

Actes de langage indirects : co-construction d'un modèle logico-psychologique*

Dominique Longin*

Dominique.Longin@irit.fr

Éric Raufaste*

raufaste@univ-tlse2.fr

* Institut de Recherche en Informatique de Toulouse – CNRS (UMR 5505)
118 Route de Narbonne, F-31062 Toulouse cedex 4

* Laboratoire Travail et Cognition (UMR 5551) – Maison de la recherche
5 Allées Antonio-Machado, F-31058 Toulouse cedex 1

Résumé :

Cette communication présente les résultats d'une recherche pluridisciplinaire entre informatique, logique, psychologie cognitive et pragmatique linguistique dont le but est d'obtenir une modélisation informatique psychologiquement plausible du processus d'interprétation d'actes de langage indirects. Nous présentons les *schémas d'indirection* (formalisés dans une logique multimodale de la croyance, de l'intention et de l'action) représentant le sens indirect que peuvent contenir certains types d'énoncés, et constituant un modèle théorique normatif permettant de contrôler la portée du modèle psychologique que l'on se propose d'implémenter sous l'architecture cognitive ACT-R (et que nous présentons également mais de façon informelle).

Mots-clés : actes de langage indirects, logique modale, psychologie cognitive, modélisation cognitive, architecture cognitive.

Abstract:

This paper presents results of a multidisciplinary research in the area of computer sciences, logic, cognitive psychology and pragmalinguistics. Our aim is to obtain a computational model of the interpretation process of indirect speech acts. This model must be psychologically plausible. We present *indirection schemes* (formalized in a multimodal logic of belief, intention and action) representing indirect meaning. These schemes constitute a normative theoretical model controlling the scope of the psychological model we want to implement with the ACT-R cognitive architecture.

Keywords: indirect speech acts, modal logic, cognitive psychology, cognitive modelling, cognitive architecture.

1 Introduction

Cet article a pour cadre une recherche pluridisciplinaire commune à l'informatique et la logique formelle, à la psychologie cognitive, à la pragmatique linguistique, et aux neurosciences. Il vise à élaborer un modèle formel psychologiquement plausible du traitement des actes de langage indirects.

L'originalité du travail présenté provient selon nous non seulement de la collaboration entre

ces différentes disciplines¹ qui traduit l'ambition explicite de donner aux formalismes développés une capacité à décrire plausiblement des processus psychologiques observables chez des sujets humains, mais également de la mise en commun de ces ressources afin de réaliser une simulation informatique.

Plus qu'une simple mise en commun de travaux, cette collaboration a pour ambition de créer une symbiose entre les recherches menées dans les différentes disciplines afin d'allier notamment rigueur formelle et réalisme cognitif. Ainsi, un premier modèle logique (issu d'un travail pluridisciplinaire entre logique, pragmatique et psycholinguistique [7]) a conduit à un premier modèle psychologique [11], qui lui-même a montré la nécessité de développer un second modèle logique, permettant à son tour d'implémenter un second modèle psychologique.

Du point de vue de l'informatique, l'intérêt d'une telle étude est à replacer dans le cadre des systèmes de dialogue homme-machine. Comme Searle le montre [15, 16] nous communiquons majoritairement de façon non littérale. À partir du moment où l'on souhaite doter un système de dialogue de la capacité à communiquer de façon naturelle (en compréhension et en production), il apparaît comme incontournable d'étudier ce mode de communication.

Du point de vue psychologique, c'est toute la problématique du traitement du langage chez l'être humain qui est en jeu : comment un individu fait-il pour communiquer sa pensée, et comprendre celle des autres ? Plus précisément, comment sont sélectionnés les éléments pertinents du contexte influençant le choix d'une interprétation plutôt que d'une autre ? Par quels processus s'effectue cette influence ? Comment l'humain construit-il le sens ? Il va de soi que

*Ce travail a été soutenu par l'Action Concertée Incitative Cognitive, thème action, sous le numéro Act 55b.

¹Cf. le bilan des collaborations interdisciplinaires dressé par C. Fuchs (directrice du programme Cognitive) : « Cognitive : quel bilan ? », colloque Cognitive, les 6 et 7 décembre 2002, La Sorbonne, Paris.

le présent travail n'ambitionne pas la résolution d'une problématique aussi vaste mais il se propose néanmoins d'y apporter des éléments de réponse originaux et pertinents. Plus précisément, l'ambition du présent travail se restreint à la modélisation du sujet adulte non pathologique interprétant des énoncés produits dans sa langue maternelle. Dans ces conditions, où la compétence linguistique et pragmatique est réputée acquise, l'enjeu est de produire un modèle normatif et un modèle computationnel à visée descriptive des processus cognitifs par lesquels l'auditeur linguistiquement compétent induit l'intention réelle du locuteur. Le modèle ne cherche donc aucunement à produire une réponse sur la base de cette interprétation.

Dans ce qui suit, nous nous focalisons sur le processus interprétatif d'acte indirects de type directif (tout portant à croire que les mécanismes mis en jeu sont similaires pour les autres types d'actes illocutoires [7]). Nous introduisons la notion d'acte de langage indirect (Sect. 2), puis nous présentons le (second) modèle logique (Sect. 3). Enfin, nous donnons un aperçu du (second) modèle psychologique après avoir présenté l'architecture ACT-R et motivé son utilisation pour le présent travail (Sect. 4).

