

# Sourcer, dater et mémoriser les informations d'origine verbale - proposition de modèle symbolique et opérationnel de l'interface langage/mémoire

---

Raoul Blin

CNRS-CRLAO, France

blin@ehess.fr

## Résumé

Nous proposons un modèle symbolique, opérationnel, capable de rendre compte de la façon dont un allocutaire, dans une interaction verbale, exploite les données qui lui sont fournies pour construire ses propres connaissances, et ses croyances sur les connaissances du locuteur («théorie de l'esprit»). Le modèle est formulé en logique classique. Nous illustrons son fonctionnement en prenant exemple sur la psychologie (fausse croyance), l'intelligence artificielle (révision des connaissances) et la linguistique (interprétation de phrases simple et avec verbes épistémiques).

## Abstract

We present a symbolic, operational model capable of accounting for the way in which a speaker, in a verbal interaction, exploits the data provided to him to construct his own knowledge, and his beliefs about the speaker's knowledge («theory of mind»). We describe the main features of the model by dealing with cases that concern psychology (false belief), artificial intelligence (knowledge revision) and linguistics (sentence interpretation with epistemic verbs).

## 1 Introduction

Dans le cadre d'une étude linguistique sur l'interprétation des expressions épistémiques verbales (verbes *savoir*, *croire*, *apprendre que*, etc.), nous recherchions un modèle symbolique capable de rendre compte des états mentaux décrits par ces expressions. Ce modèle devait pouvoir rendre compte au minimum de la distinction entre connaissances propres et connaissances attribuées (théorie de l'esprit [12] et mémoire de source [5]) et de la temporalité des connaissances. Par exemple

- (1) Hier encore, Carole ne savait pas [qu'il avait plu avant-hier]<sub>P</sub>.

indique explicitement à l'allocutaire que d'après le locuteur, la donnée P n'était pas une connaissance de Carole. Il indique implicitement que c'est une connaissance du locuteur. Bien sûr, l'énoncé ne dit rien de l'état des connaissances de l'allocutaire. Par contre, l'énoncé provoquera des modifications des connaissances de l'allocutaire. Temporellement, la date de l'événement décrit par P est distincte de la date de l'énonciation d'une part, et de la période à laquelle le locuteur estime que Carole ignore P. Le locuteur dit explicitement que cette période d'ignorance inclut hier, mais il ne dit pas si cette ignorance est toujours en cours au moment de l'énonciation. Il indique aussi implicitement que P est une donnée que lui-même tient présentement pour valide. Ces inférences ne constituent qu'une partie de tout ce qui peut être inféré par l'allocutaire sur les connaissances du locuteur. Nous en verrons bien d'autres.

Nous nous sommes naturellement tourné vers les logiques modales mais celles-ci se sont révélées frustrantes. Leur spécialisation (épistémique, déontique, temporelle, etc) ne les destine pas *a priori* à traiter en même temps deux phénomènes aussi éloignés que la théorie de l'esprit d'une part et la temporalité d'autre part. Cela ne signifie pas que c'est infaisable. Mais pour l'instant, à notre connaissance, personne n'a proposé de tel modèle modal et montré l'intérêt de l'approche modale par rapport à d'autres logiques pour traiter cette combinaison de phénomènes. Nous nous sommes plus particulièrement intéressé à la logique épistémique mais elle est à bien des égards peu réaliste d'un point de vue cognitif. Nous mentionnerons quelques problèmes dans cette présentation (ce qui ne nous empêche pas par ailleurs d'en reprendre certaines idées).

Nous avons aussi cherché du côté des modèles en intelligence artificielle symbolique. Les recherches

proposent des architectures cognitives (une revue entre autres dans [11], [3], etc.) mais pas de modèles qui manipulent ensemble les données que nous traitons, à savoir théorie de l'esprit et temporalité. Quelques études (par exemple [13]) se sont intéressées au langage mais sont difficiles à exploiter pour une étude linguistique fine.

Finalement, ne trouvant pas de modèle prêt à l'emploi, nous avons décidé de créer notre propre modèle avec le cahier des charges suivant. Le modèle doit être exploitable aussi bien en linguistique qu'en psychologie. Il doit donc couvrir une vaste plage de phénomènes et être révisable en vue d'en intégrer de nouveaux. Il doit aussi pouvoir être intégré dans des modèles de portée plus vaste encore. Pour répondre à ces contraintes, nous avons décidé de décrire le modèle à l'aide d'un langage simplement typé. Le fonctionnement du modèle, les données et les règles sont explicitement données formulées. Elles ne sont pas tenues pour des axiomes d'un «nouveau» système d'inférences et ne requiert pas de nouveau modèle sémantique.

Le présent texte décrit le modèle issu de nos observations. Tout d'abord, nous introduisons une première version du modèle, capable de distinguer les connaissances propres d'un agent et ses croyances sur les connaissances d'autrui. A l'aide de ce modèle, nous simulons l'état de fausse croyance, phénomène très étudié en psychologie. Puis nous introduisons la temporalité. Nous pouvons alors simuler l'évolution des connaissances chez un individu (enrichissement, modification). Le modèle permet aussi de décrire l'acquisition de connaissances d'origine verbale. Nous l'utiliserons pour décrire le traitement «mémoire» d'énoncés simples avec et sans expression verbale épistémique (*savoir que*).

Nous conviendrons de désigner par «expression» toute chaîne (phrase, syntagme, mot, etc) bien formée du langage naturel. Dans les exemples les expressions entre crochets {...} sont données à titre indicatif, pour faciliter la compréhension. Deux points d'interrogations ?? indiquent que l'expression est peu naturelle ou difficile à interpréter. Un «terme» désigne toute formule logique bien formée. Dans les termes, les symboles non quantifiés explicitement sont des constantes. Les connecteurs logiques sont associatifs à droite.

Nous utiliserons les termes de «mémoire» en prenant peut-être quelques libertés par rapport à l'usage qui peut en être fait dans les domaines de spécialités. Employé dans la discussion, il s'agit de l'ensemble des connaissances ou état mental d'un agent. Nous l'utiliserons aussi pour décrire la base de connaissances utilisée pour modéliser l'ensemble des connaissances d'un agent. Le terme «mémoire» est employé dans le même esprit.

## 2 La structure générale du modèle

Pour interagir verbalement avec autrui (transmettre ou recevoir des informations, provoquer chez autrui une réaction), tout agent met en oeuvre ses propres connaissances, et celles qu'il attribue à autrui. Pour décrire l'état mental d'un agent au cours d'une interaction verbale, il faut se donner un modèle de son système de connaissances et des opérations qu'il peut faire sur ses connaissances, et y intégrer la distinction entre connaissances propres et attribuées. Dans cette section, nous jetons les bases du modèle du système de connaissances et montrons comment il permet de simuler la théorie de l'esprit et le cas particulier de fausse croyance. La principale caractéristique du modèle est d'utiliser une base de connaissances stratifiée de façon récursive.

