous l'avons laissé sous-entendre au début de cette section, il existe d'importantes analyses formelles du changement de croyances, en dehors des théories intentionnelles (au sens de la théorie de l'intentionnalité) que nous venons de survoler. Les plus importantes sont la *révision* d'ALCHOURRÓN, GÄRDENFORS et MAKINSON (AGM) et les opérations de *mise-à-jour* de KATSUNO et MENDELZON (K&M) [1, 29, 28]. 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 à l'ensemble d'informations. La seconde, au contraire, décrit les changements d'un monde dynamique au fur et à mesure que ceux-ci s'opèrent [29, 20]. Ainsi, la distinction entre ces deux opérateurs est respectivement basée sur différentes hypothèses sur 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.

On peut se demander si certaines de ces opérations pourraient s'appliquer à notre exemple de dialogue. 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 *réviser* son état mental.

Il apparaît que dans beaucoup d'exemples de dialogues, l'auditeur est dans l'incapacité de décider si la nouvelle information à laquelle il est soumis correspond à un changement de monde réel [7] ou non. Dans notre exemple, après u_2, le système peut être incapable de distinguer entre le cas où l'utilisateur a changé d'avis, et celui où il a mal compris l'énoncé u_1 (cf. [28] pour une critique détaillée de l'approche de KM). Nous n'effectuons donc pas, pour ces raisons, de distinction entre ces deux notions, ainsi regroupée sous l'expression *changement de croyances*.

De plus, les opérateurs de révision et de mise-à-jour ont plusieurs propriétés communes, qui ne

6. Si la préservation de la première nous semble intuitivement 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. Nous montrons dans [27] comment être plus fin, en considérant un état transitoire d'ignorance. Afin de ne pas compliquer inutilement la sémantique de la logique, nous ne reprenons pas ici une telle subtilité.

7. Nous considérons que le monde réel inclut toute chose qui est externe à l'agent auditeur (incluant les intentions du locuteur).

sont pas appropriées dans le cadre des dialogues. En particulier, la nature surinformative de certaines informations est négligée, puisque $(K \circ A) \leftrightarrow K$ si $K \rightarrow A$ est un des postulats en question. En d'autres termes, si la nouvelle information est dérivable de l'ensemble d'informations, alors une révision (ou une mise-à-jour) de cet ensemble par la nouvelle information ne change pas l'ensemble considéré.

Enfin, aucune distinction n'est faite entre les différents niveaux de croyance (factuel, introspectif, alterné, ...).

Pour toutes ces raisons, ces approches ne nous paraîssent pas pertinentes dans le cadre de notre travail.

Dans notre approche, nous basant en cela sur des travaux précédents [14, 24, 32, 25, 26], nous implémentons un changement de croyances par l'adoption et la préservation de croyances, toutes deux basées sur les topiques. Nous commençons par introduire avant tout notre langage.

Le langage modal. Comme beaucoup, nous travaillons dans un cadre multimodal, avec des opérateurs de croyance, de croyance mutuelle, d'intention et d'action. Notre langage est celui de la logique multimodale du premier ordre sans égalité ni symbole de fonction.

Soit AGT l'ensemble des agents. Pour $i \in AGT$, $Bel_i A$ est lu « l'agent i croit que A », $Intend_i A$ est lu « l'agent i a l'intention que A ». $BelIf_i A$ est une abbréviation pour $Bel_i A \vee Bel_i \neg A$. $Bel_{i,j} A$ est lu « les agents i et j croient mutuellement que A ». Par exemple, $Bel_u Dest(\text{Lyon})$ exprime que l'agent u croit que la destination est Lyon.

Contrairement aux approches de BRATMAN, COHEN et LEVESQUE, et SADEK, l'intention est ici primitive, comme c'est le cas dans [35] et [30]. Nous avons choisi cette solution pour trois raisons.

Premièrement, construire l'intention à partir de notions primitives (telles que le but, par exemple), conduit à des notions sophistiquées et variées de l'intention, avec de subtiles différences entre elles. Nous avons choisi de ne prendre, ici, que les propriétés de l'intention communes à toutes les approches.

Deuxièmement, ces définitions étant plutôt complexes, il est difficile de leur trouver des méthodes de démonstration automatique complètes, alors que notre analyse supporte des méthodes de preuve et des techniques de complétude plus ou moins standards.

Troisièmement, il est important de souligner que, d'après nous, notre notion simplifiée d'intention est suffisante, au moins pour un large éventail d'applications. En effet, il semble que les interactions vraiment cruciales soient celles entre croyances et intentions, plus que celles entre buts et intentions.

Les actes de langage [3, 48] sont représentés par des 4-uplets de la forme $\langle \mathsf{FORCE}_{i,j}\ A \rangle$ où $\mathsf{FORCE} \in \{\mathsf{Inform}, \mathsf{Request}, \mathsf{QueryYN}, \mathsf{QueryWh}\}$ est la force illocutoire de l'acte, $i,j \in AGT$, et A est une formule du langage représentant le contenu propositionnel de l'énoncé. Par exemple, $\langle \mathsf{Inform}_{u,s}\ Dest(\text{Lyon}) \rangle$ représente un énoncé informatif de l'utilisateur u à destination du système s selon lequel la destination est Lyon; $\langle \mathsf{QueryYN}_{u,s}\ Bel_s Prix(150\,€) \rangle$ signifie « l'utilisateur demande au système s'il croit que le prix est 150 € ».