2 Les actes de langage indirects

La théorie des actes de langage est essentiellement due à Austin [4] et à Searle [14]. Ce dernier montre notamment comment tout énoncé peut se décomposer en un contenu propositionnel et une force illocutoire, et où des énoncés tels que « Jean travaille beaucoup. », « Travaille beaucoup, Jean. », « [Compte sur moi.] Jean travaillera beaucoup. », « Jean, je te demande de travailler beaucoup. », « Plaise au ciel que Jean travaille beaucoup ! » ont tous le même contenu propositionnel (*Jean travaille beaucoup*), mais des forces illocutoires différentes (respectivement : *assertive*, *directive*, *engageante*, *déclarative*, et *expressive*). Ces travaux sont formalisés principalement dans [13, 19, 20].

Dans [15, 16], Searle montre que parfois, en disant quelque chose, on peut vouloir dire autre chose. Il distingue notamment quatre cas :

- (i) en disant e , le locuteur veut dire exactement e (communication littérale) ;
- (ii) en disant e , le locuteur veut dire $e + f$ (communication indirecte) ;
- (iii) en disant e , le locuteur veut dire son contraire \bar{e} (ironie) ;

(iv) en disant e , le locuteur veut dire f (métaphore, litote...).

Les trois derniers cas forment ce que l'on appelle habituellement la communication non littérale. Dans ce travail, nous nous focalisons sur le cas (ii).

Par exemple « Peux-tu sauter de cette marche ? » peut réaliser un acte indirect de type directif qui est « Saute de cette marche » (mais pas systématiquement : le locuteur peut aussi vouloir simplement demander à son interlocuteur si celui-ci est capable de sauter de cette marche — et rien d'autre).

Searle a montré dans [16] que les énoncés susceptibles de réaliser des actes de langage indirects contiennent un certain nombre de marqueurs lexicaux et/ou syntaxiques. Ce résultat a été généralisé dans [7] en montrant qu'une telle réalisation peut se résumer à une assertion ou une question sur trois types principaux d'arguments (cas des directifs) :

- (i) des conditions de succès de l'acte indirect que l'on souhaite accomplir ;
- (ii) des raisons ou l'existence de raisons pour l'interlocuteur d'accomplir l'action qu'on souhaite qu'il accomplisse ;
- (iii) ou des préconditions de planification de cette action.

Un énoncé entrant dans l'une de ces catégories constitue ce que nous appelons une *forme d'indirection*.

Nous avons également montré que le traitement des actes de langage indirects nécessite deux étapes : d'une part reconnaître que l'énoncé est une forme d'indirection, et d'autre part établir sous forme de *critères contextuels* que le locuteur avait effectivement l'intention de parler indirectement.

Globalement, le premier modèle logique que nous avons élaboré [7] répondait à la première étape, ne suggérant que des critères locaux pour répondre à la seconde. Le présent travail constitue à la fois une refonte de ce premier modèle afin d'identifier les formes d'indirection, et un appel à des connaissances d'ordre psychologique pour identifier ces critères contextuels.

3 Modélisation logique

Fondé sur les théories philosophiques de [5] et de [17], le formalisme adopté [10, 9] s'inscrit

dans la lignée de ceux de [8, 12] et tend, comme ces derniers, à généraliser la théorie des actes de langage en une théorie de la communication, où les propriétés de ces derniers seraient dérivables de principes (plus généraux) de rationalité. Le cadre formel utilisé est celui d'une logique épistémique dynamique où les énoncés sont représentés par des actes de langage [4, 14].

Dans cette section, nous présentons le langage logique utilisé, (Sect. 3.1), la représentation des actions au travers de celle de leurs lois d'action (Sect. 3.2), puis les schémas d'indirection de notre modèle logique (Sect. 3.3).

3.1 Langage

Le langage présenté est un fragment de celui que nous utilisons habituellement pour formaliser des agents rationnels « intelligents » au sein de systèmes de dialogues ; ce fragment se limite à ce qui est nécessaire à la formalisation des schémas d'indirection (cf. Sect. 3.3)².

Le langage utilisé est celui d'une logique multimodale du premier ordre sans égalité ni symbole de fonction. La sémantique associée est définie en termes de mondes possibles et de relations d'accessibilité. Les formules atomiques sont notées p, q, \dots ou $P(t_1, \dots, t_n)$ et leur ensemble $FORMS_0$. Les formules sont notées A, B, \dots et leur ensemble $FORMS$.

Les agents sont notés i, j, k, \dots et leur ensemble $AGENTS$; les contextes c, c', \dots et leur ensemble $CONTEXTS$; les forces illocutoires F, F', \dots et leur ensemble IF . $IF \supseteq \{\text{Assert, QueryYN, QueryWhy, Request}\}$ représentant respectivement les forces d'assertion, de question dont la réponse est oui ou non, de question en pourquoi, et de requête. On note les événements e, e', \dots et leur ensemble $EVENTS$. Les événements peuvent être soit physiques, soit illocutoires (ils ont alors la forme $F(j, A)$ où j représente le destinataire de l'acte³ et A son contenu propositionnel). Une action est un couple $\langle i, e \rangle$ dont i représente son auteur. Si e est de type illocutoire, l'action est appelée *acte illocutoire* ou *acte de langage*. Les actions sont notées $\alpha, \beta, \gamma, \dots$ et leur ensemble est $ACTIONS$.

Le langage inclut des opérateurs classiques (la

²En particulier nous laissons de côté la notion d'intention et ses liens avec les autres modalités.