### 2.1 Les éléments de base

De façon très classique, nous simulons l'état des connaissances (nous simplifierons parfois en parlant de «mémoire») d'un individu, à l'aide d'une base de connaissances.

Avec le prédicat  $bc$ , on signifie qu'une donnée  $P$  (représentée par une proposition  $p$ ) figure parmi les connaissances d'un agent  $A$  (représenté par la constante individuelle  $a$ ), ou bien que cette donnée est inférable à partir des connaissances de  $A$  :  $bc(a, p)$ .  $A$  est le «propriétaire» de la connaissance. La négation  $\neg bc(a, p)$  signifie que l'agent ne possède pas l'information ni ne peut l'inférer. On prêtera attention au fait que  $bc(a, \neg p)$  signifie que  $\neg p$  est une connaissance de  $A$ . Cela ne signifie pas que  $A$  ne possède pas la connaissance  $\neg p$ .

Attribuer une connaissance  $P$  à un agent  $A$  et  $P'$  à un agent  $B$  se note simplement :  $bc(a, p) \wedge bc(b, p')$ , ou plus classiquement  $\{bc(a, p), bc(b, p')\}$ . Désormais les crochets seront omis. On glosera un terme  $bc(a, p)$  « $p$  est une connaissance de  $a$ » ou plus simplement, si aucune confusion entre glose et langue décrite n'est possible, « $A$  sait que  $P$ ».

On se dote de plusieurs règles rendant compte de raisonnements triviaux, à commencer par la «distributivité des connaissances pour la conjonction». Le dialogue suivant montre que les deux données d'une conjonction peuvent être séparées en deux connaissances.

- (2) Lucie à Alain : Carole sait qu'[il a neigé avant-hier] <sub>$p_1$</sub>  et [plu hier] <sub>$p_2$</sub>   
 Eric à Alain : Carole sait-elle qu'[il a neigé avant-hier] <sub>$p_1$</sub>  ?  
 Alain : oui  
 Eric à Alain : sait-elle qu'[il a plu hier] <sub>$p_2$</sub>  ?  
 Alain : aussi oui

Un dialogue similaire montrerait qu'à partir des connaissances séparées des deux données, on peut inférer

la connaissance de la conjonction des deux données. Nous rendons compte de cette relation entre connaissances :

- (R<sup>1</sup>) Distributivité des connaissances  
pour la conjonction (version 1)  
 $\forall x \forall p \forall p'$   
 $BC(x, p \wedge p') \equiv BC(x, p) \wedge BC(x, p')$

Il est difficile de créer ce type de dialogue pour la disjonction, ce qui nous amène à poser que la distributivité n'est pas valide pour la disjonction.

La notation n'impose aucune contrainte sur le terme  $p$ . Il est possible d'exprimer aussi bien des faits que des règles :

- (3) Alain sait qu'[il a plu hier].  
 $BC(a, pleuvoir(e) \wedge date(e) = hier)$   
Alain sait que [quand il pleut, il ne neige pas].  
 $BC(a, \forall e^1 pleuvoir(e^1) \supset \neg \exists e^2 neiger(e^2))$

Pour signifier que deux agents A et B partagent une même donnée P, il suffit de noter :  $BC(a, p) \wedge BC(b, p)$ . On pourrait envisager une notation plus compacte comme  $BC(\{a, b\}, p)$  mais ici nous n'abordons pas les questions d'optimisation de l'écriture. Pour des questions de place, nous n'abordons pas non plus le traitement des connaissances d'un groupe d'agents.

## 2.2 Théorie de l'esprit

En psychologie, la «théorie de l'esprit» désigne la capacité d'un individu à attribuer des états mentaux à autrui, parmi lesquels, des connaissances. C'est une compétence cognitive importante dans les interactions, y compris les interactions verbales ([9],[2]). Nous montrons comment notre modèle rend compte de l'attribution de connaissances.

Le fait qu'un «agent A sache qu'un agent B sait P» est rendu en jouant sur la récurrence :

$$BC(a, BC(b, p))$$

Nous utiliserons l'abréviation  $BC^n(x^n, p)$  pour désigner le terme  $BC(x^1, BC(x^2, \dots, BC(x^n, p) \dots))$  lorsque  $x^1$  désigne l'individu observé et qu'il n'est pas nécessaire de préciser l'identité de  $x^2 \dots x^{n-1}$ .  $n$  est la «profondeur». Le n-uplet  $(x^1, \dots, x^n)$  est la «chaîne de propriétaires». Nous appellerons «sous-base» tout ensemble de données  $\lambda p BC^n(x^n, p)$  pour une chaîne de propriétaires donnée.

Formellement, la notation ne restreint pas la profondeur. Il est cependant clair que dans les faits, les capacités cognitives d'un humain ne permettent pas une profondeur infinie. C'est perceptible à travers les limites imposées par la langue. Au delà de deux itérations, les phrases qui rendent compte de l'état des connaissances sur autrui deviennent difficiles à interpréter ( $k$  la constante correspondant à Carole) :

- (4) ?? Alain : ⟨je sais que⟩ Lucie sait que Carole sait que [Martine sait qu'il a plu]<sub>P</sub>.  
 $BC(a, BC(l, BC(k, BC(m, p))))$

Formellement, il n'y a pas non plus de restrictions sur les agents. Un agent peut apparaître plusieurs fois dans une chaîne de propriétaires. Ainsi, un agent A peut avoir des croyances sur les croyances d'autrui (B) sur lui-même (A) :  
« ⟨je (=A) sais que⟩ B sait que je (=A) sais que P » :

$$BC(a, BC(b, (BC(a, p))))$$

La notation permet de retranscrire l'axiome 4 ( $\Box P \supset \Box \Box P$ ), dit d'introspection positive, de la logique épistémique.

$$\text{«si X sait P, il sait qu'il sait P»}$$

$$\forall x \forall p BC(x, p) \supset BC(x, BC(x, p))$$

Pour autant, nous n'inscrirons pas cet axiome dans notre propre propre ensemble de règles car on peut douter qu'il s'applique à tout individu. Peut-on vraiment considérer qu'il est valable pour un enfant en bas âge par exemple? Par ailleurs, dans quelle mesure un adulte, même de développement typique, est systématiquement conscient de ce qu'il sait? Beaucoup de connaissances sont utilisées «inconsciemment». Ce qui soulève d'ailleurs la question de conscience : M.Jourdain «sait» faire de la prose sans en être conscient. L'axiome de la logique épistémique n'est pas claire sur ce point. Pour notre part, la présence d'une donnée dans la base signifie que cette donnée est mémorisée, pas que son propriétaire est nécessairement conscient de posséder cette donnée. Il nous semble préférable de laisser à l'observateur le soin de décider d'inscrire ou non cet axiome dans sa propre axiomatique, selon les caractéristiques de l'agent à modéliser (agent adulte de développement typique, enfant, etc.)

