HAL open science

Théorie et pratique de l'interaction humain-robot

Aurélie Clodic

▶ **To cite this version:**

Aurélie Clodic. Théorie et pratique de l'interaction humain-robot. Robotique [cs.RO]. Université de Toulouse, 2022. tel-04137174

HAL Id: tel-04137174

https://laas.hal.science/tel-04137174

Submitted on 22 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE

En vue de l'obtention de l'

HABILITATION A DIRIGER DES RECHERCHES

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *22/06/2022* **par :**

Aurélie CLODIC

Théorie et Pratique de l'Interaction Humain-Robot

JURY

Simon LACROIX	Directeur de Recherche	Président du Jury
Marie BABEL	Professeure	Rapporteure
Anne SPALANZANI	Maître de Conférences	Rapporteure
Gentiane VENTURE	Professeure	Rapporteure
Daniel SIDOBRE	Maître de Conférences	Examinateur

École doctorale et spécialité :
Ecole Doctorale Mathématiques, Informatique, Télécommunications de Toulouse (EDMITT)

Unité de Recherche :
Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS CNRS) - UPR 8001

Parrain de recherche :
Simon LACROIX

Table des matières

Présentation **1**

I Comment on a essayé de faire interagir un robot avec des humains **3**

1 De l'autonomie à l'interaction **5**
1.1 Genèse et évolution de l'interaction humain-robot au LAAS 5
1.2 L'architecture LAAS 7
1.3 Théorie de l'action 8
1.4 Théorie de l'action jointe 10
1.5 Application à la robotique 12
1.5.1 Niveau distal/décisionnel partagé 12
1.5.2 Niveau proximal/exécutif partagé 12
1.5.3 Niveau moteur/fonctionnel couplé 13
1.6 Conclusion 13
1.7 Contributions associées 13

2 Implémenter une action jointe humain-robot **15**
2.1 Introduction 15
2.2 La distinction entre soi et l'autre 16
2.3 L'attention jointe 17
2.4 La compréhension de l'action intentionnelle 23
2.5 Représentations partagées de la tâche 24
2.6 Conclusion 28
2.7 Contributions associées 28

3 Gérer ses connaissances, évaluer celles des autres **31**
3.1 Préambule 31
3.2 Acquérir et consolider les connaissances 33
3.3 Gérer les connaissances 37
3.4 Exploiter les connaissances 40
3.5 Partager la connaissance 41
3.6 Conclusions 43
3.6.1 Explorer les connaissances 43
3.6.2 Aller plus loin dans l'exploitation de l'estimation des connaissances de l'autre 44
3.7 Contributions associées 45
3.7.1 Sur les aspects évaluation de la situation 45
3.7.2 Sur les aspects ontologie 45

4 Prendre des décisions et exécuter des actions en interaction **47**
4.1 Supervision et interaction 47
4.2 Eléments nécessaires à un système de décision pour la collaboration 49
4.2.1 Faire groupe 49

4.2.2 S'engager à faire ensemble . . . 50
4.2.3 Exécuter et décider ensemble . . . 51
4.2.4 Évaluer l'interaction . . . 53
4.2.5 Communiquer avec . . . 55
4.2.6 Apprendre pendant une interaction . . . 56
4.3 Vision Système . . . 59
4.4 Contributions associées . . . 60

5 Gérer les défaillances 63
5.1 Introduction . . . 63
5.2 Ce qui peut mal se passer . . . 64
5.2.1 Défaillance Hardware . . . 64
5.2.2 Défaillance Software . . . 64
5.2.3 Le cas particulier de l'interaction . . . 65
5.2.4 Cybersécurité . . . 67
5.3 Conclusion . . . 67
5.4 Contributions Associées . . . 67

6 Imaginer le robot comme un agent social au milieu d'autres agents sociaux 69
6.1 Qu'est ce qu'un agent social ? . . . 69
6.1.1 Les conditions de la socialité . . . 69
6.1.2 Définition d'une stratégie d'interprétation sociale . . . 71
6.1.3 L'importance de la reconnaissance mutuelle . . . 72
6.2 De l'intérêt de modéliser la notion d'engagement . . . 73
6.2.1 L'approche fonctionnelle : l'engagement comme un outil pour réduire l'incertitude . . . 73
6.2.2 L'approche normative : l'aspect normatif vs descriptif des engagements 74
6.2.3 Application à la robotique . . . 75
6.3 Socialité, engagements et horizontalité : remarques de conclusion . . . 82
6.4 Contributions Associées . . . 83

7 Faire les choses bien 85
7.1 Introduction . . . 85
7.2 Pourquoi ce domaine de l'interaction personne-robot est unique ? . . . 86
7.3 Qu'est ce qu'on cherche à évaluer finalement ? . . . 86
7.4 Quels moyens pour évaluer ? . . . 87
7.5 Comment faire les choses bien ? (ou pour le mieux) . . . 88
7.6 Et alors ? . . . 90
7.7 Contributions associées . . . 90

8 Penser aux conséquences 93
8.1 Introduction . . . 93
8.2 Une responsabilité éthique en tant que développeur de systèmes . . . 94
8.3 Une responsabilité en tant que scientifique citoyen . . . 96
8.4 Une responsabilité vis-à-vis des consommateurs finaux . . . 99
8.4.1 Introduction . . . 99
8.4.2 Questions éthiques et juridiques . . . 100

8.4.3 Propositions . . . 101
8.5 Conclusion . . . 102
8.6 Contributions Associées . . . 103

9 Se confronter au monde réel 105
9.1 Le pitch de départ . . . 105
9.2 L'organisation du projet et notre implication . . . 106
9.3 Un peu de logistique pour commencer . . . 109
9.3.1 Choix d'un scénario . . . 109
9.3.2 Création d'un environnement de test accessible . . . 110
9.4 Développements . . . 110
9.4.1 Définition du scénario et conception des fonctions nécessaires (travaux communs WP1/WP5 et collaboration avec le projet JointAction4HRI) 111
9.4.2 Planificateur de perspective visuelle partagée (SVP Planner) (WP5.1) 112
9.4.3 Prise de perspective et raisonnement géométrique : UnderWorlds (avec WP5.2 avec WP2) . . . 113
9.4.4 Génération de routes et système de gestion d'ontologies travail conjoint WP5/WP4 . . . 115
9.4.5 Supervision d'une tâche interactive Humain-Robot (WP5.3 avec WP2 et WP4) . . . 116
9.4.6 Planification des mouvements, des placements et des postures (WP5.1) 118
9.5 Résultats . . . 119
9.5.1 Intégration et déploiement (WP7 avec WP5) . . . 119
9.5.2 Evaluation (WP7 et WP5) . . . 119
9.6 Conclusions . . . 120
9.7 Introspection . . . 122
9.7.1 Avec ma casquette de responsable technique . . . 122
9.7.2 Avec ma casquette de co-directrice de thèse . . . 123

II Comment on va essayer de continuer 125

1 Perspectives de recherche 127
1.1 Introduction . . . 127
1.2 La robotique sociale . . . 128
1.3 La notion de pratiques sociales . . . 130
1.4 Compétences sociales et éthiques pour un robot autonome . . . 131
1.4.1 Pratiques sociales appliquées . . . 131
1.4.2 Le *Social Practice Kernel* . . . 132
1.4.3 L'architecture . . . 133
1.4.4 Le *metadeliberation level* . . . 133
1.4.5 Le *social practices management level* . . . 134
1.4.6 Le *functional level* . . . 134
1.4.7 Gestion des connaissances et des données . . . 135
1.4.8 Évaluation de la situation, évaluation du contexte et interprétation . 136
1.4.9 Planification et monitoring . . . 136
1.4.10 Supervision . . . 136
1.4.11 Apprendre des pratiques au fil des actions . . . 137

1.5 En travaux... 137

2 Projets en cours **139**
2.1 Artificial Intelligence for Human-Robot Interaction (ai4hri) 140
2.1.1 Description 140
2.1.2 Rôle 141
2.1.3 Statut 141
2.2 EPIIC 143
2.2.1 Description 143
2.2.2 Rôle 143
2.2.3 Statut 144
2.3 ELSA 145
2.3.1 Description 145
2.3.2 Rôle 147
2.4 DISCUTER 149
2.4.1 Description 149
2.4.2 Rôle 150

III CV **153**

1 Activités **159**
1.1 Activités d'encadrement 159
1.1.1 Thèse 159
1.1.2 Master 160
1.2 Activités d'enseignement 161
1.3 Activités d'évaluation 161
1.3.1 Membre de comité de programme 161
1.3.2 Editeur associé 161
1.3.3 Jury de thèse 161
1.3.4 Jury de concours 162
1.4 Activités de diffusion de l'information scientifique 162
1.4.1 Invited talk/speaker 162
1.4.2 Dissemination grand public 162
1.5 Organisation de workshops 162
1.5.1 Description 163
1.5.2 Site web 163
1.5.3 Liste des workshops 163

2 Projets passés **165**
2.1 JointAction4HRI (French ANR project) 2017-2021 165
2.1.1 Description 165
2.1.2 Site web 165
2.1.3 Publications 165
2.2 MuMMER- MUltiModal Mall Entertainment Robot (H2020) 2016-2020 . . 167
2.2.1 Description : 167
2.2.2 Site web : 167
2.2.3 Publications : 167

2.3 RoboErgoSum - Self Aware Robot (French ANR project) - 2013-2017 169
2.3.1 Description 169
2.3.2 Site Web 169
2.3.3 Publications 169
2.4 Saphari - Safe and Autonomous Physical Human-Aware Robot Interaction (FP7) - 2011-2015 170
2.4.1 Description 170
2.4.2 Site Web 170
2.4.3 Publications 170
2.5 MarDi - Man-Robot Dialogue (French National Research Agency project) - 2012-2016 171
2.5.1 Description 171
2.5.2 Publications 171
2.6 COGNIRON (European FP6 integrated project) 2004-2008 172
2.6.1 Description 172
2.6.2 Site Web 172
2.6.3 Publications 172
2.7 HR+ (French robotics program project) 2002-2005 173
2.7.1 Description 173
2.7.2 Site Web 173
2.7.3 Publications 173

3 Publications 175
3.1 Thèse 175
3.2 Revues 175
3.3 Conférences à comité de lecture 176
3.4 Workshops 178
3.5 Autres 180

Bibliographie 181

Présentation

Ce travail synthétise mes activités de recherche au sein du groupe Robotique et InteractionS du LAAS, depuis que je l'ai intégré en tant qu'ingénieure de recherche en novembre 2012. Il est composé de trois parties :

— la première partie est une présentation originale et personnelle de mon domaine de recherche : l'interaction humain-robot ;
— la seconde partie présente mes perspectives de recherche et les projets de recherche qui y sont associés ;
— la troisième partie est constituée de mon CV.

Ce travail ne se veut pas comme un livre de recettes, mais plutôt comme un livre donnant les éléments auxquels il faut être attentifs lorsque l'on travaille dans le domaine de l'interaction humain-robot. Cela regroupe un ensemble de choses dont il faut avoir conscience ou être attentif à. À ceux qui savent, cela donnera peut-être l'impression de donner des leçons, mais ce n'est pas mon but lorsque j'écris ces lignes. Mon but est de donner modestement quelques clés de compréhension de mon domaine de recherche.

Qu'est ce que j'ai fait ?

Une de mes rapporteures m'a fait remarquer que dans ce manuscrit j'utilisais tout le temps le *«nous»*. En robotique, on ne fait jamais rien tout seul, on travaille toujours en équipe. C'est l'essence même du travail en robotique. Mais je vais quand même essayer de répondre à la question avec ce *«je»*.

La première réponse c'est de dire que je n'ai rien fait, parce qu'en fait nous savons tous que ce sont nos étudiants, nos doctorants et nos post-doctorants qui créent tous ces logiciels, qui mettent en musique toutes ces idées et MERCI à eux pour tout çà. En tant qu'ingénieure de recherche, je suis sensée me mettre au service d'un chercheur. Au fil des années, ce chercheur, Rachid Alami en l'occurrence, et toute l'équipe d'ailleurs, m'a permis de prendre part à cette recherche parce qu'en fait dans nos domaines c'est très compliqué, c'est juste impossible de ne pas le faire. Cela m'a permis de monter des projets et de développer ma propre vision sur laquelle je peux aussi travailler aujourd'hui.

La seconde réponse, c'est de dire que j'ai créé des liens. Je suis à l'origine de toutes ces collaborations que nous avons aujourd'hui avec des psychologues et des philosophes sur les différents sujets qui sont exposés dans ce manuscrit. L'une des actions emblématiques a été ce projet ANR JointAction4HRI pour lequel nous avons accueilli au LAAS pendant 18 mois un post-doctorant en philosophie et une thèse en co-encadrement entre un laboratoire de psychologie et le LAAS. Une chose remarquable dans cette création de liens, c'est qu'aujourd'hui par exemple nous sommes sollicités par des psychologues qui sont intéressés par le fait que nous ne nous arrêtons pas aux mécanismes qu'ils analysent mais à la manière dont ces mécanismes peuvent trouver écho dans notre architecture robotique et dans nos développements.

La troisième réponse, c'est de dire que j'ai mis des concepts derrière des intuitions, parce que je suis allée chercher ces travaux sur l'action jointe, sur la notion d'engagement, sur la notion de reconnaissance mutuelle, sur la notion de pratique sociale. Le but ici n'est pas

de chercher à imiter une interaction humain-humain mais d'en chercher un modèle pour le borner et le rendre compréhensible et utilisable dans un cadre d'interaction humain-robot.

Et maintenant, il est temps de dire…

Merci !

Merci à mon jury, merci de m'avoir aidé, merci d'avoir rapporté sur mon travail et d'avoir été là pour la soutenance.

J'ai fait mes premiers pas en robotique à l'iup systèmes intelligents de Toulouse, je voulais remercier ici tous mes professeurs, qui sont aujourd'hui pour certains des collègues. C'est sans nul doute grâce à vous que je suis là aujourd'hui.

Merci à tous les collaborateurs d'hier et d'aujourd'hui, et notamment tous ces étudiants, ces doctorants et ces post-doctorants qui nous ont fait confiance et qui nous font confiance et avec qui nous avons réalisé ces travaux.

Merci aux équipes du département robotique du LAAS et principalement à l'équipe RIS (Robotique et InteractionS) (et avant RIA (Robotique et Intelligence Artificielle)).

Merci à tout le monde au LAAS parce que au côté des chercheurs il faut beaucoup d'administratifs et d'ITA pour rendre tout çà possible et merci bien entendu particulièrement à mes collègues du service IDEA et de la Plateforme Robotique.

Première partie

Comment on a essayé de faire interagir un robot avec des humains

Chapitre 1

De l'autonomie à l'interaction

1.1 Genèse et évolution de l'interaction humain-robot au LAAS

La genèse et l'évolution de l'interaction humain-robot au LAAS sont intimement liées à la ménagerie des robots qui se sont succédé au fil des années et des travaux qu'ils ont permis de réaliser. Je vous propose pour commencer un tour d'horizon des robots sur lesquels j'ai travaillé.

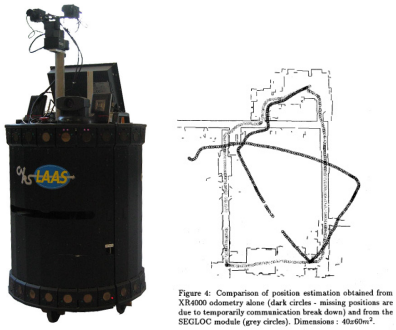


Figure 4: Comparison of position estimation obtained from XR4000 odometry alone (dark circles - missing positions are due to temporarily communication break down) and from the SEGLOC module (grey circles). Dimensions : 40x60m².

DILIGENT a officié dans les années 2000 [Alami 2000] (je vous conseille la savoureuse partie 4.8 Chronology de ce papier [Alami 2000]). Une partie *User Interaction* était déjà présente qui implémentait un mode de mouvement conforme sûr dans lequel le robot se déplace comme s'il était "poussé" par un humain ; la direction du mouvement est détectée à l'aide de l'anneau de capteurs infrarouges tandis que le capteur à ultrasons est utilisé pour créer des forces répulsives afin d'éviter le contact avec les obstacles. Cependant, l'un des principaux défis à l'époque était déjà de réussir à se localiser comme on peut le voir sur la figure de droite. C'est sur ce robot que j'ai fait tourner ma première manip' robotique qui avait été implémentée par Benoit Morisset [Morisset 2002].

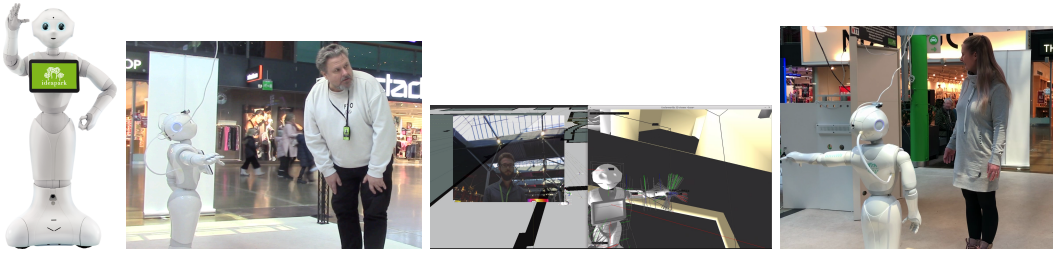
RACKHAM est l'un des deux robots sur lesquels j'ai travaillé pendant ma thèse. Il a été guide à la Cité de l'Espace pendant plusieurs semaines dans les années 2004 et 2005. Vous

pouvez voir un aperçu en consultant sa page web `https://www.laas.fr/robots/rackham/` et la publication [Clodic 2006]. Le projet correspondant est le projet HR+. Le premier défi de Rackham a aussi été la localisation (surtout quand on s'aperçoit sur le lieu d'exposition qu'une partie de l'environnement est transparent et que le moyen de localisation est un télémètre laser Sick) et la navigation. Au niveau interaction, Rackham disposait d'une interface tactile avec un clone parlant et était capable de détecter des visages avec sa caméra. J'ai implémenté la supervision qui pilotait l'ensemble des modules et permettait au robot de réaliser sa tâche de guide en interaction avec les visiteurs.

JIDO est le second robot sur lequel j'ai travaillé pendant ma thèse dans le cadre du projet Cogniron Cognitive Robot Companion `https://www.cogniron.org/final/Home.php`, il a aussi sa page web perso `https://www.laas.fr/robots/jido/`. Équipé d'un bras, Jido permettait de commencer à appréhender un échange d'objet avec une personne. Ses deux télémètres laser (il y en avait un à l'avant et un à l'arrière permettait aussi la détection des personnes sur "presque" 360 degrés). On peut remarquer la présence de deux paires de caméras stéréoscopiques, sur le bras et sur un mât, qui nous ont fait commencer à appréhender les possibles problèmes d'anthropomorphisation du robot.

PR2 MAX et BOB. Lorsque je suis revenue au LAAS en 2008, un des PR2 était déjà présent et le second arriverait peu après. À la même époque, l'équipe RIS déménageait dans le nouveau bâtiment ADREAM équipé d'une maquette d'appartement (avec meubles et murs). Nous bénéficions également aujourd'hui d'une couverture de tout cet environnement avec un système de motion capture. Bien que sur une base mobile, on peut considérer PR2

comme un robot "humanoïde" avec ses deux "bras" et sa "tête". Le robot est conçu pour interagir et ne pas être dangereux (toutes proportions gardées). Une grande partie des travaux qui vont être présentés dans ce manuscrit ont été réalisés avec ces robots.



Pepper est arrivé lors du projet MuMMER dont il sera question section 9. Toujours sur roulettes, Pepper est aussi considéré comme un robot "humanoïde". En termes de capacités de manipulation d'objet ou de navigation, on est loin des PR2, ce qui limite grandement les démos que nous pouvons réaliser. Néanmoins, on ne peut pas lui enlever son aspect sympathique et engageant.

1.2 L'architecture LAAS

Lorsque l'on parle d'architecture robotique au LAAS, on en revient souvent, toujours, au papier "originel" de Alami, Chatila, Fleury, Ghallab et Ingrand, "An Architecture for Autonomy" paru en 1997 [Alami 1998] présentant l'architecture LAAS pour "LAAS Architecture for Autonomous System". Dans l'introduction, les auteurs indiquent qu'ils décrivent une architecture intégrée (qu'il faut comprendre ici par implémentée) permettant à un robot mobile de planifier ses tâches en prenant en compte des contraintes temporelles et de domaine pour réaliser les actions correspondantes et contrôler leurs exécutions en temps réel tout en étant réactif aux événements possibles.

Ils définissent déjà un certain nombre de concepts qui restent importants et clés aujourd'hui. Ainsi, ils indiquent que l'organisation d'un système robotique détermine ses capacités à accomplir des tâches et à réagir aux événements. La structure de contrôle d'un robot autonome doit avoir à la fois des capacités décisionnelles et réactives. Les situations doivent être anticipées et les actions adéquates décidées par le robot en conséquence. Les tâches doivent être instanciées et affinées au moment de l'exécution, selon le contexte réel. Le robot doit réagir en temps opportun aux événements.

Pour répondre à ces exigences, ils définissent les propriétés suivantes :

Programmabilité : un robot ne peut pas être conçu pour un seul environnement ou une seule tâche. Il doit être capable d'accomplir plusieurs tâches dont la description doit pouvoir se faire à un certain niveau d'abstraction. Ses fonctions doivent pouvoir être facilement combinées en fonction de la tâche à exécuter.

Autonomie et adaptabilité : le robot doit être capable de réaliser ses actions et d'affiner ou de modifier la tâche et son propre comportement en fonction de l'objectif et du contexte d'exécution.

Réactivité : le robot doit être capable de prendre en compte les événements dans des délais compatibles avec la réalisation correcte et efficace de ses objectifs (en intégrant également sa propre sécurité).

Comportement cohérent : les réactions du robot aux événements doivent être guidées par les objectifs de sa tâche.

Robustesse : l'architecture de contrôle doit pouvoir exploiter la redondance des fonctions disponibles. La robustesse exigera que le contrôle soit décentralisé à certains degrés.

Extensibilité : l'intégration de nouvelles fonctions et la définition de nouvelles tâches doivent être aisées. Les capacités d'apprentissage sont importantes à considérer ici : l'architecture doit rendre l'apprentissage possible.

Les auteurs proposent ensuite un découpage en trois niveaux pour cette architecture, chacun des niveaux ayant ses propres contraintes temporelles et manipulant des données de nature différentes. Ces niveaux sont :

Le niveau fonctionnel : ce niveau comprend toutes les capacités de base du robot concernant ses actions et ses capacités de perception. On considère que ces fonctions de traitements et boucles de contrôle (traitement d'image, évitement d'obstacles, contrôle de mouvement, etc.) doivent être indépendantes autant que possible et donc portables d'un robot à un autre. Pour cela, on prévoit un niveau d'interface avec les capteurs et les actionneurs.

Le niveau de contrôle d'exécution : ce niveau contrôle et coordonne l'exécution des fonctions réparties dans les modules selon les exigences de la tâche.

Le niveau décisionnel : ce niveau comprend les capacités de production du plan et de supervision de son exécution tout en étant réactif aux événements du niveau précédent. Ce niveau peut être décomposé en deux ou plusieurs couches basées sur le même design conceptuel, mais utilisant des abstractions de représentation différentes ou des outils algorithmiques différents et ayant des propriétés temporelles différentes dépendant de l'application.

Plusieurs modèles d'architectures robotiques dites "trois niveaux" suivant plus ou moins les mêmes principes ont été développés à cette époque (e.g. [Gat 1992, Muscettola 1998, Nesnas 2003, Tambe 1997]).

1.3 Théorie de l'action

Nous allons voir maintenant que les approches philosophiques et robotiques partagent un certain nombre d'idées concernant l'architecture nécessaire pour réaliser une action individuelle.

Selon les théories philosophiques classiques de l'action individuelle, le comportement n'est qualifié d'action que si il a un certain type d'antécédent mental ou implique certains types de processus psychologiques (par exemple [Davidson 1980, Mele 1992, Searle 1983, Bratman 1999]). Typiquement, cet antécédent mental est identifié comme une intention. Les intentions sont souvent caractérisées comme des plans d'action que l'organisme choisit et dans lesquels il s'engage à la poursuite d'un objectif (par exemple [Bratman 1999]). Selon ce point de vue, les intentions sont des attitudes exécutives dont les fonctions consistent a) à mettre fin au raisonnement pratique sur les fins, b) à susciter un raisonnement pratique sur les moyens et les plans, c) à aider à coordonner le comportement de l'agent dans le temps et avec le comportement des autres agents, d) à initier et à maintenir une action intentionnelle et à surveiller et le guider jusqu'à son achèvement. Les intentions comprennent donc à la fois des représentations d'objectifs et de moyens pour atteindre ces objectifs, c'est-à-dire des plans d'action qui peuvent aller de simples représentations d'actions de base à des stratégies complexes pour atteindre des objectifs lointains. De

plus, comme le souligne Bratman [Bratman 1999], les plans d'action sont soumis à des contraintes de rationalité. Les différents éléments qui forment les briques d'un plan d'action doivent être cohérents entre eux (cohérence interne). Le plan dans son ensemble doit être cohérent avec les croyances de l'agent sur le monde et sur la réalité actuelle, y compris ses croyances sur ses propres capacités et compétences (cohérence externe). Enfin, le plan doit prendre en compte le cadre plus large des activités et des projets dans lesquels l'agent est également impliqué et être coordonné avec eux dans un plan plus global (cohérence globale). Ce bref aperçu des références issues de la philosophie fournit des définitions de travail d'une action, d'une intention, d'un but et d'un projet, des éléments qui doivent être manipulés pour permettre d'agir.

Elisabeth Pacherie [Pacherie 2008, Pacherie 2012] propose un modèle dynamique des intentions qui distingue trois étapes principales dans le processus de spécification des actions :

Un niveau d'intentions distales (D-intentions) responsable de la dynamique de prise de décision, de la flexibilité temporelle et du pilotage rationnel de haut niveau et du suivi de l'action ; elles ont pour fonctions de :

— gérer la dynamique de la prise de décision (*dynamics of decision making*), en décidant quel objectif poursuivre et en assurant : la cohérence interne (les différents éléments qui forment les blocs de construction d'un plan d'action doivent être cohérents entre eux) ; la cohérence externe (le plan dans son ensemble doit être cohérent avec les croyances de l'agent sur le monde) ; la cohérence globale (le plan doit prendre en compte le cadre plus large des activités et des projets dans lesquels l'agent est également impliqué et être coordonné avec eux dans un plan plus global)

— assurer une certaine flexibilité temporelle (*temporal flexiblity*) en donnant la possibilité pour un agent de former une D-intention d'accomplir une action d'un type donné même si sa situation présente n'est pas de nature à permettre son exécution immédiate (= détachable de la situation actuelle de l'agent)

— assurer le pilotage rationnel de haut niveau et le suivi de l'action (*high level rational guidance and monitoring of the action*). La D-intention doit être maintenue vivante dans la mémoire prospective jusqu'au moment de l'exécution. Quand c'est le cas, une fonction essentielle d'une D-intention est d'assurer le contrôle rationnel de l'action en cours : contrôle du suivi (assurer que chaque étape successive du plan d'action est réussie avant de passer à l'étape suivante) et contrôle collatéral (contrôle des effets secondaires de l'accomplissement d'une action)

Un niveau d'intentions proximales (P-intentions) qui hérite d'un plan du niveau précédent et dont le rôle est d'ancrer ce plan dans la situation d'action, cet ancrage doit s'effectuer à deux niveaux : ancrage temporel (décider quand commencer à agir) et ancrage situationnel (s'assurer que le plan d'actions correspond bien à la situation courante). Ce niveau va intégrer les informations conceptuelles sur l'action héritées de la D-intention avec les informations perceptives sur la situation actuelle et les informations en mémoire provenant de son répertoire moteur pour donner une représentation plus précise de l'action à accomplir (passer d'un schéma abstrait (D-intention) à un schéma plus concret contraint notamment par les informations perceptuelles actuelles). L'échelle de temps des P-intentions se situe au niveau du tempo de l'action.

Un niveau d'intentions motrices (M-intentions) qui encode les détails fins de l'action (correspondant à ce que les neuroscientifiques appellent les représentations motrices). Il est responsable de la précision et de la fluidité de l'exécution de l'action, et opère à une échelle de temps plus fine que les autres niveaux. Ces représentations ont trois caractéristiques importantes :

1 Les attributs des objets et des situations sont représentés dans un format utile pour la sélection immédiate des schémas moteurs appropriés (par exemple, si l'on veut saisir un objet, sa position spatiale sera représentée en termes de mouvements nécessaires pour l'atteindre en fonction de sa forme, de sa taille et du type de prise en main qu'il permet).

2 Ces représentations des mouvements à effectuer reflètent une connaissance implicite des contraintes biomécaniques et des règles cinématiques et dynamiques régissant le système moteur (par exemple, les mouvements des effecteurs seront programmés de manière à éviter les positions inconfortables ou inconfortables des membres et à minimiser les temps passés dans des angles articulaires extrêmes).

3 Une représentation motrice code normalement des mouvements transitifs, où le but de l'action détermine l'organisation globale de la séquence motrice (par exemple, le type de prise choisi pour un objet donné n'est pas seulement fonction de ses caractéristiques intrinsèques (sa forme et sa taille) mais aussi de l'usage ultérieur que l'on veut en faire ; une même tasse sera saisie de différentes manières selon qu'on veut la porter à ses lèvres ou la mettre à l'envers)

On peut alors décrire une cascade intentionnelle. Une D-intention déclenche des processus de niveau P en leur fournissant un plan d'action qui peut encore être principalement descriptif et abstrait. La tâche des processus de niveau P est alors d'ancrer ce plan d'action dans la situation actuelle et de proposer une représentation plus détaillée de l'action adaptée à la situation en cours. Autrement dit, les processus de niveau P prennent en entrée à la fois une spécification de l'action à effectuer fournie par la D-intention ainsi que des informations perceptives sur la situation, y compris des informations sur l'état propre de l'agent, et produisent en sortie une représentation plus détaillée. De même, les P-intentions déclenchent des processus de niveau M leur fournissant une représentation située de l'action, disons une cible à un certain emplacement et une action donnée à effectuer sur cette cible. Leur tâche est alors de traiter l'information sur cette cible et d'élaborer une représentation sensori-motrice de l'action qui en fixe les paramètres. Enfin, la dynamique des trois niveaux d'intentions n'implique nullement que toute action nécessite la présence de toute la cascade intentionnelle.