³Nous pouvons utiliser des ensembles d'agents. Pour simplifier l'écriture des schémas d'indirection (Sect. 3.3), nous nous limitons ici à deux agents (l'auteur et le destinataire de l'acte).

négation \neg et la disjonction \vee) et modaux : $Bel_i A$ (l'agent i croit A), $Wish_i A$ (i [souhaite/veut/a pour but] que A), $\Box_c A$ (il est obligatoire que A), $\heartsuit_c A$ (il est [bien/bon] que A), $Before_{\alpha,c} A$ (A est vrai avant toute exécution de α), $After_{\alpha,c} A$ (A est vrai après toute exécution de α).

Les autres formules classiques sont définies selon les abréviations usuelles. De plus : $\boxtimes_c A \stackrel{\text{déf.}}{=} \Box_c \neg A$ (A est interdit), $\diamond_c A \stackrel{\text{déf.}}{=} \neg \Box_c \neg A$ (A est permis), $\blacktimes_c A \stackrel{\text{déf.}}{=} \neg \heartsuit_c A$ (A est mauvais, négatif), $Done_{\alpha,c} A \stackrel{\text{déf.}}{=} \neg Before_{\alpha,c} \neg A$ (α vient d'être exécuté, avant quoi A était vrai), $Feasible_{\alpha,c} A \stackrel{\text{déf.}}{=} \neg After_{\alpha,c} \neg A$ (α est exécutable, après quoi A sera vrai).

Les opérateurs Bel_i sont définis dans une logique modale de type KD45 ; $Wish_i, \Box_c, \heartsuit_c, Before_{\alpha,c}$ et $After_{\alpha,c} A$ le sont dans KD⁴.

3.2 Lois d'action

Les lois d'action regroupent les lois d'exécutabilité et les lois d'effet.

Les lois d'exécutabilité. Elles déterminent dans quelles conditions une action peut être exécutée. De façon générale, elles sont de la forme

$$A \rightarrow Before_{\alpha,c} Precond(\alpha, c) \quad (\text{Exec}_{\alpha,\alpha})$$

qui exprime que « dans le contexte où A est vrai, avant toute exécution de α les préconditions représentées par $Precond(\alpha, c)$ doivent être vraies ». Les préconditions d'une action rendent compte du fait qu'un agent a la capacité d'accomplir cette action (*condition d'accomplissement*), souhaite accomplir cette action (*condition psychologique*) et a la permission de le faire (*condition déontique*).

DÉFINITION 1. — Si $\alpha = \langle i, e \rangle$, alors

$$Precond(\alpha, c) \stackrel{\text{déf.}}{=} Performable(\alpha, c) \wedge Wish_i Do(\alpha, c) \wedge \diamond_c Do(\alpha, c)$$

Par exemple, la condition d'accomplissement de l'action de *donner le sel à quelqu'un* (action physique) peut être : tenir le sel dans la main, être à portée de main de son destinataire...

⁴Notre logique comporte d'autres axiomes (mono et multimodaux) dont nous ne parlons pas ici.

En ce qui concerne un acte de langage, la condition d'accomplissement regroupe (1) les *conditions locutoires* LC (par exemple, le locuteur peut parler, l'auditeur peut entendre...), et (2) les *conditions illocutoires* IC (les *conditions de succès*, selon la terminologie de Searle) qui sont celles usuellement définies [13, 18], dépendantes de la force illocutoire.

DÉFINITION 2. — Si α est un acte de langage

$$\text{Performable}(\alpha, c) \stackrel{\text{déf.}}{=} LC(\alpha, c) \wedge IC(\alpha, c)$$

DÉFINITION 3. — Si $\beta = \langle j, e \rangle$, et que $\alpha = \langle i, \text{Request}(j, Do(\beta, c)) \rangle$

$$IC(\alpha, c) \stackrel{\text{déf.}}{=} \text{Wish}_i Do(\beta, c) \wedge \text{Performable}(\beta, c) \wedge \Diamond_c Do(\beta, c)$$

EXEMPLE 1. — Si α est l'acte par lequel i demande à j de lui passer le sel, et β représente cette action de passer le sel, alors :

$$\begin{cases} \beta = \langle j, \text{PasserSel}(i) \rangle \\ \alpha = \langle i, \text{Request}(j, Do(\beta, c)) \rangle \end{cases}$$

D'après la Définition 3 la condition illocutoire de α traduit le fait que : i veut que j lui passe le sel (condition d'état psychologique), j peut et a le droit de lui passer le sel (conditions préparatoires).

REMARQUE 1. — Dans l'optique qui est la nôtre (traiter les actes indirects de type directif), si α est l'acte indirect, alors il est de la forme $\langle i, \text{Request}(j, Do(\beta, c)) \rangle$ où $\beta = \langle j, e \rangle$. D'après ce qui précède, sa condition préparatoire (au sens de [14]) est $\text{Performable}(\beta, c)$.

Pour les besoins de la suite, cette condition de α (liée à β) est notée $\theta_{\alpha(\beta)}$ (en relation avec la notation de [20]). Si $\theta_{\alpha(\beta)}$ est vrai, alors cela peut impliquer logiquement d'autres propositions. Ainsi, pour reprendre l'exemple précédent, le fait que le locuteur pense que l'auditeur peut lui passer le sel implique qu'il pense que le locuteur a le sel dans la main, ou un plan pour l'attraper, qu'il est à bonne distance pour le lui passer...

Les lois d'effet. Elles expriment les postconditions d'une action, c.-à-d. ce qui est vrai après l'exécution de cette action. En toute généralité, elles sont de la forme

$$A \rightarrow \text{After}_{\alpha, c} \text{Postcond}(\alpha, c) \quad (\text{Effect}_{\alpha, \alpha})$$

qui exprime que « si A est vrai, après toute exécution de α les postconditions $\text{Postcond}(\alpha, c)$ seront vraies ».