Soit ACT l'ensemble de tous les actes de langage. À chaque $\alpha \in ACT$ nous associons un opérateur modal $Done_\alpha$. $Done_\alpha A$ est lu « α vient juste d'être accompli, avant quoi A était vrai »[8]. $Done_\alpha \top$ est lu « α vient juste d'être accompli ».

$Bel_s Dest(\text{Lyon})$ est un exemple de formule. De même $Bel_s(Dest(\text{Lyon}) \wedge Classe(1^{\text{re}}) \rightarrow Prix(150\,€))$. Cette dernière est également un axiome non logique (*i.e.* une loi du domaine) permettant au système de déduire le prix à partir des informations relatives à la destination et à la classe. Un autre exemple significatif est la formule $\neg Done_{\langle \mathsf{Inform}_{u,s}\ p \rangle} \neg Bel_u p$ exprimant la sincérité de u, et qui est également une partie de la théorie non logique : u vient d'informer s que p, juste avant quoi u croyait p.

Les formules atomiques sont notées p,q,... ou $P(t_1,...,t_n)$. ATM est l'ensemble de toutes les formules atomiques. Les formules sont notées A,B,....

8. $Done_\alpha A$ est similaire à $\langle \alpha^{-1} \rangle A$ de la logique dynamique [23].

3 Changement de croyances basés sur les topiques

Compétence des agents. Dans notre approche, contrairement à SADEK ou GALLIERS [15] qui donnent à l'auditeur la possibilité de rejeter la nouvelle information, nous l'acceptons toujours[9]. En revanche, comme nous le montrons dans ce qui suit, l'agent peut n'accepter qu'un sous-ensemble des conséquences de cette admission.

Dans ce but, nous procédons en deux étapes : l'auditeur consomme[10], *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 pas consommé à proprement parlé, mais est considéré comme tel si, à l'issue du processus de changement de croyance, il est dérivable de l'état mental de l'auditeur.

Ainsi, la *compétence du locuteur* est notre critère pour déterminer quelle part des conséquences de l'acceptation de la nouvelle information est consommée par l'auditeur. Dans notre exemple, 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 l'accomplissement de u_3 (la proposition de prix par l'utilisateur), car il considère u comme étant compétent (entre autres) sur les classes mais pas sur les prix des billets.

Influence des actes de langage. Mais que l'agent soit en mesure d'acquérir de nouvelles croyances à l'issue d'un acte n'est pas tout : nous avons montré dans la section précédente la nécessité de pouvoir préserver certaines croyances que l'agent possédait juste avant cet accomplissement (problème du décor). La question se résume alors à la suivante : quelles croyances de l'auditeur peuvent être préservées après l'accomplissement d'un acte de langage ?

Notre concept clé est ici celui d'*influence d'un acte de langage sur les croyances.* S'il existe une relation d'influence entre l'acte accompli et une croyance, cette croyance ne peut alors pas être préservée dans le nouvel état mental. Dans notre exemple, l'ancienne classe de transport ne peut pas être préservée à l'issue de l'accomplissement de 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.

Le concept d'influence (ou de dépendance) d'un acte est proche des notions qui ont été récemment étudiées dans le domaine du raisonnement sur les actions dans le but de résoudre le problème du décor (cf. par exemple l'occlusion de SANDEWALL [45], la relation d'influence de THIELSCHER [49], ou les opérateurs de changements possibles de GIUNCHIGLIA *et al.*'s [18]). De tels concepts ont en particulier été utilisés au sein de notre équipe [7].

Compétence et influence *via* les topiques. Tout cela présuppose que nous sommes capables de déterminer la compétence d'un agent et l'influence d'un acte de langage. Contrairement aux approches précédentes en raisonnement sur les actions (où le concept d'influence est primitif), le fondement de ces deux notions sera fourni ici par le concept de topique : nous partons de l'idée qu'à tout agent, acte de langage, et formule, on peut associer des ensembles de topiques.

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

9. Pour nous, la nouvelle information correspond au fait qu'un acte de langage vient juste d'être accompli, soit, formellement : $Bel_i Done_\alpha \top$ où i représente l'auditeur considéré et α l'acte venant juste d'être accompli. L'acte accepté peut ou non être l'acte réellement accompli. Bien qu'elle soit différente, cette phase correspond, dans son esprit, à la phase d'admission de SADEK [43]. Aussi, par abus de langage, parlons-nous parfois de l'*admission d'un acte* (ou *d'une nouvelle information*).

10. Ce terme est employé par SADEK [40] pour décrire l'ajout d'une formule logique représentant un effet à l'état mental de l'agent considéré. Là encore, bien que reprenant le terme, notre phase de consommation est différente de celle décrite par l'axiome du même nom chez SADEK (cf. [43] pour plus de détails).

d'un acte de langage α s'il n'y a aucun topique commun à l'ensemble des topiques associés à A et à l'ensemble des topiques associés à α.

Les topiques constituent un concept naturel et intuitif, qui va nous permettre d'appréhender finement la consommation d'un acte de langage. Cette notion est importante en linguistique [6, 16, 50, 51]. D'un point de vue logique, EPSTEIN [13, p. 68], définit le *subject matter* d'une proposition A. Il montre que nous pouvons alors définir deux propositions comme étant en relation si elles ont des *subject matter* en commun. Notre fonction $\mathfrak{subject}$ peut être vue comme une extension de cette fonction à un langage multimodal [11].