1.4 Théorie de l'action jointe

Les deux sections précédentes suggèrent une convergence remarquable entre une explication philosophique de la structure et de la dynamique de l'action humaine et une architecture de commande de robot dédiée à l'action autonome. Dès lors, il nous est apparu pertinent de nous intéresser à la théorie de l'action jointe telle qu'elle est définie en philosophie pour voir si elle pouvait nous éclairer sur la définition d'une architecture robotique pour l'interaction.

Comme indiqué par Knoblich et al. [Knoblich 2011], les philosophes s'accordent généralement à dire que "ce qui distingue les actions jointes des actions individuelles, c'est

que les actions jointes impliquent une intention partagée et que les intentions partagées sont essentielles pour comprendre la coordination dans l'action jointe". Tomasello et al. [Tomasello 2005] ne disent rien d'autre lorsqu'ils affirment que "comprendre les actions intentionnelles et la perception d'autrui n'est pas suffisant en soi pour produire des activités sociales ou culturelles à l'image de l'humain. Il faut quelque chose de plus. Notre hypothèse pour ce "quelque chose de plus" est l'intentionnalité partagée". Et ils ajoutent : "l'intentionnalité partagée fait référence à des interactions collaboratives dans lesquelles les participants ont un objectif commun (engagement partagé) et un rôle dans la coordination des actions pour poursuivre cet objectif commun". Une action jointe réussie dépend de la coordination efficace des objectifs, des intentions, des plans et des actions des agents participants. En d'autres termes, dans l'action jointe, il ne suffit pas que les agents contrôlent leurs propres actions, c'est-à-dire prédisent correctement leurs effets, surveillent leur exécution et fassent des ajustements si nécessaire. Ils doivent également coordonner leurs intentions et leurs actions avec celles de leurs co-agents afin d'atteindre leur objectif commun.

Elisabeth Pacherie définit là encore trois niveaux d'intentions [Pacherie 2012, Pacherie 2011, Pacherie 2007] qu'elle décrit comme partagées pour les deux premières et comme couplée pour la dernière car aucune preuve empirique ne permet d'affirmer l'existence de représentations motrices de buts communs :

— Un niveau d'intentions distales partagées (SD-intentions) (*Shared Distal intentions*) À ce niveau, les agents participants (1) représentent l'objectif global, mais n'ont pas besoin de représenter l'ensemble du plan, mais seulement leurs propres sous-plans et les parties des sous-plans des autres dans lesquels ils sont impliqués et (2) certaines parties de ce qu'ils représentent doit être exécutées par d'autres. Les agents doivent être capables de gérer l'ajustement triadique et dyadique à ce niveau.

— Un niveau d'intentions distales partagées (SD-intentions) (*Shared Distal intentions*) C'est à ce niveau que s'effectue l'ancrage situationnel et temporel de l'action, ce qui signifie que le plan d'action hérité du niveau distal/décisionnel doit encore être affiné et ajusté à la situation en présence pour que l'action soit lancée et son déroulement surveillé et contrôlé. À ce niveau, les agents ont besoin de pouvoir partager des représentations (dans le meilleur des cas conjointement) et de coordonner leurs perceptions (pour partager une attention jointe) afin de coordonner leurs actions et éventuellement de réaliser des ajustements (dyadiques, triadiques et collaboratifs) en fonction du contexte courant. Pour que les agents partagent une intention proximale, les éléments suivants doivent être respectés :

(1) les agents représentent chacun leurs propres actions et leurs conséquences prévues dans la situation actuelle (autoprédictions),

(2) les agents représentent chacun les actions, buts, intentions motrices et proximales de leurs co-agents et leurs conséquences (prédictions sur les autres),

(3) les agents représentent chacun comment ce qu'ils font affecte ce que font les autres et vice-versa et ajustent leurs actions en conséquence (ajustement dyadique),

(4) les agents ont chacun une représentation (qui peut n'être que partielle) de la hiérarchie des buts situés et des états souhaités aboutissant au but commun global (plan d'action commun),

(5) les agents prédisent chacun les effets conjoints de leurs propres actions et des

actions des autres (prédictions jointes), et

(6) les agents utilisent chacune des prédictions jointes pour surveiller les progrès vers l'objectif commun et décider de leurs prochains mouvements, y compris les mouvements qui peuvent impliquer d'aider les autres à réaliser leurs contributions à l'objectif commun (ajustement triadique).

— Un niveau d'intentions motrices couplées (CM-intentions) (*Coupled-Motor intentions*) Enfin, pour que certaines actions jointes aboutissent, il ne suffit pas que les coacteurs partagent une représentation d'une hiérarchie d'objectifs situés et convergent sur la répartition des rôles ; il faut aussi que leurs actions soient très précisément coordonnées dans le temps et dans l'espace.

1.5 Application à la robotique

Il faut noter qu'il existe des travaux fondateurs sur la modélisation du travail d'équipe en informatique tels que les travaux Cohen et Levesque sur la théorie de l'intention jointe (*Joint Intention Theory*) [Cohen 1990] et Grosz et Kraus sur les plans partagés (*Shared Plans*) [Grosz 1996]. Nous proposons ici d'analyser la théorie de l'action jointe telle que proposée par Elisabeth Pacherie [Pacherie 2012, Pacherie 2011, Pacherie 2007], qui comme nous l'avons vu considère également trois niveaux, et de voir quelle signification on pourrait donner à chacun de ces niveaux dans une architecture robotique.

1.5.1 Niveau distal/décisionnel partagé

À ce niveau, agissant seul, le robot gère son objectif, son plan et sa prise de décision ; tous les éléments qu'il représente seraient réalisés par lui-même. Agissant conjointement, le robot doit être capable de gérer un objectif commun, une représentation de plan et d'action et éventuellement une prise de décision coopérative (y compris, par exemple, des capacités de planification jointe). Il représentera non seulement ce qui serait réalisé par lui-même, mais aussi par l'autre (avec des niveaux de granularité et d'exhaustivité potentiellement différents). De plus, une surveillance de haut niveau comprendrait non seulement la surveillance par le robot de ses propres actions et objectifs, mais aussi plus généralement la surveillance de l'objectif commun et, par conséquent, la surveillance des actions des autres également.

1.5.2 Niveau proximal/exécutif partagé

À ce niveau, le robot doit être capable de gérer : sa représentation du monde, une représentation du monde de l'humain avec lequel il interagit (potentiellement limitée à la tâche à accomplir), l'effet possible de ses actions sur les actions humaines (et vice versa), la représentation de leur objectif commun et de leur plan d'action, une prévision de leurs actions, un moyen de suivre les progrès vers l'objectif commun (et éventuellement un moyen de réviser le plan commun en cours). Un ajustement triadique signifie que le robot et l'humain peuvent adapter leur comportement vers l'objectif commun. Cela implique, par exemple, que si l'humain laisse tomber un objet qu'il doit manipuler dans le cadre de son action dans l'espace atteignable par le robot, le robot peut terminer l'action de l'humain pour faire avancer la tâche globale. S'il avait fait un ajustement dyadique, il aurait rendu l'objet accessible à l'humain pour lui permettre de terminer son action.

Un ajustement dyadique signifie que le robot et l'humain peuvent adapter leur comportement aux actions de l'autre (et non vers l'objectif commun). Fait intéressant, Tomasello et al. [Tomasello 2005] ont proposé que la capacité d'engagement triadique présente deux phases dans le cours du développement humain. Vers l'âge de 9 à 12 mois, les nourrissons commencent à interagir avec un agent orienté vers un objectif commun. Ce faisant, les deux surveillent le comportement et les perceptions de leurs partenaires par rapport à cet objectif commun. Cependant, ce n'est que vers l'âge de 12 à 15 mois qu'ils commencent à s'engager dans une action jointe coordonnée, comprenant non seulement l'objectif commun, mais aussi la complémentarité entre leurs propres plans d'action spécifiques et ceux de leur partenaire. Comme le soulignent Tomasello et ses collègues : "Cela signifie, par exemple, que l'enfant comprend qu'en poursuivant l'objectif commun de construire une tour de blocs, l'adulte maintient l'édifice stable pendant qu'elle, l'enfant, place des blocs. Les nourrissons de cet âge partagent non seulement des objectifs, mais coordonnent également des rôles" [Tomasello 2005]. Cette compréhension rend ainsi possibles des processus d'ajustement triadique plus souples, comme inverser les rôles avec un partenaire ou aider le partenaire à jouer son rôle. Ces mécanismes d'ajustement exploitent des représentations de tâches partagées qui non seulement spécifient à l'avance quelles sont les tâches respectives de chacun des co-agents, mais fournissent également des structures de contrôle qui permettent une coordination flexible.

1.5.3 Niveau moteur/fonctionnel couplé

Ce niveau correspond au comportement sensori-moteur du robot qui permettrait d'obtenir une interaction niveau moteur avec le partenaire humain. Un exemple pourrait être l'échange d'un objet avec un humain et les processus de retour de force associés. Dans de telles situations impliquant une coordination précise entre les acteurs, le paramétrage du niveau fonctionnel doit être couplé avec celui de l'autre acteur. Cela signifie que les boucles de contrôle du robot seraient directement adaptées en fonction des mouvements ou des actions de l'autre acteur.

1.6 Conclusion

C'est équipé de cette connaissance de notre architecture LAAS à trois niveaux et des exigences pour le déroulement d'une action jointe proposées par l'approche philosophique et de ce qu'elles peuvent nous apporter pour la définition d'une architecture robotique que nous commençons. Ce travail va donner un cadre, un fil d'Ariane à la plupart des travaux que nous présenterons ensuite.

1.7 Contributions associées

[Clodic 2017] Aurélie Clodic, Elisabeth Pacherie, Rachid Alami, and Raja Chatila. Key Elements for HumanRobot Joint Action. In Sociality and Normativity for Robots Philosophical Inquiries into Human-Robot Interactions, Studies in the Philosophy of Sociality, pages 159–177. Springer, 2017. doi : 10.1007/978-3-319-53133-5_8 URL : `https://hal.laas.fr/ijn_03084126/`

Chapitre 2

Implémenter une action jointe humain-robot

2.1 Introduction

Une manière d'appréhender ce que signifie modéliser une interaction est d'y faire attention quand on y est confronté. Donnez-vous l'occasion de le faire sur un exemple simple. L'un des exercices que je donne à mes étudiants lors de mon cours d'initiation à l'interaction humain-robot est le suivant :

— Prenez chacun 2 crayons, mettez-vous en binôme, et construisez ensemble un carré avec.

— Notez tout ce que vous avez dû faire.

Il y a déjà le problème de définition du sujet, certains cherchent par exemple à construire un carré en 3D (se disant que la tâche que je leur demande en elle-même est trop simple sinon) alors que je ne demande qu'à mettre les quatre crayons en ordre pour faire un carré. Ensuite, il y a la définition d'un plan ou non. Certains vont discuter d'abord qui pose quoi et le faire ensuite. D'autres vont se placer dans un processus leader-follower, l'un d'eux va commencer en mettant ses deux crayons et l'autre va poursuivre ensuite. Mais dans tous les cas, pour faire cela, ils vont s'échanger des signaux verbaux et/ou non verbaux.

On voit alors que même sur un exemple simple, se posent des problèmes de prise de décision et d'échanges de signaux et d'information. C'est le fondement de ce que l'on va chercher à modéliser en interaction et pour cela nous nous basons sur les travaux réalisés dans le domaine de l'action jointe.

Une action jointe est une interaction sociale où deux individus ou plus coordonnent leurs actions dans l'espace et dans le temps pour amener un changement dans l'environnement [Sebanz 2006] (*Joint action is "a social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment." [Sebanz 2006]*). Une action jointe fait intervenir un certain nombre de processus de coordination parmi lesquels [Sebanz 2006] :

— la distinction entre soi et l'autre (*self-other distinction*),

— l'attention jointe (*joint attention*),

— la compréhension de l'action intentionnelle (*understanding of intentional action*),

— la représentation partagée de la tâche (*shared task representation*).

En prenant comme fil rouge l'exemple décrit figure 2.1 où une personne et un robot ont pour objectif commun de construire une pile de quatre cubes, nous allons expliquer comment ces processus peuvent être interprétés dans la mise en œuvre d'un tel scénario.

À ce stade, il faut remarquer que d'un point de vue philosophique, on nous a appris que certains philosophes tels que Seibt [Seibt 2014] ont souligné que le vocabulaire intentionnaliste robotique que nous utilisons est considéré comme problématique, en particulier

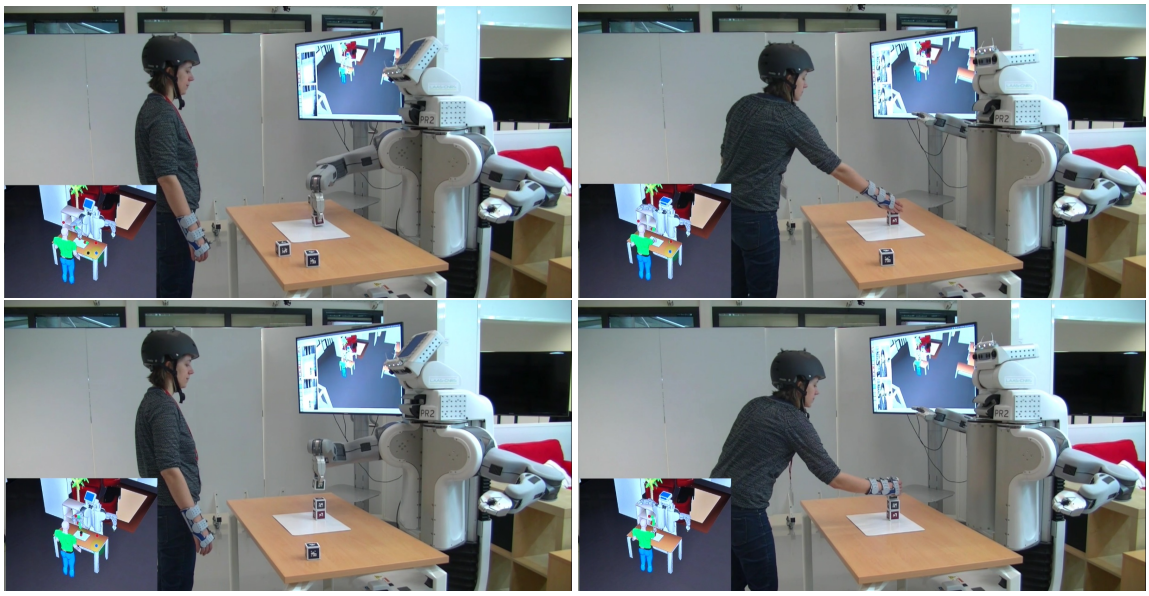


FIGURE 2.1 – Un scénario d'interaction simple : une personne et un robot ont pour objectif commun de construire une pile de quatre cubes. Ils doivent empiler les cubes dans un ordre précis (1, 2, 3, 4). Chaque agent participe à la tâche en plaçant ses cubes sur la pile. Les actions disponibles pour chaque agent sont les suivantes : prendre un cube sur la table, mettre un cube sur la pile, retirer un cube de la pile, placer un cube sur la table et donner un cube à l'autre agent. De plus, l'humain et le robot s'observent.

lorsque les robots sont placés dans un cadre d'interaction humain-robot. Dans ce qui suit, nous utiliserons ce vocabulaire intentionnaliste pour décrire les fonctionnalités du robot, telles que « croire » et « répondre », car c'est ainsi que nous décrivons notre travail dans notre communauté. Cependant, pour tenir compte de la préoccupation philosophique, nous aimerions noter que cela peut être considéré comme un raccourci pour "le robot simule la croyance", "le robot simule une réponse", etc. Ainsi, chaque fois que le comportement robotique est décrit avec un verbe qui caractérise normalement une action humaine, ces passages peuvent être lus comme une référence à la simulation par le robot de l'action concernée.

2.2 La distinction entre soi et l'autre

Le premier processus de coordination est la nécessaire distinction entre soi et l'autre, ainsi "pour que les représentations partagées des actions et des tâches favorisent la coordination plutôt qu'elles ne créent de la confusion, il est important que les agents soient capables de séparer les représentations de leurs propres actions et intentions de celles des autres" [Pacherie 2012].

Concernant notre exemple, cela signifie que chaque agent doit être capable de créer et de maintenir une représentation du monde pour lui-même, mais aussi du point de vue de l'autre agent. Dans ce qui suit, nous expliquerons ce que le robot peut faire pour construire ce type de représentation. La manière dont un humain peut construire une telle représentation pour l'agent robot (et sur quelle base) est encore une question ouverte à mon avis.

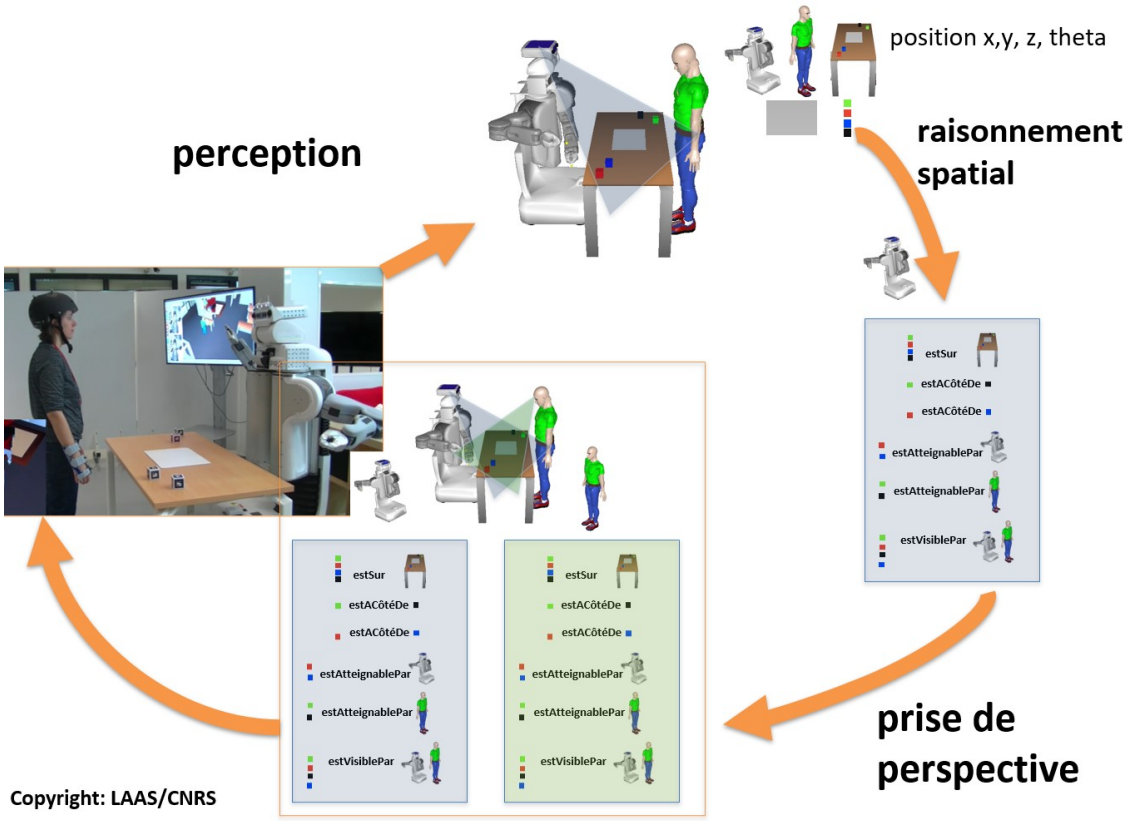


FIGURE 2.2 – Évaluation de la situation : le robot perçoit son environnement, en construit un modèle et calcule des faits par un raisonnement spatial pour pouvoir partager des informations avec la personne à un niveau d'abstraction adapté et réalise une mise en perspective pour pouvoir calculer les mêmes faits du point de vue de la personne avec laquelle il interagit.

2.3 L'attention jointe

Le second processus de coordination est celui de l'attention jointe. L'attention est un processus cognitif qui consiste à se concentrer sélectivement sur un aspect de l'environnement. L'attention jointe ne se réduit pas au fait que deux personnes portent leur attention sur un même objet ou événement, au moins deux conditions supplémentaires doivent être satisfaites [Pacherie 2012]. Premièrement, il doit y avoir une raison qui fait que les deux agents s'intéressent ensemble à ce même objet ou événement (e.g. nous réalisons une tâche ensemble). Deuxièmement, le fait que tous deux portent leur attention sur le même objet ou événement doit leur être mutuellement manifeste ; autrement dit, chacun doit avoir une forme de conscience que l'objet de son attention est un objet présent à tous les deux.

Côté robot, cela signifie qu'il doit être capable de détecter et de représenter ce qui est présent dans l'espace d'attention jointe. Il doit donc être doté de capacités d'évaluation de la situation [Lemaignan 2018, Milliez 2014]. Dans l'exemple illustré par la figure 2.2, il s'agit d'une action jointe d'empilement de cubes, le robot doit alors :

— être capable d'estimer sa propre position dans l'environnement (dans cet exemple, en effectuant des mesures à l'aide de son laser pour se localiser sur la carte de

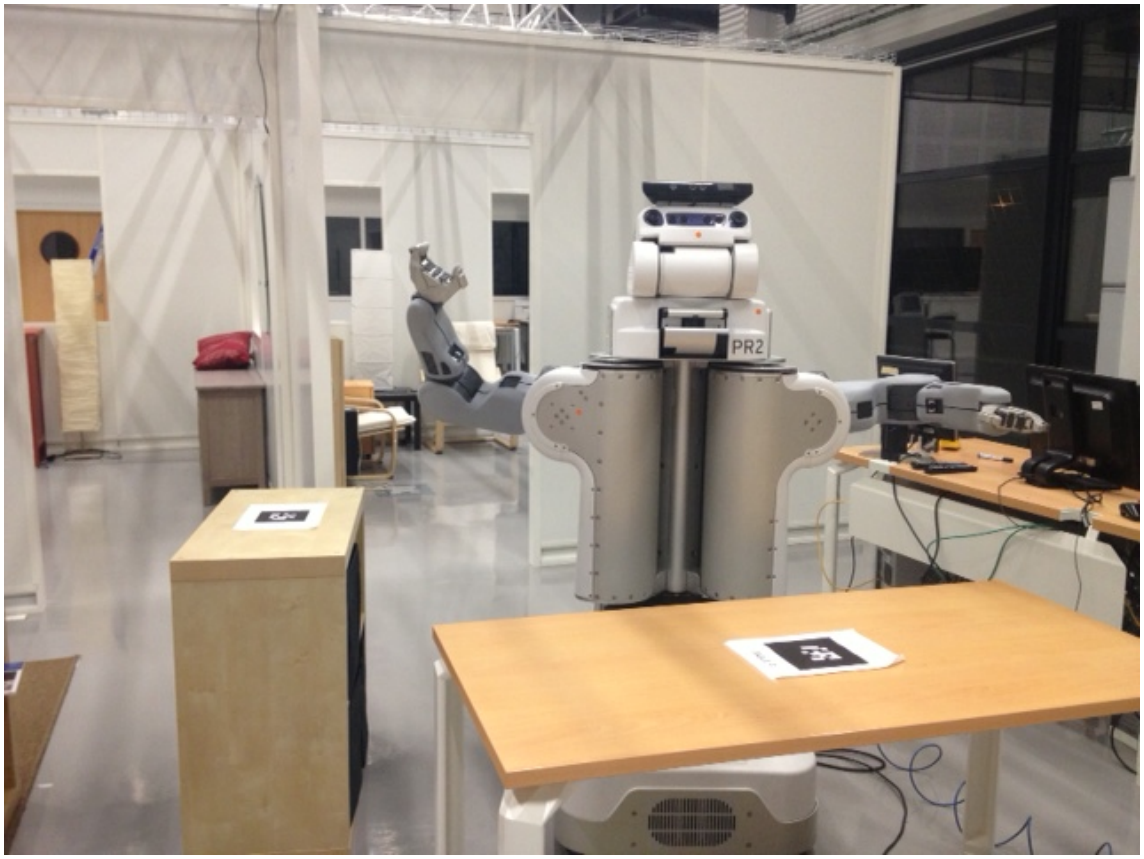
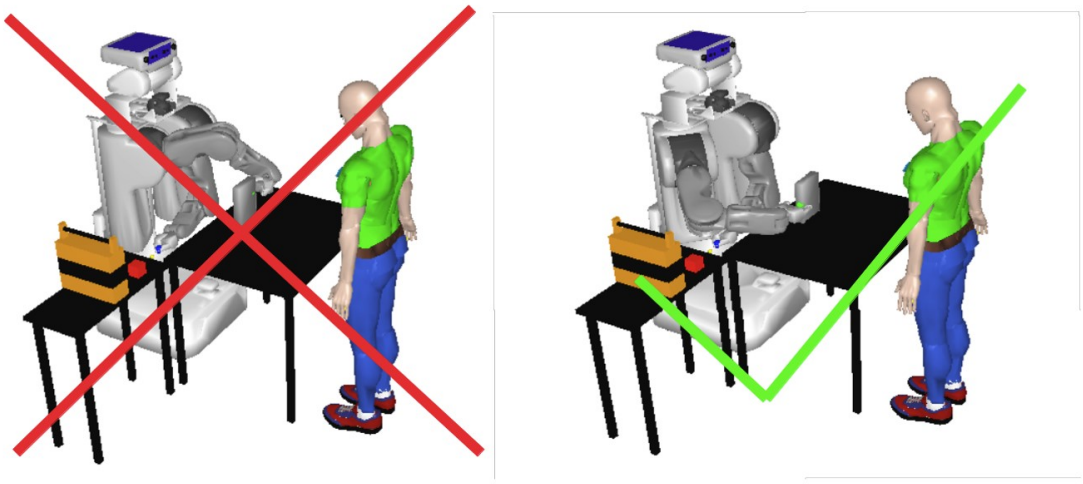


FIGURE 2.3 – Que pouvons-nous en déduire en regardant ce robot ? Il n'y a pas (au jour d'aujourd'hui) d'interface standard pour les robots donc il est difficile, voire impossible, de déduire ce que ce robot est capable de faire et ce qu'il est capable de percevoir (de son environnement, mais aussi de la personne avec laquelle il interagit).

l'environnement qu'il aura lui-même construite [1]) ainsi que la position de chacune des ses articulations (qu'il peut récupérer à partir de son propre modèle),
— la position de la personne avec qui il interagit (par exemple, ici, la personne est suivie grâce à l'utilisation d'un système de capture de mouvements [2], c'est pourquoi la personne porte un casque et une attelle de poignet. Donc plus précisément, dans cet exemple, le robot a accès à la position de la tête et à la position de la main droite),
— la position des objets dans l'environnement (par exemple, ici, un QR-code [3] a été collé sur chaque face des cubes. Ces codes, et donc, les cubes sont suivis avec l'une des caméras du robot. Nous obtenons la position 3D de chaque cube dans l'environnement [4]))

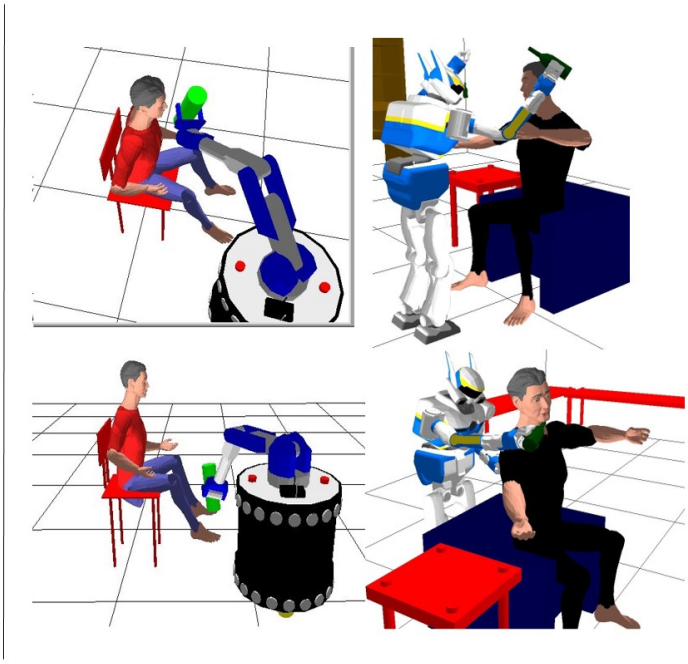
Nous ne pouvons pas imaginer utiliser de telles informations pour élaborer une interaction verbale avec l'humain : "veuillez prendre le cube à la position x = 7,5 mètres, y = 3,0 mètres, Z = 1,0 mètre, thêta = 3,0 radians dans le repère de la salle". Pour dépasser cette limite, nous devons transformer chaque position en une information compréhensible par (et donc partageable avec) l'humain, par exemple *RedBlock isOn table*. Nous pouvons également calculer des informations supplémentaires telles que *GreenBlock isVisibleBy Human* ou *BlueBlock isReachableBy Robot*. C'est ce que nous appelons le raisonnement spatial. Enfin, le robot doit également avoir les moyens d'inférer que les informations dont dispose la personne peuvent être différentes de celles auxquelles il a accès, par ex. un obstacle sur la table peut l'empêcher de voir ce qu'il y a sur la table. Pour déduire cette connaissance, on calcule toutes les informations non seulement du point de vue du robot, mais aussi du point de vue de la personne avec laquelle il interagit, pour cela le robot utilise la prise de perspective [Alami 2011, Warnier 2012, Milliez 2014].

Du côté de la personne, nous savons qu'elle dispose également d'un ensemble similaire d'informations à partir de la situation. Mais l'attention jointe va au-delà. Chaque sujet doit être conscient dans un certain sens, de l'objet en tant qu'objet présent aux deux ; en d'autres termes, le fait que les deux agents se préoccupent du même objet ou du même événement doit être montré clairement et perçu de manière manifeste (ce qu'elle nomme "mutual manifestness") [Pacherie 2012]. Cela soulève plusieurs questions. Comment un robot peut-il manifester une attention jointe ? Quels signaux, quels indices, le robot doit-il envoyer pour permettre au sujet de déduire que l'attention jointe est partagée ? Comment un robot peut-il savoir que la personne avec laquelle il interagit est réellement impliquée dans la tâche jointe ? Quels sont les indices qui doivent être collectés par le robot pour déduire l'existence d'une attention jointe ? Ces questions sont encore des questions ouvertes. Pour y répondre, il faut travailler particulièrement sur la façon de rendre le robot plus compréhensible et plus lisible. Par exemple, en regardant ce robot sur la figure 2.3, que peut-on en déduire de ses capacités ?



Copyright: LAAS/CNRS

FIGURE 2.4 – Deux positions finales du bras du robot pour saisir un objet. Celle de droite est meilleure du point de vue de l'interaction, puisque facilement compréhensible par la personne. Cependant, du point de vue du calcul, de l'informatique (et même d'un point de vue efficacité si l'on ne considère que l'action doit être effectuée par le robot), elles sont équivalentes. Par conséquent, nous devons tenir compte explicitement de ces caractéristiques lors de la planification des mouvements du robot. C'est ce que vise la planification de mouvement prenant en compte l'humain (human-aware motion planning).