Nous avons montré dans [7] que si α est une forme d'indirection permettant d'accomplir indirectement un acte β , alors on peut représenter cette indirection en écrivant plusieurs lois d'effets pour α . La loi $(\text{Effect}_{\alpha, \alpha})$ illustre alors les effets directs de α , tandis que

$$A' \rightarrow \text{After}_{\alpha, c} \text{Postcond}(\beta, c) \quad (\text{Effect}_{\alpha, \beta})$$

illustre ses effets indirects (identiques à ceux obtenus si β avait été accompli).

Nous supposons que pour tout α $(\text{Effect}_{\alpha, \alpha})$ est défini. Un schéma d'indirection rend compte de l'existence de lois d'effets du type de $(\text{Effect}_{\alpha, \beta})$ et est noté :

$$\alpha, C \Rightarrow \beta \quad (\text{Ind}_{\alpha, \beta})$$

où C (appelé *contexte minimal d'indirection*) dérive de A' dans $(\text{Effect}_{\alpha, \beta})$ et constitue le contexte minimal pour que les effets de β soient dérivables de l'exécution de α . Quand il est vide il est omis.

3.3 Schémas d'indirection

Les schémas d'indirection sont présentés selon la classification des différentes manières d'accomplir des actes de langage indirects élaborée dans [7], ce qui nous garantit une adéquation de ces schémas à l'ensemble des formes d'indirection possibles. Notre but ici n'est pas de justifier cette classification, mais simplement de l'utiliser.

Un *contexte minimal* est une condition nécessaire mais non suffisante pour garantir que l'interprétation doit être indirecte : c'est le rôle majeur de la psychologie que de déterminer quels sont les facteurs contextuels (représentés par l'index c dans les formules) orientant l'interprétation (cf. Sect. 4).

DÉFINITION 4. — Le prédicat $Do(\langle i, e \rangle, c)$ est vrai si et seulement si i exécute e dans le contexte c .

Afin de rendre plus concise l'écriture des schémas d'indirection, nous définissons la notation suivante (valable dans tout ce qui suit) :

NOTATION 1. — $\beta = \langle j, e \rangle$.

Assertion ou question sur des conditions de succès de l'acte indirect visé. Par exemple, « Peux-tu faire β » est une question sur la condition préparatoire de l'acte de langage produit par l'énoncé « Fais β » : il en découle que le premier énoncé peut être une façon d'accomplir le second. Les Schémas 1 rendent compte de ces cas.

SCHÉMAS D'INDIRECTION 1. — Si $A \in \{\diamond_c Do(\beta, c), Performable(\beta, c), Wish_i Do(\beta, c), Later Do(\beta, c), Do(\beta, c)\}$, et $A' \in \{Performable(\beta, c), Wish_j Do(\beta, c), \diamond_c Do(\beta, c), Later Do(\beta, c)\}$

$$\langle i, Assert(j, A) \rangle \Rightarrow \langle i, Request(j, Do(\beta, c)) \rangle \quad (1)$$

$$\langle i, QueryYN(j, A') \rangle \Rightarrow \langle i, Request(j, Do(\beta, c)) \rangle \quad (2)$$

Ces schémas permettent de traiter des énoncés respectivement de la forme « Il t'est permis de faire β », « Tu [peux/est capable de] faire β », « Je veux que tu fasses β », « Tu fais β », « Tu vas faire β », « Peux-tu faire β ? », « Veux-tu faire β ? », « T'est-il permis de faire β ? », « Vas-tu faire β ? ». Ces énoncés peuvent tous être utilisés pour signifier « Fais β ».

Assertion ou question sur l'existence de raisons de (ne pas) faire l'action demandée indirectement. Par exemple « Tu devrais faire β » (sans que l'on dise pourquoi) suppose l'existence d'une raison (non explicitée) de faire β et constitue à ce titre un moyen de demander à son interlocuteur de faire β .

SCHÉMAS D'INDIRECTION 2. — Si $\chi \in \{\square_c, \heartsuit_c\}$ et $A \in \{Do(\beta, c), \neg Do(\beta, c)\}$

$$\langle i, Assert(j, \chi A) \rangle \Rightarrow \langle i, Request(j, A) \rangle \quad (3)$$

$$\langle i, Assert(j, \neg \chi A) \rangle \Rightarrow \langle i, Request(j, \neg A) \rangle \quad (4)$$

$$\langle i, QueryYN(j, \neg \chi A) \rangle \Rightarrow \langle i, Request(j, A) \rangle \quad (5)$$

$$\langle i, QueryYN(j, \chi A) \rangle \Rightarrow \langle i, Request(j, \neg A) \rangle \quad (6)$$

$$\langle i, QueryWhy(j, A) \rangle \Rightarrow \langle i, Request(j, \neg A) \rangle \quad (7)$$

(3) permet de traiter des énoncés du type « Tu [dois/devrais/es obligé de/...] faire l'action β », « Il faudrait faire β », « Ce serait [bon/bien] de faire β » pour signifier « Fais β » et du type « Il faudrait ne pas faire β », « Ce serait [bon/bien] de ne pas faire β »... pour signifier « Ne fais pas β ».

(4) permet de traiter leur négation : « Tu ne [dois/devrais/...] pas faire β », « Ce ne serait pas [bon/bien] de faire β »... pour signifier « Ne fais

pas β », et « Tu ne [dois/devrais/...] pas ne pas faire β », « Ce ne serait pas [bon/bien] de ne pas faire β »... pour signifier « Fais β ».