Dans le reste de cette section, nous présentons notre théorie métalinguistique des topiques.

Thèmes et topiques. Un *thème* représente ce à propos de quoi est quelque chose. Nous supposons que l'ensemble de thèmes est non vide. Dans notre exemple courant, les thèmes sont : le destination, la classe, et le prix.

Pour $i \in AGT$, ma_i est appelé *contexte atomique*. ma_i est mis pour « l'attitude mentale de l'agent i ». Un *contexte* est une séquence potentiellement vide de contextes atomiques, notée $ma_{i_1}{:}ma_{i_2}{:}\ldots{:}ma_{i_n}$. Le contexte vide est noté λ.

Un thème t associé à un contexte c forme un *topique d'information*, dénoté par $c{:}t$. Par exemple, $ma_u{:}prix$ est un topique relatif à une attitude mentale de l'utilisateur à propos du prix ; $ma_s{:}ma_u{:}prix$ est un topique relatif à une attitude mentale du système à propos d'une attitude mentale de l'utilisateur sur les prix.

Par convention, $\lambda{:}c = c{:}\lambda = c$, et $\lambda{:}t = t$. En outre, des principes d'introspection motivent l'identité :

$$ma_i{:}ma_i = ma_i. \tag{1}$$

Étant donnés un ensemble de thèmes et un ensemble d'agents, nous notons $\mathbb{T}$ l'ensemble des topiques associés. $\mathbb{T}_n$ est l'ensemble des topiques dont le contexte a une longueur au plus égale à n. Ainsi, $\mathbb{T}_0$ est l'ensemble des thèmes. Ici, pour des raisons techniques (et aussi pour des raisons d'économie de représentation), nous supposerons que la longueur de chaque contexte est au plus égale à 2. Ainsi, nous restreignons $\mathbb{T}$ à $\mathbb{T}_2$.

Le sujet d'une formule. Nous appelons *sujet d'une formule* un ensemble de topiques à propos desquels est une formule. Cette notion est formalisée par une fonction $\mathfrak{subject}(A) \subseteq \mathbb{T}$. Dans notre exemple courant, $\mathfrak{subject}(Classe(1^{\text{re}})) = \{classe\}$, $\mathfrak{subject}(Dest(\text{Lyon})) = \{destination\}$, et $\mathfrak{subject}(Bel_s Bel_u Prix(80\,€) \wedge Bel_s Prix(100\,€)) = \{ma_s{:}ma_u{:}prix, ma_s{:}prix\}$. Nous donnons les axiomes suivants :

$$\mathfrak{subject}(p) \subseteq \mathbb{T}_0 \text{ et } \mathfrak{subject}(p) \neq \emptyset \text{ si } p \in ATM \tag{2}$$

$$\mathfrak{subject}(\top) = \emptyset \tag{3}$$

$$\mathfrak{subject}(\neg A) = \mathfrak{subject}(A) \tag{4}$$

$$\mathfrak{subject}(A \wedge B) = \mathfrak{subject}(A) \cup \mathfrak{subject}(B) \tag{5}$$

$$\mathfrak{subject}(Bel_i A) = \{ma_i{:}c{:}t \mid c{:}t \in \mathfrak{subject}(A)\} \tag{6}$$

$$\mathfrak{subject}(Intend_i A) = \mathfrak{subject}(Bel_i A) \tag{7}$$

$$\mathfrak{subject}(Done_{\langle \mathsf{FORCE}_{i,j}\ A' \rangle} A) = \mathfrak{subject}(A) \cup \mathfrak{subject}(A') \tag{8}$$

$$\mathfrak{subject}(\forall x A) = \mathfrak{subject}(A) \tag{9}$$

$$\mathfrak{subject}(A[t/x]) \subseteq \mathfrak{subject}(A). \tag{10}$$

(3) dit que la vérité est à propos de rien. (4) stipule que la négation d'une formule ne modifie pas ce à propos de quoi est cette formule, et (5) que la fonction $\mathfrak{subject}$ est transparente aux opérateurs de la logique classique. Nous soulignons que dans (6) c peut d'une part être le contexte vide, et d'autre part, en toute généralité, un contexte de longueur quelconque. Cet axiome entraîne, *via* l'identité (1) cidessus, que $\mathfrak{subject}(Bel_i \ldots Bel_i A) = \mathfrak{subject}(Bel_i A)$. (7) exprime que les contextes atomiques font abstraction de la nature (croyance ou intention) des attitudes mentales. (8) entraîne par exemple que $\mathfrak{subject}(Done_{\langle \mathsf{Inform}_{u,s}\ Prix(100\,€)\rangle} Bel_s Prix(150\,€)) = \{prix, ma_s{:}prix\}$. (9) et (10) expriment que les formules ouvertes sont considérées comme étant universellement quantifiées. Finalement, notons

11. D'autres études de cette notion de topique existent dans la littérature logico-philosophique, en particulier [31, 19, 12].

que notre fonction sujet n'est pas extensionnelle : des formules logiquement équivalentes peuvent avoir des topiques différents (cf. la discussion en Sect. 5).