Copyright: LAAS/CNRS

FIGURE 2.5 – Positions du robot qui ne sont pas adaptées à l'humain. Plusieurs critères sont à prendre en compte comme la sécurité, le confort et la visibilité. Il s'agit de la position de transfert, mais aussi de la position globale du robot lui-même.

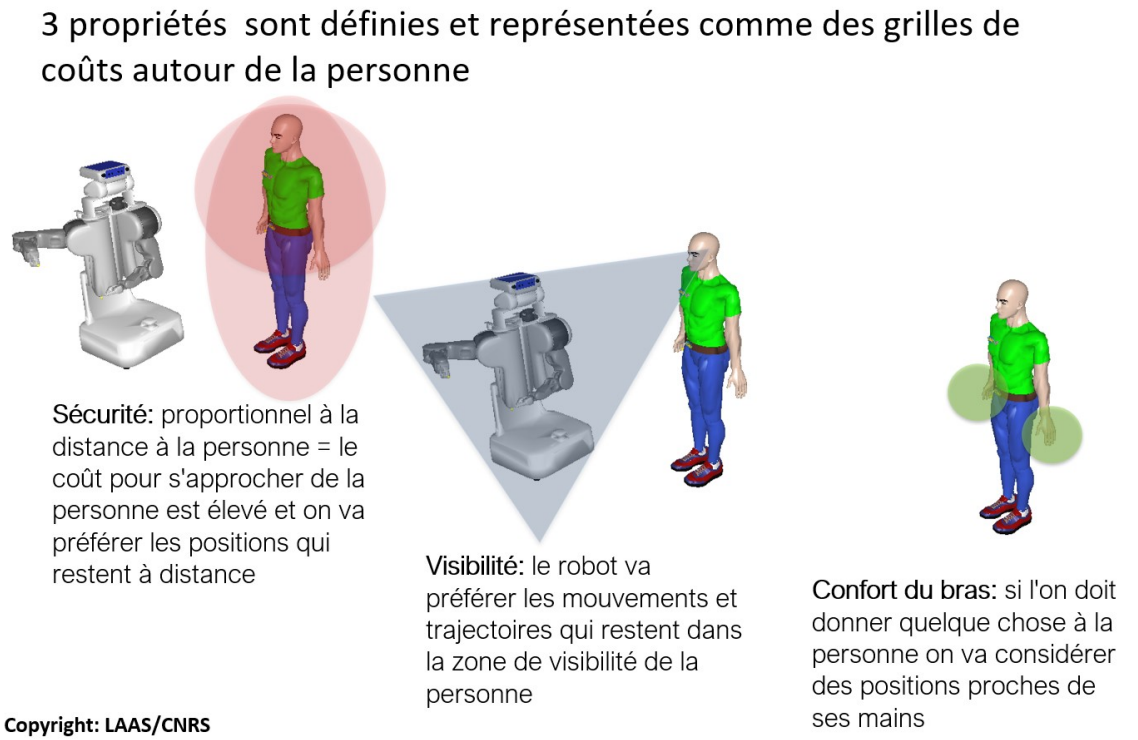


FIGURE 2.6 – Trois critères définis pour un algorithme de planification de mouvement prenant en compte l'humain : sécurité de l'humain, visibilité du robot par l'humain et confort de l'humain. Les trois critères peuvent être pondérés en fonction de leur importance auprès d'une personne donnée, à un endroit ou à un moment particulier de la tâche.

2.4 La compréhension de l'action intentionnelle

"La compréhension des intentions est fondamentale, car elle fournit la matrice d'interprétation pour décider précisément ce que quelqu'un fait en premier lieu. Ainsi, exactement le même mouvement physique peut être considéré comme donner un objet, le partager, le prêter, le déplacer, s'en débarrasser, le rendre, l'échanger, le vendre, et ainsi de suite, selon les objectifs et les intentions de l'acteur" [Tomasello 2005] (*"Understanding intentions is foundational because it provides the interpretive matrix for deciding precisely what it is that someone is doing in the first place. Thus, the exact same physical movement may be seen as giving an object, sharing it, loaning it, moving it, getting rid of it, returning it, trading it, selling it, and on and on—depending on the goals and intentions of the actor" [Tomasello 2005]*). La compréhension de l'action intentionnelle est un élément constitutif de la compréhension des intentions, cela signifie que chaque agent devrait être capable de lire les actions de son partenaire. Pour comprendre une action intentionnelle, un agent doit, lorsqu'il observe l'action ou le cours des actions d'un partenaire, être capable de déduire l'intention de son partenaire. Ici, lorsque nous parlons de l'intention du partenaire, nous entendons son objectif et son plan. Elle est liée à la prédiction de l'objectif à partir de l'action (*action to goal prediction*) (c'est-à-dire, en visualisant et en comprenant l'action en cours, l'agent est capable de déduire l'objectif sous-jacent) et à la prédiction de l'action à partir de l'objectif (*goal to action prediction*) (c'est-à-dire, en connaissant l'objectif, l'agent est capable de déduire ce que seront les actions nécessaires pour y parvenir).

Du côté du robot, cela signifie qu'il doit être capable de comprendre ce que l'humain est en train de faire et de pouvoir prédire les résultats de ses actions, par exemple, il doit être équipé de capacités de reconnaissance d'actions. La difficulté ici est de définir ce qui devrait et peut être reconnu puisque le spectre est vaste en ce qui concerne ce que l'humain est capable de faire. Une façon d'y parvenir est de choisir de ne considérer qu'un ensemble d'actions lié à une tâche particulière.

De l'autre côté, la personne doit être capable de comprendre ce que fait le robot, être capable de déduire l'objectif et de prédire les résultats des actions du robot. Cela signifie qu'en voyant un mouvement, l'humain doit être capable de déduire quelle est l'action du robot. Cela signifie que le robot doit effectuer un mouvement lisible par l'homme. Avant de faire un mouvement, le robot doit le calculer, c'est la planification du mouvement. La planification du mouvement prend en entrée une configuration initiale et une configuration finale (pour la manipulation, c'est la position des bras ; pour la navigation, c'est la position de la base du robot). La planification de mouvement calcule un chemin ou une trajectoire depuis la configuration initiale jusqu'à la configuration finale. Ce chemin pourrait être possible, mais non compréhensible et/ou lisible et/ou prévisible pour l'humain. Par exemple, sur la figure 2.4, le déplacement à gauche est possible, mais celui de droite devra être préféré.

Certains déplacements peuvent également être dangereux et/ou peu confortables pour l'humain, comme illustré à la Figure 2.5. La planification de mouvement prenant en compte l'humain (Human-Aware motion planning) [Sisbot 2007, Kruse 2013, Khambhaita 2017b, Khambhaita 2017a] a été développée pour permettre au robot de gérer le choix d'un chemin

1. par exemple en utilisant amcl (`http://wiki.ros.org/amcl`) et gmapping (`http://wiki.ros.org/gmapping`))
2. ici de type optitrack `https://optitrack.com/`
3. `https://en.wikipedia.org/wiki/QR_code`
4. par exemple, avec `http://wiki.ros.org/ar_track_alvar`

qui est acceptable, prévisible et confortable pour la personne avec laquelle il interagit. La figure 2.6 montre un exemple de critères utilisés pour l'implémentation d'un tel algorithme de planification de mouvement [Sisbot 2007, Sisbot 2010, Sisbot 2012]. Cet algorithme prend en compte la sécurité, la visibilité et le confort de l'humain. De plus, il est capable de calculer un chemin à la fois pour le robot et pour l'humain, ce qui peut résoudre une situation où une action de l'humain est nécessaire ou peut être utilisée pour équilibrer l'effort entre les deux agents.

Cependant, ce n'est pas suffisant. Lorsqu'un robot est équipé de quelque chose qui ressemble à une tête, par exemple, les gens ont tendance à considérer que cet élément devrait agir comme une tête parce que les gens anthropomorphisent. Cela signifie que nous devons considérer l'ensemble du corps du robot et pas seulement la base ou les bras du robot pour le mouvement même s'il n'est pas nécessaire pour réaliser l'action (par exemple, pour un mouvement de bras ou de la base du robot, la tête n'a pas nécessairement besoin de bouger, mais avoir une coordination du mouvement des deux éléments peut améliorer la compréhension [Gharbi 2015, Khambhaita 2016]). Cela pourrait être lié au concept de "coordination smoother" qui est défini comme tout type de modulation de ses propres mouvements qui a pour effet de fiabiliser la compréhension et simplifier la coordination [Vesper 2010].

2.5 Représentations partagées de la tâche

Le dernier processus de coordination est celui de la représentation partagée des tâches. Comme l'ont souligné Knoblich et ses collègues [Knoblich 2011], la représentation partagée des tâches joue un rôle important dans la coordination. À mon sens, partager des représentations peut être considéré comme une mise en perspective de tous les processus déjà décrits, par exemple, savoir que le robot et l'humain suivent le même cube dans la scène d'interaction grâce à l'attention jointe et que le robot est en train de déplacer ce cube vers la pile à l'aide de la compréhension de l'action intentionnelle. La compréhension prend sens dans le contexte où le robot et l'humain construisent ensemble une pile dans le cadre d'une action jointe.

Pour pouvoir partager des représentations de tâches, nous devons avoir les mêmes (ou un moyen de les comprendre). On peut par exemple utiliser des représentations telles que celles développées par le planificateur HATP Human-Aware Task Planner basé sur une représentation à base de Hierarchical Task Network (HTN) [Alami 2006], [Montreuil 2007], [Alili 2009a], [Clodic 2009], [Lallement 2014]. La représentation du domaine est illustrée figure 2.7, elle est composée d'un ensemble d'actions (par exemple, *placerCube*) et d'un ensemble de tâches (par exemple, *construirePile*) qui combinent action(s) et tâche(s). L'un des avantages d'une telle représentation est qu'elle est lisible par l'humain. Ici, *placerCube(Agent R, Cube C, Zone A)* signifie que pour un Agent R, pour placer le Cube C dans la Zone A, la condition préalable est que R ait en main le Cube C et les effets de l'action sont que R n'a plus le cube C en main, mais l'objet C est sur la pile de la zone A. Il est possible d'ajouter un coût et une durée à chaque action si on veut peser l'influence de chacune des actions.

D'autre part, *construirePile* se fait en ajoutant un cube *ajouterCube* puis en continuant à construire la pile *construirePile*. Ensuite, chaque tâche est également affinée jusqu'à ce que nous obtenions une action. HATP calcule un plan à la fois pour le robot et pour l'humain (ou les humains) avec lequel il interagit, comme illustré à la figure 2.8.

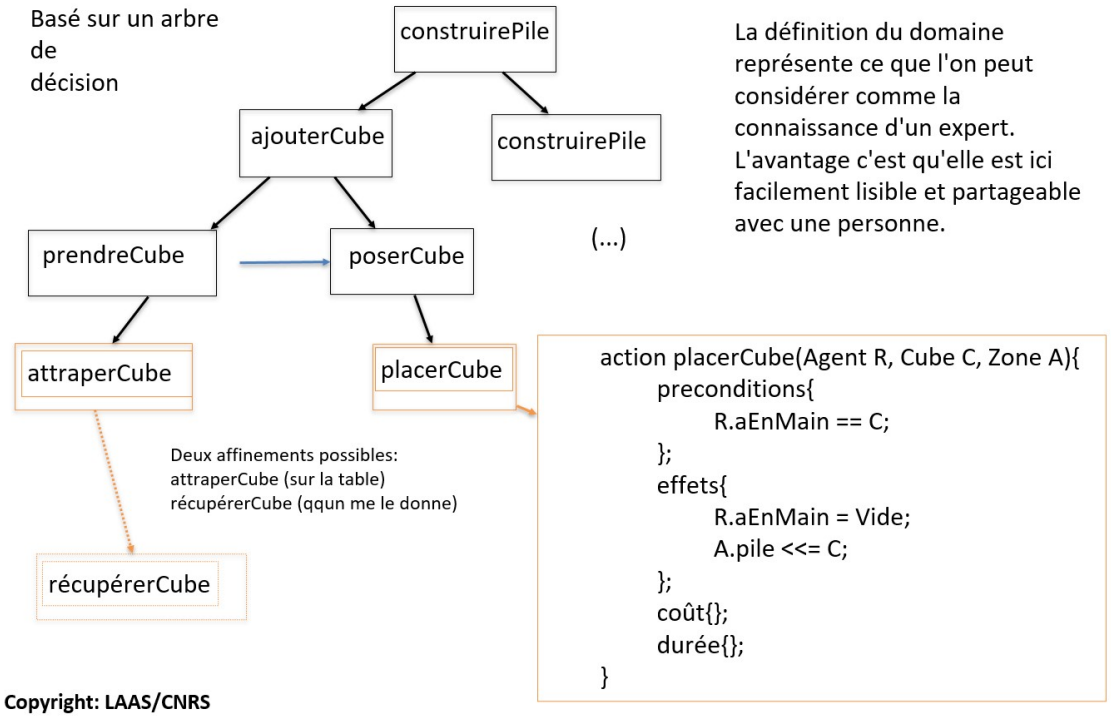


FIGURE 2.7 – Définition du domaine HATP pour la tâche jointe construirePile et définition de l'action placerCube : L'action placerCube pour un Agent R, un Cube C dans une Zone A, pourrait être définie comme suit. La condition préalable est que l'agent R ait le cube C en main avant l'action, l'effet de l'action est que l'agent R n'a plus le cube C et que le cube C est sur la pile dans la zone A. La tâche construirePile combine ajouterCube et construirePile. La tâche ajouterCube combine prendreCube et poserCube. La tâche prendreCube peut être effectuée soit en attrapant le cube sur la table, soit quelqu'un vous le donne à la main.

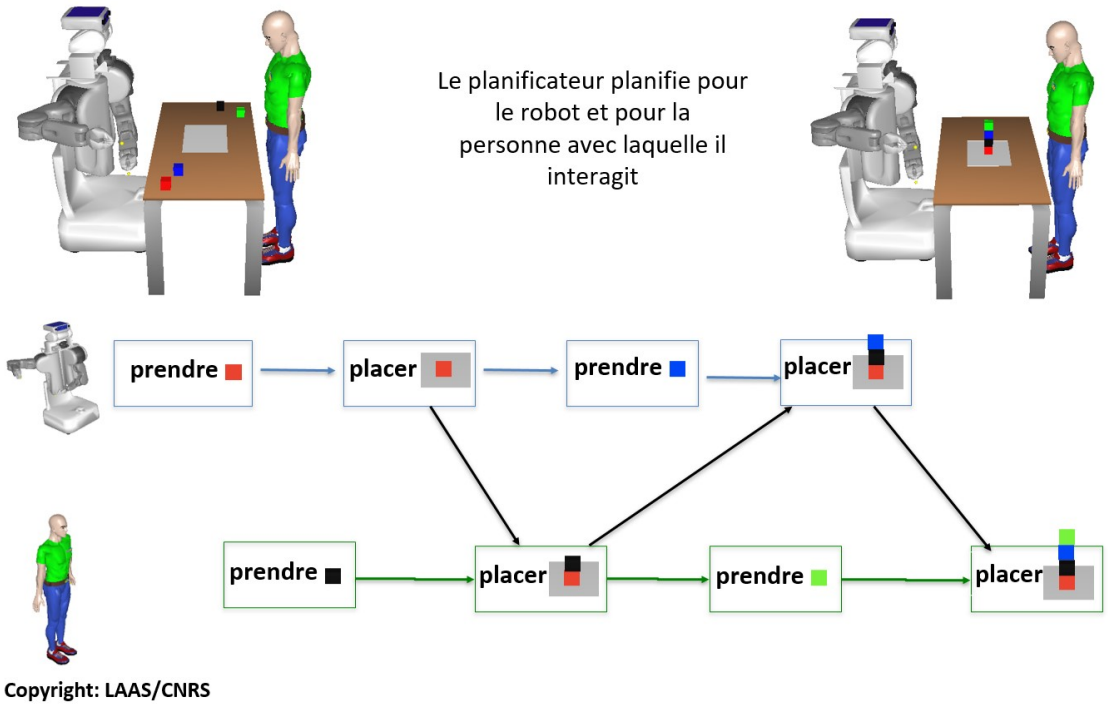


FIGURE 2.8 – Exemple de plan partagé HATP (humain et robot) pour l'exemple de pile de cubes.

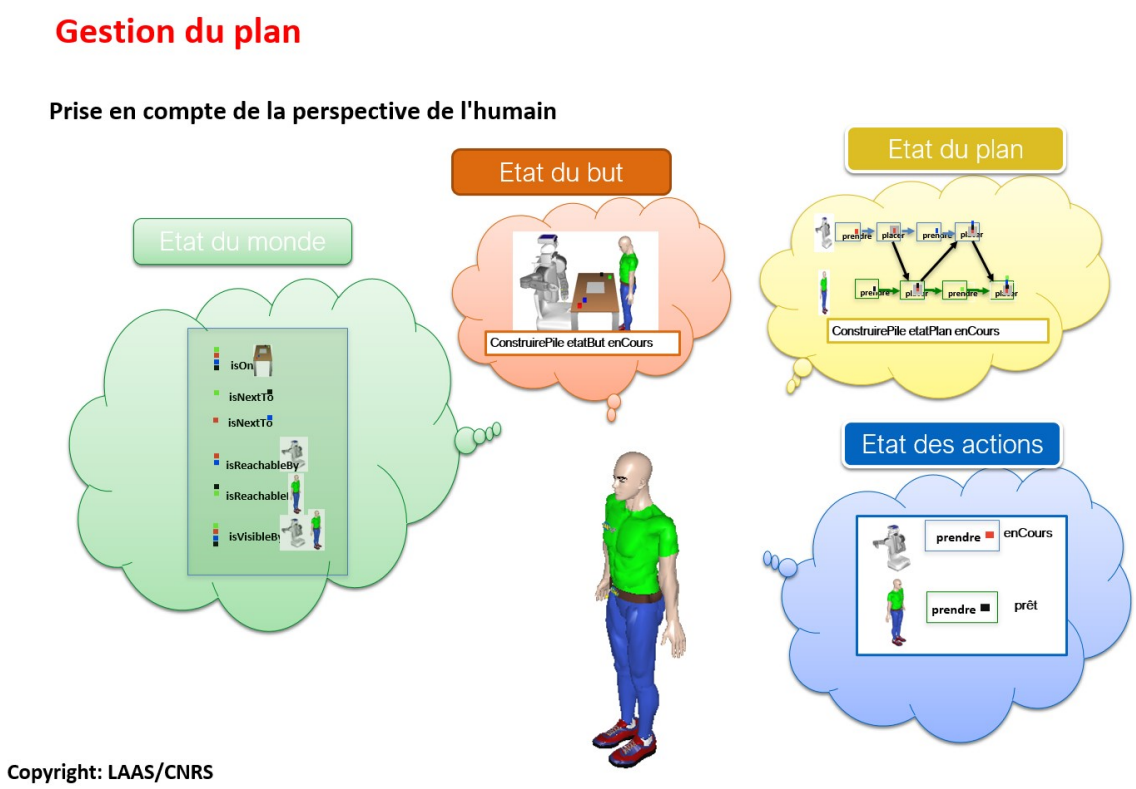


FIGURE 2.9 – Surveillance de l'exécution du plan : outre l'état du monde, le robot calcule l'état des objectifs à atteindre, l'état des plans en cours ainsi que de chaque action. Cela se fait non seulement de son point de vue, mais aussi du point de vue de l'humain.

Avec un système de planification comme HATP, nous avons un plan à la fois pour le robot et pour l'humain avec lequel il interagit, mais cela ne suffit pas. Si nous suivons l'idée de Knoblich et ses collègues [Knoblich 2011], les représentations partagées des tâches ne précisent pas seulement à l'avance quelles sont les tâches respectives de chacun des agents, elles fournissent également des structures de contrôle qui permettent aux agents de surveiller et de prédire ce que leurs partenaires font, permettant ainsi une coordination interpersonnelle en temps réel. Cela signifie que le robot n'a pas seulement besoin du plan, mais aussi des moyens de surveiller ce plan. Outre l'état du monde (cf. section 2.2 concernant l'évaluation de la situation) et le plan, nous avons développé un système de surveillance qui permet au robot d'inférer l'état du plan et l'état de l'action à la fois de son point de vue et du point de vue de l'humain comme illustré figure 2.9 [Devin 2016a, Devin 2017b]. Grâce à ces informations, le robot est capable d'adapter son exécution en temps réel. Par exemple, il peut y avoir un décalage entre l'état de l'action côté robot et côté humain (par exemple, le robot attend une action de l'humain). Équipé de cette surveillance, le robot peut détecter le problème et avertir. Le problème peut se situer au niveau de l'état du plan, par exemple, le robot considérant que le plan n'est plus réalisable alors qu'il détecte que l'humain continue d'agir.

2.6 Conclusion

Dans ce chapitre, nous avons présenté quatre processus de coordination nécessaires à la réalisation d'une action jointe.

L'exécution d'une action jointe nécessite non seulement que le robot puisse accomplir sa part dans la tâche mais qu'il l'accomplisse d'une manière compréhensible pour l'humain avec lequel il interagit (et qu'il tienne compte de la réaction de l'humain le cas échéant). Mélanger l'exécution et la surveillance nécessite de faire certains choix à un moment donné, par exemple, si la caméra est nécessaire pour effectuer une action, le robot ne peut pas l'utiliser pour surveiller l'humain s'il n'est pas dans le même champ de vision. Ces choix sont effectués par le système de supervision qui gère l'exécution globale des tâches depuis la planification des tâches jusqu'à l'exécution des actions de bas niveau.

Nous avons aussi parlé un peu de la façon dont l'humain gérait ces différents processus de coordination dans un cadre d'interaction humain-robot et du fait qu'il y avait encore une certaine incertitude à ce sujet.

2.7 Contributions associées

[Clodic 2017] Aurélie Clodic, Elisabeth Pacherie, Rachid Alami, and Raja Chatila. Key Elements for HumanRobot Joint Action. In Sociality and Normativity for Robots Philosophical Inquiries into Human-Robot Interactions, Studies in the Philosophy of Sociality, pages 159–177. Springer, 2017. doi : 10.1007/978-3-319-53133-5_8 URL : `https://hal.laas.fr/ijn_03084126/`

— Un dialogue à plusieurs voix pour analyser comment les théories de l'action jointe peuvent apporter au domaine de l'interaction humain-robot.

[Clodic 2021] Aurélie Clodic and Rachid Alami. What Is It to Implement a Human-Robot Joint Action ? In Robotics, AI, and Humanity, pages 229–238. Springer International Publishing, February 2021. DOI : 10.1007/978-3-030-54173-6_19 URL : `https://hal.laas.fr/hal-03184727v1`

— Suite à une invitation pour faire une présentation, j'ai essayé de mettre en perspective les travaux du groupe Robotique et InteractionS avec les processus décrits dans le cadre d'une action jointe. L'idée était en partant d'un exemple simple d'expliquer l'ensemble des éléments nécessaires au niveau implémentation.

Chapitre 3

Gérer ses connaissances, évaluer celles des autres

Les données et les connaissances que le robot doit manipuler dans un cadre d'interaction humain-robot sont de nature très diverses. Elles peuvent venir de sources différentes que ce soit une information brute de perception, un modèle 3D pour représenter l'environnement, une ontologie pour un domaine donnée récupérée sur le web, etc. Elles peuvent être traitées par le robot pour en faire une information compréhensible par l'humain avec lequel il interagit ou pour permettre la mise en route d'une action. Les données et les connaissances sont partout (la figure 3.1 représente quelques-unes de ces données), elles évoluent tout le temps et c'est un défi d'en garder une vision cohérente et juste ainsi que de les stocker et les mémoriser lorsque c'est nécessaire.

Dans ce chapitre, nous allons aborder la gestion des connaissances. Dans un premier temps, nous allons analyser comment acquérir et consolider ces connaissances (d'un point de vue géométrique principalement), ensuite nous expliquerons le système de gestion des connaissances sur lequel nous avons travaillé et comment exploiter cette connaissance. Enfin, nous analyserons la possibilité de mettre la connaissance au cœur du système pour faciliter la gestion des données.

3.1 Préambule

Commençons par réfléchir à ce que signifie l'interprétation des données et la signification que l'on va donner dans notre cas.

Si vous regardez les images de la figure 3.2, il existe de nombreuses interprétations possibles. Sur l'image de gauche, si l'on considère l'image renvoyée par la caméra du robot :

- avons-nous "seulement" accès à l'image ?
- avons-nous des informations sur chaque objet : une table, deux cubes (peut-être avec leurs caractéristiques comme un cube bleu et un cube jaune), une petite voiture rouge, un lego gris ?
- avons-nous accès aux relations entre chaque objet : tous les objets sont sur la table, le cube jaune est sur le cube bleu, les cubes sont à gauche de la voiture, la voiture est à gauche du lego gris ?
- avons-nous accès à l'information selon laquelle un cube est fixé à l'autre et que cela forme un nouvel objet ?
- avons-nous accès à l'information que ces objets sont visibles par le robot à un instant donné ?
- avons-nous accès à la relation liée aux capacités du robot : le robot peut attraper les deux objets cubes ?

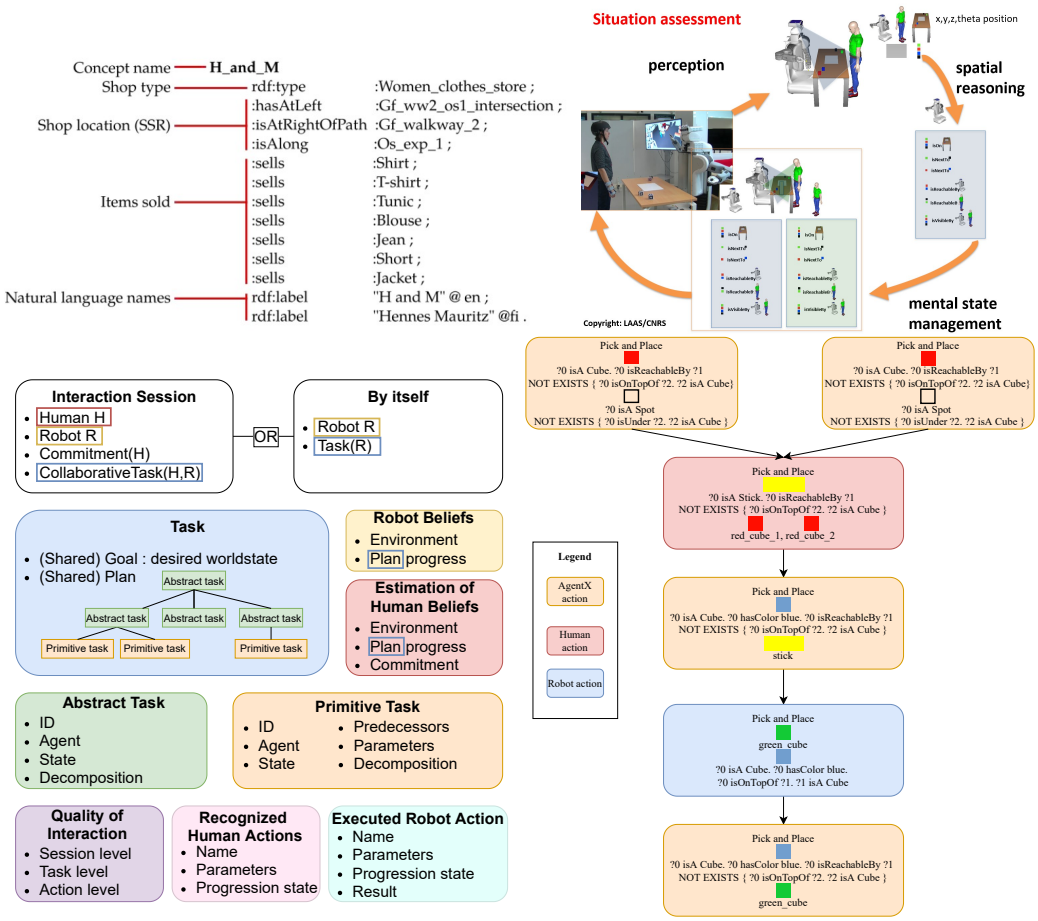


FIGURE 3.1 – Aperçu (non-exhaustif) des données manipulées (et de quelques structures associées) : en haut à gauche la représentation d'un magasin et des produits qu'il vend dans une ontologie [Sarthou 2021a] ; en haut à droite les données géométriques manipulées (les positions mais aussi les représentations 3D des différents éléments) [Milliez 2014] ; en bas à gauche les données manipulées par le système de supervision [Mayima 2021a] ; en bas à droite les données manipulées par la planification de tâches [Mayima 2021a].

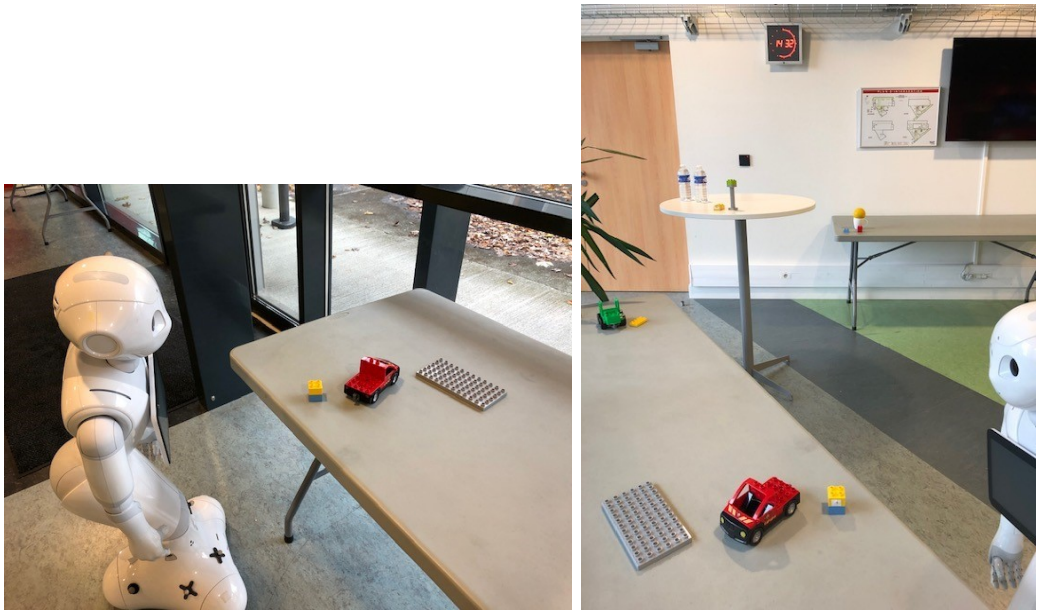


FIGURE 3.2 – Que signifie l'interprétation des données ? Quelles données sont acquises et que peut interpréter le robot dans ces différents cas ?