(5) permet de traiter des énoncés du type « Ne [dois/devrais/...] tu pas faire β ? », « Ne serait-il pas [bon/bien] de faire β ? »... pour signifier « Fais β », et du type « Ne [dois/devrais/...] tu pas ne pas faire β ? », « Ne serait-il pas [bon/bien] de ne pas faire β ? »... pour signifier « Ne fais pas β ».

(6) permet de traiter des énoncés du type « [Dois/devrais/es]-tu obligé de faire l'action β ? », « [Faut-il/est-il nécessaire de] faire β ? », « Serait-il [bon/bien] de faire β ? » pour signifier « Ne fais pas β », ainsi que les négations non modales de ces énoncés : « [Dois/devrais/es]-tu obligé de ne pas faire l'action β ? », « [Faut-il/est-il nécessaire de] ne pas faire β ? », « Serait-il [bon/bien] de ne pas faire β ? » pour signifier « Fais β ».

(7) permet de traiter des énoncés du type « Pourquoi fais-tu β ? » (respectivement « Pourquoi ne fais-tu pas β ? ») pour signifier « Ne fait pas β » (respectivement « Fait β »).

REMARQUE 2. — Les opérateurs \diamond_c et \boxtimes_c étant définis à partir de \square_c , et \heartsuit_c l'étant à partir de \heartsuit_c , tous les énoncés faisant appel à ces opérateurs sont pris en compte pas les schémas précédents. Par exemple « Il t'est interdit de manger avec les doigts » est logiquement équivalent à « Tu dois ne pas manger avec les doigts », i.e. $\langle i, Assert(j, \square_c \neg Do(\beta, c)) \rangle$, qui est traité par (3) pour signifier « Ne mange pas avec les doigts ».

Assertion ou question sur des raisons de (ne pas) faire l'action demandée indirectement. On précise ici explicitement des raisons (au travers d'une description de l'état présent ou futur du monde) de (ne pas) faire une certaine action dans le but de demander indirectement de (ne pas) faire cette action. Nous distinguons le cas où il est fait référence à une action (reconnue comme étant mauvaise ou interdite) du cas où il est fait référence à un état (reconnu comme mauvais ou interdit). Dans le premier cas l'acte indirect est une demande pour que cette action ne soit pas (plus) faite, alors que dans le second on demande qu'une action capable de remédier à l'état en question soit exécutée.

Par exemple « Tu parles fort » (en supposant qu'il est interdit ou mauvais de parler dans ce

contexte) peut signifier « Ne parle pas fort ». De même, « Tu as l'air fatigué » dénote d'un état présent du monde constituant une raison pour se reposer, et peut donc signifier « Repose-toi », ou « Arrête de travailler »...

DÉFINITION 5. — *La formule atomique $State(k, t, c)$ est vrai si et seulement si dans le contexte c l'agent k est dans l'état t (terme du langage).*

Il est important de remarquer que dans la définition précédente, k est une variable d'agent pouvant en particulier (mais pas nécessairement) dénoter le locuteur ou l'auditeur.

DÉFINITION 6. — *Par définition,*

$$Prevent(\beta, A, c) \stackrel{\text{déf.}}{=} After_{\beta, c} \neg A$$

signifie qu'exécuter β permet de se prémunir contre A .

SCHÉMAS D'INDIRECTION 3. — *Si $\chi \in \{\boxtimes_c, \mathbf{X}_c\}$ et si $A \in \{Do(\beta, c), \neg Do(\beta, c)\}$, et $A' \in \{State(x, t, c), \neg State(x, t, c)\}$*

$$\langle i, Assert(j, A) \rangle, \chi A \Rightarrow \langle i, Request(j, \neg A) \rangle \quad (8)$$

$$\langle i, QueryYN(j, \neg A) \rangle, \chi A \Rightarrow \langle i, Request(j, \neg A) \rangle \quad (9)$$

$$\langle i, Assert(j, A') \rangle \wedge \chi A' \wedge Prevent(\beta, A', c) \Rightarrow \langle i, Request(j, Do(\beta, c)) \rangle \quad (10)$$

$$\langle i, QueryYN(j, \neg A') \rangle \wedge \chi A' \wedge Prevent(\beta, A', c) \Rightarrow \langle i, Request(j, Do(\beta, c)) \rangle \quad (11)$$

(8) permet de traiter des énoncés du type « Tu fais β » (respectivement « Tu ne fais pas β ») dans un contexte (minimal) où le fait de faire (respectivement : ne pas faire) β possède une connotation négative : ces énoncés servent donc à signifier « Ne fais pas β » (« Fais respectivement β »).

(9) traite les questions du type : « Ne fais-tu pas β ? » (respectivement « Fais-tu β ? ») dans un contexte (minimal) où le fait de faire (respectivement : ne pas faire) β possède une connotation négative : ces énoncés servent donc à signifier « Ne fais pas β » (respectivement « Fais β »).

(10) permet de traiter des énoncés contenant des verbes statifs⁵, dont les états auxquels il est fait référence sont reconnus comme étant interdits ou mauvais, et auxquels il peut être remédié

⁵Ces verbes décrivent non pas une action, mais un état de faits. Cela concerne notamment quatre grandes familles de verbes : être (au sens de être dans l'état de), paraître, avoir (au sens de posséder), et devenir.

en exécutant β . Cela concerne des énoncés tels que « Tu as l'air fatigué » (pour signifier par exemple « Repose-toi »), « Je suis las de marcher », (pour « Arrête-toi de marcher »), « Il fait chaud ici » (pour « Ouvre la fenêtre »)... ou des formes négatives telles « Tu n'as pas l'air dans ton assiette » (pour signifier par exemple « Allonge-toi »)...