Il découle de nos axiomes que le sujet d'une formule arbitraire est complètement déterminé par les sujets de ses formules atomiques. C'est intéressant du point de vue de la représentation. La même motivation nous a conduit à restreindre les topiques à $\mathbb{T}_2$. En raison de cette restriction, nous supposons que :

$$ma_i{:}ma_j{:}c{:}t = ma_i{:}ma_j{:}t. \tag{11}$$

La fonction sujet correspondante peut être obtenue en réduisant, dans un premier temps, $\mathfrak{subject}(A)$ par les équations ci-dessus ($\lambda{:}t = t$, etc.), et en réduisant, dans un second temps, les topiques n'appartenant pas à $\mathbb{T}_2$. Pour ce faire, il suffit de tronquer la partie droite de leur contexte comme indiqué par (11).

La compétence d'un agent. La compétence d'un agent traduit les domaines sur lesquels les croyances de cet agent prédominent sur celles des autres. Ainsi, dans notre exemple, l'utilisateur est compétent sur les destinations et les classes, mais pas sur les prix.

Il est à noter que la compétence d'un agent n'est pas forcément absolue : elle peut être relative aux croyances d'un autre agent. Ainsi, deux agents peuvent être en désaccord sur la compétence d'un troisième agent à propos d'un thème t. La compétence devrait alors être une fonction binaire. Comme nous ne considérons que deux interlocuteurs dans notre exemple, nous avons supprimé le second argument pour simplifier.

Nous appelons *compétence d'un agent* i l'ensemble des topiques à propos desquels i est compétent, *i.e.* $\mathfrak{competence}(i) \subseteq \mathbb{T}$. Nous supposons ici que $\mathfrak{competence}(i) \subseteq \mathbb{T}_0$, *i.e.* nous ne considérons pas ici la compétence des agents sur leurs propres attitudes mentales ou sur celles des autres [12].

Le scope d'un acte. Le *scope d'un acte de langage* est la fonction qui détermine quelles attitudes mentales d'un agent sont remises en cause par l'accomplissement de cet acte. Nous supposons qu'un acte influence toujours (au moins en partie) les croyances des agents. Le scope d'un acte est un ensemble de topiques, *i.e.* $\mathfrak{scope}(\alpha) \subseteq \mathbb{T}$. Comme nous l'avons dit précédemment, nous supposons dans ce qui suit que $\mathfrak{scope}(\alpha) \subseteq \mathbb{T}_2$.

Nous avons dit dans la section précédente que dans la formalisation des actes de langage, la force illocutoire détermine un ensemble de formules schématiques (les préconditions et les effets de cet acte) instanciées par le contenu propositionnel (*via* des lois du domaine). Par exemple, dans le cadre de l'acte $\alpha = \langle \mathsf{Inform}_{u,s}\ Dest(\text{Lyon})\rangle$, la loi du domaine $Done_\alpha \top \rightarrow Done_\alpha Bel_u Dest(\text{Lyon})$ indique que si u vient d'informer s que la destination est Lyon, alors juste avant de l'en informer, u croyait effectivement que la destination était Lyon (expression de la précondition de sincérité d'un acte d'information).

Grossièrement, (en relation avec la remarque ci-dessus), les thèmes d'un acte de langage sont déterminés par le contenu propositionnel de ce dernier, alors que les contextes le sont par la force illocutoire. Ainsi, les contextes nous indiquent quelles attitudes mentales sont susceptibles de changer. Dans l'exemple précédent, quand s va admettre que α vient d'être accompli, il va présupposer que u est sincère. On peut donc prévoir que ses croyances à propos de la croyance de u sur la destination vont être influencées par α.

Un acte influence toujours les croyances de l'auditeur à propos de l'attitude du locuteur envers son contenu propositionnel. Formellement, pour toute force illocutoire FORCE:

$$\begin{aligned}&\mathfrak{scope}(\langle \mathsf{FORCE}_{i,j}\ A\rangle) \supseteq \\ &\qquad \{ma_j{:}ma_i{:}t \mid t \in \mathfrak{subject}(A)\}\end{aligned} \tag{12}$$

Par exemple, si l'utilisateur informe le système à propos du prix du billet, alors l'acte de langage correspondant influence les croyances du système sur les croyances de l'utilisateur à propos du prix. Ainsi, $ma_s{:}ma_u{:}prix \in \mathfrak{scope}(\langle \mathsf{Inform}_{u,s}\ Prix(150\ €)\rangle)$.

12. Notons que la compétence de i sur ses propres croyances et intentions sera de toute façon assurée par les axiomes standards d'introspection (cf. [25]).

Interactions. Une question qui se pose est alors de savoir s'il existe un lien entre ces fonctions. La réponse est positive, et pour s'en convaincre, il suffit de considérer que l'état mental issu d'un changement de croyance doit être maximalement consistant (idéalement, toute croyance que l'agent avait avant le changement, et qui est consistante avec le nouvel état, devrait être préservée). L'idée sous-jacente est que si aucune nouvelle croyance n'est adoptée sur un thème donné (*via* la compétence), alors toute formule relative à ce thème doit être préservée.