De la même façon, sur l'image de droite, il y a d'autres éléments de l'environnement qui ne sont pas visibles à un instant donné (mais peut-être déjà vus ?) comment stocke-t-on ces informations ? peut-être faudrait-il qu'on dise au robot de regarder à gauche ou à droite pour détecter ces autres éléments de son environnement. Comment gérer une question ou une demande à propos de la voiture verte ?

Nous allons expliquer dans ce chapitre une manière, notre manière, de modéliser les données et de les transformer en informations utilisables par nos systèmes. Notre manière de faire est une possibilité, il en existe d'autres bien entendu. Dans un cadre d'interaction, l'un des défis va consister également à mettre en cohérence les informations modélisées par le robot et les informations disponibles à la personne avec laquelle il interagit, et s'assurer que c'est le cas, c'est-à-dire assurer le *grounding* [Clark 1996a].

3.2 Acquérir et consolider les connaissances

Nous allons ici parler d'un travail débuté par les travaux de Amit Kumar Pandey, Matthieu Warnier et Rachid Alami, poursuivi par Grégoire Milliez et enfin par Séverin Lemaignan et Yoan Sallami.

L'ancrage des perceptions dans un modèle symbolique propice à la prise de décision nécessite des capacités de perception et leur interprétation symbolique. C'est ce que nous appelons évaluation de la situation. De nombreuses approches existent, comme les proxys amodaux [Jacobsson 2008], les représentations amodales ancrées [Mavridis 2006], les cartes sémantiques [Nüchter 2008, Galindo 2008, Blodow 2011] ou la planification et la classification d'objets basées sur les affordances [Lörken 2008, Varadarajan 2011].

Le premier rôle d'un système d'évaluation de la situation est de rendre disponible un état du monde afin de fournir une base au robot pour planifier, agir, réagir et interagir. Cet état du monde correspond à la position des différents agents, objets, éléments de l'environnement ; au calcul de faits symboliques tels que (objet estSur Table) ou (ob-

jet estAtteignablePar Robot) ou (objet estVisiblePar Robot) puisque c'est à ce niveau d'abstraction qu'est défini le domaine de planification et que va travailler le système de supervision.

Dans le cadre d'une interaction humain-robot, ce système va aussi avoir pour rôle de simuler la perspective de la personne avec qui le robot interagit pour in fine calculer le même type de faits - bel correspond à belief - (bel Robot (bel Human (objet estSur Table))) ou (bel Robot (bel Human (Human regarde Table))). Cela peut permettre de déterminer par exemple une différence entre les connaissances du robot et les connaissances supposées de la personne avec qui il interagit par exemple.

Au LAAS, plusieurs systèmes ont été développés à commencer par SPARK (SPAtial Reasoning and Knowledge) [Milliez 2014], puis TOASTER (Tracking Of Agents and Spatio-TEmporal Reasoning) [Milliez 2016a] et enfin plus récemment Underworlds [Lemaignan 2018, Sallami 2021].

Underworlds [Lemaignan 2018] permet, en temps réel, de construire et d'affiner des modèles spatio-temporels de l'environnement entourant le robot. Underworlds permet de partager ces modèles de monde entre les composants logiciels exécutés sur le robot. De plus, Underworlds permet aux utilisateurs de représenter et de manipuler de multiples alternatives au modèle du monde actuel perçu et ceci de manière distribuée. Par exemple, le monde avec certains objets filtrés ; le monde « vu » du point de vue d'un autre agent ; un monde hypothétique résultant de l'application simulée d'un plan, etc. Les figures 3.3 et 3.4 décrivent comment les mondes et les différents clients peuvent être définis et utilisés.

Underworlds nous a permis de développer un certain nombre d'expérimentations très intéressantes [Sallami 2019, Foster 2019] et est un système qui prend déjà en compte un certain nombre de contraintes de notre domaine. Cependant, on peut quand même citer quelques difficultés liées au développement et à l'utilisation d'un tel système et auxquelles nous serons attentifs pour la suite.

La fusion des données Bien entendu, la fusion de données est un domaine de recherche en soi et on ne va pas s'étendre, car cela sort de mon champ d'expertise, mais voici mon expérience avec l'utilisation de ce genre de système. Si vous percevez à un moment deux personnes et que vous vous apercevez par la suite que ce n'est qu'une seule et même personne, il faut que votre système soit capable de le prendre en compte. Idem pour les objets. Cela signifie que ce n'est pas une perception qui crée une instance, mais qu'une instance est créée à partir d'un certain nombre de suppositions. Idem avec les objets. C'est vraiment très compliqué dans un contexte d'interaction humain-robot, car déterminer la personne qui est en face ou l'objet précis qui est au centre de l'attention (et pas un humain ou un quelconque objet) est important. De même si l'on envisage de faire de la manipulation, il faut pouvoir avoir une position précise.

La gestion des ressources Une autre difficulté concerne la gestion de la ressource capteurs. Le robot est capable de percevoir en général avec les capteurs qui sont "sur sa tête" à la place de ses "yeux". Lorsque le robot a besoin d'attraper un objet, il va "tourner sa tête" vers cet objet, pour en obtenir une position plus précise. Le premier problème vient du fait qu'il ne faut pas prendre en compte les données pendant que la "tête" et donc que les capteurs qui y sont situés bougent, car ces données ne sont pas précises et peuvent venir fausser tout le reste. Il faut donc avoir cette possibilité de désactiver la prise en compte de l'acquisition à des moments précis. Le second problème vient du fait que

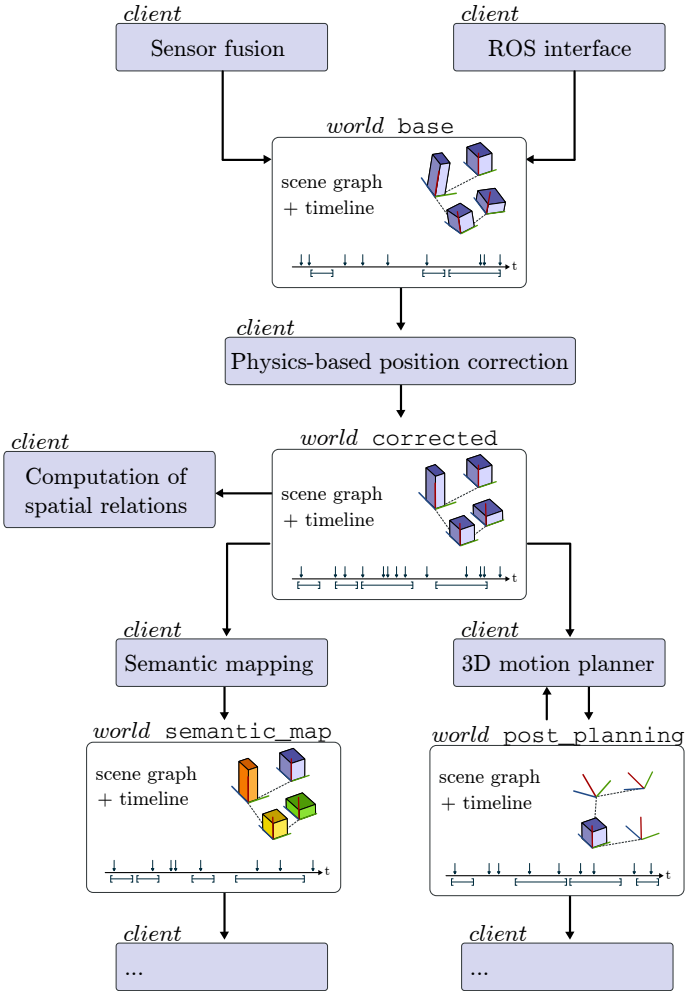


FIGURE 3.3 – Schéma d'un réseau possible : huit *clients* partagent des modèles d'environnement à travers quatre *mondes* indépendants (constitués de modèles spatiaux et temporels conjoints). Cette architecture permet des raffinements successifs et modulaires des modèles, adaptés aux besoins de chaque client [Lemaignan 2018]

lorsque la "tête" du robot s'est déplacée, l'ensemble des éléments qui sont perçus vont être modifiés (si le robot regardait la personne en face de lui en utilisant de la détection de visages, il va probablement ne plus la percevoir...).

Cela implique des choix. Tous ces choix doivent être documentés et discutés.

Parler de la même chose La définition que l'on peut donner aux choses peut être différente selon le niveau où l'on se place. Par exemple, si l'on calcule le fait (RedCube isVisibleBy Human), qu'est-ce que cela veut dire ? Est-ce que le cube est visible maintenant, est-il dans le champ de vue, a été vu, pendant combien de temps ? etc.

De même si l'on veut détecter l'attention de l'humain avec lequel on interagit, cela peut être fortement lié à la tâche et pas seulement au fait qu'il regarde le robot. L'humain qui regarde dans la direction où le robot guide lui dit de regarder est attentif...

Là encore, les choix faits doivent être documentés et discutés.

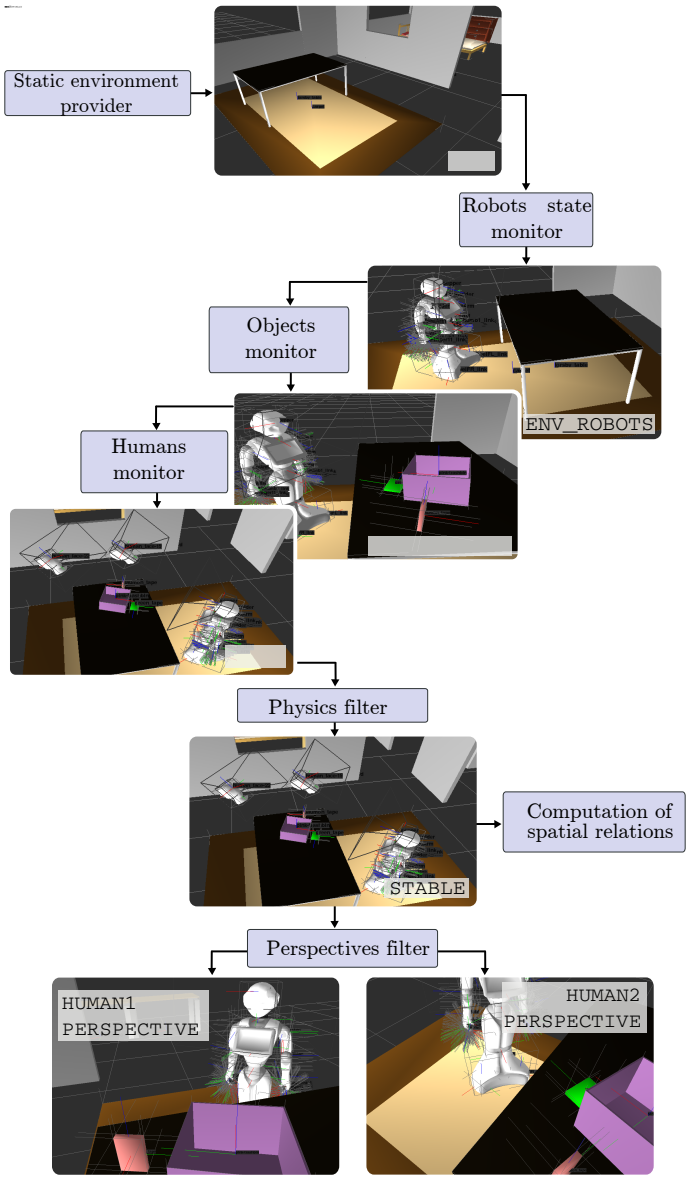


FIGURE 3.4 – Schéma d'une architecture Underworld. Les clients lisent et génèrent les mondes `ENV` → `ENV_ROBOT` → …→ `HUMAN*_PERSPECTIVE`. Les deux derniers mondes `HUMAN{1,2}_PERSPECTIVE` représentent la perspective visuelle immédiate de chacun des humains, ainsi que leurs perceptions visuelles passées. En tant que tels, ce sont les *mémoires visuelles* des humains, sur lesquelles le robot peut s'appuyer pour prendre des décisions [Lemaignan 2018]

La difficulté des hypothèses Il est souvent nécessaire de faire des hypothèses au sein de système d'évaluation de la situation. Ainsi si le robot a vu un objet sur la table à un moment donné et qu'il ne l'a pas vu ailleurs, on peut faire l'hypothèse que l'objet est toujours là. De même, il est possible d'utiliser un moteur physique.

Cependant, il faut là encore que ces hypothèses soient clairement définies et inscrites dans un contexte particulier (où elles forment un ensemble d'hypothèses raisonnable pour comprendre la situation actuelle). Idéalement, il faudrait pouvoir définir un ensemble d'hypothèses pour un contexte.

Là encore, les choix faits doivent être documentés et discutés.

Donner des éléments de configuration Le niveau décisionnel dispose d'un certain nombre d'informations qui peuvent influer sur l'évaluation de la situation. Il pourrait être utile de donner des moyens au système de supervision de venir configurer l'évaluation en fonction du contexte.

Lorsque l'on fait de la manipulation, on a besoin d'avoir une position très précise des objets, mais cela veut aussi dire qu'à d'autres moments "seulement" savoir que l'objet est sur la table suffit (sans qu'il soit nécessaire de calculer une position précise). Cela pourrait être un élément de configuration.

Lorsque l'on utilise de la reconnaissance de visages, dans certains contextes on pourrait indiquer quel ensemble de personnes on souhaite reconnaître (et pas tout ceux qui peuvent se trouver dans la base). Cela peut avoir une importance non seulement pour simplifier le travail de la reconnaissance de visages, mais également à plus long terme pour des raisons de confidentialité.

Toujours pour la reconnaissance de visages, le système de décision peut savoir à un moment donné qu'une personne ne va plus regarder le robot, typiquement quand un robot guide demande à la personne avec qui il interagit de regarder la direction qu'il pointe. Être capable de partager cette information pourrait être intéressant pour éviter au système de reconnaissance de continuer à chercher pendant ce temps-là.

Continuité dans les développements Le dernier souci est bien entendu la continuité dans les développements. Au jour d'aujourd'hui, nous n'avons plus de capacités de développement pour ce système au laboratoire.

3.3 Gérer les connaissances

Il n'y avait malheureusement pas eu de poursuite du travail réalisé par Séverin Lemaignan sur les ontologies avec ORO [Lemaignan 2010]. Nous avons eu l'occasion de renouveler ce travail avec la thèse de Guillaume Sarthou [Sarthou 2021a].

La première question que l'on peut se poser c'est pourquoi un système à base d'ontologie(s) est pertinent en robotique en général et dans le domaine de l'interaction humain-robot en particulier.

C'est en fait un constat partagé au niveau du domaine, ainsi Prasad et al. [Prasad 2020] indiquent qu'un graphe sémantique ou graphe de connaissances, est aujourd'hui fréquemment utilisé comme représentation des connaissances dans les applications robotiques (e.g. [Bálint-Benczédi 2018], [Tenorth 2017], [Paulius 2018]). Cela permet de représenter entre autres :

— les catégories d'entités à différents niveaux d'abstraction, par ex. une poignée est un objet physique
— les caractéristiques des entités, par ex. un frigo a une poignée
— la fonction ou le but, les affordances des entités, par ex. la poignée du frigo permet d'ouvrir le frigo
— la localisation d'une entité par rapport à une autre (c'est-à-dire les relations spatiales), par ex. la bouteille de lait est au frigo

L'utilisation des ontologies en robotique fait l'objet de trois types de contributions. La création d'ontologie, le stockage de ces ontologies et les mécanismes d'inférence associés, les frameworks utilisant les ontologies.

La création d'ontologies se concentre sur la définition du vocabulaire et des règles. Les ontologies créées sont souvent divisées en quatre catégories : niveau supérieur, référence, domaine et application. Les ontologies de niveau supérieur définissent des concepts largement applicables, transversaux à plusieurs disciplines. Les plus utilisées sont DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering) [Masolo 2003], Cyc [Lenat 1989], ou SUMO (Ontologie supérieure fusionnée suggérée) [Niles 2001]. Les ontologies de référence sont basées sur l'ontologie supérieure et décrivent le vocabulaire d'une discipline, comme l'ingénierie ou la médecine. Par exemple, [Schlenoff 2015] couvre la robotique et l'automatisation, décrivant le robot d'un point de vue technique, avec ses capteurs, ses processus et sa position. Les ontologies de domaine visent à affiner une discipline, en se concentrant sur un domaine plus restreint, comme l'ontologie SOMA [Beßler 2020] se concentrant sur la représentation des activités. Enfin, les ontologies d'applications étendent les ontologies de domaine pour des applications précises. La conception de l'ontologie utilise des langages tels que RDF (Resources Description Framework), FOL (First Order Logic) ou OWL (Web Ontology Language).

À l'aide de ces ontologies, un certain nombre de cadres ont été développés pour prendre en charge les processus de raisonnement de haut niveau dans les applications robotiques. Parmi eux, on trouve des systèmes à base d'ontologies utilisés pour la prise de décision, l'évaluation de la situation, la planification ou le maintien des croyances. Une revue complète de ces capacités de raisonnement et des cadres correspondants est disponible dans [Olivares-Alarcos 2019], on peut citer KnowRob [Tenorth 2013], ROSETTA [Stenmark 2013], CARESSES [Bruno 2017], RoboBrain [Saxena 2014], ou encore ORO [Lemaignan 2010].

Le dernier type de contribution est le stockage et l'inférence à base d'ontologies pour former une base de connaissances sur laquelle le cadre peut s'appuyer pour effectuer un raisonnement de plus haut niveau. KnowRob utilise Prolog [Wielemaker 2003] avec une bibliothèque pour gérer les triplets RDF. ROSETTA utilise Sesame pour stocker et interroger des triplets RDF [Broekstra 2002]. ORO utilise Jena en plus du raisonneur Pellet [Sirin 2007]. Enfin, RoboBrain utilise une base de données de graphes personnalisée.

En résumé, une Base de Connaissances, sous forme d'ontologie, peut être vue comme un graphe sémantique reposant sur une spécification formelle et explicite du sens partagé des concepts utilisés pour représenter une situation donnée. Afin d'être utilisée dans un framework pour effectuer un raisonnement de haut niveau, l'ontologie doit être stockée pour exposer les connaissances qu'elle contient à d'autres composants.

Aucune solution de stockage standard n'a émergé jusqu'à présent, car cela dépend principalement des besoins de l'application et c'est ce sur quoi nous allons revenir maintenant. Quelles sont les fonctionnalités souhaitées pour un système à base d'ontologies pour notre

domaine ?

Travailler en tant que serveur Dans une architecture robotique, on trouve différents composants qui acquièrent, traitent et produisent des données à différents niveaux d'abstraction comme nous l'avons vu avec la figure 3.1. Un problème commun avec une telle architecture est que chaque composant possède une partie des connaissances générales, ce qui peut conduire à des incohérences. Par conséquent, la base de connaissances que nous voulons doit être un serveur, accessible par requête et mise à jour par n'importe quel composant. Ainsi, elle peut fournir une connaissance uniforme à toute l'architecture.

Prise en charge de plusieurs instances Dans le domaine de l'interaction humain-robot, on veut pouvoir modéliser la connaissance du robot, mais aussi être capable de représenter une estimation des connaissances des autres agents. Certains systèmes le font en choisissant d'intégrer l'ensemble des données dans une seule base. Cependant, nous pensons que séparer les connaissances (et les estimations des connaissances) des différents agents dans des bases différentes peut faciliter la prise en charge de la mise à jour, la recherche de similarités ou de différences, etc. De plus, si l'on envisage aussi bien des interactions court-terme (par ex, d'un robot-employé avec un client dans un magasin) et des interactions long-terme (par ex, d'un robot-employé avec un autre employé du magasin), il faut être capable de pouvoir mémoriser facilement une interaction, mais peut-être aussi de l'effacer. La possibilité d'avoir plusieurs bases peut aussi permettre d'envisager de se projeter dans un état futur (e.g. quel sera l'état de ma base si je réalise ce plan) et de le comparer à l'état présent. Le système que nous avons développé permet de gérer plusieurs bases, mais pas encore de gérer les aspects mémorisation, projection et adéquation au contexte, mais ce sont des choses sur lesquelles nous souhaitons travailler pour la suite.

Interrogation au niveau sémantique Comme expliqué dans [Broekstra 2002], pour interroger des graphes RDF, trois niveaux sont possibles. Considérant une ontologie écrite en XML, le niveau syntaxique interrogerait la structure XML. Par conséquent, il interrogerait un arbre plutôt qu'un graphe. Une requête serait "donnez-moi les ressources imbriquées dans un élément Description ayant l'attribut about avec la valeur X". Il est donc dépendant de la langue et nécessite de connaître les mots clés utilisés. Le niveau structurel, tel que disponible dans [Lassila 1998], fournit une abstraction au langage. Il permet une requête du type "donnez-moi tous les éléments de type A" quelle que soit la manière dont il est représenté dans le langage. Cependant, le niveau structurel ne regarde que les triplets explicites. Si B est un sous-type de A et que l'élément x est décrit comme étant de type B, la demande des éléments de type A, x ne sera pas renvoyé. Le dernier niveau de requête est le niveau sémantique qui ne se limite pas aux connaissances explicites. À ce niveau, en demandant les éléments de type A, x serait renvoyé, car x est de type B et B est une spécialisation de A. Pour supporter ce niveau de requête, deux solutions peuvent être utilisées : calculer la fermeture du graphe (ajouter A comme type de x et le stocker), ou l'inférer à la requête. Alors que la première solution supprime une partie de la sémantique en aplatissant la hiérarchie (on perd le fait que x est de type A parce qu'il est de type B), la seconde peut être chronophage pour les requêtes.

Requêtes spécifiques Notre objectif est de fournir un accès fin aux connaissances stockées à haute fréquence. À la différence de Prolog qui possède son propre algorithme de

recherche, on préférera des requêtes de niveau inférieur telles que : « quels sont les types directs de X ? », « quelles sont les propriétés inverses de Y ? », « quelles propriétés lient C et D ? , ou « E est-il dans le domaine d'application de F ? », et cela au niveau sémantique. Néanmoins, pour les requêtes rapides, nous devons faire attention au nombre d'inférences nécessaires au moment de la requête.

Thread safe Étant donné que le logiciel doit fonctionner comme un serveur, nous souhaitons qu'il prenne en charge plusieurs requêtes en parallèle, mais qu'il soit sécurisé lors de la mise à jour.

Niveau d'expressivité Considérant le Web Ontology Language (OWL), il existe différents niveaux d'expressivité en fonction des besoins. La plus expressive est OWL-full, cependant, on dit qu'elle n'est pas computationnelle, ce qui signifie que les inférences qu'elle permet ne peuvent pas être résolues au moment de la requête. Ensuite, OWL-DL (Description Logic) prend en charge la hiérarchie des propriétés et des classes, la restriction de valeur d'énumération, les propriétés inverses ou la restriction de cardinalité. Enfin, OWL-lite est le moins expressif, ne prenant pas en charge la restriction de cardinalité et la restriction de valeur. Même si OWL-DL conviendrait, dans une première version OWL-lite pourrait suffire. C'est pourtant l'expressivité minimale à supporter dans son ensemble.

Raisonnement personnalisé À des fins de recherche, nous ne voulons pas nous limiter au raisonnement de la logique du premier ordre. Nous voulons pouvoir intégrer des processus de raisonnement dépendant de l'application ayant un accès direct à la structure interne. Une telle fonctionnalité pourrait être fournie par le support de plugins par exemple. L'avantage des plugins est que n'importe qui peut ajouter une capacité de raisonnement sans modifier le cœur du logiciel ou posséder une version personnalisée.

Au vu de ces fonctionnalités souhaitées, du nombre d'outils disponibles, plus maintenus ou sans documentation, et de notre besoin de pouvoir mettre en œuvre de nouvelles capacités en fonction de nos besoins de recherche évoluant dans le temps, Guillaume Sarthou a fait le choix de développer un nouveau logiciel à partir de zéro. Le logiciel résultant est Ontologenius `https://sarthou.github.io/ontologenius/`. Ontologenius s'inscrit dans la continuité du logiciel ORO mais en proposant une gestion multi-instances plus poussée et un système de requêtes plus adapté à l'intégration dans des algorithmes.

3.4 Exploiter les connaissances

L'une des premières utilisations d'Ontologenius a été une tâche de calcul d'itinéraires dans le cadre du projet MuMMER. Ce travail a été effectué avec Guillaume Sarthou et Rachid Alami et a fait l'objet de la publication Semantic Spatial Representation : a unique representation of an environment based on an ontology for robotic applications [Sarthou 2019a].

Équipé d'une représentation de son environnement, le robot doit trouver un chemin non pour lui-même, mais pour l'humain qui lui a demandé son chemin. Le chemin calculé doit être adapté aux connaissances et aux capacités de l'humain. Ainsi, dans un centre commercial, nous n'indiquerons pas un chemin avec un escalier à une personne à mobilité réduite ou à une personne munie d'un caddie. Une fois calculé, le robot doit expliquer

l'itinéraire. Là où les cartes interactives n'ont qu'à mettre en évidence un chemin, ici, le robot doit générer une phrase que l'humain pourra mémoriser. Bien sûr, le robot ne va pas guider l'humain avec des phrases du style "marcher 30 mètres, tourner à -90 degrés". Ce ne serait pas adapté. L'usage de l'orientation et de références aux éléments de l'environnement s'imposera à travers une phrase du type « marcher jusqu'au fleuriste puis tourner à gauche ». On veut ainsi que le robot génère des chemins compréhensibles et exécutables par l'humain. Pour cela avait été créée une représentation unifiée d'un environnement intérieur à l'aide d'une ontologie, pour inclure à la fois des connaissances topologiques et sémantiques. Puis, sur la base de cette représentation, nous avons élaboré un premier algorithme pour trouver un itinéraire adapté à expliquer à un humain et un deuxième algorithme pour le verbaliser. La figure 3.5 offre une illustration du système.

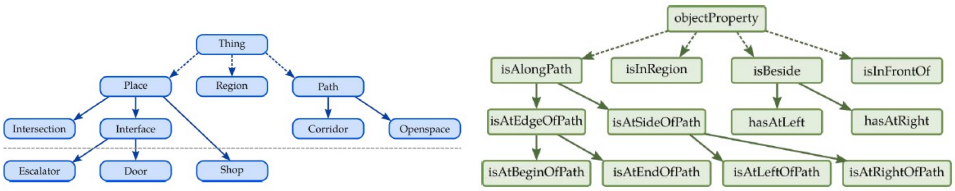
3.5 Partager la connaissance

Ce travail a été effectué avec Guilhem Buisan, Guillaume Sarthou, Arthur Bit-Monnot et Rachid Alami et fait l'objet de la publication Efficient, situated and ontology based referring expression generation for human-robot collaboration [Buisan 2020b].

Il ne suffit pas d'avoir toutes les informations, encore faut-il pouvoir les utiliser pour communiquer de manière efficace. Prenons l'exemple de la figure 3.6, imaginez que c'est votre point de vue et que vous êtes face au robot. Maintenant, considérons que le robot doit trouver le moyen de vous demander le stylo rouge qui n'est pas accessible par lui, mais qui est visible et accessible par vous. Lorsqu'un seul stylo est présent (Fig. 3.6 (a)), la référence est évidente à produire. Si par contre, il y a deux stylos dans deux pots de couleurs différentes (Fig. 3.6 (f)), le robot doit générer une expression faisant référence à l'un des pots pour se référer au stylo cible.

De plus, jusqu'à présent, nous considérions que le robot connaissait les notions de stylo et de boîte à crayons ainsi que leurs noms en langage naturel pour en parler. Cependant, si notre robot doit parler anglais et ne connaît pas la traduction de pot à crayons, il devra recourir à un terme plus générique, tel que « conteneur ». Cela peut soulever de nouvelles ambiguïtés qui devront être résolues, par ex. dans le cas où d'autres contenants sont visibles par l'humain. Le robot doit également faire attention aux relations qu'il utilise pour se référer à une entité. Par exemple, utiliser le poids exact d'un objet ne serait pas utile et utiliser la couleur ne serait pas la propriété la plus appropriée pour une personne daltonienne. Cela signifie que le robot ne doit utiliser que des relations connues et perçues par la personne avec qui il interagit. Cela se fait en prenant en compte les connaissances de l'autre et donc en générant l'expression à partir de l'estimation par le robot de la base de connaissances du partenaire humain.

Le processus sous-jacent à tous les exemples précédents est ce que l'on appelle communément une tâche de génération d'expression de référence *Referring Expression Generation* (REG). Elle est souvent composée de deux sous-tâches : la détermination du contenu (déterminer les relations à utiliser) et la réalisation linguistique (choisir les mots à utiliser pour communiquer le contenu) [Krahmer 2012]. Il est souvent impossible de considérer ces deux sous-tâches de manière totalement indépendante. En effet, la génération et la valorisation du contenu dépendent fortement des concepts utilisables en langage naturel. Cela peut être assuré en utilisant une base de connaissances dédiée, contenant uniquement des concepts verbalisables, mais de telles bases de connaissances peuvent être difficiles à maintenir pendant l'interaction. Dans un scénario coopératif, il faut également s'assurer



Concept name	**Burger_King**	
Shop type	rdf:type	:Fast_Food ;
Shop location (SSR)	:hasAtLeft	:escalator_kompassikatu ;
	:hasAtRight	:Rax ;
	:isAlong	:Keskuspuisto ;
Items sold	:sells	:Hamburger ;
	:sells	:Soda ;
	:sells	:Fries ;
	:sells	:Salad ;
Natural language names	rdf:label	"burger king" @ en ;
	rdf:label	"burger king" @fi .