Remarquons ici qu'en ne recourant pas à une logique modale donnée avec son axiomatique propre, le modèle est d'un usage plus souple. On peut modéliser le comportement de deux agents de capacités différentes sans avoir à adopter un système logique pour l'un (avec axiome 4 par exemple) et un autre système (sans l'axiome) pour l'autre agent. Dans notre cas, le système axiomatique est unique et commun aux agents. Le «réglage individuel» se fait par adjonction de règles faciles à cibler.

Bien que ce soit techniquement faisable, nous n'adoptons pas l'axiome 5, dit d'introspection négative, de la logique épistémique.

$$\text{«si X ne sait pas que P, il sait qu'il ne sait pas que P»}$$

$$\forall x \forall p \neg BC(x, p) \supset BC(x, \neg BC(x, p))$$

Nous postulons que cet axiome n'est cognitivement pas réaliste. Prenons un exemple. Le concept «La physique quantique est hilarante» est absente de l'esprit d'un enfant de quatre ans. Un enfant de cet âge n'a jamais entendu parler

de physique quantique, il n'y a jamais pensé, et a fortiori il n'a jamais conçu l'idée que cette «chose» pouvait être hilarante. Il n'est à aucun moment en mesure de savoir qu'il ne sait pas.

Insistons enfin sur le cloisonnement des données. La théorie de l'esprit nécessite de bien séparer les connaissances propres et attribuées. Avec notre notation, les connaissances de A ( $\lambda p \text{ BC}(a, p)$ ) sont strictement séparées des connaissances que A attribue à B ( $\lambda p \text{ BC}(a, \text{BC}(b, p))$ ). Néanmoins, le partage de données entre sous-bases n'est pas interdit. Simplement, il ne peut se faire que lorsqu'il est explicitement autorisé. Ainsi, lorsqu'un agent A accorde du crédit à un autre agent B, il peut faire siennes les données de l'agent B. Nous parlerons «d'appropriation des connaissances». Si par exemple on voulait rendre compte du fait que les enfants adoptent les connaissances de leurs parents, nous pourrions écrire :

$$\forall x \forall y \text{ enfant}(y) \wedge \text{parente}(x, y) \\ \supset (\forall p \text{ BC}(y, \text{BC}(x, p)) \supset \text{BC}(y, p))$$

La théorie de l'esprit ne concerne pas seulement des connaissances statiques. Elle touche aussi les raisonnements : un agent doué de théorie de l'esprit distingue ses propres raisonnements des raisonnements qu'il attribue à autrui. La règle suivante rend compte de cette faculté en permettant des raisonnements (modus ponens) à l'intérieur d'une sous-base. Cette règle correspond à l'axiome **K** dit «de distribution» de la logique épistémique.

$$(R^2) \text{ Règle de distribution (ou } \mathcal{MP}), \text{ version 1} \\ \frac{\text{BC}^n(x^n, \alpha) \quad \text{BC}^n(x^n, \alpha \supset \beta)}{\text{BC}^n(x^n, \beta)}$$

### 3 Fausse croyance

Le phénomène de fausse croyance ([14]) joue un rôle essentiel dans les études sur la théorie de l'esprit, à travers les épreuves dites de fausse croyance ([1]).

La fausse croyance désigne une situation où un agent observé (A) considère qu'un agent (autre que lui-même) (B) a une connaissance erronée, c'est à dire une connaissance en contradiction avec ses (=A) propres connaissances. Ce que l'agent A pourrait décrire <sup>1</sup> «*je sais que*» B sait que P tandis que moi-même je sais que non-P» :

$$\text{BC}(a, p) \wedge \text{BC}(a, \text{BC}(b, p))$$

Le phénomène peut avoir lieu à n'importe quel niveau d'itération, comme le montre la phrase suivante. Alain sait que Lucie attribue à Carole une fausse croyance, donc une connaissance qu'il sait erronée aux yeux de Lucie.

$$(5) \text{ Alain, Lucie sait que Carole croit à tort que P.} \\ \text{BC}(a, \text{BC}(l, \neg p) \wedge \text{BC}(l, \text{BC}(k, p)))$$

1. Rappelons et insistons sur le fait que dans les gloses, «savoir que P» et «croire que P» signifient «P est une connaissance» ou «P est dans la base de connaissance».

Voici une formulation très générale, indépendante du niveau d'itération, de la situation où un agent  $x^n$  attribue une fausse croyance  $p$  à un agent  $x^{n+1}$  : Cette représentation ne dit pas si les autres agents de la chaîne de propriétaires (en particulier l'agent  $x^1$ ) accordent ou non du crédit à la croyance P.

$$\text{(Hyp) Fausse croyance attribuée par } x^n \text{ à } x^{n+1} \\ \text{BC}^n(x^n, \text{BC}(x^{n+1}, p)) \wedge \text{BC}^n(x^n, \neg p)$$

Voici à titre indicatif la notation pour un cas particulier de fausse croyance où un agent A sait qu'un agent B lui (=A) attribue une croyance erronée, c'est à dire en contradiction ce que A sait :

$$(6) \text{ Alain}^A : \text{Lucie}^B \text{ croit (à tort) que je}^A \text{ crois que P} \\ \text{(alors qu'en réalité, je sais que non-P)} \\ \text{BC}(a, \neg p) \wedge \text{BC}(a, \text{BC}(b, \text{BC}(a, p)))$$

### 4 Détection d'un mensonge

On peut aussi modéliser la capacité à détecter un mensonge. Considérons le propos suivant :

$$(7) \text{ Lucie}^B \text{ à Alain}^A : \text{Il a plu hier.} \quad (\text{P})$$

Pour qu'un agent A considère P comme un mensonge, il faut qu'il sache lui-même que P n'est pas vrai ( $c^1$  dans l'hypothèse ci-dessous). Cela ne suffit pas pour distinguer une simple erreur et un mensonge. Il faut en plus que l'allocutaire détecte une intentionnalité de la part du locuteur. «L'intentionnalité» n'est pas un concept aisé à définir et nous nous garderons de le faire. Par contre, on peut utiliser des indices. A peut suspecter le locuteur B de mentir si il sait que B tient P pour faux ( $c^2$ ) («*je sais que*» B dit P alors qu'il sait non-P»). Enfin, on ment en général à quelqu'un qui ignore la vérité ( $c^3$ ). L'allocutaire peut donc être alerté si il sait que le locuteur le croit ignorant («le locuteur croit que j'ignore la vérité, c'est-à-dire qu'il croit que je ne sais pas non-P, ou il croit que je crois P»).

En résumé :

(Hyp) Mensonge

$$\text{Un agent A peut soupçonner un agent B de mentir} \\ \text{si B tient le propos P et que A est dans l'état mental} \\ \text{BC}(a, \neg p) \quad (c^1) \\ \text{BC}(a, \text{BC}(b, \neg p)) \quad (c^2) \\ \text{BC}(a, \text{BC}(b, \neg \text{BC}(a, \neg p))) \vee \text{BC}(a, p)) \quad (c^3)$$

### 5 Datation des connaissances

Les connaissances des individus évoluent avec le temps. Un individu de développement typique est capable de mémoriser une bonne partie de ses connaissances passées, sans les confondre avec celles présentes. Le langage naturel met d'ailleurs à disposition de nombreuses expressions pour rendre compte de l'état des connaissances passées :

- (8) Hier je pensais qu'[avant hier il avait plu]<sub>P</sub> <mais> j'ai appris ce matin qu'en fait, [il avait neigé]<sub>P</sub>. Je sais donc désormais qu' [il n'avait pas plu]<sub>non-P</sub>.