Nous proposons à cet effet l'axiome suivant pour les actes de type informatif.

$$\begin{array}{l}\text{Si } \alpha = \langle \mathsf{Inform}_{i,j}\ A\rangle \text{ et} \\ \quad t \in \mathfrak{subject}(A) \cap \mathfrak{competence}(i) \\ \text{alors } t \in \mathfrak{scope}(\alpha) \text{ et } ma_j{:}t \in \mathfrak{scope}(\alpha).\end{array} \quad (13)$$

Par exemple, si l'utilisateur informe le système à propos de la destination de son voyage, et comme l'utilisateur est compétent sur les destinations, l'acte engendré influence les croyances factuelles (*i.e.* celles portant sur des faits, représentés par des formules de la logique classique) du système à propos de la destination du voyage. Du fait des lois statiques, le prix, du point de vue du système, va également être modifié (à nouvelle destination, nouveau prix) : l'acte doit également influencer les croyances factuelles du systèmes à propos du prix. Soit : $\mathfrak{scope}(\langle \mathsf{Inform}_{u,s}\ Dest(\text{Lyon})\rangle)$ contient $destination$, $prix$, $ma_s{:}destination$ et $ma_s{:}prix$. Nous avons par ailleurs défini d'autres postulats pour d'autres forces illocutoires [32].

Nous montrons dans qui suit comment les topiques nous permettent de décrire formellement l'évolution des croyances des agents par le biais de deux principes : l'*adoption* et la *préservation.*

4 Axiomatique et sémantique

Axiomes d'adoption de croyance.

$$Bel_i A \rightarrow A \text{ si } \mathfrak{subject}(A) \subseteq \mathfrak{competence}(i). \quad (14)$$

Ce schéma exprime que si i croit que A, et est compétent sur A, alors A est vrai.

Par exemple, $Bel_u Dest(\text{Lyon}) \rightarrow Dest(\text{Lyon})$ parce que $\mathfrak{subject}(Dest(\text{Lyon})) \subseteq \mathfrak{competence}(u)$.

De cette formule, $Bel_s Bel_u Dest(\text{Lyon}) \rightarrow Bel_s Dest(\text{Lyon})$ peut être prouvé en utilisant la nécessitation et les K-principes standards pour Bel_s. À l'opposé, $Bel_u Prix(80\,€) \rightarrow Prix(80\,€)$ n'est pas une instance de notre schéma d'axiomes, car $\mathfrak{subject}(Prix(80\,€)) \not\subseteq \mathfrak{competence}(u)$.

Axiomes de préservation.

$$\begin{array}{l}A \rightarrow \neg Done_\alpha \neg A \\ \quad \text{si } \left\{\begin{array}{l}\mathfrak{scope}(\alpha) \cap \mathfrak{subject}(A) = \emptyset \text{ et} \\ Done_\beta \text{ n'apparaît pas dans } A.\end{array}\right.\end{array} \quad (15)$$

$$\begin{array}{l}Intend_i A \rightarrow \neg Done_\alpha \neg Intend_i A \\ \quad \text{si } \left\{\begin{array}{l}\mathfrak{scope}(\alpha) \cap \mathfrak{subject}(Intend_i A) \cap \mathbb{T}_1 = \emptyset \\ \text{ et } Done_\beta \text{ n'apparaît pas dans } A.\end{array}\right.\end{array} \quad (16)$$

La restriction selon laquelle aucun opérateur $Done_\beta$ ne doit apparaître dans A est nécessaire, car notre lecture de $Done_\beta$ est que β vient juste d'être accompli (et non à quelque instant arbitraire dans le passé).

Le second axiome de préservation des intentions est un renforcement du premier, car l'indépendance de α et A est restreinte à $\mathbb{T}_1$.

Sémantique. La sémantique est en termes de *modèles de mondes possibles* $\langle W,\mathcal{S},D,V\rangle$, où : (1) W est un ensemble de mondes ; (2) D est le domaine ; (3) V est une valuation qui associe aux symboles de constante et variable des éléments de D, et associant à chaque monde possible $w \in W$ une interprétation V_w des symboles de prédicats ; (4) $\mathcal{S}$ est une collection de structures sur W, constituée : de relations $\mathcal{D}_\alpha$ et $\mathcal{B}_i$ pour tout $\alpha \in ACT$ et tout $i \in AGT$, vues comme des fonctions associant W à 2^W, et de fonctions $\mathcal{I}_i$ pour tout $i \in AGT$, associant W à 2^{2^W}.

$\mathcal{D}_\alpha$ et $\mathcal{B}_i$ sont des relations d'accessibilité au sens habituel. $\mathcal{D}_\alpha(w)$ représente les résultats possibles d'un acte α. Nous parlons de $\mathcal{B}_i(w)$ comme étant l'état de croyance de l'agent i. $\mathcal{I}_i$ sont des fonctions de voisinage [8, Chap. 7]. Tout ensemble de mondes $U \in \mathcal{I}_i(w)$ correspond à une intention de i.

La relation de satisfaction $\Vdash$ est définie de manière usuelle. Une abréviation utile est $[\![A]\!] = \{w \in W : w \Vdash A\}$, appelée *l'extension de la formule* A. Alors :

- $w \Vdash P(t_1, \ldots, t_n)$ si $\langle V_w(t_1), \ldots, V_w(t_n)\rangle \in V_w(P)$;
- $w \Vdash Done_\alpha A$ si $\mathcal{D}_\alpha(w) \cap [\![A]\!] \neq \emptyset$;
- $w \Vdash Bel_i A$ si $\mathcal{B}_i(w) \subseteq [\![A]\!]$;
- $w \Vdash Intend_i A$ si $[\![A]\!] \in \mathcal{I}_i(w)$.

Pour la suite, nous avons besoin de la définition suivante : $Atm_W(w,T) = \{p \in ATM \mid w \Vdash p$ et $\mathfrak{subject}(p) \subseteq T\}$.