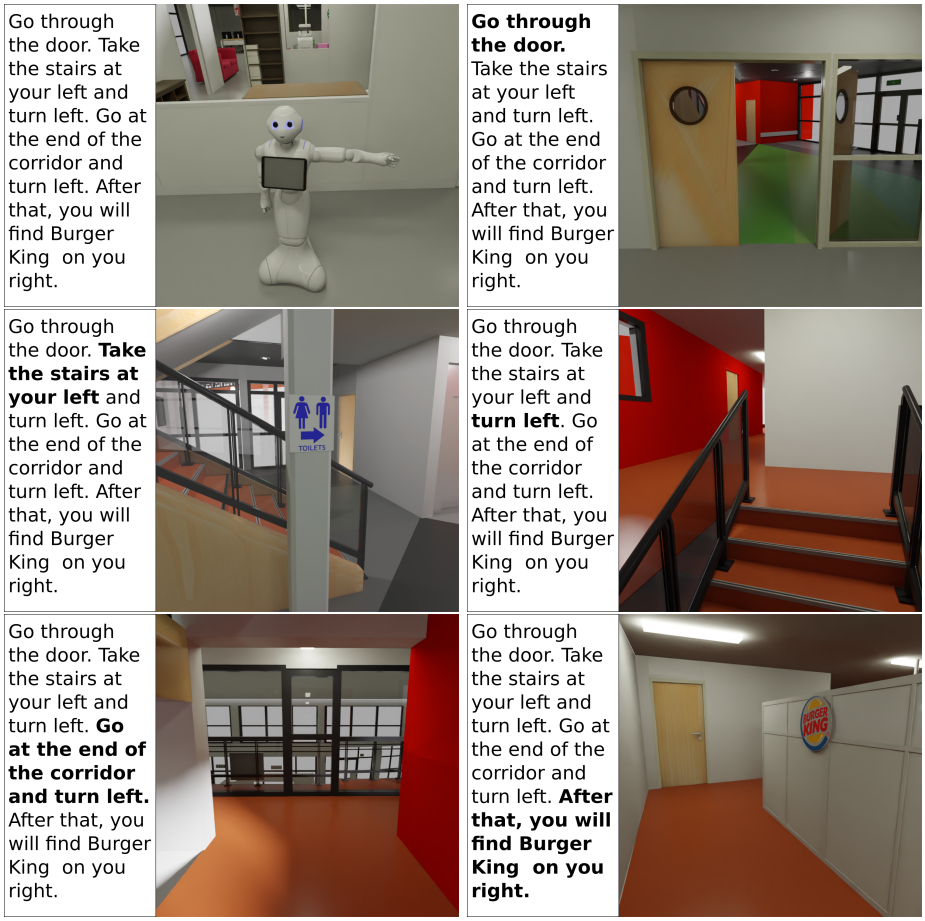


FIGURE 3.5 – Exemple d'utilisation d'Ontologenius pour le calcul de route et la Représentation Spatiale Sémantique associée. En haut, en bleu, on va retrouver les classes utilisées pour décrire la topologie d'un environnement intérieur. En haut, en vert, on va retrouver les propriétés définies. En dessous, on retrouve la représentation de la connaissance concernant un magasin stocké dans l'ontologie. On voit que cette représentation contient des informations sémantiques, mais aussi des informations topologiques.
En bas de la figure, on trouve un exemple de phrase générée si la personne demande d'aller au Burger King et une représentation du véritable chemin pour y arriver, montrant que la phrase générée permet bien d'atteindre le but.

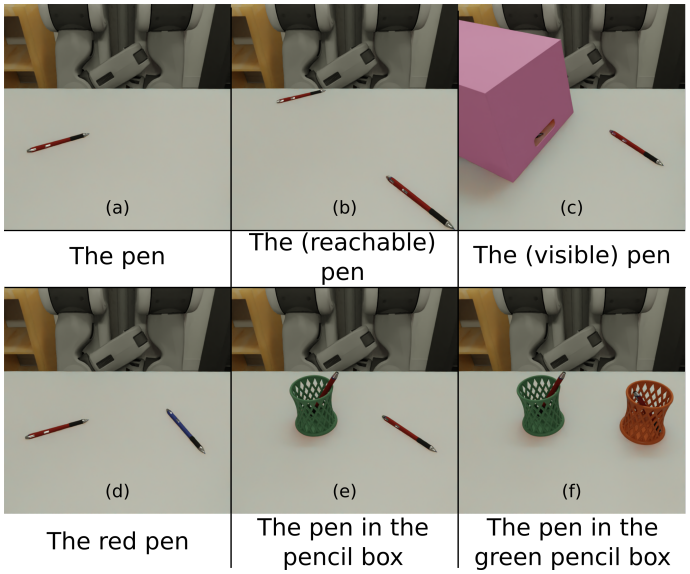


FIGURE 3.6 – Six situations où se référer à un stylo conduisent à différentes réponses pour faire face aux ambiguïtés. Les vues sont affichées du côté de l'humain, le robot est placé de l'autre côté de la table.

que l'expression de référence ne repose sur aucun fait inconnu du partenaire humain.

Nous ne développerons pas plus ici, mais nous avons proposé un algorithme basé sur l'ontologie et indépendant du domaine pour la génération d'expressions de référence pour apporter une réponse à ce problème. À l'aide d'une fonction de coût personnalisable estimant la charge cognitive nécessaire à un humain pour interpréter l'expression de référence, l'algorithme produit l'ensemble optimal d'assertions verbalisables permettant de discriminer sans ambiguïté une entité donnée référencée par le robot.

3.6 Conclusions

3.6.1 Explorer les connaissances

Nous n'avons fait que commencer à travailler sur la structuration de la base de connaissances, sur l'apport des ontologies, sur la nécessaire mise en place d'un processus de mémorisation etc. mais le travail déjà réalisé nous amène à faire les remarques suivantes :

- Écrire et manipuler des ontologies çà s'apprend.
  - Cela peut paraître simple, lorsque l'on ouvre un fichier owl de le modifier pour ajouter des propriétés ou des éléments dans l'ontologie mais c'est un leurre. Il faut être très attentif lorsque l'on fait évoluer une ontologie car on a vite fait de "casser" quelque chose et de perdre alors les capacités de raisonnement associées. Il faut apprendre, prendre le temps de faire les tutos de Protégé (`https://protege.stanford.edu/`) par exemple pour savoir ce qu'on fait.
- Mettre la connaissance au coeur du système
  - Comme nous l'avons vu, il y a énormément de connaissances autour du robot et c'est compliqué d'avoir une vue globale de cet ensemble. L'un des soucis auquel on est confronté, c'est que les fonctionnalités sont souvent implémentées de manière découplée, par des gens venant de communautés différentes (mo-

tion planning, vision, task planning) et (voulant souvent bien faire) prenant en compte par eux-mêmes un maximum de choses. Ainsi, de nombreux choix sont cachés et rendus indisponibles et surtout invisibles pour le système décisionnel. C'est, je pense, une erreur, un maximum de paramètres devrait être rendu disponible. Le système décisionnel pourrait alors faire des propositions (qui seraient ensuite prises en compte ou pas au niveau fonctionnel).
  - Nous sommes sur la première marche de ce travail de mise à disposition des connaissances pour une utilisation (raisonnée) et partagée et nous allons continuer d'y travailler car nous pensons que c'est un aspect clé pour une architecture cognitive.
- Apprentissage, données incomplètes et/ou incertaines
  - Nous n'avons pas abordé ici l'aspect apprentissage, ni la prise en compte de données incomplètes et/ou incertaines, mais bien entendu la connaissance doit pouvoir évoluer au cours du temps et l'on doit se poser la question du comment. Nous allons notamment envisager comment le dialogue peut permettre de faire évoluer la connaissance.

3.6.2 Aller plus loin dans l'exploitation de l'estimation des connaissances de l'autre

Alors que nous n'avons pas encore fini d'utiliser les possibilités actuelles d'Ontologenius, nous réfléchissons déjà aux éléments qu'il serait intéressant de pouvoir prendre en compte. Nous pensons, par exemple, que le lien entre le dialogue (et les structures de modélisation associées [Asher 2016, Asher 2020, Hunter 2018, Hunter ming]) et nos modèles d'interaction pourraient bénéficier l'un de l'autre. C'est ce que nous allons aborder dans le cadre du projet ANR ASTRID DISCUTER Dialogue Interactif Structuré, Consolidé et Unifié pour la réalisation de Tâches En Robotique (cf section 2.4).

Une partie des questions que nous espérons pouvoir étudier dans le cadre de cette collaboration sont les suivantes. Tout d'abord, être capable de déterminer quelle est l'information pertinente à un moment donné. Ainsi, si l'on envisage une base de connaissances conséquente, cela peut devenir d'une part compliqué de faire de l'inférence sur la base entière, mais cela n'a pas forcément de sens non plus. Il est important de pouvoir savoir quel est l'espace d'information pertinent à un moment donné et d'être capable de se focaliser sur cette information pertinente pour décider. On pourrait rapprocher cela de l'idée d'un processus d'attention non plus seulement perceptuelle, mais aussi conceptuelle. Ensuite, être capable de déterminer qu'une information est manquante, imprécise ou incertaine (et peut-être même être capable de déterminer et modéliser que quelqu'un d'autre peut avoir cette information). Enfin, être capable de générer le dialogue pour avoir cette information, de mettre à jour la base de connaissances aussi bien avec des actions ou par le dialogue et de discuter à propos d'un plan et si besoin le corriger ou en proposer des modifications. Pour permettre la prise en compte de tout cela, il faudra également être capable (entre autres) de déterminer ce qui fait partie du dialogue et ce qui n'en fait pas partie, être capable de prendre en compte l'aspect temporel et le fait que l'interaction a un but : la réalisation d'une tâche, être capable de remettre en cause une information et de savoir que sa valeur peut changer au cours du temps.

3.7 Contributions associées

3.7.1 Sur les aspects évaluation de la situation

[Sallami 2019] Yoan Sallami, Séverin Lemaignan, Aurélie Clodic, Rachid Alami. Simulation-based physics reasoning for consistent scene estimation in an HRI context. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Nov 2019, Macau, China. pp.7834-7841, 10.1109/IROS40897.2019.8968106 . hal-02925945

[Lemaignan 2018] Séverin Lemaignan, Yoan Sallami, Christopher Wallbridge, Aurélie Clodic, Tony Belpaeme, et al.. UNDERWORLDS : Cascading Situation Assessment for Robots. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018), Oct 2018, Madrid, Spain. 10.1109/IROS.2018.8594094 . hal-01943917

[Milliez 2014] Grégoire Milliez, Matthieu Warnier, Aurélie Clodic, Rachid Alami. A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. IEEE RO-MAN 2014, Aug 2014, Edinburgh, United Kingdom. pp.FrAT2.4, 10.1109/ROMAN.2014.6926399 . hal-01064546

3.7.2 Sur les aspects ontologie

[Sarthou 2021b] Guillaume Sarthou, Amandine Mayima, Guilhem Buisan, Kathleen Belhassein, Aurélie Clodic. The Director Task : a Psychology-Inspired Task to Assess Cognitive and Interactive Robot Architectures. 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2021), Aug 2021, Vancouver (Virtual), Canada. doi : 10.1109/RO-MAN50785.2021.9515543 . URL : `https://hal.science/hal-03327222/`

[Buisan 2020b] Guilhem Buisan, Guillaume Sarthou, Arthur Bit-Monnot, Aurélie Clodic, Rachid Alami. Efficient, Situated and Ontology based Referring Expression Generation for Human-Robot collaboration. The 29th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), Aug 2020, Naples (on line), Italy. pp.349-356, doi : 10.1109/RO-MAN47096.2020.9223485 . URL : `https://hal.science/hal-02922098v2`

[Sarthou 2019b] Guillaume Sarthou, Aurélie Clodic, Rachid Alami. Ontologenius : A long-term semantic memory for robotic agents. IEEE International Conference on Robot & Human Interactive Communication (IEEE Ro-MAN 2019), Oct 2019, New Delhi, India. URL : `https://hal.science/hal-02283342`

[Sarthou 2019a] Guillaume Sarthou, Rachid Alami, Aurélie Clodic. Semantic Spatial Representation : a unique representation of an environment based on an ontology for robotic applications. Combined Workshop on Spatial Language Understanding (SpLU) and Grounded Communication for Robotics (RoboNLP), Jun 2019, Minneapolis, United States. URL : `https://hal.science/hal-02177140`

[Clodic 2007b] Aurélie Clodic, Maxime Ransan, Rachid Alami, Vincent Montreuil. A management of mutual belief for human-robot interaction. 2007 IEEE International Conference on Systems, Man and Cybernetics, Oct 2007, Montreal, Canada. URL : `https://hal.science/hal-01977577`

Chapitre 4

Prendre des décisions et exécuter des actions en interaction

Nous avons abordé dans le chapitre 1 et le chapitre 2 les différents éléments qu'il fallait considérer dans le cadre de l'exécution d'une interaction humain-robot. Nous allons aborder ici comment ces différents éléments peuvent être mis en musique.

Les travaux présentés dans ce chapitre sont pour la plupart issus des travaux réalisés par Sandra Devin [Devin 2017a] et Amandine Mayima [Mayima 2021a].

4.1 Supervision et interaction

Mon travail de thèse portait sur la supervision, le titre de ma thèse était : Supervision pour un robot interactif : action et interaction pour un robot autonome en environnement humain [Clodic 2007a]. Je crois que je n'ai pas à faire beaucoup évoluer l'explication que je donnais à l'époque pour expliquer ce qu'est la supervision.

La réalisation d'une tâche en robotique est dépendante d'un contexte. Ce contexte et sa perpétuelle évolution nécessitent une adaptation constante du comportement du robot. Superviser une tâche robotique c'est potentiellement déterminer toutes les issues possibles des actions, tous les contextes envisageables et prévoir le traitement à réaliser dans chacun des cas. Le but est de rendre le robot autonome, i.e. si l'on demande au robot de réaliser une tâche, il l'exécute et entreprend si besoin toutes les procédures d'adaptation nécessaires. Le résultat pourra être un succès si la tâche à accomplir est terminée à l'issue des actions mises en route par le robot ou un échec si le robot conclut, selon ses capacités et l'état de l'environnement que la tâche est irréalisable ou encore si un cas non prévu ou ingérable se produit. Dans tous les cas, le système de supervision devra, dans la mesure du possible, s'assurer de laisser le système dans un état dénué de danger pour le robot lui-même et les personnes qui l'entourent, e.g. arrêter les mouvements d'un bras manipulateur et le remettre en position sûre, arrêter les moteurs et mettre les freins si l'on doit arrêter la base.
Lorsque la tâche est de nature collaborative, c'est-à-dire qu'elle engage non seulement le robot mais aussi une autre personne (ou même un autre robot), de nouvelles exigences apparaissent. En effet, à côté de la réalisation de la tâche elle-même, vont devoir se produire un ensemble d'actions reliées à la mise en place de la tâche, à sa conduite et à sa terminaison. Prenons par exemple la tâche GiveObject(ObjectName,PersonName), tâche pour laquelle le robot doit donner l'objet ObjectName à la personne PersonName se trouvant à sa proximité. Nous pouvons décrire cette tâche d'une manière "robot-centrique" où le robot va devoir suivre un ensemble prédéfini d'étapes : tendre l'objet à la personne, attendre que la personne prenne l'objet, remettre le bras à sa position initiale. Réalisant cette suite d'actions, la tâche GiveObject serait réalisée. Cependant si la tâche est définie

de cette manière, que se passe-t-il si l'humain ne suit pas exactement le protocole défini par l'enchaînement des tâches, e.g. si la personne ne prend pas l'objet, le robot attend-il ou abandonne-t-il la tâche ? Qui impose le rythme ? Comment l'humain est-il averti étape par étape des intentions du robot ? Le robot exhibe-t-il de telles intentions ? Dans ce cas, la tâche est réalisée en boucle ouverte sans que l'on donne au robot la possibilité d'avoir ou de percevoir un retour d'information de la part de l'humain et de manière symétrique de donner des informations en ce sens à l'humain. Ainsi, à côté de la gestion du processus d'évolution de la tâche, il faut gérer le processus d'évolution de l'interaction. Ces processus ne sont pas disjoints. Par exemple, prendre en compte la position, l'attitude de l'humain lors de la planification de la trajectoire du bras, et adapter constamment la vitesse et l'accélération lors de sa réalisation participent de ces deux composantes.
(...)
La théorie de l'intention jointe énonce que deux agents ont un but commun persistant (ou joint persistent goal) si ils ont la croyance mutuelle : que le but n'est pas encore réalisé, qu'ils souhaitent que le but soit réalisé, que tant qu'ils ne possèdent pas la croyance commune que le but est terminé (achieved), irréalisable (unachievable) ou n'est plus pertinent (irrelevant), le but ne peut être abandonné. Cette dernière condition indique qu'il est de la responsabilité des agents engagés dans le but d'assurer la communication pour atteindre la croyance commune concernant l'état de la tâche. Étant donné que nous ne pouvons nous engager que sur les actes effectués par le robot, i.e. nous ne pouvons nous engager sur l'exécution d'un acte par un autre agent, nous analyserons ce que le robot doit faire pour satisfaire cette définition. Concernant notre tâche GiveObject, le minimum de communication qu'il faut réaliser va consister à s'assurer que la personne est prête à interagir et à la prévenir lorsque la tâche est terminée. Ainsi lors de l'exécution d'une tâche en collaboration avec l'humain, le robot doit non seulement avoir les moyens de réaliser la tâche, mais aussi de gérer le flot d'interaction pour communiquer ses intentions et ses croyances et de monitorer l'activité de son partenaire. Ceci de manière analogue au cas du dialogue [Clark 1996a] où l'on peut distinguer le dialogue au sein de la tâche elle-même et la communication nécessaire à sa gestion. Communiquer c'est partager l'information, obtenir des croyances partagées va conduire à la communication par l'intermédiaire d'actes de communication. La communication va pouvoir se faire par le dialogue, i.e. la communication verbale, mais également par l'intermédiaire de toutes les possibilités de communication disponibles : gestes, mouvements, affichage, etc. De plus décrit par Clark [Clark 1996a], contribuer à la communication ne doit pas se restreindre à la negative evidence - évidence que nous n'avons pas été entendus ou compris, mais aussi à la positive evidence par l'obtention d'accusé de réception (acknowledgement), de (relevant next turn) ou par une attention continue (continued attention). La communication est un processus bilatéral. Dans notre cas, elle va considérer deux acteurs : le robot et l'humain. La communication peut être à l'instigation de l'un ou de l'autre de manière indifférente.
(...)
Le contexte de l'interaction humain-robot nécessite la prise en compte explicite de l'humain à tous les niveaux. Il va notamment impliquer des communications entre l'humain et le robot pour partager les connaissances concernant la tâche à réaliser, en cours de réalisation ou réalisée. Pour exécuter une tâche collaborative, notre système de supervision va devoir réaliser non seulement la tâche elle-même, mais devra également monitorer l'activité de l'humain pour pouvoir détecter l'apparition d'actes de communication et produire la réponse appropriée. Par exemple, dans notre tâche GiveObject, si lorsque le robot tend le

bras pour donner l'objet à l'humain, si celui-ci détourne la tête (e.g. commençant à parler à une autre personne) le robot doit pouvoir interpréter cette perte d'attention comme une communication (en l'occurrence une perte de communication) et la traduire par exemple comme une suspension de la tâche en cours et agir en conséquence en arrêtant le mouvement du bras tant que l'humain ne regarde pas en direction du robot. Il est donc nécessaire de prendre en compte l'humain au niveau décisionnel. Nous ne pouvons restreindre la communication à un schéma prédéfini et statique. Par exemple, la communication peut être plus ou moins équilibrée entre les interlocuteurs selon qu'il y ait ou non ou leader à la tâche, ou elle peut être plus ou moins nécessaire selon les connaissances qu'ils ont l'un de l'autre.
Et ainsi, dans ma thèse j'avais proposé :
la définition d'un tel schéma de manière adaptée à la tâche (et à chacune de ses sous-tâches), à l'agent avec lequel va avoir lieu la communication et au contexte courant de la tâche. Nous proposons un moyen de définir ces schémas de communication et nous proposons un schéma basé sur notre traduction de la notion de croyance mutuelle. Cependant, nous restreignons notre étude aux communications concernant le déroulement d'une tâche : sa mise en route, sa suspension, sa reprise, son arrêt. Notre système permet la gestion d'un ensemble de tâches, défini par un ensemble partiellement ordonné de sous-tâches auxquelles nous rattachons des schémas de communication adaptés.

Mise à part la manière originale de considérer la communication comme une action à part entière dans mon travail de thèse, chose que nous n'avons pas reconsidérée pour la suite (en tous les cas pas comme çà) et qu'il serait pertinent de réintroduire, le reste montre bien le cadre dans lequel vont s'introduire les travaux dont nous allons parler. L'interaction nécessite de doter le robot des capacités de prendre en compte l'autre.

4.2 Eléments nécessaires à un système de décision pour la collaboration

4.2.1 Faire groupe

Une interaction ne se résume pas à la réalisation d'une tâche. Par exemple, pour un robot guide dans un centre commercial, l'interaction commence lorsqu'une personne s'approche du robot, lui parle ou le regarde. Le robot peut alors engager l'interaction et proposer ses services. L'interaction peut donner lieu à plusieurs tâches réalisées les unes à la suite des autres. L'interaction se termine lorsque la personne part (en disant au revoir ou non).

Pour définir cette période pendant laquelle le robot et la personne sont en interaction, Amandine Mayima a proposé la notion de session d'interaction. Elle est divisée en trois parties, suivant la structure proposée par [Robinson 2012] : les salutations, le corps de l'interaction et les adieux. Premièrement, *les salutations* correspond à la période où un agent démarre une interaction en l'initiant avec un autre agent. La session d'interaction dure tant que l'interaction est maintenue soit par la conversation soit par l'exécution de tâches collaboratives qui correspondent au *corps de l'interaction*. Enfin elle se termine lorsqu'au moins un des agents impliqués dans l'interaction se dégage, soit en mettant fin brusquement à l'interaction, soit en fermant l'interaction comme décrit par [Schegloff 1973], cela correspond aux "au revoir".

La figure 4.1 représente la structure hiérarchique d'une session d'interaction.

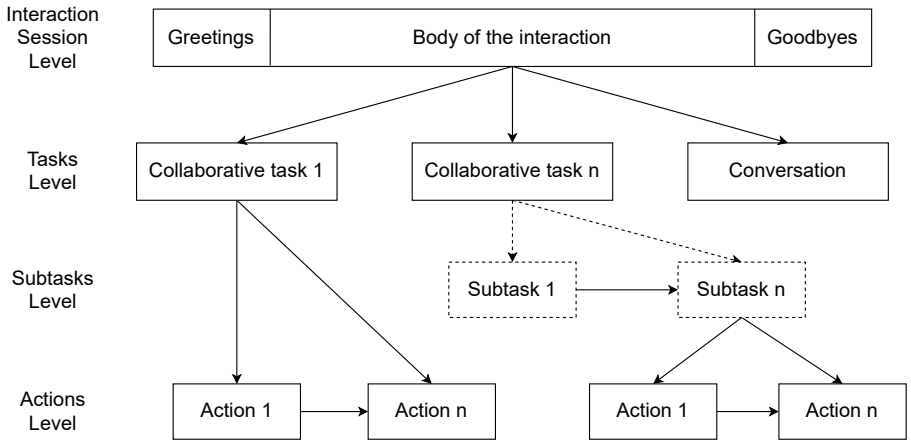


FIGURE 4.1 – La structure hiérarchique d'une session d'interaction. Le niveau le plus élevé est la session d'interaction. Le deuxième niveau est composé des tâches. Elles sont incluses dans le corps d'interaction de la session d'interaction. Deux types de tâches sont considérées et peuvent se chevaucher, les tâches collaboratives et conversationnelles. Avec cette représentation, une tâche peut être raffinée récursivement en sous-tâches jusqu'à atteindre le dernier niveau, le niveau des actions, qui est considéré comme atomique.

4.2.2 S'engager à faire ensemble

Cette notion de session d'interaction est reliée à la notion d'engagement. Sidner et al. proposent que "L'engagement est le processus par lequel deux (ou plusieurs) participants établissent, maintiennent et mettent fin à leur connexion perçue lors d'interactions qu'ils entreprennent conjointement" [Sidner 2003]. Le robot doit être capable de manifester son engagement (ainsi que son désengagement) et aussi de les évaluer par rapport à son partenaire humain.

On peut définir plusieurs états possibles au sein d'une session d'interaction pour la partie corps de l'interaction, e.g. :

- conversation : un bavardage social ou une négociation d'objectif, sans aucune action physique effectuée à l'exception des gestes de communication
- tâche collaborative : les deux agents exécutent des actions afin d'atteindre un objectif commun
- phases d'inactivité : les agents ne discutent pas ou n'exécutent pas une tâche collaborative ensemble, mais restent engagés dans la session d'interaction, cela se produit entre les phases d'interaction actives

Pour chacun de ces trois états, la manière de manifester l'engagement varie (par exemple, dans une conversation, un agent regardant son partenaire affiche son engagement ; lors d'une tâche, un agent exécutant correctement son action est une manière de démontrer son engagement). C'est pourquoi il est nécessaire de définir quel comportement le robot doit présenter dans chaque état et quel comportement il doit attendre de l'humain dans chaque état, car ces comportements sont généralement très spécifiques (par exemple, dans une tâche de guidage, le robot garde la tête orientée vers le visage de son partenaire pour démontrer son engagement dans la conversation et les contextes inactifs et lorsqu'il donne une direction, il s'attend à ce que l'humain regarde la direction qu'il montre ; dans une tâche de réalisation d'une pile de cubes, lorsque le robot donne une instruction il

attend de l'humain qu'il prenne le cube donné).

4.2.3 Exécuter et décider ensemble

Dans la suite, nous considérons que l'humain et le robot disposent d'un plan partagé. Pour générer ce plan partagé, au LAAS, nous utilisons le planificateur HATP Human-Aware Task Planner [Montreuil 2007, Alili 2009b, Silva 2015, Lallement 2018]. Ce planificateur renvoie un plan partagé avec d'une part les actions du robot et d'autre part les actions de l'humain pour pouvoir atteindre le but. La figure 4.4 page 57 montre un exemple de plan. Le système planifie donc pour deux, le robot et la personne avec laquelle il interagit.

Nous avons déjà évoqué dans le chapitre 3 page 31, la modélisation par le robot des connaissances que la personne a concernant son environnement. Lorsqu'il s'agit d'exécuter une action, il s'agit également de modéliser les connaissances des autres agents à propos du but, de l'objectif courant, de l'état du plan et/ou du monde. Sandra Devin [Devin 2017a] a proposé une modélisation des connaissances sur le plan et sur l'avancée dans l'exécution des actions. Le robot maintient non seulement sa propre connaissance sur l'avancement du plan commun et de ses actions, mais également une estimation de la connaissance de l'humain sur l'avancement du plan commun et de ses propres actions. Ceci est bien entendu limité aux éléments perceptibles par le robot ou qu'il est possible d'acquérir par le dialogue. Le système considère que les deux agents ont connaissance du plan partagé. Nous n'allons pas expliquer ici l'ensemble des processus permettant d'effectuer ces calculs, mais ce qu'ils permettent :

Achever un but en commun : Si le robot arrive à la conclusion que le but courant de la tâche est atteint, ou est devenu inatteignable ou annulé et qu'il a calculé que l'humain n'avait pas la même information, il peut se donner pour but que l'humain obtienne cette information (si l'on suit la définition de *weak achievement goal* de Cohen Levesque [Cohen 1991]).

Avant une action de l'humain : Le robot peut aussi s'assurer que l'humain a l'information quand il doit réaliser une action. Pour cela, à chaque fois que le robot considère que l'humain doit réaliser une action, il vérifie que l'humain a bien l'information. Il peut y avoir plusieurs raisons au fait que le robot et l'humain n'aient pas la même information concernant l'action à réaliser. Soit l'humain manque d'informations sur les actions réalisées antérieurement pour savoir que c'est son action qui doit être réalisée maintenant, le robot peut alors vérifier son estimation de la connaissance de l'humain par rapport à ces actions passées et identifier si il y a des différences entre ses connaissances et son estimation des connaissances de l'humain et donner l'information correspondante à l'humain. Soit l'humain manque d'informations sur l'état du monde pour savoir que son action est désormais possible. Dans ce cas, le robot peut vérifier les conditions préalables à l'action et vérifier quelle information est incorrecte ou manquante.

Prévenir les erreurs : Il peut arriver que l'humain fasse une action alors que ce n'était pas l'action prévue dans le plan à ce moment-là. Pour éviter cela, le robot peut vérifier que ses connaissances sont cohérentes avec son estimation des connaissances de l'humain. Il peut arriver que l'humain pense que l'action précédent sa propre action a été réalisée avec succès alors que ce n'est pas le cas. Il peut arriver que l'humain ait une croyance erronée sur l'état du monde, car il n'a pas perçu quelque

chose et là encore le robot peut l'estimer. S'il s'aperçoit d'une divergence, le robot peut être pro-actif et la communiquer à l'humain.

Signaler ses actions : Lorsque le robot est sur le point d'effectuer une action, il peut calculer si sur la base de ses connaissances l'humain est au courant de l'imminence de son action. Si l'humain n'est pas au courant, le robot peut alors signaler qu'il va agir.

Inaction et incertitude : Même si le robot estime que l'humain doit agir et qu'il le sait, il est possible que l'humain n'exécute toujours pas son ou ses actions. Le robot est capable d'estimer si l'humain est déjà occupé ou s'il n'est pas actuellement engagé dans la tâche, et dans ce cas il attend que l'humain soit disponible. Si ce n'est pas le cas, le robot considère que son estimation de la connaissance de l'humain sur la tâche est peut-être erronée et lui demande spécifiquement d'effectuer l'action. Si l'humain n'agit toujours pas, le robot considère que l'action a échoué, abandonne le plan en cours et essaie de trouver un plan alternatif excluant cette action.

Avec ces éléments, le robot est capable de communiquer et d'agir efficacement dans le cadre de l'exécution d'un plan partagé calculé précédemment comme ceux proposés par HATP.

Cependant, utiliser un tel plan a aussi des limites dans un cadre d'interaction. En effet, plutôt qu'un plan où toutes les actions et leurs paramètres sont calculés dès le départ, on pourrait vouloir laisser certaines décisions lors de l'exécution. Pour cela, Sandra Devin a travaillé à permettre la création d'un plan gardant une certaine souplesse pour pouvoir prendre des décisions au plus près de l'exécution et plus précisément permettre de répondre à deux questions au plus près de la réalisation de l'action :

1 *Qui doit effectuer une action ?* Cette décision est quelquefois implicite quand une seule personne est capable d'exécuter une action, mais peut, dans certains cas, demander une négociation ou une adaptation de la part du robot. Dans les versions précédentes d'HATP, toutes ces décisions étaient prises à l'élaboration du plan. Ici on rend possible le report de la décision à l'exécution quand elle n'est pas implicite afin de gagner en fluidité et en adaptabilité, en attendant, l'action est allouée à **l'agent X**. C'est-à-dire que lors de la planification on considère que cette action peut être réalisée par l'un ou l'autre des agents - sans choisir.
   - A l'exécution, quand le robot doit choisir qui doit réaliser une action allouée à **l'agent X**, il vérifiera dans un premier temps quels sont les agents réellement disponibles pour effectuer cette action. Si le robot est le seul à pouvoir réaliser l'action (e.g. la personne est déjà en train d'effectuer une autre action), il prendra l'initiative de réaliser l'action. Si les deux peuvent réaliser l'action, le robot aura alors deux différents modes possibles :
     - Le mode négociation : où le robot demande à l'humain si il veut réaliser l'action et agit (ou non) en fonction de sa réponse.
     - Le mode adaptation : où le robot attend un temps court de voir si l'humain prend l'initiative de réaliser l'action, et, si ce n'est pas le cas, prend lui-même l'initiative de la réaliser.