Cet énoncé peut se gloser «P était une de mes connaissances jusqu'à ce matin. Depuis ce matin, P n'est plus une connaissance valide mais elle ne disparaît pas de ma mémoire pour autant. Par ailleurs, ce matin, j'ai acquis une nouvelle connaissance P'. Cette connaissance est toujours valide présentement. Depuis l'instant où P' est devenue valide, je suis capable d'inférer non-P.»

Mémoriser la chronologie des connaissances étant nécessaire pour une cognition efficiente ([6]), il faut l'intégrer à notre modèle. Dans un premier temps, nous décrivons le dispositif formel. Puis nous montrons comment la datation offre une solution simple pour gérer les mises à jour des (sous-)bases de connaissances et pour rendre compte de l'évolution des connaissances.

### 5.1 Modéliser la datation des connaissances

On adopte une indexation des connaissances en s'inspirant de la notation néo-davidsonienne ([10]) utilisée en sémantique des langues. Cela nous permet d'utiliser un langage commun pour décrire les connaissances et les représentations sémantiques. L'interfaçage entre données linguistiques et cognition en est facilité.

Désormais,  $bc$  est un prédicat à trois arguments. Le terme  $bc(e, a, p)$  signifie qu'une donnée P figure dans les connaissances de l'agent A.  $e$  est l'index de cette connaissance. L'itération reste bien sûr possible :  $bc(e^1, x^1, bc(e^2, x^2, \dots bc(e^n, x^n, p)))$ .  $(e^1, \dots, e^n)$  est une «chaîne d'index».

L'index obéit à la règle d'unicité (règle  $R^3$ ), classique dans la notation néo-davidsonienne. Nous la formulons informellement :

( $R^3$ ) Unicité de l'index

Chaque connaissance, quel que soit le niveau d'itération, dispose d'un index (symbole individuel  $e^i$ ) qui lui est propre.

Cela signifie que pour une chaîne d'index, une chaîne de propriétaires et une donnée particulières, l'index est unique. A l'inverse, à un index ne peut correspondre qu'un seul triplet chaîne d'index, chaîne de propriétaires et donnée.

On impose à toute connaissance d'avoir une date (*date*) «d'activité» (terme que l'on préférera à celui de «validité», susceptible d'entraîner des confusions).

( $R^4$ ) Obligation de temporalité

$$\forall e \forall x \forall p \ bc^n(e^n, x^n, p) \supset \exists t \ date(e) = t$$

Toute date d'activité a un début ( $deb(e)$ ) et une fin ( $fin(e)$ ). La fonction  $maint()$  décrit «l'instant présent». L'abréviation *actuel* signifie qu'une connaissance est active à l'instant présent :

$$actuel = \lambda e \ maint() \in date(e)$$

Par exemple, rendons compte d'une connaissance de P que Lucie peut gloser comme suit, à la date  $maint() = t$ .

- (9) Lucie : Je sais depuis ce matin qu'[il a neigé avant-hier]<sub>P</sub>

$$neiger(e^1) \wedge date(e^1) = avthier \quad (p)$$

$$bc(l, e^2, p) \wedge deb(e^2) = cematine \quad (c^1)$$

$$actuel(e^2) \quad (c^2)$$

Le terme  $c^1$  rend compte de la présence de P dans les connaissances de Lucie et indique le début d'activité de cette connaissance.  $c^2$  rend compte du fait que la connaissance est toujours active au moment présent. La règle  $R^4$  dit que cette connaissance aura une fin mais celle-ci n'est pas spécifiée. Une règle pourrait identifier par défaut la fin de toute connaissance et la disparition (physique) du propriétaire. Mais ce genre de considération dépasse le cadre de notre étude.

Il n'existe pas de contraintes sur la relation temporelle entre les dates  $td$  «contenues» dans une donnée et la date d'activité  $ta$  de la connaissance de cette donnée. Ainsi, on peut acquérir aujourd'hui ( $ta$ ) une donnée relative à un événement qui a eu lieu à la date antérieure  $td$ . C'est le cas de l'exemple ex.9 où l'événement décrit a lieu avant hier, et sa connaissance débute ce matin.

La datation permet de faire coexister chez un agent des connaissances contradictoires dès lors qu'elles sont actives à des dates différentes. Par exemple, (1) «jusqu'à hier soir, je croyais qu'[il avait neigé avant-hier]<sub>P</sub>.» (nous reprenons la représentation précédente de P). Puis (2) «Et soudainement, hier soir, j'ai réalisé qu'[il n'avait pas du tout neigé]<sub>non-P</sub>». Autrement-dit, jusqu'à hier soir, ma connaissance de P était active. Depuis hier soir, elle est désactivée, et c'est ma connaissance de non-P qui est active au moment où je parle.

$$1) \ bc(l, e^1, a, p) \wedge fin(e^1) \in hiersoir$$

$$2) \ bc(l, e^2, a, \neg p) \wedge deb(e^1) = fin(e^1) \wedge actuel(e^2)$$

Il faut adapter les règles d'inférence pour prendre en compte la temporalité. Pour appliquer le modus ponens sur deux connaissances, il faut, comme auparavant, que ces connaissances aient la même profondeur (nombre d'itérations) et la même chaîne de propriétaires. Nous imposons en plus que les connaissances aient une date commune. La connaissance inférée sera active durant la date commune. Par exemple, je connais depuis mon enfance la règle que si il pleut il ne neige pas et vice versa (P). Par ailleurs, j'ai appris ce matin (P') qu'il avait plu avant hier. Je peux donc, depuis ce matin, c'est à dire dès que les deux connaissances sont actives en même temps, inférer (P'') qu'il n'a pas neigé avant hier. Je n'aurai pas été capable de faire cette inférence avant l'acquisition de la connaissance de P'. On rend compte

formellement des contraintes temporelles en n'activant un modus ponens que pour l'intervalle d'activité commun aux deux connaissances manipulées. Automatiquement, le modus ponens sera empêché si l'intersection des dates est nulle. Cet empêchement est déclenché par la règle (R<sup>4</sup>), qui impose à toute connaissance d'avoir une date d'activité. Voici la reformulation de la règle de distribution (R<sup>2</sup>).