Ainsi $Atm_W(w,\mathfrak{competence}(i))$ est la partie du monde réel w sur laquelle i est compétent, et $Atm_W(w,\mathbb{T}_0 \setminus \mathfrak{scope}(\alpha))$ est la partie de w indépendante de α.

La CONTRAINTE D'ADOPTION de croyance stipule qu'en cas de compétence, la croyance devrait s'assimiler à la connaissance.

- Pour tout $w \in W$ et tout agent i, il existe $v \in \mathcal{B}_i(w)$ tel que $Atm_W(w,\mathfrak{competence}(i)) = Atm_W(v,\mathfrak{competence}(i))$.

Cela signifie que dans l'état de croyance de i, il y a un « monde témoin » reflétant la partie du monde actuel sur laquelle i est compétent.

Les CONTRAINTES DE PRÉSERVATION s'établissent comme suit. Quel est l'effet d'un acte α sur le monde actuel ? Si un atome est indépendant de α, alors sa valeur de vérité devrait être préservée.

- Si $w' \in \mathcal{D}_\alpha(w)$, $Atm_W(w,\mathbb{T}_0 \setminus \mathfrak{scope}(\alpha)) = Atm_W(w',\mathbb{T}_0 \setminus \mathfrak{scope}(\alpha))$.

Quel est l'effet de α sur les croyances ? En accord avec [29, 20], nous supposons que quand un agent i apprend que α a été exécutée, alors il met à jour son état de croyance concernant le monde, en faisant correspondre chaque $v \in \mathcal{B}_i(w)$ à un nouvel ensemble de mondes $\mathcal{D}_{i:\alpha}(v)$ [13] :

- Si $w' \in \mathcal{D}_\alpha(w)$ alors $\mathcal{B}_i(w') = \bigcup_{v \in \mathcal{B}_i(w)} \mathcal{D}_{i:\alpha}(v)$

Nous devons ainsi introduire des actes contextuels dans notre langage. Nous étendons notre définition des actes de façon récursive :

$i{:}\alpha$ est un acte si i est un agent et α est un acte.

Nous parlons de $i{:}\alpha$ comme l'acte mental associé à α. Techniquement, ces actes peuvent être comparés à des fonctions de SKOLEM, qui sont des procédés pour obtenir la complétude : ce sera la même chose ici.

Comment α et $i{:}\alpha$ sont-ils reliés ? $i{:}\alpha$ étant l'image mentale de α, son scope est obtenu en enlevant ma_i de $\mathfrak{scope}(\alpha)$, *i.e.*

$$\mathfrak{scope}(i{:}\alpha) = \{c{:}t \mid ma_i{:}c{:}t \in \mathfrak{scope}(\alpha)\}.$$

Ainsi, notre fonction $\mathfrak{scope}$ est complètement définie par la partie non contextuelle de son domaine. En particulier, nous avons : $\mathfrak{scope}(i{:}i{:}\alpha) = \mathfrak{scope}(i{:}\alpha)$. Comme ici nous avons restreint le contexte d'un thème t à une longueur au plus égale à deux, nous supposons en conséquence que

$$\mathfrak{scope}(i{:}j{:}k{:}\alpha) = \mathfrak{scope}(i{:}j{:}\alpha).$$

Quel est l'effet des actes sur les intentions ? Nous ne considérons pas ici la génération de nouvelles intentions, parce que nous considérons que cela peut être fait lors d'une étape à venir séparée. Nous nous concentrons sur la préservation des intentions *via* l'indépendance [14]. À cette fin, étant donné un ensemble de mondes $U \subseteq W$, nous définissons

$$\mathfrak{subject}(U) = \bigcup_{\substack{\{p \in ATM \mid p \text{ apparaît} \\ \text{dans l'ensemble des} \\ \text{modèles minimaux de } U\}}} \mathfrak{subject}(p)$$

13. On pourrait penser au premier abord que i exécute mentalement, simplement le même acte α dans tous les mondes possibles. Cela ne peut pas être le cas. En effet, on peut penser que les actes de langage modifient toujours l'état de croyance des agents, alors que le monde physique reste inchangé. En conséquence, nous avons besoin d'un acte différent $i{:}\alpha$, dépendant de i et α, et reflétant l'effet de α sur l'état de croyance de i.

14. Il est déjà pris en compte au sein même de la sémantique que, parfois, des intentions sont abandonnées à cause de nouvelles croyances. Supposons par exemple $w \Vdash Intend_i A$, *i.e.* $[\![A]\!] \in \mathcal{I}_i(w)$. Soit $w' \in \mathcal{D}_\alpha(w)$ et supposons qu'il existe $v' \in \mathcal{B}_i(w')$ tel que $v' \Vdash A$. Alors, nous ne pouvons avoir $[\![A]\!] \in \mathcal{I}_i(w')$, car la contrainte liant $\mathcal{B}_i$ et $\mathcal{I}_i$ requiert que $[\![A]\!] \cap \mathcal{B}_i(w) = \emptyset$.

En accord avec [4, 10], une autre condition déterminant l'abandon d'intentions est la croyance selon laquelle ces intentions ne pourront jamais être satisfaites. Nous ne traitons pas ce cas ici, car il requiert un opérateur temporel que nous n'avons pas introduit ici dans le but de simplifier le présent exposé. Un tel opérateur modal est intégré dans notre cadre formel dans [32].