     Une fois une action allouée par le robot, il calcule un nouveau plan pour vérifier que cette allocation n'a pas d'autres implications dans le plan.

2 *Avec quels objets effectuer une action ?* Il peut arriver que deux objets soient fonctionnellement équivalents dans le cadre de la tâche. Dans ce cas, le robot prenait auparavant à l'élaboration du plan une décision arbitraire quant à l'objet à utiliser

durant une action. Ici on rend possible le report de cette décision à l'exécution, là encore pour gagner en fluidité et en adaptabilité.

4.2.4 Évaluer l'interaction

Nous allons étudier ici la possibilité pour le robot, pendant l'exécution de sa tâche, d'évaluer si l'interaction se passe bien ou pas. L'idée est que le robot dispose d'un moyen d'estimer comment se déroule l'interaction. Nous nommons cette capacité "la mesure de la Qualité d'Interaction du point de vue du robot" (Quality of Interaction (QoI)). Ceci est un travail réalisé lors de la thèse d'Amandine Mayima [Mayima 2021a].

Si on dispose d'une telle mesure et que l'on considère qu'une QoI élevée est le signe que l'interaction se passe bien alors qu'une QoI basse est le signe que l'interaction ne se déroule pas bien. Alors par exemple, si le robot détecte que la QoI commence à baisser, il peut prendre une décision basée sur cette information et agir pour essayer d'améliorer la qualité de l'interaction (e.g. il peut choisir de changer certaines modalités comme la langue dans laquelle il communique avec l'humain, le volume de ses haut-parleurs, ou les paramètres de ses planificateurs). Au contraire, lorsque la QoI est élevée, le robot peut décider de continuer l'interaction comme prévu même si peut-être le plan suivi n'est pas optimal (e.g. la trajectoire choisie peut ne plus être la plus adéquate, mais si tout se passe bien peut-être est-il préférable de ne pas en changer et de risquer de modifier la qualité d'interaction). Bien entendu, les décisions doivent être pondérées, mais nous pensons que donner au robot la possibilité d'évaluer lui-même, en temps réel, sa perception de la qualité de l'interaction qui est en train de se dérouler peut aider et améliorer ses capacités décisionnelles. De plus, au long cours, cela peut permettre d'analyser si certains comportements conduisent invariablement à une baisse de la qualité d'interaction et donc remettre en cause certains choix (cela pouvant se faire off-line). Nous avons pour le moment travaillé à mettre en place une mesure de la QoI et sommes en train d'analyser plus finement comment l'utiliser efficacement au niveau décisionnel.

La question est : sur quelle base mesurer la QoI ? quels sont les éléments à prendre en compte ? Nous avons pour le moment défini un certain nombre de métriques et nous sommes focalisés sur deux concepts : la mesure de l'engagement de l'humain et la mesure de l'efficacité de l'exécution des tâches collaboratives. Cependant, ceci n'est en aucun cas exhaustif, et d'autres métriques et paramètres pourraient (et devraient) être ajoutés ultérieurement. Notre travail peut être vu comme une boîte à outils parmi lesquels il est possible de piocher les métriques souhaitées en fonction des tâches ou des contextes. Nous proposons un moyen d'agréger ces métriques, produisant la QoI. L'évaluation de la QoI est effectuée à trois niveaux d'abstraction différents : le niveau de la session d'interaction, le niveau de la tâche et le niveau de l'action. Le tableau 4.1 montre un exemple des différentes métriques que nous avons définies.

		Metric names	Measures	Illustration	Session	Task	Action
Effectiveness	Progress towards goal	Distance-to-Goal	Geometric distance	Planned / Execution		x	x
		Time-to-Goal	Time	Planned / 5 sec / Execution / 5 sec		x	x
		Steps-to-Goal	Number of executed actions/substasks	Planned / Execution t=5 sec / t=8 sec		x	
	Deviation from standard duration		Time	5 Standard / 10 Over standard duration		x	x
Engagement	Fulfilling robot expectations about social interaction		*e.g.* attention ratio, with-meness,...		x	x	x
	Human contribution to the goal		*e.g.* number of repeated instructions, number of successful human actions,...	x ? / x ? / x ?		x	x

TABLE 4.1 – Tableau résumant les métriques définies

4.2.5 Communiquer avec

La communication est un processus un peu à part et pourtant central. Par exemple, dans sa thèse Amandine Mayima [Mayima 2021a] explique les différents types de communication verbale que son système est capable de gérer :

1. Ouverture et fermeture de la session d'interaction (par exemple, "Bonjour"),
2. Donner des informations sur ses capacités/rôle (par exemple, "Je suis là pour vous aider à trouver votre chemin"),
3. Initier un objectif partagé (par exemple, "Construisons cette pile de cubes."),
4. Donner des informations sur la tâche en cours (par exemple, "J'ai nettoyé la table."),
5. Demander des informations sur la tâche en cours (par exemple, "Avez-vous nettoyé la table ?"),
6. Demander à l'humain d'effectuer une action (par exemple, "Pouvez-vous nettoyer la table ?"),
7. Abandonner l'objectif commun (par exemple, "Mon action a échoué, j'abandonne la tâche.")

On le voit, ces différents types de communication ont différents buts, certaines font intégralement partie de la tâche ou de l'action en cours (e.g. "Où désirez-vous aller ?"), d'autres sont plus générales et sont en lien avec le déroulement général d'une tâche en commun (e.g. "J'ai réalisé mon action, c'est à votre tour"), d'autres ont à voir avec la gestion de l'interaction elle-même (e.g. "Bonjour"). Avoir une gestion cohérente de la communication dans le cadre d'une interaction humain-robot est un défi en soi.

L'une des difficultés provient aussi du fait qu'une communication, ce n'est pas forcément une communication verbale, un geste peut aussi avoir un rôle dans la communication. C'est quelque chose que nous avions envisagé dans ma thèse en intégrant l'action et la communication au sein de la définition des tâches [Clodic 2007a]. Si l'on reprend le célèbre adage de Paul Watzlawick, "On ne peut pas ne pas communiquer" et dans le cas d'un robot c'est aussi vrai et le fait que l'ensemble des fonctionnalités d'un robot soit découpé en une multitude de disciplines est là encore un élément à prendre en compte. Par exemple, lorsque l'on va calculer pour le robot un mouvement pour donner un objet à une personne, on va faire appel à la planification de mouvement qui va nous renvoyer un mouvement pour le bras du robot. Cependant, pour toute plateforme comme Pepper ou PR2 équipée d'une "tête", si le mouvement du bras est correct, mais que la tête du robot n'agit pas en cohérence (e.g. le robot regarde à l'opposé) la personne en face va être décontenancée. Idéalement, il faudrait que les algorithmes de planification de mouvement pour le bras prennent en compte aussi le mouvement de la tête. Par exemple, nous avions mené une petite étude pour analyser où devait regarder la tête dans le cadre d'un geste de don [Gharbi 2015] et avions trouvé un modèle où à la fin du mouvement, le robot devrait regarder l'objet puis la personne à qui l'objet était destiné pour que la personne en face considère le geste plus écologique.

Dans le cadre du projet JointAction4HRI (cf section 2.1), nous avons étudié comment les signaux de communication pouvaient être utilisés pour améliorer l'interaction. Dans Addressing joint action challenges in HRI : Insights from psychology and philosophy [Belhassein 2022], nous relevions notamment l'importance de la reconnaissance mutuelle et le fait de considérer l'ensemble du robot quand on considère une tâche. Nous reviendrons plus avant sur ces aspects dans le chapitre 6 où nous aborderons la prise en compte du contexte social en robotique.

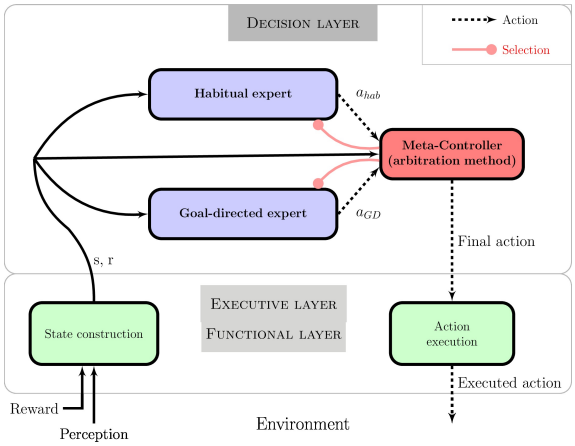


Figure 4.2 – Architecture du système développé : Dans cette figure, (Habitual expert = model free expert = système d'apprentissage) et (Goal-directed expert = model-based = HATP).

4.2.6 Apprendre pendant une interaction

Ceci est un travail qui a été réalisé dans le cadre du projet ANR Roboergosum avec Sandra Devin au LAAS et Erwan Renaudo de l'ISIR pour les aspects expérimentations dont je vais parler ici et qui sont un peu plus détaillés dans l'article Learning to interact with humans using goal-directed and habitual behaviors [Renaudo 2015] ainsi que dans les thèses d'Erwan Renaudo [Renaudo 2016] et Sandra Devin [Devin 2017a]. Les autres partenaires du projet sont signataires de l'article issu du projet Toward Self-Aware Robots [Chatila 2018].

La question ici est la suivante, comment est-il possible d'envisager l'utilisation d'apprentissage et ici, de l'apprentissage par renforcement [Renaudo 2014] [Sutton 2018] dans un cadre d'interaction humain-robot. Si l'on se place au niveau décisionnel, le robot en phase d'apprentissage va tester l'ensemble des actions disponibles et il "suffit" alors de trouver la bonne fonction de récompense pour atteindre le but. Cependant, dans un cadre de collaboration, il est contre-intuitif que le robot commence par faire n'importe quelle action disponible (potentiellement "n'importe quoi"), même si c'est "pour apprendre". Imaginez-vous dans la même situation, quelqu'un en face de vous fait n'importe quoi, comment répondre ? Comment mettre en place une interaction ? Il paraît alors évident que l'apprentissage d'une tâche en interaction ne peut pas partir de zéro.

Dans ce projet, ce que nous avons fait c'est utiliser notre système de planification HATP [Alili 2009b] comme amorce du processus d'apprentissage. En effet, avec HATP on dispose d'un domaine réalisé par un expert et qui permet de générer les différentes solutions possibles tant du point de vue du robot que du point de vue de l'humain. L'idée était d'utiliser ce plan HATP comme un fil d'Ariane, comme une base logique de ce qu'il faut réaliser. Une fois la "manière de faire" d'HATP apprise, on pouvait laisser le système d'apprentissage faire des digressions pour se sortir de situations qui n'étaient pas prévues ou pour apprendre de nouvelles possibilités.

Pour réaliser cela, nous avons mis en compétition chacune de nos briques logicielles (les actions disponibles étaient identiques) :

HATP (pour le LAAS) : un planificateur HTN, disposant d'un modèle du monde et

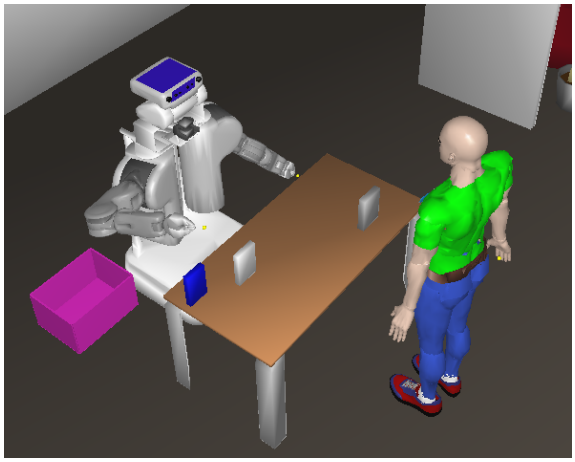


FIGURE 4.3 – Représentation de la simulation réalisée : un ensemble de cassettes est sur la table, certaines sont accessibles à l'agent humain, d'autres à l'agent robot, certaines aux deux. Les deux agents doivent nettoyer la table et à la fin de la tâche, toutes les cassettes doivent être déposées dans une poubelle (accessible à l'un ou l'autre des protagonistes).

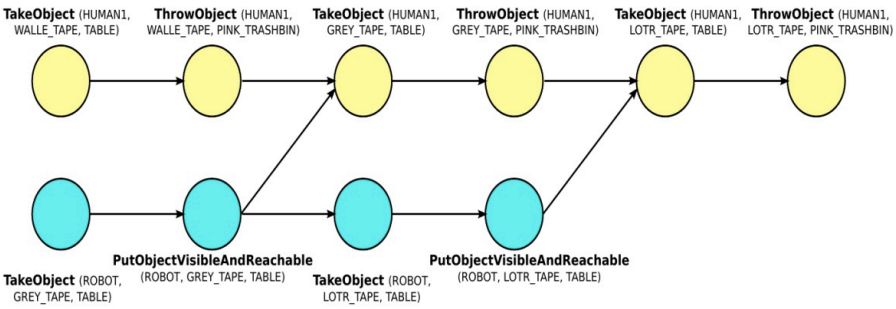


FIGURE 4.4 – Exemple de plan HATP produit pour cette tâche

capable de calculer un plan à partir de l'état courant fourni par le système (lorsque l'état initial le permet) comprenant l'ensemble des actions pour réaliser la tâche, cependant dans notre cas, HATP ne fournit que la première action du plan [Alili 2009b],

un algorithme d'apprentissage model-free (pour l'ISIR) : capable également de proposer la prochaine action à exécuter [Renaudo 2014].

A chaque itération, le système a donc accès à deux actions possibles. Pour choisir entre ces deux actions, nous avons conçu un MetaContrôleur qui choisit entre les deux options à chaque pas d'exécution selon le schéma présenté figure 4.2.

Notre intuition était que les deux experts pouvaient avoir des rôles complémentaires :

- HATP :
  - — dispose de connaissances à priori, stockées dans son modèle, qui lui permet de fournir un plan complet, il ne nécessite pas de phase d'apprentissage, mais il se peut qu'il soit incapable de proposer une nouvelle action (si son modèle du monde lui indique qu'aucun plan ne permet de réaliser la tâche)
  - — le temps de calcul d'un plan est fonction du nombre d'états et d'actions envisagées
  - — le système calcule un plan complet jusqu'au but
- l'algorithme d'apprentissage model-free :
  - — nécessite une phase d'apprentissage qui peut être longue, mais proposera toujours une action à réaliser.
  - — il ne vérifie pas si l'action est possible, il l'essaie

  => le MetaContrôleur pourrait guider l'apprentissage de l'algorithme model-free en se basant sur les actions proposées par HATP. Une fois cet apprentissage réalisé, l'algorithme model-free pourrait être privilégié par le MetaContrôleur.

Nous avons envisagé trois cas particuliers pour lequel la complémentarité de nos systèmes pourrait s'avérer pertinente :

Cas 1 : L'environnement ne change pas, l'espace d'état permet au planificateur HATP de trouver un plan valide. Dans ce cas, HATP est l'Expert qui est le mieux à même de résoudre la tâche, puisque sa planification lui permet de trouver un chemin vers l'état but en exploitant ses connaissances. Le MetaContrôleur doit donc le sélectionner préférentiellement lors des premières exécutions de la tâche. L'Expert habituel apprend en parallèle des actions choisies par HATP, et prend la main lorsque son apprentissage est suffisant pour lui faire choisir les bonnes actions à un coût moindre.

Cas 2 : L'environnement change, HATP est capable de trouver un plan. On se place après apprentissage d'une habitude comportementale par l'Expert habituel, et on change les conditions dans lesquelles le robot agit (i.e. la poubelle change de côté). La tâche est la même, mais l'habitude n'est plus pertinente dans ces conditions, le MetaContrôleur doit donc redonner la main à HATP pour retrouver un plan qui permette de résoudre la tâche. En conservant ces nouvelles conditions environnementales, l'Expert habituel peut réapprendre une habitude (avec les défauts de plasticité de cet algorithme, ou éventuellement avec stockage de l'habitude précédente et apprentissage à partir de zéro).

Cas 3 : HATP ne peut pas trouver un plan. En partant d'un système qui n'a pas encore appris, nous plaçons le robot dans un cas pour lequel HATP ne peut trouver de

plan valide. Malgré l'échec de la planification, l'approche essais-erreurs de l'Expert habituel permet au robot d'agir et au planificateur de trouver un état dans lequel il peut réussir à planifier.

Je ne vais pas détailler ici les résultats qui sont restés très préliminaires (on pourra se référer aux chapitres correspondant dans la thèse d'Erwan Renaudo [Renaudo 2016] et de Sandra Devin [Devin 2017a]). Cependant, ce travail a permis la mise en place d'une première preuve de concept permettant d'utiliser un système de planification basé modèle (ici HATP) comme amorce pour un système d'apprentissage model-free (i.e. on est pas obligé de partir "from scratch", on peut faire apprendre au système à partir de connaissances données par un expert). L'autre grand avantage, c'est qu'une fois que cet amorçage est réalisé, on a un système qui ne reste pas bloqué lorsqu'aucun plan n'est possible, il va essayer de faire une de ses actions disponibles (et nous avions aussi imaginé pouvoir revenir sur le domaine HATP quand on découvrait ce genre de choses).

Nous n'avons malheureusement pas eu le temps d'affiner plus avant toutes les expérimentations que nous avions en tête mais fort heureusement, à la faveur d'un nouveau projet ANR PRCI ELSA (Effective Learning of Social Affordances cf section 2.3) nous allons pouvoir continuer à explorer ces différentes pistes et même de nouvelles.

4.3 Vision Système

Nous avons abordé dans ce chapitre une partie des éléments à prendre en compte lorsqu'il s'agit de gérer l'exécution d'une interaction humain-robot. Gérer l'ensemble de ces processus de manière cohérente pour permettre la réalisation de l'interaction demande la mise en place d'un système de supervision. Dans sa thèse, Amandine Mayima [Mayima 2021a] proposait un ensemble d'exigences pour le développement d'un système de supervision dédié à la gestion de l'interaction humain-robot, nous la reprenons ici en l'adaptant un peu :

Être générique. Les objectifs développés dans la suite de cette liste sont valables pour la plupart des tâches collaboratives. Ainsi, il nous a semblé essentiel de développer un logiciel non dédié à une tâche humain-robot particulière mais capable de gérer des plans pour diverses tâches.

Prendre en compte le partenaire humain Cela signifie prendre en compte ses capacités, ses connaissances et sa perspective. Cela devrait inclure aussi les connaissances de l'humain sur le robot.

Garder de la souplesse et laisser la possibilité à l'humain de décider. Dans certains cas, il n'est pas utile, voire contre-productif que le robot planifie tout en amont surtout si le plan met en jeu de l'interaction (et des actions de l'humain). Décider qui fait quoi et dans quel ordre peut être décidé au moment de l'exécution.

Reconnaître les actions humaines. Pour surveiller l'avancement du plan, le robot doit être capable de surveiller l'humain en reconnaissant ses actions mais aussi en étant capable de dire s'il est inactif ou qu'il n'est plus engagé dans la tâche.

Gérer les imprévus. Le plan partagé généré par le robot au début de l'interaction doit représenter une trame idéale mais il faut être capable de gérer des déviations à ce plan et de gérer les imprévus.

Gérer l'ensemble des communications de manière pertinente. La communication est l'une des clés de la collaboration. Il est donc important de doter le robot de la capacité à gérer des actions de communication pertinentes, verbales et

non verbales.

Considérer l'interaction en soi et pas seulement la tâche. Nous proposons de considérer ce fait en définissant ce que nous avons appelé session d'interaction qui donne un cadre à l'interaction.

S'adapter à l'humain S'adapter à l'expérience, aux capacités ou aux préférences de la personne avec qui le robot interagit.

On peut voir dans cette liste une extension des propriétés définies pour un robot autonome : programmabilité, autonomie et adaptabilité, réactivité, comportement cohérent, robustesse, extensibilité (cf section 1.2).

Pour répondre à ces exigences, Amandine Mayima a développé JAHRVIS : Joint Action-based Human-aware supeRVISor. JAHRVIS est un système de supervision, c'est-à-dire qu'il intègre les décisions de haut niveau du robot et contrôle son comportement en tenant compte de l'humain avec lequel le robot interagit. Ainsi, JAHRVIS se différencie des systèmes de supervision dédiés au contrôle robotique ou multi-robots dans la mesure où l'humain est pris en compte. JAHRVIS interroge, gère et exécute des plans (partagés) qui sont un ensemble (partiellement) ordonné d'actions à effectuer par des agents humains et robots afin d'atteindre un objectif (partagé). La gestion du plan est basée sur l'estimation des connaissances de l'humain, sa connaissance de l'état actuel de l'environnement et des actions humaines reconnues. Le système est capable de gérer différents types de plans. La gestion du plan repose, entre autres, sur la reconnaissance des actions de l'humain avec qui le robot interagit. JAHRVIS intègre ses propres processus de suivi d'action, c'est-à-dire de sélection du point d'intérêt du robot (permettant l'attention conjointe), et de reconnaissance d'actions. Ce dernier processus est basé sur un modèle et a été conçu pour être robuste à une perception potentiellement peu fiable de l'humain. Ensuite, comme il y a des actions du plan à exécuter par le robot, JAHRVIS a besoin d'interfaces avec les contrôleurs du robot. De plus, les actions peuvent être de deux types, physiques et communicatives, et nécessitent donc une gestion différenciée. Enfin, une caractéristique importante est la capacité de communiquer verbalement avec l'humain. En effet, lors d'une tâche collaborative, communiquer peut être nécessaire, entre autres, pour informer le partenaire d'une action effectuée, ou pour lui demander d'en effectuer une.

La figure 4.5 représente d'une part l'ensemble des données gérées par le système JAHRVIS et d'autre part l'ensemble des processus associés et leurs liens avec les différents composants.

4.4 Contributions associées

[Fiore 2014] Michelangelo Fiore, Aurélie Clodic, and Rachid Alami. On Planning and Task achievement Modalities for Human-Robot Collaboration. In International Symposium on Experimental Robotics (ISER 2014), page 15p., Marrakech, Morocco, June 2014. URL `https://hal.archives-ouvertes.fr/hal-01149109`.

[Gharbi 2015] Gharbi, Paubel, Clodic, Carreras, Alami, and Cellier] Mamoun Gharbi, Pierre-Vincent Paubel, Aurélie Clodic, Ophélie Carreras, Rachid Alami, and Jean-Marie Cellier. Toward a better understanding of the communication cues involved in a human-robot object transfer. In 24th International Symposium on Robot and Human Interactive Communication, IEEE RO-MAN, pages 319–324, Kobe, Japan, August 2015. doi : 10.1109/ROMAN.2015.7333626. URL `https://hal.archives-ouvertes.fr/hal-01628249`.

[Renaudo 2015] Erwan Renaudo, Sandra Devin, Benoît Girard, Raja Chatila, Rachid Alami, Mehdi Khamassi, and Aurélie Clodic. Learning to interact with humans using goal-directed and habitual behaviors. In Ro-Man 2015, Workshop on Learning for Human-Robot Collaboration, Kobe, Japan, August 2015. URL https://hal.laas.fr/hal-01944380.

[Devin 2017b] Sandra Devin, Aurélie Clodic, and Rachid Alami. About Decisions During Human-Robot Shared Plan Achievement : Who Should Act and How ? In Social Robotics,9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings, volume 10652 of Lecture Notes in Computer Science book series, pages 453–463. springer, October 2017. URL https://hal.laas.fr/hal-01943959.

[Devin 2018] Sandra Devin, Camille Vrignaud, Kathleen Belhassein, Aurélie Clodic, Ophélie Carreras, and Rachid Alami. Evaluating the Pertinence of Robot Decisions in a Human-Robot Joint Action Context : The PeRDITA Questionnaire. In 27th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2018), Nanjing, China, August 2018. doi : 10.1109/RO-MAN.2018.8525785. URL https://hal.laas.fr/hal-01906101.

[Chatila 2018] Raja Chatila, Erwan Renaudo, Mihai Andries, Omar Ricardo Chavez-Garcia, Pierre Luce-Vayrac, Raphaël Gottstein, Rachid Alami, Aurélie Clodic, Sandra Devin, Benoît Girard, and Mehdi Khamassi. Toward Self-Aware Robots. Frontiers in Robotics and AI, 5 :88, August 2018. doi : 10.3389/frobt.2018.00088. URL https://hal.archives-ouvertes.fr/hal-01856931.

[Mayima 2019] Amandine Mayima, Aurélie Clodic, and Rachid Alami. Evaluation of the Quality of Interaction from the robot point of view in Human-Robot Interactions. In 1st Edition of Quality of Interaction in Socially Assistive Robots (QISAR) Workshop, Madrid, Spain, November 2019. The 11th International Conference on Social Robotics (ICSR 2019). URL https://hal.laas.fr/hal-02403081.

[Mayima 2021b] Amandine Mayima, Aurélie Clodic, and Rachid Alami. Towards Robots able to Measure in Real-time the Quality of Interaction in HRI Contexts. International Journal of Social Robotics, 2021. doi : 10.1007/s12369-021-00814-5. URL https://hal.laas.fr/hal-03326954.

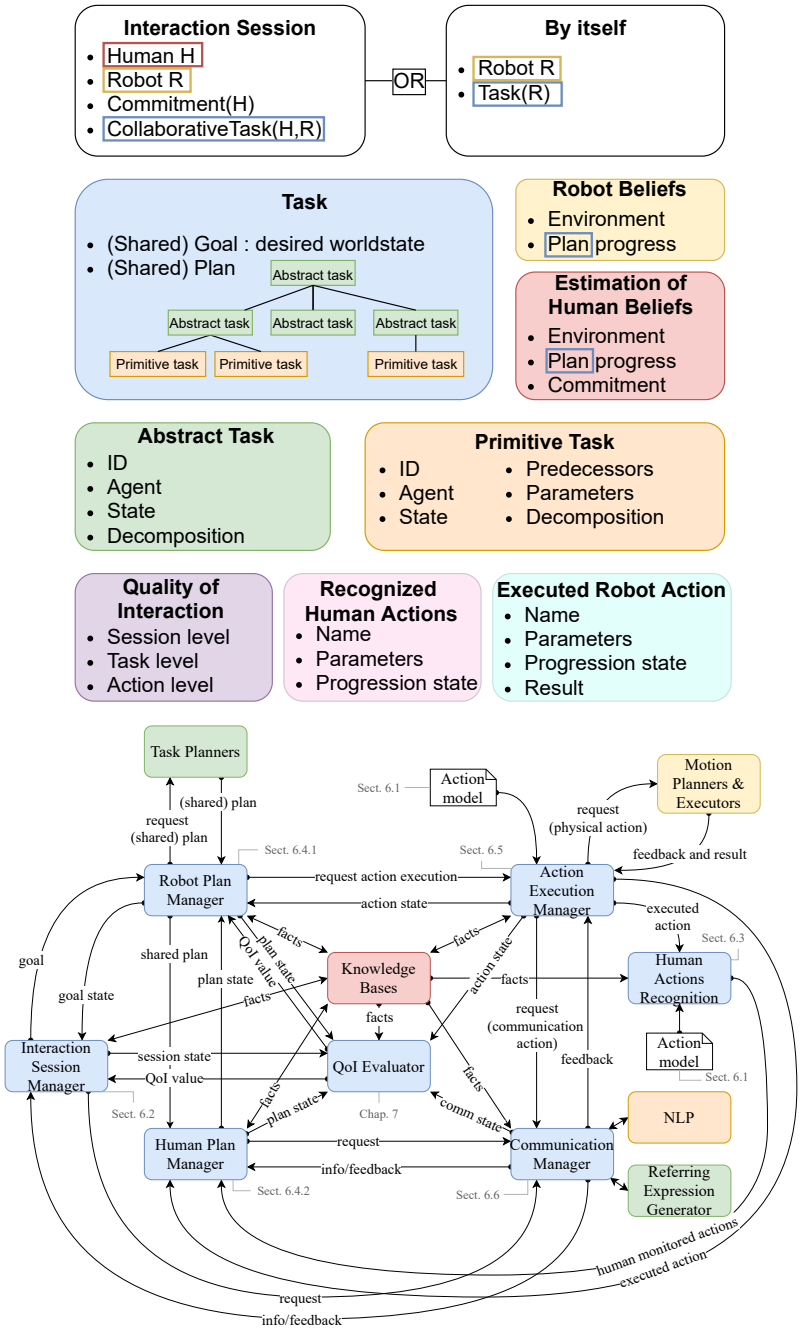


FIGURE 4.5 – Une vue des éléments nécessaires au fonctionnement du système de supervision JAHRVIS [Mayima 2021a]. En haut, l'ensemble des données considérées pour la gestion du système de supervision. Le robot peut soit être en session d'interaction avec un humain, soit être seul. Lorsqu'il s'agit d'une tâche, la progression dans le plan est calculée (des deux points de vue). Un plan est composé de tâches abstraites et primitives dont l'état évolue au fur et à mesure du déroulement de la tâche. La qualité de l'interaction (QoI) et la reconnaissance des actions de l'humain sont deux éléments continuellement calculés. Lorsque le robot exécute une action, l'état de celle-ci est mis à jour ainsi que son résultat (succès ou échec). En bas, l'ensemble des processus de JAHRVIS (en bleu) et leurs interactions avec les autres éléments de l'architecture.

Chapitre 5

Gérer les défaillances

5.1 Introduction

J'aimerais aborder ici la problématique de la prise en compte des aspects sécurité dans le cadre de développement robotique en général et dans un cadre d'interaction humain-robot en particulier. Il est à noter que j'ai effectué mon stage de DEA en collaboration entre l'équipe BIP (Eric Rutten et Karine Altisen) et l'équipe Verimag (Florence Maraninchi) sur le sujet de la synthèse de contrôleur discret pour la vérification d'un système robotique (ce travail a fait l'objet de la publication [Altisen 2003]).

Si l'on envisage un robot évoluant au milieu des humains, la sécurité en premier lieu des humains évoluant près du ou des robots et également du robot lui-même est primordiale. J'ai été invité à participer au séminaire Dagstuhl "Verification and Synthesis of Human-Robot Interaction" (`https://drops.dagstuhl.de/opus/volltexte/2019/10861/`). Ce séminaire avait pour but de faire se rencontrer les communautés de la vérification formelle en robotique et celle de l'interaction humain-robot. Lors d'une discussion, un des chercheurs nous a fait part d'une anecdote. Lors du déploiement d'un robot dans le cadre d'un projet de recherche dans l'équivalent d'un EPAHD, le robot était parti à fond et avait fini sa course dans l'escalier. Il n'y avait pas eu de victimes, à part le robot lui-même. C'était une de nos craintes lors de l'expérimentation à la Cité de l'Espace et c'est pourquoi un agent de la Cité de l'Espace était toujours présent lorsque le robot était en service, et c'est ce que nous avons fait également lors du déploiement du robot MuMMER en Finlande. Nous n'avons aujourd'hui pas les capacités, dans le cadre d'un projet de recherche, d'assurer qu'un bug n'est pas présent dans nos systèmes et qu'il ne pourrait pas engendrer une situation dangereuse pour les personnes autour du robot, c'est un risque mesuré certes, mais il faut en avoir conscience et le prendre en compte.