(11) permet de traiter le même type d'énoncés que le cas précédent, mais formulés avec des questions, telles que « N'es-tu pas fatigué ? » (pour signifier par exemple « Repose-toi »), « Ne fait-il pas chaud ici ? » (pour signifier par exemple « Ouvre la fenêtre »)... ou des formes positives tel « Te sens-tu bien ? » (pour « Allonge-toi »)...

Assertion ou question sur des conditions d'accomplissement de l'action qu'il est demandée indirectement d'accomplir, ou ses sous-buts.

Dans l'Exemple 1 c'est l'action β qu'il est indirectement demandé d'accomplir. Une des conditions d'accomplissement de passer le sel à quelqu'un est par exemple de tenir le sel dans la main, et si cette condition n'est pas satisfaite, cela devient sous-un but à réaliser. Ainsi, demander le sel à quelqu'un peut se faire en disant « As-tu le sel en main ? » (question sur une condition d'accomplissement de l'action), ou « Attrape le sel » (requête sur un des sous-but à réaliser pour pouvoir accomplir l'action).

Ce sous-but (en tant que but) peut lui aussi avoir des préconditions, elles-mêmes utilisables pour l'accomplissement d'actes indirects. Par exemple, si une personne veut passer le sel à quelqu'un mais qu'elle ne l'a pas en main, elle va planifier l'action d'attraper le sel, action qui a ses propres conditions d'accomplissement ; par exemple : voir le sel. Une manière de demander le sel est alors de poser une question sur cette sous-conditions : « Vois-tu le sel ? ».

Dans notre exemple, voir le sel est une condition du sous-but de saisir le sel dont la réalisation correspond à une condition pour être en mesure de le passer. En résumé, toutes ces conditions sont dérivables de la condition d'accomplissement de l'action de passer le sel, et plus précisément de la condition préparatoire $\theta_{\alpha(\beta)}$ de l'acte α (cf. Remarque 1).

SCHÉMAS D'INDIRECTION 4. — *Si $\alpha = \langle i, Request(j, Do(\beta, c)) \rangle$*

$$\langle i, QueryYN(j, A) \rangle, (\theta_{\alpha(\beta)} \rightarrow A) \Rightarrow \alpha \quad (12)$$

$$\langle i, QueryYN(j, \neg A) \rangle, (\theta_{\alpha(\beta)} \rightarrow A) \Rightarrow \alpha \quad (13)$$

Ces schémas permettent de traiter des énoncés tels que « As-tu le sel ? », « Vois-tu le sel ? »... ainsi que leur négation « N'as-tu pas le sel ? », « Ne vois-tu pas le sel ? »... Les contenus propositionnel de tels énoncés sont tous dérivables de la condition préparatoire $\theta_{\alpha(\beta)}$ de l'acte α constituant une demande à quelqu'un de passer le sel. À ce titre, ils constituent tous une manière de dire « Passe-moi le sel » en vertu du fait qu'ils constituent des questions sur des faits dérivables de la condition préparatoire de cet énoncé.

4 Modélisation cognitive

Disposer d'un ensemble de schémas d'indirection ne constitue qu'une première étape pour la compréhension du processus d'interprétation des actes de langage indirects : si on se fixe un objectif de simulation de ce processus, voire d'utilisation réelle au sein de systèmes de dialogue, encore faut-il disposer d'éléments contextuels permettant non seulement de choisir un schéma plutôt qu'un autre, mais également d'expliquer ce choix afin de répliquer ce mécanisme au sein d'architectures d'agents rationnels « intelligents » telles que nous les concevons. Nous pensons qu'une bonne façon d'atteindre ces deux objectifs est de les considérer sous l'angle de la psychologie cognitive.

Dans cette section, nous présentons l'architecture cognitive sous laquelle nous avons partiellement implémenté le modèle précédent (Sect. 4.1), puis nous expliquons comment, en nous appuyant sur une démarche d'ordre psychologique, nous comptons fournir un modèle autonome de traitement d'actes indirects (Sect. 4.2).

4.1 L'architecture cognitive ACT-R

ACT-R est une théorie implémentée de l'architecture cognitive, simulant les principes de base du système cognitif humain. ACT signifie « Adaptive Control of Thought ». Le R renvoie à la rationalité de l'architecture cognitive humaine, rationalité issue de millions d'années d'évolution [2]. Il est important de noter que cette architecture a reçu de très nombreuses validations expérimentales et est largement compatible avec les acquis de la neurophysiologie de la cognition humaine [3]. ACT-R est une architecture hybride, à la fois symbolique et subsymbolique.

Au niveau symbolique on peut décrire les connaissances explicites, appelées *connaissances déclaratives* (les souvenirs et les connaissances générales de l'agent modélisé). Les connaissances déclaratives constituent un réseau sémantique. Les briques de ce réseau (appelées *chunks*) se composent chacune d'un symbole central (l'étiquette du chunk) et d'une liste de propriétés pouvant être instanciées par d'autres chunks. Au niveau subsymbolique, chaque chunk est associé à une « quantité d'activation » qui traduit l'accessibilité à la conscience de cette connaissance : plus un chunk est activé, et plus il est probable que cette connaissance devienne consciente et soit capable de déclencher des règles de raisonnement. Réciproquement, lorsque le sujet admet un chunk comme but conscient, celui-ci et tous les chunks qu'il incorpore deviennent des sources d'activation. Les chunks étant liés entre eux, l'activation circule depuis certains chunks « sources » (ceux qui sont actuellement dans le focus attentionnel) vers l'ensemble des chunks de la mémoire déclarative sémantiquement liés. Par exemple, si je pense consciemment au concept « chat », ce concept devient une source d'activation à partir de laquelle une certaine quantité d'activation va se propager au concept « animal ». Celui-ci se trouvera par là même plus facilement accessible à la conscience. Ce mécanisme de diffusion a pour conséquence que les informations les plus probablement utiles dans le contexte seront aussi les plus probablement sélectionnées pour un traitement élaboré.