(R<sup>5</sup>) Règle de distribution (ou  $\mathcal{MP}$ ) (version 2)

$$\frac{\text{BC}^n(x^n, e^1, \alpha) \quad \text{BC}^n(x^n, e^2, \alpha \supset \beta)}{\text{BC}^n(x^n, e^3, \beta) \wedge \text{date}(e^3) = \text{date}(e^1) \cap \text{date}(e^2)}$$

Il faut noter qu'avec cette règle, la date de réalisation d'un raisonnement est indépendante de la date d'activité des connaissances de ces données. On peut ainsi reconstituer des raisonnements passés, mais sans en tirer des connaissances actives au moment où est réalisé le raisonnement. Pour en rendre compte, il faut que les connaissances inférées soient activées à la même date que les connaissances utilisées, et non pas à la date de réalisation du raisonnement. Il n'y a donc pas de risque d'anachronisme. Dans l'exemple qui suit, la connaissance P<sup>1</sup> n'est désormais plus active. Le locuteur peut l'exploiter mais la période d'activité de sa conclusion P<sup>3</sup> sera identique à celle de la connaissance de P<sup>1</sup>. Il n'y a pas de conflit entre P<sup>3</sup> et P<sup>4</sup> parce qu'au moment (actuel) où P<sup>4</sup> est active, P<sup>3</sup> ne l'est plus.

(10) Avant, [lorsque je voyais un sol mouillé, je pensais que c'était à cause de la pluie]<sub>P<sup>1</sup></sub>. C'est pourquoi [en voyant ta terrasse mouillée]<sub>P<sup>2</sup></sub>, j'avais pensé qu'[il avait plu]<sub>P<sup>3</sup></sub>. Mais je sais désormais que [c'est mouillé par les arrosoirs]<sub>P<sup>4</sup></sub>.

Il faut aussi adapter la règle de distributivité par rapport à la conjonction (R<sup>6</sup>). A partir des deux connaissances suivantes, acquises à des dates différentes mais toujours en vigueur au moment de l'énonciation :

(11) Alain : J'ai su avant hier qu'[il était parti en mars]<sub>P<sup>1</sup></sub> et aujourd'hui qu'[il est revenu en février]<sub>P<sup>2</sup></sub>.  
 $\text{BC}(a, e^1, p^1) \wedge \text{BC}(a, e^2, p^2)$   
 $\wedge \text{deb}(e^1) = \text{avthier} \wedge \text{deb}(e^2) = \text{aujourd'hui}$   
 $\wedge \text{actuel}(e^1) \wedge \text{actuel}(e^2)$

on peut inférer

$\Rightarrow$  présentement, je sais [qu'il est parti en mars et revenu en février].  
 $\text{BC}(a, e^3, p^1 \wedge p^2) \wedge \text{actuel}(e^3)$

Pour réécrire la règle, il faut déterminer la date d'activité de la connaissance de la conjonction des données. On serait tenté d'identifier son début et celui de la connaissance acquise en dernier. Nous n'avons pas trouvé de tests capables de confirmer cette hypothèse. L'hypothèse la plus prudente nous semble au final d'identifier la date d'activité de la connaissance d'une conjonction de connaissances à

l'intersection des dates des connaissances d'origine. Cela nous semble valide aussi pour la datation de la connaissance de la conjonction, inférée à partir des connaissances distinctes des données.

(R<sup>6</sup>) Distributivité de la conjonction

par rapport aux connaissances (version 2)

$$\forall x \forall p \forall p' (\exists e^1 \text{BC}^n(x^n, e^1, p \wedge p')) \\ \equiv \exists e^2 \exists e^3 \text{BC}^n(x^n, e^2, p) \wedge \text{BC}(x^n, e^3, p') \\ \wedge \text{date}(e^1) = \text{date}(e^2) \cap \text{date}(e^3)$$

R<sup>6</sup> permet fusionner des informations redondantes. Lorsqu'une information est acquise deux fois ou plus (1), on ne peut pas pour autant considérer qu'elle génère deux connaissances distinctes. La règle permet de fusionner les données (2) :

- (1)  $\text{BC}(x, e^1, p) \wedge \text{BC}(x, e^2, p)$   
(2) (1),  $\text{Ex.6} \vdash \exists e^3 \text{BC}(x, e^3, p \wedge p)$

## 6 Mise à jour des connaissances

Avec la datation, mettre à jour les connaissances (expansion, révision) consiste à clôturer et/ou faire débiter la date d'activité d'une connaissance.

L'expansion consiste à ajouter une nouvelle connaissance et en faire débiter l'activité à la date d'acquisition. Par exemple, pour rendre compte du fait qu'à la date  $t$ , Lucie apprend P, on note :

$$\text{BC}(l, e, p) \wedge \text{deb}(e) = t$$

Une connaissance est révisée lorsqu'une donnée est changée en partie ou complètement. Ce qui est important, c'est que la connaissance de la donnée reste dans la base et qu'il sera toujours possible d'y faire référence («Autrefois, je croyais que ...»). Une révision complète consiste à clôturer l'ancienne connaissance et à insérer la nouvelle connaissance. Par exemple, Lucie ( $c^1$ ) croyait que Martine et Alain étaient ses voisins. A la date  $t$  elle apprend que ce n'est pas le cas. On ( $c^2$ ) clôture donc  $c^1$  et ajoute ( $c^3$ ) la nouvelle connaissance que l'on fait débiter à  $t$ .

$$\begin{array}{ll} \text{BC}(l, e^1, \text{voisins}(l) = \{m, a\}) & (c^1) \\ \text{fin}(e^1) = t & (c^2) \\ \text{BC}(l, e^2, \neg(\text{voisin}(l) = \{m, a\}) \wedge \text{deb}(e^2) = t & (c^3) \end{array}$$

L'exemple (ex.8) provoque aussi une révision complète.

Les révisions partielles sont traitées comme des révisions complètes. Par exemple, Lucie pensait jusqu'à la date  $t$  que ses voisins étaient Martine et Alain. Finalement elle apprend à la date  $t$  que ses voisins sont en fait Martine et Carole. La donnée a partiellement changé puisque Carole remplace Alain, mais Martine ne change pas. Comme précédemment, la connaissance antérieure est clôturée et la nouvelle connaissance est ajoutée et démarre au même moment.

$$\begin{aligned} \text{bc}(l, e^1, \text{voisins}(l) = \{m, a\}) & \quad (c^1) \\ \text{fin}(e^1) = t & \quad (c^2) \\ \text{bc}(l, e^2, \neg(\text{voisin}(l) = \{m, k\}) \wedge \text{deb}(e^2) = t & \quad (c^3) \end{aligned}$$

En définitive, pour l'expansion comme pour la révision, il faut ajouter et démarrer une nouvelle connaissance. La révision nécessite en plus de clôturer une connaissance active et d'identifier la date de fin de l'ancienne connaissance avec la date de début de la nouvelle. Pour des raisons de clarté, nous avons volontairement distingué les deux modes de modification de base mais ce n'est peut-être pas pertinent. En effet, tout ajout de connaissance est susceptible d'entraîner une révision.

Même si la mise à jour de la base est grandement facilitée par le jeu sur les datations, l'opération n'est probablement pas simple à implémenter. La principale difficulté que nous voyons est de savoir quelles connaissances clôturer lors d'une révision. En effet, lorsqu'une nouvelle information est fournie, celle-ci n'est pas livrée avec la liste des informations à clôturer. Il faut donc disposer d'un mécanisme de détermination des connaissances à clôturer.