J'ai travaillé dans le secteur de l'aéronautique, régi par des normes très strictes (à cette époque c'était la norme DO-178B, niveau C dans mon cas : "Un défaut du système ou sous-système étudié peut provoquer un problème majeur entraînant un dysfonctionnement des équipements vitaux de l'appareil"). Dans ce secteur, c'est en quelque sorte le donneur d'ordre (par exemple Airbus) qui donne le "la". Ainsi, la norme est traduite par le donneur d'ordre et les sous-traitants doivent dériver chacune des exigences dans leur logiciel. Cela demande un travail de traçabilité important accompagné d'un travail de tests et de contrôle qualité considérable. Cela permet, in fine, que chacune des lignes du logiciel soit tracée par une ou des exigences. Ainsi, toutes les lignes de codes ont une nécessité et toutes les exigences doivent être testées.

Dans le cadre automobile, avant l'arrivée de la norme ISO-26262, la problématique était traitée de manière différente, la méthode AMDEC (Analyse des modes de défaillance, de leurs effets et de leur criticité) était utilisée pour réfléchir aux différentes défaillances qui pouvaient se produire, comment les pallier si possible ou comment réagir à leurs conséquences.

En robotique, l'une des premières choses que l'on apprend lorsque l'on utilise un robot, c'est avoir l'arrêt d'urgence dans les mains lorsque le robot fonctionne, "au cas où". Cependant, il faut bien avoir en tête que ce que le système doit faire lorsque l'on presse l'arrêt d'urgence n'a rien d'évident. Quel sens donner à cet arrêt pour un robot comme Pepper ou un robot humanoïde qui doit "tenir" debout. Couper les moteurs ? Mais alors que se passe-t-il ? Le robot tombe ? Que faire pour un drone ? Que faire pour un véhicule autonome inséré au milieu de la circulation ? Pour un robot circulant sur une pente ?

Nous allons ici survoler sans rentrer dans les détails, mais nous souhaitons expliquer l'importance de plus en plus grande que va prendre, que doit prendre, cette considération.

5.2 Ce qui peut mal se passer

5.2.1 Défaillance Hardware

Un robot c'est un (ou plusieurs) PC, sur roues, sur pattes, à hélices... avec des capteurs et des actionneurs. Au niveau hardware, on peut penser à un certain nombre de défaillances possibles. Tout d'abord au niveau des capteurs et des actionneurs. Une défaillance peut venir d'une panne du capteur/actionneur lui-même (e.g. un odomètre qui renvoie des données erronées), mais aussi d'un cas d'utilisation limite ou imprévue (e.g. fumée devant un capteur de vision). En fait, n'importe quel composant peut tomber en panne, par exemple il peut se produire un problème mémoire sur une carte électronique. D'autres types de problèmes hardware peuvent survenir, par exemple, la batterie peut être déchargée ou faible, comme pour l'arrêt d'urgence, il faut réfléchir prudemment à la procédure qui doit être mise en place (e.g. le drone il se pose où quand il n'a presque plus d'énergie ?).

5.2.2 Défaillance Software

Un robot exécute du code, il peut aussi se produire des défaillances software, il peut y avoir des bugs et ces bugs peuvent se produire à de nombreux niveaux et à de nombreux endroits. Nous allons en donner ici une liste (non exhaustive) pour permettre d'appréhender ce qu'il est nécessaire de considérer : bug dans une fonction capteur, bug dans une fonction actionneur, bug dans une fonction de planification et d'exécution de trajectoire, bug dans une fonction de planification de tâches, bug dans le système de supervision, bug car une donnée de l'ontologie est incorrecte, bug car une fonction de perception n'est pas assez robuste, etc.

Bien sûr, c'est notre travail de programmeur en informatique de nous poser ces questions lorsque l'on code les fonctions. Mais le fait est qu'aujourd'hui on utilise de plus en plus de fonctions en "boite noire" dont on ne maîtrise pas forcément tout le code et ensuite, nous n'avons pas les moyens de lancer des procédures de vérification comme celles que j'ai connues dans l'aéronautique. Il est possible d'utiliser des techniques de synthèse de contrôleur pour générer un système "safe" ou des techniques de model-checking pour vérifier qu'il respecte certaines règles, mais l'intégration de ces techniques est encore coûteuse en termes d'investissement. De plus, l'intégration de techniques de deep-learning rend là encore les possibilités de vérification compliquées.

5.2.3 Le cas particulier de l'interaction

Après, rassurons-nous, ce ne sont pas nos robots de recherche qui doivent se retrouver dans nos maisons, mais des robots développés par des entreprises qui elles devront certifier leur système. Ainsi, il existe aujourd'hui un certain nombre de normes qui s'appliquent aux robots qui sont au contact des personnes telles que ISO 13855 :2010 Sécurité des machines – Positionnement des protections par rapport aux vitesses d'approche des parties du corps humain (`https://www.iso.org/standard/42845.html`), ISO/TS 15066 :2016 Robots et dispositifs robotiques – Robots collaboratifs (`https://www.iso.org/standard/62996.html`) et ISO 13482 :2014 Robots et dispositifs robotiques – Exigences de sécurité pour les robots de soins personnels (`https://www.iso.org/standard/53820.html`). Un projet comme le projet COVR (`https://safearoundrobots.com/`) propose une boîte à outils d'évaluation des risques qui fournit des lignes directrices pour gérer ces normes. Il propose une méthode d'évaluation des risques basée sur l'analyse des cas d'utilisation qui est l'approche de base pour traiter la certification.

Mais posons-nous quand même la question de la possible survenue de ces bugs lorsque le robot est en interaction :

— bug dans une fonction capteur (une des difficultés peut être de savoir si c'est un bug ou si c'est l'utilisateur qui veut "tester" le robot, ainsi à la Cité de l'Espace, les gens posaient leurs mains sur les capteurs ultra-sons pour tenter de "tester" le robot)
— bug dans une fonction actionneur (une division par zéro et votre robot part à fond, car la consigne envoie une valeur infinie)
— bug dans une fonction de planification et d'exécution de trajectoire (le robot fait trois tours sur lui-même avant de démarrer, cela n'a pas forcément une incidence sur la sécurité, mais c'est compliqué à comprendre)
— bug dans une fonction de planification de tâches
— bug dans le système de supervision
— bug car une donnée de l'ontologie est incorrecte
— bug car une fonction de perception n'est pas assez robuste (il y avait une personne, mais la perception ne l'a pas vu, si la perception ne l'a pas vu aucun autre algorithme ne va pouvoir le prendre en compte...)

Nous l'avons déjà dit l'une des premières choses c'est qu'au départ, l'utilisateur va tenter de "tester" le robot, l'utilisateur va penser à beaucoup de choses que vous n'aviez même pas imaginées (e.g. j'ai appris à mes dépens que l'écran tactile du robot que nous utilisions à la Cité de l'Espace était multi-touch... plusieurs enfants ont réussi à appuyer sur plusieurs boutons de l'interface en même temps...), mais il va aussi des fois ne pas penser que quelque chose est possible (e.g. par contre les grands-parents des enfants n'avaient pas compris au départ que l'écran était tactile et leur disaient de ne surtout pas y toucher).

Nous pensons qu'il y a aussi un élément à prendre en compte à ce niveau. C'est la connaissance des robots par les personnes qui vont les utiliser et les rencontrer. On a tous appris à utiliser les différents outils techniques qui nous entourent et nous en sommes devenus familiers au fur et à mesure. Mais la plupart du temps pour cela nous avons eu un mode d'emploi. Pour l'automobile, nous avons même passé un permis de conduire pour apprendre les règles, le code de la route, connues de tous et qui fait que nous faisons énormément de prédiction lorsque nous sommes dans un environnement routier. En voiture, vous savez qu'une voiture est sensée s'arrêter lorsque vous attendez à un passage piéton

FIGURE 5.1 – A gauche, le panneau français "zone de rencontre". A droite, notre proposition de création d'une telle zone pour les robots naviguant parmi les humains (et autres robots). Au milieu, la façon dont les rues étaient partagées avant le code de la route, il n'y avait aucune organisation et par exemple les gens traversaient partout. Aujourd'hui, le code de la route aide les personnes et les autres véhicules à partager l'espace, nous proposons la création de règles de circulation pour aider les personnes et les robots à faire de même.

(en France vous savez aussi attendre d'être sûr que la voiture s'arrête...), vous savez qu'il y a la priorité à droite, etc.

Lors du montage d'un projet européen portant sur le développement d'un système de navigation pour robot autonome, nous avions imaginé la création d'un code de la route pour les robots et les piétons. Si l'on imagine un robot circulant au milieu des piétons. Il est nécessaire à notre sens que les piétons puissent faire des prédictions sur ce que va faire un robot et de la même façon que le robot puisse faire et utiliser des prédictions lorsqu'il navigue auprès des piétons (mais il faut que ces capacités soient connues des humains). Les zones où les robots partagent l'espace et naviguent parmi les piétons n'existent pas encore. Dans plusieurs pays européens, il existe un panneau de signalisation correspondant à une zone de rencontre ou partagée, tel que représenté sur le côté gauche de la Fig. 5.1. Dans ces zones, un certain nombre de règles s'appliquent comme le fait que les piétons ont toujours la priorité (quoi qu'il arrive) et que la vitesse est limitée. Fait intéressant, il est aussi dit que "le partage de la voirie se fait par la cohabitation entre piétons et véhicules à basse vitesse au centre de la rue. Dès que la largeur de la rue le permet, il est important de garder un espace continu réservé aux piétons. Le terme "rencontre" souligne que les conflits doivent être gérés, non par un rapport de force, mais plutôt par courtoisie au profit des plus vulnérables." (`http://www.creuse.gouv.fr/content/download/4936/35503/file/pdf_fiche_zone_rencontre_cle768b71.pdf`) Cette dernière phrase signifie beaucoup en termes d'attentes sur ce qui pourrait arriver et aide à la formation des attentes mutuelles entre piétons et conducteur dans ce cas. Nous pensons que le déploiement des robots auprès des piétons devrait s'accompagner du développement de telles zones et de telles signalisations ainsi que de règles de circulation dédiées qui permettront aux piétons de se forger des attentes sur les comportements des robots et d'aider à la résolution des situations où la négociation est nécessaire. Bien sûr, en fin de compte, le piéton aura toujours la priorité donc la sécurité et l'évaluation des risques sont au premier plan des priorités, le robot ne doit blesser personne quoi qu'il arrive et doit pouvoir s'arrêter (ou être placé dans une configuration sécurisée pour les humains autour) chaque fois que nécessaire.

5.2.4 Cybersécurité

En général, on communique avec le robot à travers un réseau wifi. Que se passe-t-il si quelqu'un réussissait à pirater le réseau ? Est ce qu'il est possible de prendre le contrôle du robot ? Si l'on pense à avoir des robots grands publics dans le futur, il faut penser aux éventuelles menaces. A côté des problèmes éthiques que cela poseraient, une personne qui réussirait à récupérer le "flux vidéo" issue d'un drone ou d'un robot intervenant sur un site sensible pourrait être problématique. De même une personne qui prendrait le contrôle du haut-parleur d'un robot pour diffuser de la propagande.

5.3 Conclusion

Dans ce chapitre, nous avons voulu donner des pistes de compréhension et de réflexion sur les défaillances qui peuvent se produire. Nous n'avons pas de travaux sur la question mais nous pensons que c'est un sujet qu'il faudrait pouvoir aborder de façon globale pour permettre une prise en charge plus générale et pas au cas par cas pour chaque application. Un développeur d'un système particulier pourrait tirer bénéfice d'avoir un certain nombre de règles ou d'exigences à suivre.

5.4 Contributions Associées

[Altisen 2003] Karine Altisen, Aurélie Clodic, Florence Maraninchi, and Eric Rutten. 2003. Using controller-synthesis techniques to build property-enforcing layers. In Proceedings of the 12th European conference on Programming (ESOP'03). Springer-Verlag, Berlin, Heidelberg, 174–188. URL : `https://team.inria.fr/ctrl-a/files/2015/02/Esop03.pdf`

— Cet article fait suite aux travaux que j'ai réalisé avec Karine Altisen dans le cadre de mon DEA en co-encadrement entre Florence Marininchi du laboratoire Verimag et Eric Rutten de l'INRIA Rhône-Alpes. Dans les systèmes complexes, comme les usines de robots, les applications sont construites sur un ensemble de composants ou dispositifs. Chacun d'eux a des contraintes individuelles particulières, et il existe également des contraintes logiques sur leurs interactions, liées par exemple aux caractéristiques mécaniques ou à l'accès aux ressources partagées. La gestion de ces contraintes peut être séparée de l'application, et réalisée par une couche intermédiaire. Nous montrons comment construire de telles couches de propriétés, dans un style mixte impératif/déclaratif : 1) les contraintes intrinsèques à un composant sont modélisées par un automate ; le produit de ces automates est une première approximation de l'ensemble des contraintes à respecter ; 2) les contraintes qui impliquent plusieurs composants sont exprimées en propriétés logiques temporelles de ce produit ; 3) nous utilisons des techniques et des outils généraux de synthèse de contrôleurs afin de combiner l'ensemble des automates parallèles communicants avec la contrainte globale.

Chapitre 6

Imaginer le robot comme un agent social au milieu d'autres agents sociaux

"Que les compétences d'interaction sociale soient essentielles au bon fonctionnement des robots sociaux n'implique pas nécessairement que les robots sociaux doivent être ou peuvent être comme les humains. En fin de compte, le type de compétence dont un robot donné a besoin pour fonctionner correctement est déterminé par son environnement de tâche ou le paramètre pour lequel il est programmé. Pour illustrer, les robots nettoyeurs demandent d'autres compétences sociales à l'hôtel qu'à la maison, un robot réceptionniste fait appel à d'autres compétences à l'hôtel qu'aux urgences, et les robots soignants demandent des compétences différentes des robots nettoyeurs." [Brinck 2018]

6.1 Qu'est ce qu'un agent social ?

Nous allons présenter ici plusieurs travaux que nous avons menés dans un cadre interdisciplinaire avec des psychologues et des philosophes sur la notion de socialité. Le premier est issu d'un travail avec Víctor Fernández Castro et Raul Hakli, What does it take to be a social agent ? [Fernández Castro 2020], le second est issu d'un travail avec Kathleen Belhassein, Víctor Fernández-Castro, Amandine Mayima, Elisabeth Pacherie, Michèle Guidetti, Rachid Alami et Hélène Cochet, Addressing joint action challenges in HRI : Insights from psychology and philosophy [Belhassein 2022].

6.1.1 Les conditions de la socialité

Lorsqu'un robot nous accueille d'une voix joyeuse, cela peut induire en nous une réaction émotionnelle ou une tentative de contact visuel qui sont des réactions typiques de l'interaction sociale humaine. Si le robot ne peut pas répondre de manière réciproque et socialement appropriée, nos attentes ne sont pas satisfaites et nous pouvons nous sentir déçus ou frustrés. Nous avons tendance à "socialiser" le robot de manière irréaliste, c'est-à-dire à lui attribuer un certain nombre de capacités que l'on retrouve chez les êtres sociaux que nous rencontrons et avec qui nous interagissons habituellement [Seibt 2014, Young 2010]. À ce problème s'ajoute le fait que nous ne disposons pas d'un modèle complet de la liste des capacités que nous associons à cette tendance à socialiser, et ne pouvons donc pas évaluer de manière fiable les conséquences de l'utilisation des robots sociaux pour notre culture et nos valeurs. Nous proposons que l'identification d'un ensemble de caractéristiques nécessaires à la socialité puisse atténuer les éventuels impacts négatifs de l'utilisation des robots sociaux, par exemple en permettant aux robots d'avoir un comportement alternatif socialement efficace, mais peu invasif, évitant ainsi d'exploiter les émotions et les sentiments humains et de nuire à l'intégrité ou à l'autonomie humaine.

Pour entamer cette réflexion, nous nous sommes inspirés d'une méthode initiée par Daniel Dennett qui a suggéré que les êtres humains ont une tendance naturelle à se comprendre en tant qu'agents intentionnels. Selon Dennett, lorsque nous essayons de comprendre le comportement des autres, nous adoptons une position intentionnelle envers eux, ce qui signifie que nous les comprenons comme des agents qui ont des états intentionnels, essentiellement des croyances et des désirs, et qui sont rationnels dans le sens où ils sélectionnent des actions qui, selon leurs croyances, servent au mieux la satisfaction de leurs désirs [Dennett 2009]. Dennett a proposé trois positions que l'on peut adopter lorsque l'on rencontre un nouvel objet et qu'on essaie de comprendre et/ou de prédire son comportement : la "stratégie du physicien" (*physical stance*), la "stratégie du technicien" (*design stance*), la "stratégie interprétative" (*intentional stance*) [1].

Prenons quelques instants pour expliquer ces trois stratégies. Si nous suivons la "stratégie du physicien", nous essayons de comprendre un objet en fonction de ses propriétés physiques. Ainsi, si l'on examine une vieille branche sèche tombée d'un arbre, nous pouvons prédire, simplement sur la base des propriétés matérielles du morceau de bois, qu'il se brisera si on lui applique une force et brûlera si l'on y met le feu. Si nous suivons la "stratégie du technicien", nous essayons de comprendre un objet en fonction de ce pour quoi il est conçu, de son but. Ainsi, si nous sommes confrontés à un ascenseur, nous allons considérer que c'est un appareil conçu dans un but précis, i.e. transporter des personnes et des objets entre les étages d'un bâtiment. Nous allons en déduire qu'il va nous permettre de nous rendre à un autre étage. S'il ne commence pas à bouger immédiatement, nous allons nous demander comment indiquer à quel étage se rendre, et utiliser à nouveau la "stratégie du technicien" pour en déduire que les boutons numérotés sur le mur sont probablement là à cette fin. Bien sûr, un ingénieur spécialisé dans les ascenseurs sera peut-être capable d'expliquer le fonctionnement en suivant la "stratégie du physicien", mais pour le commun des mortels, il est plus aisé d'en appréhender le fonctionnement en faisant référence à son objectif et aux buts (et éventuels sous-buts) qu'il permet d'atteindre. Lorsque nous rencontrons des animaux ou d'autres personnes, aucune de ces deux stratégies n'est véritablement adaptée et l'on suit plutôt une "stratégie interprétative". Avec quelques connaissances de base et un minimum d'observations, nous les comprenons comme des agents, comme des êtres intentionnels qui ont des croyances et des désirs en relation avec le monde. En adoptant une telle position intentionnelle, nous pouvons donner un sens à leurs comportements : ce lapin me fuit parce qu'il pense que je suis une menace et qu'il veut éviter les menaces ; cette personne entre dans cet immeuble parce qu'elle veut rentrer chez elle et elle vit dans cet immeuble. Attribuer de tels états intentionnels aux animaux et aux êtres humains est logique, car cette façon de penser rend leur comportement compréhensible et plus facile à prévoir (que si nous adoptions les autres stratégies). Cependant, parfois, expliquer ou prédire les actions à partir de cette "stratégie interprétative" peut également devenir fastidieux. Pourquoi cette personne se dépêche-t-elle d'arriver au bureau si tôt le matin ? Est-ce que c'est vraiment son intention alors qu'elle a l'air fatiguée ? Pourquoi s'arrête-t-elle au feu rouge si elle veut arriver le plus tôt possible ? A-t-elle vraiment envie de rester là tout d'un coup ? En fait, pourquoi tout le monde s'arrête aux feux rouges ? Est-ce qu'ils aiment regarder la couleur rouge ou quoi ?

Lorsque nous observons le comportement des êtres humains, la "stratégie interprétative" basée sur les intentions ne semble pas suffisante. La personne se dépêche parce

1. pour la traduction, j'adopte la terminologie proposée sur la page wikipédia de Daniel Dennett `https://fr.wikipedia.org/wiki/Daniel_Dennett`

qu'elle travaille dans ce bâtiment et que dans ce rôle, elle est censée être là à 8 heures du matin, et c'est une norme dans son entreprise. Mais elle s'arrête au feu rouge (comme tout le monde), car c'est une règle dans la société de s'arrêter aux feux rouges. On voit donc apparaître des références aux normes sociales, aux rôles sociaux et aux institutions sociales et nous nommerons cette stratégie, "stratégie d'interprétation sociale" en regard de la "stratégie d'interprétation intentionnelle" que nous avons exposée précédemment. Ce concept établirait un ensemble de fonctionnalités que les roboticiens pourraient utiliser comme point de départ dans la conception de robots sociaux avant de passer à des ensembles plus riches de fonctionnalités de socialité. Cela serait également utile pour les utilisateurs, car les informer sur ces caractéristiques de socialité qui restent en deçà de la personnalité les aiderait à éviter de faire des hypothèses irréalistes sur les capacités des robots en termes de conscience, d'émotions ou de moralité. Adopter une compréhension instrumentaliste permet en outre de parler légitimement de « robots sociaux » sans avoir à s'engager dans l'idée que les robots sont « vraiment » sociaux au sens où le sont les humains, ou partagent avec nous des mécanismes similaires qui produisent les types de comportements que nous considérons comme sociaux.

6.1.2 Définition d'une stratégie d'interprétation sociale

Dans [Fernández Castro 2020], avec Victor Fernandez Castro et Raul Hakli, nous avons proposé un certain nombre d'exigences que les agents doivent être capables d'exhiber pour pouvoir être considérés comme des agents sociaux. Par symétrie, ces exigences nous donnent des pistes pour développer une stratégie d'interprétation sociale :

Exigence de Comportement Rationnel *(Intentional Stance Requirement)* Pour être considérée comme un agent, la cible de l'observation doit se comporter de manière à ce que son comportement puisse être interprété comme étant rationnellement motivé par des croyances, des désirs et des intentions.
- On pourrait objecter que cette exigence nécessite l'existence d'états mentaux "réels" et considérer que cela n'a pas de sens pour un robot. Cependant, nous considérons que cette exigence est neutre par rapport à cela, l'important ici est que le comportement puisse être interprété et compris en postulant des croyances, des désirs et des intentions, pas que ces croyances existent.

Exigence de Sensibilité aux autres agents *(Sensitivity to Other Agents)* Pour être social, un agent doit avoir la capacité d'identifier d'autres agents et être sensible à leur présence en adaptant son comportement en conséquence.
- Ici nous déterminons qu'un agent social est différent d'un objet et qu'un agent social est capable de faire la différence entre un objet et un autre agent.
- Cette sensibilité aux autres agents est nécessaire à l'interaction sociale. L'interaction sociale nécessite d'être capable de prédire les actions des autres et de coordonner ses propres actions avec les leurs.

Exigence de compréhension normative *(Normative Understanding Requirement)* Les agents sociaux doivent être capables de comprendre leur comportement et celui des autres agents à la lumière des attentes générées par les normes, règles et conventions sociales, et être en mesure de modifier leur comportement pour s'accorder avec eux.
- En d'autres termes, cette exigence demande simplement à un agent social d'être capable de reconnaître des actions ou des comportements comme conformes

à une norme (c'est-à-dire que l'action répond aux attentes générées par une norme), et de réguler son comportement en conséquence.

Exigences normatives d'évaluation et de régulation *(Normative evaluation and regulation requirements)* Les agents sociaux doivent être capables d'évaluer positivement et négativement un plan d'action particulier conformément à une norme et de déclencher différents comportements régulateurs appropriés comme sanctionner des réponses, demander des raisons, blâmer, manifester une approbation ou justifier l'action.

Nous proposons ces exigences comme un fil d'Ariane à suivre lors de la conception d'un robot amené à opérer dans un contexte social. Ces exigences nous semblent fournir des conditions suffisantes et nécessaires pour l'action sociale dans un sens minimal. Nous croyons qu'il existe une classe d'agents sociaux qui sont capables d'une interaction sociale rudimentaire qui est encore loin d'une interaction sociale complète qui suppose des capacités symétriques des participants et c'est ce que nous avons essayé de caractériser.

6.1.3 L'importance de la reconnaissance mutuelle

Nous pensons que la prise en compte de ces exigences dans la conception d'un robot destiné à agir/opérer en société peut permettre notamment d'améliorer la reconnaissance mutuelle (*mutual recognition*).

La reconnaissance mutuelle [Brandom 2007, and 2014, Scanlon 2000] permet aux individus de s'identifier et de s'accepter en tant qu'agents sociaux. La reconnaissance mutuelle exige que chaque individu identifie l'autre comme un partenaire possible pour l'interaction. Ce partenaire sera capable de générer des attentes et d'anticiper en fonction de différentes caractéristiques physiques et sociales, en fonction des informations concernant les interactions précédentes ou des structures sociales telles que des normes ou des conventions. De plus, la reconnaissance mutuelle requiert la confirmation implicite ou explicite que l'autre, en tant que partenaire social, accepte d'une manière ou d'une autre l'interaction. L'un des moyens les plus sûrs d'établir cette reconnaissance est de s'appuyer sur des stratégies de communication et c'est que nous avons étudié dans [Belhassein 2022].

La reconnaissance de l'autre comme partenaire potentiel d'une interaction peut se faire par des indices de communication verbaux et/ou non verbaux, qui peuvent être plus ou moins explicites à différentes étapes de l'interaction. Fait intéressant, les humains établissent souvent des stratégies de communication pour faciliter l'échange d'informations avant l'action jointe elle-même, même dans des situations où des normes sociales, des conventions ou des scripts sont disponibles pour réguler nos interactions sociales [Andrews 2012, Castro 2020, Schank 1977, Zawidzki 2013]. Par exemple, en tant que clients, nous savons généralement comment interagir avec un serveur dans un restaurant, car toutes les parties impliquées connaissent les règles claires d'étiquette et de normes sociales. C'est cette connaissance de la manière de procéder qui régule l'interaction pour atteindre l'objectif commun de prendre un repas. Cependant, même lorsque ces règles et normes existent, les interactions humaines nécessitent de signaler et de communiquer différents types d'informations concernant le lancement, le maintien ou la sortie d'une action conjointe, la reconnaissance de l'attribution des rôles ou des spécificités concernant les préférences, les objectifs et les sous-tâches. Ainsi, la mise en place d'une reconnaissance mutuelle est fondamentale pour le démarrage de l'action jointe, mais influence aussi fortement son déploiement. Par exemple, l'établissement d'une reconnaissance mutuelle

facilite l'attribution des rôles, ce qui détermine également les stratégies de communication utilisées lors de l'exécution de l'action.

Nous sommes en accord avec Ingar Brink et al. [Brinck 2018] qui propose que cette reconnaissance est la clé d'une coopération réussie avec les robots car une fois la reconnaissance mutuelle établie entre les agents : (1) les autres deviennent pour moi des ressources d'une manière qui favorise l'égalité et le partage, ce qui permet à son tour (2) d'obtenir et de partager plus d'informations au cours de l'interaction humain-robot.

Il faut bien comprendre que cette compréhension mutuelle est un processus double. D'une part cela nous donne des pistes sur ce que nous devons intégrer au robot pour pouvoir gérer une interaction mais il induit aussi le fait que l'on donne ces clés aux personnes avec qui le robot va interagir. Quels sont les éléments auxquels le robot doit savoir répondre et quels sont les éléments pour lesquels il ne peut pas ou n'a pas à réagir. Cela donne des clés de compréhension dans les deux sens. C'est en cohérence avec ce que proposait Cakmak et Takayama [Cakmak 2014] sur l'intérêt d'apprendre aux gens à utiliser les robots avant de les utiliser.

6.2 De l'intérêt de modéliser la notion d'engagement

Dans cette section, nous allons tout d'abord présenter deux conceptions philosophiques de la notion d'engagement dans le cadre d'une interaction humain-humain. Nous expliquerons ensuite comment ces modèles peuvent être une source d'inspiration pour le domaine de l'interaction humain-robot et proposerons un premier modèle d'opérationalisation de la gestion des engagements.

Ces travaux sont issus d'une collaboration avec Victor Fernandez Castro, Elisabeth Pacherie et Rachid Alami "Commitments in Human-Robot Interaction" [Fernandez Castro 2019, Clodic 2019] qui s'est poursuivi ensuite toujours avec Victor Fernandez Castro, Amandine Mayima et Kathleen Belhassein "The Role of Commitments in Socially Appropriate Robotics" (à venir) [Fernández Castro].

Dans la suite du document nous traduirons *commitment* par engagement et *expectation* par attente.

La notion d'engagement joue un rôle central dans la compréhension de l'action conjointe ([Cohen 1990, Bratman 1992]). Nous discutons de deux propositions spécifiques aux engagements : l'approche fonctionnelle, qui soutient que la fonction centrale des engagements est de réduire différentes formes d'incertitude ; et l'approche normative, qui soutient que les engagements créent des obligations vis-à-vis de ses co-mandataires qui sont en droit d'exiger que ces obligations soient remplies.

6.2.1 L'approche fonctionnelle : l'engagement comme un outil pour réduire l'incertitude

On peut soutenir que ce qui distingue une action conjointe (par exemple marcher ensemble) de la simple coordination du comportement (deux personnes marchant au même rythme au hasard) est que les partenaires s'engagent à atteindre un objectif ensemble [Gilbert 2013]. De plus, un tel engagement participatif s'accompagne souvent d'autres engagements contralatéraux [Roth 2004] associés à l'objectif général ou au plan nécessaire pour l'atteindre (par exemple ne pas accélérer, ou attendre son partenaire s'il est laissé pour compte). La nécessité d'établir ces différents engagements vient du fait que même

en supposant que les partenaires souhaitent arriver au résultat, différentes sources d'incertitude pourraient empêcher la réalisation de l'action commune. Comme le soutiennent Michael et Pacherie [Michael 2014], les participants peuvent être confrontés à trois sources d'incertitude lors d'une action conjointe :

1. une incertitude motivationnelle (*motivational uncertainty*) : les participants peuvent ne pas savoir si le partenaire est motivé ou non pour s'engager dans l'action conjointe. Ainsi, ils ne peuvent pas être sûrs de la convergence de leurs intérêts respectifs ou s'ils partagent ou non un objectif.
2. une incertitude instrumentale (*instrumental uncertainty*) : les participants peuvent ne pas savoir s'ils sont d'accord ou non sur la manière de procéder ; par exemple, comment les participants doivent répartir leurs rôles ou quand et où ils doivent agir.
3. une incertitude sur le terrain d'entente (*common ground uncertainty*) : il peut arriver que les croyances et les motivations instrumentales ne se manifestent pas mutuellement. Ainsi, même si les participants partagent un objectif ou s'entendent sur la manière de procéder, ils peuvent ne pas savoir que c'est le cas.