L'autre type de connaissances modélisé correspond aux savoir-faire de l'agent : les *connaissances procédurales*. Au niveau symbolique, une inférence élémentaire est ainsi modélisée comme l'exécution d'une règle de type <si condition, alors action>. La partie condition de chaque règle se compose d'une série de définitions des propriétés que doivent posséder les chunks de la mémoire déclarative (ou des buffers d'entrées sensorielles) qui peuvent déclencher l'exécution de la règle. Une règle ne peut s'exécuter que si les chunks qui possèdent les propriétés satisfaisantes pour remplir ses conditions sont suffisamment activés. L'action peut consister à créer un nouveau chunk en mémoire déclarative, à opérer une action motrice, à déclencher une récupération en mémoire déclarative, etc. Au niveau subsymbolique les connaissances procédurales sont associées à des quantités d'activation dont la valeur dépend notamment de la qualité de l'apparie-

ment de la partie condition avec les chunks activés en mémoire déclarative. Par exemple si un sujet a le but actuel d'effectuer une addition, toutes les règles dont la partie condition spécifique que le but est de faire une addition reçoivent de l'activation. En cas de conflit, c'est-à-dire lorsque plusieurs règles ont toutes leurs conditions satisfaites, un mécanisme de résolution de conflit choisit la règle qui sera exécutée selon une fonction probabiliste de la quantité d'activation (plus la règle est activée, plus sa probabilité d'exécution augmente).

Sur ces bases, construire un modèle ACT-R revient à décrire et éventuellement à paramétrer l'ensemble des connaissances déclaratives et procédurales que les sujets sont censés avoir acquis lorsqu'ils possèdent la compétence à modéliser. ACT-R permet aussi de modéliser l'apprentissage de ces connaissances mais ce n'est pas notre objectif ici. Comme nous l'avons signalé en introduction de l'article, nous traitons en effet du cas du sujet adulte déjà pleinement compétent pour le traitement des actes de langage directs et indirects.

En résumé, en utilisant l'architecture ACT-R on peut décrire des connaissances et des règles de déduction de façon similaire à ce qu'on fait en logique, mais le contrôle de la série des opérations peut aussi faire intervenir des paramètres d'ordre subsymboliques comme l'expérience de l'agent, l'état du contexte etc. Ces propriétés subsymboliques permettent, entre autres, de modéliser des comportements partiellement stochastiques, et donc d'expliquer certaines erreurs. L'emploi d'une modélisation ACT-R a donc essentiellement une visée descriptive, non normative.

4.2 Le traitement du contexte

À l'heure actuelle, notre modèle ACT-R comprend les schémas d'indirection (niveau symbolique), ainsi qu'une représentation et une gestion de ce que nous avons appelé précédemment des *contextes minima*. Sans une représentation et une utilisation correcte du contexte d'énonciation, toute entreprise de simulation du processus d'interprétation des actes indirects est vouée à l'échec. De ce point de vue, le mécanisme de diffusion d'activation dans ACT-R permet de rendre compte des effets de contexte, en particulier comment la prise en compte du contexte permet d'activer en priorité les connaissances les plus probablement utiles

dans ce contexte [1] (facilitant ainsi l'interprétation pertinente des énoncés). Dans notre modèle ACT-R, les interactions entre les connaissances du sujet et les propriétés du contexte expliquent comment l'individu déclenche une série d'inférences élémentaires conduisant finalement à l'interprétation littérale ou à une interprétation non littérale.

La validation expérimentale du présent travail requerra de soumettre des sujets humains à un certain nombre d'énoncés présentés dans des contextes judicieusement choisis pour susciter une variété potentielle d'interprétations, puis de recueillir l'interprétation que les sujets font effectivement (directe ou indirecte). Outre l'interprétation des énoncés, nous recueillerons aussi divers indices comportementaux comme le temps de réponse. Sachant que les opérations mentales au niveau des règles de raisonnement ont des durées connues (en général de l'ordre de 50 ms pour une opération mentale élémentaire), il est possible de modéliser non seulement la réponse produite mais aussi le temps de réponse des sujets. De même l'architecture ACT-R permet de prédire la distribution des erreurs d'interprétation. On peut alors comparer la performance du modèle avec la performance des sujets humains. Les résultats issus de l'expérimentation devraient donc nous conduire à obtenir un modèle psychologiquement plausible (en terme de nature des réponses données, de leur fréquence, de leur motifs, et de leurs temps de production). Bien entendu, le test du modèle à ce niveau impliquera une modélisation, même minimaliste, de la production de réponse. Afin de réduire l'impact de cet aspect qui, rappelons-le, échappe à notre objectif principal, les situations expérimentales proposeront toujours aux participants de sélectionner l'interprétation qu'ils jugent la plus plausible parmi un ensemble d'options préalablement établies par le chercheur.