Cela nous permet de calculer le sujet de l'extension $[\![A]\!]$ d'une formule A. Nous sommes maintenant prêt pour définir la préservation des intentions :

– Si $w' \in \mathcal{D}_\alpha(w)$, $U \in \mathcal{I}_i(w)$, et $\mathfrak{scope}(i{:}\alpha) \cap \mathfrak{subject}(U) = \emptyset$ alors $U \in \mathcal{I}_i(w')$.

Ainsi, les intentions sont préservées si leur sujet n'est pas dans le scope de l'acte mental $i{:}\alpha$ associé à α.

5 Discussion

Nous avons esquissé une théorie du changement de croyances dans le contexte du dialogue. Elle est basée sur la notion de topique d'information, qui est exploitée au travers d'axiomes (basés sur les topiques) d'adoption et de préservation de croyance. PERRAULT et APPELT et KONOLIGE avaient argumenté que les défauts étaient des éléments cruciaux dans une théorie des actes de langage. En un sens, ce que nous réalisons est de transférer cette tâche dans les relations métalinguistiques de compétence et de scope. Cela permet de conserver un cadre de travail monotone.

Nous avons supposé que l'ensemble de topiques associé à une formule est déterminé par les formules atomiques apparaissant dans cette dernière. C'est très certainement un choix contestable. Celui-ci a été principalement motivé par une économie de représentation. Néanmoins, la voie utilisée permet à la préservation d'être saine.

Note fonction $\mathfrak{subject}$ est sensible à la syntaxe, dans le sens où des formules logiquement équivalentes n'ont pas forcément le même sujet. Cependant, il est important de noter que malgré cette sensibilité à la syntaxe au niveau de notre théorie métalinguistique des topiques, notre logique (et en particulier notre principe de préservation) ne l'est pas. En effet, on peut montrer que si $A \leftrightarrow A'$ et $Done_\alpha A \rightarrow A$, alors $Done_\alpha A' \rightarrow A'$.

Supposons par exemple que $\mathfrak{subject}(p) = \{t\}$, $\mathfrak{subject}(q) = \{t'\}$, et $\mathfrak{scope}(\alpha) = \{t'\}$. Alors, p et $p \wedge (q \vee \neg q)$ n'ont pas le même sujet, et $Done_\alpha p \rightarrow p$ est une instance de l'axiome de préservation, alors que $Done_\alpha(p \wedge (q \vee \neg q)) \rightarrow (p \wedge (q \vee \neg q))$ n'en est pas une. Néanmoins, cette dernière formule peut être déduite de la précédente par les principes standards de la logique modale : comme $p \leftrightarrow p \wedge (q \vee \neg q)$ nous avons $Done_\alpha p \leftrightarrow Done_\alpha(p \wedge (q \vee \neg q))$. Ainsi, $Done_\alpha p \rightarrow p$ est équivalent à $Done_\alpha(p \wedge (q \vee \neg q)) \rightarrow (p \wedge (q \vee \neg q))$.

Nous n'avons pas formulé d'axiomes de compositionalité si fort pour la fonction $\mathfrak{scope}$. La raison est qu'un acte de langage peut influencer plus que les topiques de son contenu propositionnel. Par exemple, le scope de $\langle \mathsf{Inform}_{u,s}\ Classe(1^{\mathrm{re}}) \rangle$ contient non seulement $ma_u{:}ma_s{:}class$ mais aussi $ma_u{:}ma_s{:}prix$. Notre hypothèse ici est que le scope d'un acte de langage est déterminé par le sujet de son contenu propositionnel et des contraintes d'intégrité (celles, par exemple, liant destinations, classes et prix). C'est l'objet de futures recherches.

Références

[1] Carlos E. Alchourrón, Peter Gärdenfors et David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(2), 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 Second Int. Workshop on Non-Monotonic Reasoning*, volume 346 of *LNAI*. Springer-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 de doctorat, Université Paris Nord, 1995.

[6] Daniel Büring. Topic. In Peter Bosch et Rob van der Sandt, éditeurs, *The Focus Book*. Cambridge Univ. Press, 1995.

[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] B. F. Chellas. *Modal Logic: an introduction*. Cambridge Univ. Press, 1980.

[9] Philip R. Cohen et Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2–3), 1990.

[10] Philip R. Cohen et Hector J. Levesque. Persistence, intentions, and commitment. In Philip R. Cohen, Jerry Morgan et Martha E. Pollack, éditeurs, *Intentions in Communication*. MIT Press, 1990.

[11] 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.

[12] 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.

[13] R. L. Epstein. *The Semantic Foundations of Logic Volume 1: Propositional Logic*. Kluwer Academic Publishers, 1990.

[14] Luis Fariñas del Cerro, Andreas Herzig, Dominique Longin et Omar Rifi. Belief reconstruction in cooperative dialogues. In Fausto Giunchiglia, éditeur, *Proc. of AIMSA'98*, volume 1480 of *LNAI*. Springer-Verlag, 1998.

[15] Julia Rose Galliers. Autonomous belief revision and communication. In Peter Gärdenfors, éditeur, *Belief Revision*. Cambridge Univ. Press, 1992.

[16] J. Ginzburg. Resolving questions I,II. *Linguistics and Philosophy*, 18, 1995.

[17] S. Gitton. *Mise en oeuvre d'un système expérimental de dialogue oral et modèle formel de traitements d'erreurs.* PhD thesis, Université de Rennes I, 1995.