Se pose aussi la question de savoir à quel moment s'opère la clôture lors d'une révision : au moment de l'acquisition de la nouvelle connaissance, ou à mesure que l'on rencontre des données qui sont contredites par la nouvelle information. Dans la réalité, les conséquences d'une nouvelle connaissances peuvent apparaître bien après l'acquisition. Il paraît donc plus vraisemblable que la clôture s'étale dans le temps et qu'une distinction se fasse selon la « familiarité » des connaissances : une connaissance très sollicitée sera peut-être vérifiée plus rapidement que d'autres moins sollicitées. Une étude cognitive reste à réaliser avant d'implémenter une procédure dans le modèle.

Enfin, il est un cas que nous ne travaillerons pas ici, c'est l'oubli. Intuitivement, il s'agirait d'un cas de rétractation pure et simple. Nous postulons que la modélisation du phénomène d'oubli est différent de ce que nous venons de voir, puisque typiquement, on ne garde pas trace de la connaissance ancienne. Cela a certainement des conséquences sur les connaissances inférées. C'est là un sujet délicat dont l'analyse dépasse le présent projet de présentation du modèle. Nous laissons la modélisation de l'oubli à des études ultérieures.

## 7 Datation des connaissances et temps réel

Un agent est plongé dans le temps, et quand bien même il ne fait rien (ce qui en soi est quelque chose, que l'on peut même désigner verbalement (dans un langage relâché) : « je glande depuis 10 mn »), sa base de connaissances est mise à jour en permanence (« je ne fais rien depuis 1mn, je ne fais rien depuis 2mn, ... »). Il faut supposer qu'à chaque unité de temps, la mémoire est mise à jour et que toutes les données qui n'ont pas été clôturées doivent être signalées comme actives.

Il y a là une difficulté notoire. Dans un modèle « à l'échelle », il est hors de question de passer en revue les milliards de connaissances et de notifier expressément la prolongation de leur activité pour chacune d'elles. On devine par ailleurs que c'est une procédure qui se fait naturellement chez l'humain. Mais avec la logique classique que nous nous sommes donnée, il n'y a pas moyen d'indiquer que « par défaut » les connaissances qui n'ont pas été clôturées sont toujours actives. Nous ne voyons pas d'autre solution ici que de recourir à une technique ad hoc. Celle-ci consisterait à introduire de la non-monotonie, et à accepter un raisonnement par défaut :

(R<sup>7</sup>) Si une fin d'activité pour une connaissance n'est pas inférable, alors la connaissance est actuelle.

## 8 Langage

Jusqu'à présent, nous avons observé l'état des connaissances et le mécanisme de mise à jour. Nous nous intéressons maintenant à la relation entre langage, état des connaissances et opérations sur les connaissances. Il faut distinguer l'interprétation et la génération. Faute de place, nous nous en tiendrons ici à l'interprétation et à la gestion des connaissances de l'allocutaire.

En interprétation, le traitement d'un énoncé se déroule en deux temps : l'interprétation proprement dite, qui produit une représentation sémantique, puis le post-traitement. L'interprétation est classiquement modélisée par un passage à l'aide d'une grammaire, comme la grammaire de Montague. Mais contrairement à ce qui se pratique avec ce type de grammaire l'objectif n'est pas de traduire un énoncé en une valeur de vérité. « Notre » grammaire s'entient à produire une représentation sémantique.<sup>2</sup>

La représentation est ensuite traitée de différentes manières. Si l'énoncé est une interrogative, l'allocutaire peut être amené à déclencher une recherche d'informations dans ses connaissances et à formuler une réponse. L'allocutaire peut aussi modifier les connaissances qu'il attribue à autrui (ex : « si le locuteur me pose cette question, c'est qu'il n'a pas telle connaissance. »).

Avec des phrases simplement affirmatives comme celles que nous étudions ici, l'énoncé est mémorisé et les connaissances mises à jour. L'allocutaire réalise une expansion ou une révision de ses connaissances propres, et modifie les connaissances qu'il attribue à autrui, avec là encore expansion ou révision. Nous proposons ici

2. Nous considérons en effet que le « sens » de l'énoncé réside dans sa relation aux connaissances existantes de l'allocutaire. La valeur de vérité n'est qu'un des aspects du sens. Par exemple la formule « E=mc<sup>2</sup> » a beau être vraie, elle n'a aucun sens pour un agent qui n'a pas les moyens de la « raccrocher » à son réseau de connaissances (à quoi font référence les composants de la formule, que représentent les opérations décrites par la multiplication, qu'est-ce que cela implique, etc.).

plusieurs exemples en nous appuyant sur notre modèle des connaissances.

Dans les exemples de la section, nous présentons un énoncé, sa représentation sémantique, et les modifications déclenchées dans la base de connaissances. Pour signifier qu'un énoncé  $E$  entraîne une modification  $M$  dans les connaissances, nous écrivons  $E \Rightarrow M$ . Le symbole est à néanmois la valeur  $\supset$ .

Jusqu'ici, nous avons utilisé le terme «savoir» comme l'abréviation de «posséder une donnée». Désormais, nous ne l'utilisons que comme objet linguistique, en tant que verbe épistémique dont on étudie le sens.

### 8.1 Mémorisation d'un énoncé simple (sans verbe épistémique)

Considérons un énoncé simple, sans verbe épistémique, dans un dialogue :

- (12) Lucie à Alain : Il a plu avant hier (P)  
 $pleuvoir_i(e) \wedge date(e) = avthier$  (p)

Le fait que Lucie énonce P est considéré par l'allocutaire comme une preuve que P est une connaissance de Lucie («dire P revient à dire qu'on possède la connaissance de P»). Cela rejoint la règle «de nécessité» de la logique épistémique<sup>3</sup>. À la date  $maint()$  (qui vaut ici la date d'énonciation), l'allocutaire prend donc connaissance de la connaissance de la locutrice. Il n'est pas en mesure d'établir sa date d'activité. Il peut seulement dire que la connaissance est active présentement. L'activité de la connaissance de l'allocutaire (sur la connaissance de la locutrice) débute à la date de l'énonciation  $deb(e^1) = maint()$ . Après interprétation et mémorisation, l'état de la mémoire de l'allocutaire est la suivante.

$$\text{ex.12} \Rightarrow \text{bc}(a, e^1, \text{bc}(l, e^2, p) \wedge \text{actuel}(e^2)) \\ \wedge deb(e^1) = maint()$$

L'allocutaire met en plus à jour ses connaissances (voir section 6). Ce traitement s'applique à tout énoncé sans verbe épistémique, aussi bien pour la forme affirmative que négative, et quelle que soit la date. On peut donc se donner une règle de gestion d'un énoncé sans verbe épistémique ni expression modale (ex : «peut-être», etc.).  $maint()$  vaut la date de l'énonciation.