Secondairement, nous nous attendons à ce que ces données expérimentales soient elles-mêmes sources d'enrichissement du modèle formel.

5 Conclusions et perspectives

Dans ce qui précède nous avons présenté un modèle formel d'inférence d'actes de langage indirects de type directif. Comme nous l'avons signalé, ce travail s'appuie sur des travaux antérieurs [7], et notamment sur une classification à vocation exhaustive des différentes façon d'ac-

complir un acte indirect de type directif, ce qui nous assure une adéquation des schémas proposés avec notre modèle issu de la pragmatique linguistique.

Nous avons également montré tout l'intérêt de considérer les aspects psychologiques du problème, notamment afin d'isoler et d'utiliser des éléments contextuels d'interprétation dans le but d'obtenir une procédure de décision sur le système de règles proposé. En particulier, un tel travail nécessite de s'appuyer sur des expérimentations car le modèle psychologique obtenu en implémentant le modèle formel dans ACT-R ne produira a priori pas les mêmes résultats que ce dernier. Pour s'en convaincre il suffit de remarquer que le système cognitif humain incorpore constitutionnellement une composante stochastique qui échappe complètement au modèle normatif : l'humain fait des erreurs qu'il faut aussi modéliser.

Le modèle ACT-R n'est cependant pas encore terminé. En effet, de récents travaux en psychologie cognitive montrent l'émergence de nouveaux modèles reposant sur une construction incrémentale du sens d'un énoncé (pour une revue voir [6]). Ces modèles fonctionnent sur la base de patterns d'interprétation prédéfinis, dont l'utilisation est réalisée dynamiquement au fur et à mesure de la perception des termes de l'énoncé. Par exemple, si à table on entend le début d'énoncé « Peux-tu me passer... », on anticipe déjà, avant d'entendre la suite de la phrase, que l'intention du locuteur est probablement que nous lui faisons passer quelque chose (interprétation indirecte) et pas de savoir si oui ou non nous sommes capables de le faire (interprétation directe). Restructurer le modèle de façon à le rendre incrémental devrait nous permettre d'optimiser l'ajustement entre les temps de traitements empiriquement observés et les temps de traitements modélisés.

Au final, c'est un travail à vocation hautement pluridisciplinaire où il y a interpénétration des savoirs : l'informatique et la logique y puisent des méthodes pour résoudre des problèmes pour lesquels aucune solution n'existe encore, la pragmatique linguistique et la psychologie cognitive utilisent les modèles formels comme modèles normatifs de la performance, faisant ainsi à leur tour évoluer leur propres modèles.

Références

- [1] J.R. ANDERSON. *Rules of the Mind*. LEA, NJ, 1993.
- [2] J.R. ANDERSON et C. LEBIERE. *The Atomic Components of Thought*. LEA, Mahwah, NJ, 1998.
- [3] J.R. ANDERSON et C. LEBIERE. *The Newell test for a theory of Mind*. Behavioral and Brain Sciences, 1998. In press.
- [4] John L. AUSTIN. *Quand dire, c'est faire*. Éditions du Seuil, France, 1970.
- [5] Michael E. BRATMAN. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.
- [6] R. BUDIU. *The role of background knowledge in sentence processing*. Thèse de doctorat, Carnegie-Mellon University, USA, 2001.
- [7] Maud CHAMPAGNE, Andreas HERZIG, Dominique LONGIN, Jean-Luc NESPOULOUS, et Jacques VIRBEL. Formalisation pluridisciplinaire de l'inférence d'actes de langage non littéraires. *Information, Interaction, Intelligence*, Hors série :197–225, 2002.
- [8] Philip R. COHEN et Hector J. LEVESQUE. Intention is choice with commitment. *Artificial Intelligence Journal*, 42(2–3), 1990.
- [9] Andreas HERZIG et Dominique LONGIN. Belief dynamics in cooperative dialogues. *Journal of Semantics*, 17(2) :91–118, 2000.
- [10] 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, Toulouse, France, novembre 1999. www.irit.fr.
- [11] Dominique LONGIN et Éric RAUFASTE. Interprétation d'actes de langage indirects de type directif : formalisation logique & modèle psychologique ACT-R. Rapport technique 2003-04-R, Institut de Recherche en Informatique de Toulouse, www.irit.fr, janvier 2003. 81 pages.
- [12] M. D. SADEK. *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication*. Thèse de doctorat, Université de Rennes I, Rennes, France, 1991.
- [13] J. R. SEARLE et D. VANDERVEKEN. *Foundation of illocutionary logic*. Cambridge University Press, 1985.

- [14] John R. SEARLE. *Les actes de langage*. Hermann, Paris, 1972.
- [15] John R. SEARLE. Indirect speech acts. Dans J. P. COLE et J. L. MORGAN, éditeurs, *Syntaxe and Semantics*, volume 3, *Speech Acts*. Academic Press, 1975.
- [16] John R. SEARLE. *Sens et expression*. Éditions de Minuit, 1982.
- [17] John R. SEARLE. *L'Intentionnalité : Essai de philosophie des états mentaux*. Les Éditions de Minuit, 1985.
- [18] D. VANDERVEKEN. *Les actes de discours : essai de philosophie du langage et de l'esprit sur la signification des énonciations*. Mardaga, Liège – Bruxelles, 1988.
- [19] D. VANDERVEKEN. *Meaning and Speech Acts*, volume 1, *Principles of language use*. Cambridge University Press, 1990.
- [20] D. VANDERVEKEN. *Meaning and Speech Acts*, volume 2, *Formal semantics of success and satisfaction*. Cambridge University Press, 1991.