[18] E. Giunchiglia, G. N. Kartha et V. Lifschitz. Representing action: indeterminacy and ramifications. *Artificial Intelligence*, 95, 1997.

[19] N. Goodman. About. *Mind*, LXX(277), 1961.

[20] Gösta Grahne. Updates and counterfactuals. In J. A. Allen, R. Fikes et E. Sandewall, éditeurs, *Proc. of KR'91*. Morgan Kaufmann Pub. , 1991. extended version to appear in the J. of Logic and Computation.

[21] H. Paul Grice. Logic and conversation. In J. P. Cole et J. L. Morgan, éditeurs, *Syntaxe and Semantics: Speech acts*. Academic Press, 1975. Également disponible dans [22, Chapitre 2].

[22] H. Paul Grice. *Studies in the way of words*. Harvard Univ. Press, 3e edition, 1989.

[23] David Harel. Dynamic logic. In D. Gabbay et F. Guenthner, éditeurs, *Handbook of Philosophical Logic*, volume II. D. Reidel Publishing Company, 1984.

[24] A. Herzig et D. Longin. Belief dynamics in cooperative dialogues. In Jan van Kuppevelt, Noor van Leusen, Robert van Rooy et Henk Zeevat, éditeurs, *Proc. of the Third Int. Workshop on the Semantics and Pragmatics of Dialogue (Amstelogue'99)*, 1999.

[25] A. Herzig et D. Longin. Belief dynamics in cooperative dialogues. *Journal of Semantics*, 2000. À paraître.

[26] A. Herzig et D. Longin. Beliefs, intentions, speech acts and topics. Technical Report 00-08-R, Institut de Recherche en Informatique de Toulouse (IRIT), Mars 2000. Avalaible on `http://www.irit.fr/ACTIVITES/EQ_ALG`.

[27] A. Herzig et D. Longin. Towards an analysis of dialogue acts and indirect speech acts in a BDI framework. In Massimo Poesio et David Traum, éditeurs, *Proc. of the Fourth Int. Workshop on the Semantics and Pragmatics of Dialogue (Götalog-2000)*, Juin 2000.

[28] Andreas Herzig et Omar Rifi. Propositional belief base update and minimal change. *Artificial Intelligence*, 115(1), 1999.

[29] 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.

[30] Kurt Konolige et Martha E. Pollack. A representationalist theory of intention. In *Proc. of IJCAI'93*. Morgan Kaufmann Pub. , 1993.

[31] D.K. Lewis. General semantics. In D. Davidson et G. Harman, éditeurs, *Semantics of natural language*. D. Reidel Publishing Company, 1972.

[32] Dominique Longin. *Interaction rationnelle et évolution des croyances dans le dialogue : une logique basée sur la notion de topique*. Thèse de doctorat, Université Paul Sabatier, 1999.

[33] J. McCarthy et P. Hayes. Some philisophical problems from the standpoint of artificial intelligence. In B. Meltzer et D. Michie, éditeurs, *Machine Intelligence*, volume 4. Edinburgh University Press, 1969.

[34] 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.

[35] 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. of KR'91*. Morgan Kaufmann Pub. , 1991.

[36] Anand S. Rao et Michael P. Georgeff. An abstract architecture for rational agents. In Bernhard Nebel, Charles Rich et William Swartout, éditeurs, *Proc. of KR'92*. Morgan Kaufmann Pub. , 1992.

[37] Anand S. Rao et Michael P. Georgeff. Decision procedures for BDI logics. *Journal of Logic and Computation*, 8(3), Juin 1998. Special Issue: Computational and Logical Aspects of Multiagent Systems, Oxford univerty Press.

[38] Ray Reiter. A logic for default reasoning. *Artificial Intelligence*, 13, 1980.

[39] David Sadek, Philippe Bretier et Franck Panaget. ARTIMIS: Natural dialogue meets rational agency. In Martha E. Pollack, éditeur, *Proc. of IJCAI'97*, volume 2. Morgan Kaufmann Pub. , Août 1997.

[40] M. D. Sadek. *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication*. Thèse de doctorat, Université de Rennes I, Rennes, France, Juin 1991.

[41] M. D. Sadek. Dialogue acts are rational plans. In *Proc. of ESCA/ETRW, Workshop on The Structure of Multimodal Dialogue (Venaco II)*, Septembre 1991.

[42] M. D. Sadek. A study in the logic of intention. In Bernhard Nebel, Charles Rich et William Swartout, éditeurs, *Proc. of KR'92*. Morgan Kaufmann Pub. , Octobre 1992.

[43] 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.

[44] 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. of ICSLP'96 Int. Conf. on Spoken Language Processing*, Octobre 1996.

[45] Erik Sandewall. *Features and Fluents*. Oxford Univ. Press, 1994.

[46] J. R. Searle. *Intentionality: An essay in the philosophy of mind*. Cambridge Univ. Press, 1983.

[47] J. R. Searle et D. Vanderveken. *Foundation of illocutionary logic*. Cambridge Univ. Press, 1985.

[48] John R. Searle. *Speech acts: An essay in the philosophy of language*. Cambridge Univ. Press, 1969.

[49] Michael Thielscher. Computing ramifications by postprocessing. In *Proc. of IJCAI'95*, 1995.

[50] J. van Kuppevelt. *Topic en Comment*. Thèse de doctorat, University of Nijmegen, 1991.

[51] J. van Kuppevelt. Discourse structure, topicality and questioning. *Linguistics*, 31, 1995.