(R<sup>8</sup>) Attribution de croyances au locuteur :

$$\text{pour tout énoncé simple } p \\ \exists e^1 \exists e^2 \text{bc}(\text{alloc}(), e^1, \text{bc}(\text{loc}(), e^2, p) \\ \wedge \text{actuel}(e^2)) \\ \wedge deb(e^1) = maint()$$

L'énonciation apporte aussi à l'allocutaire une information sur les connaissances que le locuteur lui

attribue désormais à lui, l'allocutaire : «[Lucie sait désormais que [je sais qu'[elle sait P]<sub>c<sup>3</sup></sub>]<sub>c<sup>2</sup></sub>]<sub>c<sup>1</sup></sub>...». L'ajout de cette connaissance étant systématique, il peut être déclenché par une règle. Pour plus de lisibilité, nous la décomposons.

(R<sup>9</sup>) Auto-attribution de connaissances

Pour tout énoncé simple  $p$

$$\exists e \text{bc}(\text{alloc}(), e, c^1) \wedge deb(e) = maint()$$

$$\exists e^1 \text{bc}(\text{loc}(), e^1, c^2) \wedge deb(e^1) = maint() \quad (c^1)$$

$$\exists e^2 \text{bc}(\text{alloc}(), e^2, c^3) \wedge deb(e^2) = maint() \quad (c^2)$$

$$\exists e^3 \text{bc}(\text{loc}(), e^3, p) \wedge deb(e^3) = maint() \quad (c^3)$$

Ajoutons pour finir que l'allocutaire peut s'approprier cette connaissance. Cela n'étant pas systématique (dépend du crédit accordé au locuteur ou encore à la compatibilité de la nouvelle information avec les connaissances propres existantes de l'allocutaire), ça n'est pas géré par une règle.

### 8.2 Énoncé direct simple avec verbe modal savoir

Simulons l'acquisition d'un énoncé direct simple<sup>4</sup> en *savoir que*.

- (13) Lucie : Je sais qu'il a plu hier.  
 $pleuvoir(e) \wedge date(e) = hier$  (p)  
 $savoir(l, e^1, p) \wedge \text{actuel}(e^1)$

Les effets sur la mémorisation seront identiques à ceux provoqués par un énoncé simple sans verbe épistémique. La procédure peut être «automatisée» en réutilisant les règles de la section précédente. Il faut prêter attention au fait que la règle qui suit n'est valable que lorsque le sujet de *savoir* est le locuteur (première personne).

(R<sup>10</sup>) Mémorisation d'un énoncé direct simple avec verbe épistémique  
 $\forall e \forall p \text{savoir}(\text{loc}(), e, p) \wedge \text{actuel}(e) \\ \Rightarrow \text{appliquer } R^8 \text{ et } R^9 \text{ sur } p.$

### 8.3 Acquisition d'un énoncé oblique en savoir

Observons cette fois-ci une phrase où le sujet du verbe *savoir que* est autrui (*Carole*) :

- (14) Lucie à Alain : Carole sait qu'[il a plu]<sub>p</sub>.  
 $pleuvoir(e) \wedge date(e) < maint()$  (p)  
 $savoir(k, e^1, p) \wedge maint() \in date(e^1)$

L'énoncé rend compte de ( $c^2$ ) l'attribution d'une connaissance  $c^1$  (par la locutrice à autrui). Le tout bien sûr devient une connaissance ( $c^3$ ) de l'allocutaire. Elle démarre au moment de l'énonciation. Par hypothèse, nous appliquons aux connaissances  $c^1$  et  $c^2$  la date du verbe *savoir* de l'énoncé :

4. Le terme est emprunté à Gosselin [7]; il désigne le présent à la première personne.

3.  $\frac{\alpha}{K_i(\alpha)}$

$$\begin{aligned}
(\text{ex.14}) \Rightarrow & \\
\text{BC}(k, e^1, p) \wedge \text{date}(e^1) = \text{date}(e) & \quad (c^1) \\
\text{BC}(l, e^2, c^1) \wedge \text{date}(e^2) = \text{date}(e^1) & \quad (c^2) \\
\text{BC}(a, e^3, c^2) \wedge \text{deb}(e^3) = \text{maint}() & \quad (c^3)
\end{aligned}$$

Il faut ajouter à cela l'auto-attribution d'une croyance ( $R^9$ ) : «la locutrice sait désormais que je sais qu'elle sait  $c^2$  et  $c^3$ ». On peut généraliser ces inférences pour tous les énoncés en *savoir que* où le sujet n'est pas le locuteur.

Il faut en plus rendre compte de la règle mentionnée dans la littérature ([8]), selon laquelle affirmer que «X (autre que moi) *sait que* P.» sous-entend que «moi (locuteur) aussi je *sais* que P.». C'est une différence notoire avec d'autres expressions en *penser que* ou *savoir si*. Bien sûr la date de la connaissance attribuée égale celle du verbe savoir. La connaissance implicite du locuteur est active présentement. Le tout doit être énoncé du point de vue de l'allocutaire. La connaissance de l'allocutaire débute à la date de l'énonciation. On obtient donc :

$$\begin{aligned}
(R^{11}) \text{ Descente du savoir attribué vers le locuteur} & \\
\forall p \forall x \forall e^1 \text{ savoir}(x, e^1, p) & \\
\Rightarrow \exists e^2 \exists e^3 \text{ BC}(\text{alloc}(), e^2, \text{BC}(\text{loc}(), e^3, p) & \\
\wedge \text{date}(e^3) = \text{date}(e^1)) & \\
\wedge \text{deb}(e^2) = \text{main}() &
\end{aligned}$$

#### 8.4 *savoir*, première personne, négatif au passé

Nous ne nous sommes jusqu'ici intéressé qu'aux connaissances «qui existent». Le modèle permet de traiter aussi l'absence de connaissances

Considérons la phrase négative, au passé, à la première personne :

$$\begin{aligned}
(15) \text{ Lucie à Alain : } \langle \text{hier} \rangle \text{ je ne savais pas qu'il avait plu} & \\
\text{avant hier} & \\
\text{pleuvoir}(e) \wedge \text{date}(e) = \text{avthier} & \quad (p) \\
\neg(\exists e^1 \text{ savoir}(l, e^1, p) \wedge \text{date}(e^1) = \text{hier}) &
\end{aligned}$$

Une telle phrase à la forme négative déclenche une première inférence : «P n'a pas appartenu à la base de la locutrice hier» :

$$\Rightarrow \neg \exists e^2 \text{ BC}(l, e^2, p) \wedge \text{date}(e^2) = \text{hier}$$

Une deuxième inférence est que désormais, Lucie sait qu'il a plu. Si ce n'était pas le cas, Lucie aurait par exemple utilisé la forme «je ne sais pas si P».

$$\Rightarrow \text{BC}(l, e^3, p) \wedge \text{actuel}(e^3)$$

Puisque ces inférences sont systématiquement déclenchées par une phrase épistémique négative avec le locuteur comme sujet, il est possible d'en tirer une règle. Il faut bien sûr adopter le point de vue de l'allocutaire : l'énoncé «je ne savais pas que P» déclenche «l'allocutaire a désormais connaissance que le locuteur a connaissance

que P n'était pas dans ses (=locuteur) connaissances ( $c^1$ ) mais qu'actuellement P y est ( $c^2$ ). ».

#### ( $R^{12}$ ) Mémorisation de phrases épistémiques négatives

avec locuteur comme sujet :

$$\begin{aligned}
\forall p \forall x \forall e \neg \text{savoir}(\text{loc}(), e, p) & \\
\Rightarrow \exists e^1 \text{ BC}(\text{alloc}(), e^1, c^1 \wedge c^2) & \\
\wedge \text{deb}(e^1) = \text{maint}() & \\
\neg \exists e^2 \text{ BC}(x, e^2, p) \wedge \text{date}(e^2) = \text{date}(e) & \quad (c^1) \\
\exists e^3 \text{ BC}(\text{loc}(), e^3, p) \wedge \text{actuel}(e^3) & \quad (c^2) \\
\Rightarrow \text{Appliquer } R^9 &
\end{aligned}$$

Notons que l'ignorance du locuteur est délimitée dans le temps ( $\text{date}(e)$ ). Cela n'empêche formellement pas d'avoir eu cette connaissance à une date antérieure. C'est possible dans le cas d'un oubli. Comme pour tout énoncé, «le locuteur sait désormais que l'allocutaire sait que le locuteur sait» ( $R^9$ ). Enfin, une appropriation de la connaissance par l'allocutaire est possible.

## 9 Conclusion

La présentation avait pour objectif d'introduire un modèle symbolique du processus de mémorisation des données verbalement acquises. Ce modèle est destiné à «expliquer» le processus et non simplement produire un résultat identique. Il avait aussi pour contrainte d'être cognitivement réaliste et pour cela de prendre en compte la théorie de l'esprit et la temporalité des connaissances.

Le cahier des charges nous amène à nous démarquer des travaux des logiciens. En particulier, nous avons clairement pris le parti de rendre compte des processus de mémorisation du point de vue d'un agent (l'allocutaire dans un dialogue) et non pas d'un observateur omniscient. Cela nous amène à renoncer à certains partis pris de la logique modale, et notamment la logique épistémique. Nous constatons que la notation utilisée, basée sur la logique classique (théorie des types simples) suffit pour décrire les données (représentations sémantiques, état de la mémoire) et les opérations à réaliser sur les données pour les mettre en mémoire.

Nous avons montré sur des exemples concrets que ce modèle est exploitable dans plusieurs disciplines. En psychologie et cognition, il offre une description du phénomène de fausse croyance et peut aider à décrire les mécanismes de mémorisation des sources par exemple. Il peut aussi aider à décrire les déficiences. On peut ainsi réfléchir aux conséquences d'une oblitération pure et simple d'une connaissance. En gestion des données, le marquage temporel des connaissances est une solution simple pour les à jour une base de connaissances.

En linguistique, pour des raisons de place, nous avons limité nos observations aux phrases simples non modales et à la seule expression verbale épistémique *savoir que*. Mais nous savons, pour l'avoir déjà implémenté, que le

modèle permet de rendre compte de la distinction entre les expressions *savoir que* et *savoir si* ([4] etc.). Pour avoir commencé le travail dessus, nous avons de bonnes raisons de croire que le modèle est aussi efficace pour décrire les effets d'autres expressions épistémiques : verbales (*croire, penser que* etc.), adverbiales (*peut-être* ou autres). Il faut pour cela introduire des nouveaux prédicats qui prendront la place dans certains de BC.

Un point a été délibérément mis de côté, c'est l'efficience pratique, à l'échelle, de notre modèle. La théorie des types simples n'est pas le système de calcul le plus efficace qui soit et nombre de règles que nous avons produites sont certainement difficilement calculables. Cette dimension applicative reste à étudier.

## Références

- [1] Baron-Cohen, Simon, Alan M Leslie et Uta Frith : *Does the autistic child have a theory of mind?* Cognition, 21(1) :37–46, 1985.
- [2] Blin, Raoul : *Résolution de l'ambiguïté sémantique des noms propres par utilisation des croyances sur les connaissances d'autrui - application au prénom.* *Linguisticae Investigationes*, 40(2) :200 – 227, 2017.
- [3] Chong, Hui Qing, Ah Hwee Tan et Gee Wah Ng : *Integrated cognitive architectures : a survey.* *Artificial Intelligence Review*, 28 :103–130, 2007.
- [4] Egré, Paul : *Question-Embedding and Factivity.* *Grazer Philosophische Studien*, 77(1) :85–125, 2008.
- [5] Ferchiou, Abdelaziz, Franck Schürhoff, E Bulzacka, M Mahbouli, Marion Leboyer et Andrei Szöke : *Mémoire de source-présentation générale et revue des études dans la schizophrénie.* *L'Encéphale*, 36(4) :326–333, 2010.
- [6] Ferchiou, Abdelaziz, Franck Schürhoff, E Bulzacka, M Mahbouli, Marion Leboyer et Andrei Szöke : *Mémoire de source-présentation générale et revue des études dans la schizophrénie.* *L'Encéphale*, 36(4) :326–333, 2010.
- [7] Gosselin, Laurent : *Marqueurs de modalité épistémique et calcul des valeurs modales : sémantique de savoir que.* 2013.
- [8] Martin, Robert : *Langage et croyance.* Mardaga, Bruxelles, 2023.
- [9] Norimatsu, H., R. Blin, K. Hashiya, Ch Sorsana et H. Kobayashi : *Understanding of others' knowledge in French and Japanese children : a comparative study with a disambiguation task on 16-38-month-olds.* *Infant Behavior & Development*, 37(4) :632–643, novembre 2014, ISSN 1934-8800.
- [10] Parsons, Terence : *Events in the semantics of English.* Numéro 19 dans *Current in linguistics series.* ITMIT press édition, 1990.
- [11] Pellier, Damien, Carole Adam, Wafa Johal, Humbert Fiorino et Sylvie Pesty : *Une architecture cognitive et affective orientée interaction.* Dans *Workshop on Affect, Compagnon Artificiel and Interaction*, Brest, France, juin 2016. <https://hal.science/hal-01365355>.
- [12] Premack, David et Guy Woodruff : *Does the chimpanzee have a theory of mind?* *Behavioral and Brain Sciences*, 1(04) :515, 1978, ISSN 0140-525X, 1469-1825.
- [13] Sabouret, nicolas et Manween Belkaid : *Un modèle logique de théorie de l'esprit pour un agent virtuel dans le contexte de simulation d'entretien d'embauche.* Dans *Proc. Workshop Affect, Compagnon Artificiel, Interaction*, Rouen, 2014.
- [14] Wimmer, Heintz et Josef Perner : *Beliefs about beliefs : Representation and constraining function of wrong beliefs in young children's understanding of deception.* *Cognition*, 13(1) :103–128, jan 1983, ISSN 00100277.