Comment les engagements réduisent-ils ces sources d'incertitude ? Selon l'approche fonctionnelle soutenue par Michael et Pacherie [Michael 2014], générer des engagements réduit l'incertitude en stabilisant les attentes concernant les actions, les croyances et les motivations et ainsi, en facilitant la prédiction des autres co-acteurs, ce qui rend l'action conjointe plus fluide et efficace. À leur avis, les principaux aspects des engagements sont liés à la manière dont les attentes sont générées et, par conséquent, à la manière dont les engagements sont établis. Par exemple, comment les co-agents génèrent des attentes par la communication implicite et explicite [Austin 1962, Siposova 2018] ou la répétition ([Gilbert 2013, Michael 2016]).

6.2.2 L'approche normative : l'aspect normatif vs descriptif des engagements

Fernandez Castro et Pacherie [Pacherie 2021] se sont concentrés sur un aspect différent des engagements, à savoir comment ils créent des obligations envers ses co-agents qui sont en droit d'exiger que ces obligations soient satisfaites, ce qui fait naître des attentes que l'agent agira comme engagé ou que, si ce n'est pas le cas, les co-agents exigeront qu'il le fasse. Cette vision partage avec l'approche fonctionnelle l'idée que les attentes sont l'élément central des engagements. Cependant, ils soutiennent que les attentes générées par les engagements sont d'un type particulier : les attentes normatives. En philosophie de l'esprit, plusieurs auteurs ([Greenspan 1978, Paprzycka 1999, Wallace 1998]) ont souligné qu'on peut distinguer deux types d'attentes :

1. les attentes descriptives : attentes dont la violation ou la frustration ne déclenche pas nécessairement des attitudes réactives. Ces attentes sont liées aux prévisions. Par exemple, vous pouvez vous attendre à ce que votre amie boive une bière parce que c'est ce qu'elle fait toujours, mais si elle ne le fait pas, cela peut vous surprendre, mais pas vous déranger.
2. les attentes normatives : attentes dont la violation ou la frustration déclenche des attitudes réactives : blâme (n.b. la notion de blâme doit s'entendre ici en termes très généraux et sans charge morale ; c'est-à-dire comme une réaction qui sert à sanctionner négativement une conduite, que cette conduite soit ou non une violation

d'une norme morale), demande de justification ou de sanction. Autrement dit, la réponse normale à sa frustration est d'imposer une réaction négative à l'autre agent pour ne pas avoir agi comme prévu. De telles réactions négatives sont plus chargées émotionnellement et visent à réguler le comportement de l'autre. Par exemple, vous pouvez vous sentir en droit de sanctionner votre ami lorsqu'il contrecarre votre attente de céder sa place à une personne âgée dans le métro.

Du point de vue normatif, les engagements servent à créer pour soi-même, ou pour l'autre agent, des attentes normatives, de sorte que nous pouvons nous assurer que chaque participant se comporte comme prévu. Dans cette optique, l'aspect clé des engagements est la manière dont nous affichons différentes stratégies de communication et de comportement pour rendre explicites nos obligations envers notre partenaire et leurs propres obligations envers nous ; par exemple, en blâmant, en faisant prendre conscience à l'autre de ce qui est attendu ou en s'excusant quand quelque chose ne va pas. À ce stade, il convient de noter que les deux approches ne sont pas nécessairement incompatibles. En fait, dans la mesure où elles mettent l'accent sur différents aspects du traitement des engagements par les agents, elles peuvent être considérées comme complémentaires. Selon nous, la gestion des engagements implique à la fois de générer des attentes qui facilitent la réduction de l'incertitude et d'afficher des stratégies de réparation et de régulation lorsque ces attentes sont déçues.

6.2.3 Application à la robotique

En robotique, on trouve plusieurs implémentations concernant les engagements appliqués aux conversations collaboratives [Sidner 2003, Sidner 2005] mais aussi aux paramètres d'interactions comme dans Human-Robot Interaction Operating System [Fong 2006], dans Robot Behavior Toolkit [Huang 2013], dans Robot Tutelage [Breazeal 2004] ou lié à la supervision et aux actes de communication [Clodic 2007a].

Pour illustrer comment la notion d'engagement pourrait être instanciée et bénéficier à une architecture robotique, nous allons reprendre l'exemple de Rackham guide de musée [Clodic 2006]. Rackham est le robot sur lequel j'ai réalisé une de mes manips de thèse, il a officié comme guide à la Cité de l'espace pendant quelques semaines entre 2002 et 2005 (cf. `http://www.laas.fr/robots/rackham`). Un scénario basique était le suivant : un visiteur de la Cité arrivait près du robot, le robot l'invitait à choisir une destination, la personne choisissait la destination sur l'écran tactile du robot, le robot calculait alors le chemin et guidait le visiteur jusqu'à la destination, une fois arrivée à la destination, le robot donnait quelques explications sur le lieu d'arrivée et prenait congé. Lors de la phase de navigation, le robot vérifiait régulièrement que la personne était toujours derrière lui. Le robot devait aussi gérer les autres visiteurs et possibles attroupements (en attendant, en demandant le passage ou en recalculant une nouvelle trajectoire). Voyons si l'on essaie d'appliquer un schéma de prise en compte des engagements sur cet exemple. Je tiens à préciser que cet exemple tel qu'il est expliqué en suivant n'a pas été implémenté. Il s'agit d'une réflexion sur le possible apport des engagements à une architecture robotique.

Dans cet exemple, on considère un engagement conjoint entre le visiteur et Rackham pour aller ensemble vers une destination particulière, où le robot doit guider le visiteur tandis que le visiteur doit le suivre. Dans ce contexte, les engagements conjoints génèrent deux engagements subsidiaires qui génèrent les attentes suivantes : 1) le robot attend du visiteur qu'il le suive jusqu'à destination ; 2) le visiteur s'attend à ce que le robot le guide.

Considérons les trois types d'incertitudes que nous avons spécifiés. Cette partie sera illustrée par la figure 6.1.

6.2.3.1 L'incertitude motivationnelle

La première source d'incertitude est liée à la motivation. Comme le disent Michael et Pacherie : "nous ne pouvons pas être sûrs de la convergence des intérêts d'un partenaire potentiel avec nos propres intérêts et donc ne pas savoir s'il existe des objectifs que nous pouvons partager et promouvoir ensemble. De plus, même si nous savons quelles sont leurs préférences actuelles et qu'elles correspondent aux nôtres, nous pouvons ne pas être sûrs de la stabilité de ces préférences." [Michael 2014]. En d'autres termes, le robot et l'humain doivent aligner leurs préférences et motivations pour effectuer la tâche conjointe. En ce sens, les engagements conjoints peuvent jouer un tel rôle en imposant des obligations d'agir à chaque agent.

Dans notre exemple, Rackham va s'occuper de guider pendant que le visiteur suit, mais tant que l'engagement n'est pas établi les deux partenaires ne peuvent pas être sûrs que l'autre partie va faire sa part. On peut considérer qu'une fois que les agents se sont engagés à effectuer la tâche, deux aspects sont à prendre en compte du côté de chaque agent. Premièrement, chaque agent doit honorer ses obligations (par exemple ici, le visiteur doit suivre et Rackham doit le guider). Deuxièmement, chaque agent doit former des attentes concernant les obligations de l'autre agent (par exemple ici : pour le visiteur, que Rackham guide et pour le robot, que l'utilisateur suive). Sur la figure est représenté le fait que Rackham et le visiteur ont un plan commun, et en orange sont représentés les obligations et attentes correspondantes liées à la motivation.

Si quelque chose change concernant ces obligations et attentes, l'agent devrait être en mesure d'afficher des signaux pour avertir l'autre des changements pertinents. En effet, si l'on considère qu'une fois engagés, les agents s'engagent à baisser les incertitudes. Si quelque chose change concernant ces obligations et les attentes associées, il faut avertir l'autre. Ce point a également été souligné par la théorie de l'intention conjointe [Cohen 1990] qui introduit la notion d'engagement par rapport au but de la tâche collaborative : "si une équipe est engagée vers un but, si l'un des membres de l'équipe vient à penser que le but est atteint alors, tant que tous les membres de l'équipe ne sont pas au courant, il garde comme but de faire connaître cet état à tous les membres de l'équipe". En d'autres termes, les engagements conjoints imposent aux deux agents l'obligation d'accomplir l'objectif général ou d'informer lorsque l'objectif est atteint ou qu'il n'est plus possible de l'accomplir.

Des incertitudes concernant la motivation peuvent également survenir lors de l'exécution de la tâche. Par exemple, côté robot, il peut arriver que Rackham décide d'annuler sa participation. Dans ce cas, le robot doit signaler la décision au visiteur, afin que le visiteur suspende ses attentes concernant la participation de Rackham. Dans un autre scénario, le robot pourrait détecter que le visiteur ne s'attend plus à être guidé (par exemple, une autre personne a interrompu la tâche de guidage). Si Rackham a toujours l'engagement de guider, le robot doit signaler au visiteur qu'il est toujours engagé. Il est également possible que nous ne considérions pas une valeur oui/non mais un intervalle (par exemple, [0..1]) ou une mesure de la qualité de l'interaction [Mayima 2021b]. Dans ce cas, nous analyserions une tendance et nous devrions essayer de conserver un certain niveau pour cette valeur, par exemple, concernant la durée de l'interruption, le besoin de donner des informations

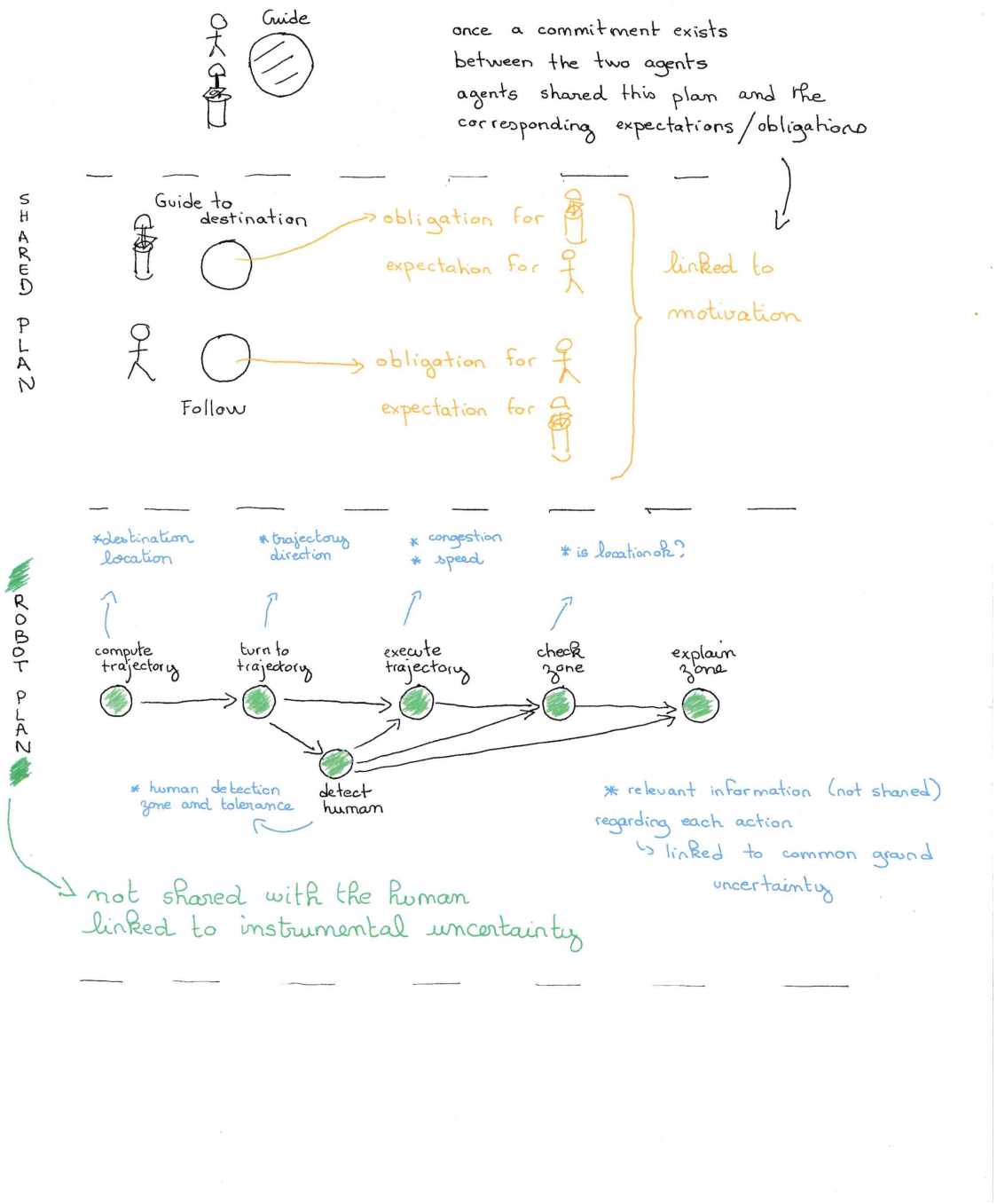


FIGURE 6.1 – Représentation de la gestion des engagements dans le cas de Rackham guide à la cité de l'espace

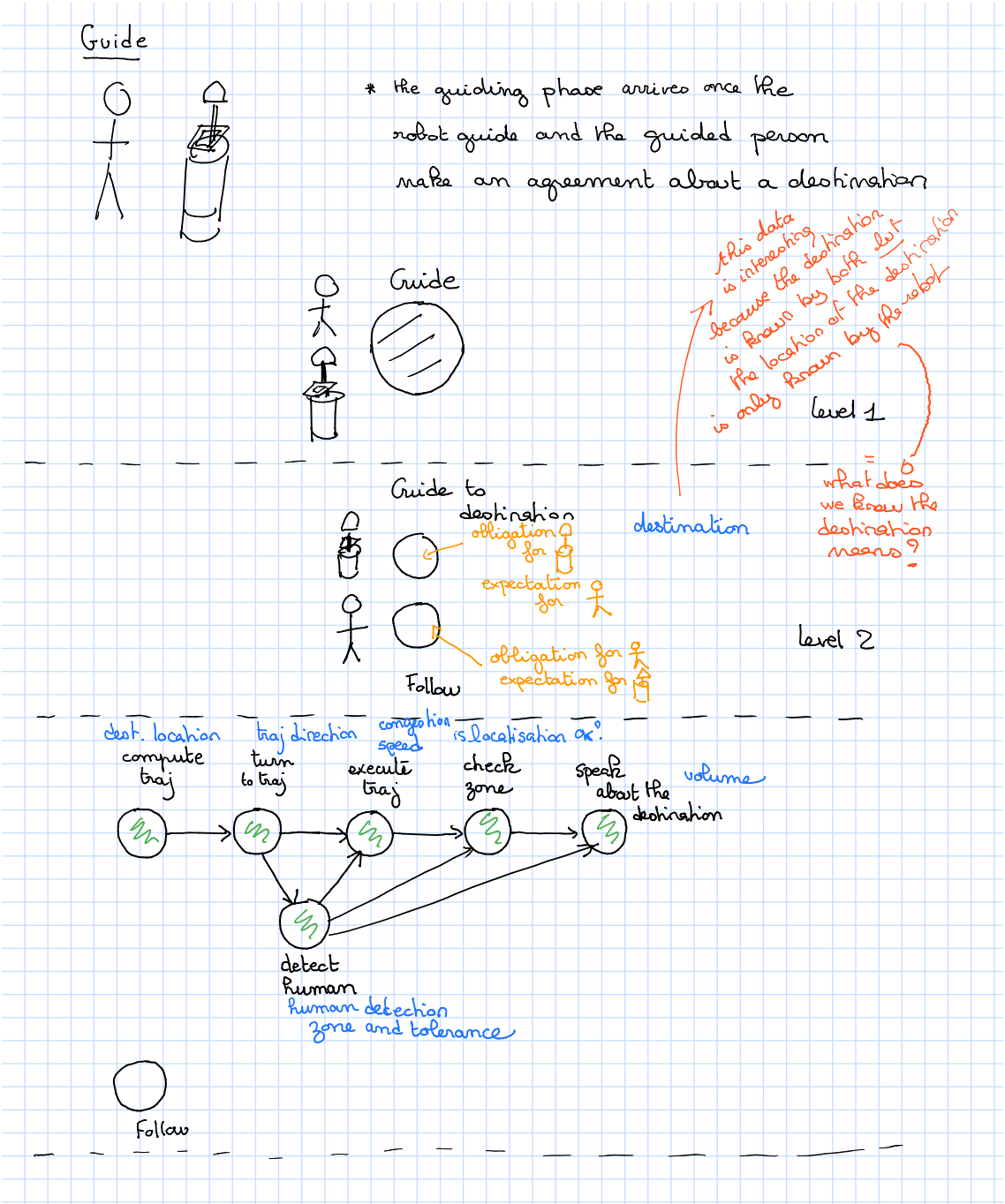


FIGURE 6.2 – Représentation de la gestion des engagements dans le cas de Rackham guide à la cité de l'espace

sur l'engagement du robot envers le visiteur serait différent.

Concernant l'engagement du visiteur à faire sa part, il peut arriver que le visiteur communique à Rackham qu'il ne fera plus la tâche. Le robot doit au moins reconnaître cette information pour mettre fin à l'engagement. Il se peut aussi que l'attente du robot sur le fait que le visiteur suive soit rompue (si le robot ne détecte plus le visiteur par exemple). Dans ce cas, le robot doit s'assurer, si possible, que l'engagement se termine également du côté du visiteur (par exemple, qu'il ne s'agit pas seulement d'un problème de perception/détection).

On peut considérer que le fait que le robot guide pendant que le visiteur suit - et les attentes correspondantes : pour le visiteur, le robot guide et pour le robot, le visiteur suit - forme un plan partagé qui est connu des deux agents (tant que l'engagement commun est en cours). Le plan pourrait être plus précis de chaque côté (par exemple, du côté de Rackham, guider quelqu'un signifierait par exemple détecter le visiteur, calculer une trajectoire, exécuter la trajectoire etc.), cependant le plan n'est partagé que dans une certaine mesure. Nous soutenons que la motivation des deux agents ne peut être remise en cause que par rapport à un objectif et un plan partagés. Par exemple, cela ne sera pas possible sur des sous-plans comme le plan du robot dans la figure 6.1, puisque ce plan n'est pas partagé avec le visiteur.

L'incertitude instrumentale Le deuxième type d'incertitude est lié aux connaissances instrumentales concernant la manière d'exécuter le plan. Nous devons remarquer que même lorsque les deux agents ont des capacités différentes pour s'assurer que les objectifs et les motivations sont correctement partagés, ils peuvent toujours être incertains de la meilleure façon de procéder ou d'exécuter les plans et les étapes nécessaires pour réaliser l'objectif [Michael 2014].

Dans notre exemple, cela se traduit par le fait que, bien que Rackham et le visiteur soient avertis du rôle de l'autre, il existe toujours une source d'incertitude au niveau du partage du plan. Le visiteur n'est pas au courant (sauf s'il s'agit du programmeur du robot mais nous ne considérerons pas ce cas ici) de la façon dont Rackham s'y prendra pour faire sa part de la tâche.

Pour faire face à cette source d'incertitude, le robot doit modéliser le fait qu'il ne partage pas avec le visiteur ses propres plans ou du moins son plan à un niveau donné. Par exemple sur la figure 6.1, le plan du robot comprend : le calcul de trajectoire, la détection de la personne, etc, et ce niveau n'est pas partagé avec le visiteur, le robot doit en tenir compte. Cela signifie que si une panne se produit à ce niveau particulier, étant donné qu'elle n'est pas partagée, le robot doit donner une explication précise de l'échec pour obtenir de l'aide ou déclencher des actions de réparation de la part du visiteur. Par exemple, un scénario possible pourrait être celui où, avant le début de la navigation, Rackham peut ne pas détecter le visiteur, il doit donc d'abord signaler ce qu'il essaie de faire afin de l'avertir. À ce titre, compte tenu de l'engagement du visiteur, on peut s'attendre à ce qu'il l'aide (répare). Une autre possibilité pourrait être que lors du calcul de la trajectoire, le calcul ait échoué. Dans ce cas, l'explication pourrait être que l'échec est sur le calcul de la trajectoire - pas sur toute la tâche - donc cela pourrait être peut-être plus acceptable/compréhensible pour le visiteur (qui serait mis dans une position où il peut partager la reconnaissance mutuelle concernant ce qui se passe).

Une autre source possible d'incertitude instrumentale peut concerner le timing, plutôt que l'exécution ou la planification en soi. Dans ce cas, si le robot dispose d'une durée

acceptable pour toutes ces actions (cette durée acceptable pourrait dépendre du visiteur), par exemple, il peut arriver que le calcul de la trajectoire soit long. Si tel est le cas, le robot devrait être capable d'envoyer un signal pour communiquer que tout va bien et que cela va juste prendre plus de temps que d'habitude.

Comme dans le cas de l'incertitude motivationnelle, nous pouvons trouver des précédents dans la littérature informatique sur la façon dont les agents pourraient gérer le type d'incertitudes générées par l'ignorance des plans du partenaire. Par exemple, la théorie des plans partagés [Grosz 1996] utilise des intentions individuelles pour établir l'engagement des collaborateurs dans leur activité commune. Il est intéressant de noter que la théorie des plans partagés propose également la notion d'engagement de l'agent envers les capacités de ses partenaires collaborateurs, ainsi que la notion de comportement utile et d'action contractuelle.

6.2.3.2 L'incertitude concernant le terrain d'entente

La troisième source d'incertitude est liée au terrain d'entente (*common ground*). Michael et Pacherie [Michael 2014] soutiennent qu'"il ne suffit pas d'assurer la coordination que nous soyons réellement motivés pour poursuivre les mêmes objectifs et que nous ayons des croyances et des plans instrumentaux suffisamment similaires concernant la manière dont ces objectifs doivent être atteints. Si vous ou moi ne pensons pas que c'est le cas, nous n'allons pas nous engager dans une action commune". À ce titre, l'action commune nécessite de s'assurer que son partenaire a les mêmes motivations et les mêmes projets que soi, mais aussi, qu'il est conscient de notre prise de conscience commune à ce sujet, de sorte que l'alignement des motivations et des plans est transparent pour les uns et les autres.

Nous avons déjà examiné l'incertitude du terrain d'entente dans la mesure où nous avons examiné comment l'expression des obligations et le partage des engagements avec les plans peuvent faciliter la réduction des incertitudes motivationnelles et instrumentales. Ici, cependant, nous aimerions également considérer une autre dimension : l'incertitude particulière du terrain d'entente qui survient lorsque le fonctionnement du robot est inconnu du visiteur ou que certains modèles de comportement du visiteur ne sont pas connus du robot. Ainsi, lorsqu'une telle incertitude commune produit une violation ou une défaillance, le robot et le visiteur peuvent réparer la situation en partageant des informations qui corrigent le problème.

En ce sens, les engagements conjoints imposent l'exigence de mettre à disposition du partenaire l'information lorsqu'elle est pertinente, et donc, elle implique différentes "variables" qui pourraient impacter la manière dont l'action est menée et la tâche accomplie. Par exemple dans la figure 6.1, chaque action du plan du robot implique des paramètres, par exemple, pour la détection de la personne, la zone de détection et la tolérance. Du côté du visiteur, toutes ces variables sont cachées mais il pourrait peut-être être intéressant, à un moment donné, de les partager chaque fois que cela est nécessaire ou possible en cas de problème. En effet, par exemple, si lors de la navigation le visiteur ne suit pas, cela peut être à cause de la vitesse du robot ou de la forme de sa trajectoire, ainsi, avant de considérer qu'une action a échoué, le robot pourrait interroger la personne sur ces paramètres (et décider avec elle s'il est nécessaire de les régler différemment). Là encore, le partage d'informations permettrait d'entrer dans une reconnaissance mutuelle de ce qui se passe. Cependant, nous considérons que ces informations n'ont pas besoin d'être partagées dès

le départ ou dans le cas où tout va bien pour ne pas surcharger l'interaction.

L'accent mis sur la compréhension d'un terrain d'entente en termes d'engagements et son importance pour la communication se retrouve également dans les travaux d'Herbert H. Clark [Clark 1996b, Clark 1991, Clark 2006]. Comme il conclut avec sa collègue Susan E. Brennan "Une fois que nous avons formulé un message, nous devons faire plus que simplement l'envoyer. Nous devons nous assurer qu'il a été compris comme nous l'avions prévu. Sinon, nous avons peu d'assurance que le discours auquel nous participons se déroulera de manière ordonnée" [Clark 1991]. Par ailleurs, il propose également de considérer d'un côté l'activité conjointe de base ou l'activité conjointe proprement dite et d'autre part les actions conjointes de coordination basées sur des actes de communication [Clark 2006]. En d'autres termes, il distingue entre la coordination et les signaux communicatifs qui explicitent les engagements impliqués dans le terrain d'entente nécessaire à la réalisation de l'action commune.

6.2.3.3 Modélisation des aspects normatifs et descriptifs

L'analyse philosophique ci-dessus a souligné qu'un aspect fondamental des engagements est son aspect normatif. Dans les interactions humaines, cet aspect normatif de l'engagement motive de nombreux signaux réglementaires et des stratégies de réparation qui facilitent l'exécution de l'objectif commun. Nous allons analyser comment différents aspects normatifs (et descriptifs) des engagements sont modélisés dans notre cadre.

Dans les sections précédentes, nous avons vu comment le partage de l'information est fondamental pour une exécution réussie de l'action conjointe - par exemple, l'objectif commun et le plan partagé -, tandis que d'autres informations - par exemple, la partie individuelle du plan - ne le sont pas. Les informations partagées ou non sont particulièrement importantes dans l'action conjointe humain-robot, compte tenu des particularités des incertitudes générées par l'opacité du comportement du robot. Une conclusion similaire peut être tirée des conséquences normatives de ces informations et de la manière dont elles déclenchent des signaux et des réactions de régulation du robot.

À notre avis, une hypothèse par défaut raisonnable serait que si les informations partagées doivent être considérées comme impliquant des engagements, et donc les attentes générées par ces engagements traitées comme normatives, les informations qui ne sont pas partagées, mais qui peuvent générer certaines attentes doivent être traitées comme descriptives. En d'autres termes, le robot doit réagir à la frustration face aux attentes générées par les informations partagées - associées à l'objectif commun et au plan partagé - en utilisant des signaux réglementaires ou des stratégies de réparation alors qu'il ne doit pas réagir de la même manière face aux attentes générées par des informations qui ne sont pas partagées.

Dans notre cadre, quelle serait la différence entre faire face à une attente descriptive ou normative ? Dans l'exemple de la figure 6.1, le plan du robot pour remplir sa part de tâche n'est pas partagé avec le visiteur. Imaginez qu'à la fin de l'interaction, Rackham donne des explications sur la destination, cela génère l'attente que le visiteur restera pour écouter l'explication. Le plan n'étant pas partagé, l'attente que le visiteur écoute est considérée comme descriptive. En conséquence, Rackham ne produirait aucun signal réglementaire si le visiteur partait ; c'est-à-dire qu'il ne considérerait pas négativement le fait que l'attente n'a pas été satisfaite. D'un autre côté, si le visiteur reste et écoute, Rackham devrait être capable de gérer l'interaction jusqu'à la fin et considérer positivement le fait que l'attente

a été remplie.

En résumé, face à une attente descriptive, le robot doit être capable d'envisager les conséquences de l'attente lorsqu'elle se réalise alors qu'il ne doit pas réagir négativement, et donc, ni produire de signaux régulateurs ni réparer lorsque l'attente est frustrée. En revanche, face à une attente normative, le robot doit être capable de produire des signaux réglementaires et entrer en mode réparation si besoin. De plus, il faut remarquer que la même différence s'applique à chaque fois que nous considérons une attente du robot concernant l'action du visiteur et si nous considérons une attente du visiteur qui devrait être remplie par le robot.

6.3 Socialité, engagements et horizontalité : remarques de conclusion

Dans ce chapitre, dans une première partie, nous avons donné des pistes pour mettre en place une stratégie interprétative "minimale" de la socialité et avons insisté sur l'importance de la reconnaissance mutuelle dans ce processus. Ensuite, nous avons expliqué comment nous pensons que la prise en compte de la notion d'engagement ainsi que les attentes et les signaux qui y sont associés pourrait fournir un cadre explicatif dans une architecture robotique dédiée à l'interaction. En effet, par exemple, lorsque nous avions déployé le robot en 2002, nous avions intégré un comportement où lorsque le chemin était long, le robot se retournait et s'arrêtait pour vérifier que l'humain était toujours présent. Notre cadre donne une explication à ce comportement que nous avions codé à la suite de nos expérimentations. Il n'est pas nécessaire d'attendre qu'une attente soit frustrée pour devoir générer un signal. Cela peut avoir un sens de le faire si une action est plus longue que prévue.

Le rôle central des engagements n'est pas seulement fonctionnel, dans le sens où ils aident et/ou facilitent la coordination entre les individus mais aussi, ils motivent un ensemble d'outils communicatifs qui aident les individus à savoir à quoi s'attendre les uns par rapport aux autres dans le cadre d'une action jointe. Ce sont précisément ces types de signaux sociaux et de comportements liés aux obligations sociales et aux réactions aux réussites ou aux échecs qui font des engagements un concept intéressant pour la conception de robots socialement appropriés.

Il y a deux idées importantes que nous aimerions introduire ici et qui, à notre avis, sont fondamentales pour de futures recherches.

Premièrement, comme toute autre capacité, la gestion aussi bien du contexte social que des engagements dans le domaine de l'interaction humain-robot nécessite une perspective intégrative. La façon dont nous comprenons comment aborder de telles expériences holistiques et contextuelles passe par le concept d'horizontalité qui renvoie à l'idée de considérer l'action jointe dans son ensemble lors de la modélisation et de la conception des capacités robotiques. Comme exprimé par Belhassein, Fernandez-Castro et Mayima, avoir une approche horizontale signifie que "les actions jointes en HRI ne doivent pas être abstraites de l'ensemble du contexte et de l'arrière-plan où la tâche collaborative particulière est intégrée et située dans… [Nous devons] prendre en considération la façon dont les différentes informations contextuelles, les capacités robotiques et humaines et le contexte interagissent de manière complexe avec les dispositifs [robotiques] en question" [Belhassein 2020].

Deuxièmement, on ne peut pas faire l'économie d'intégrer l'utilisateur dans la boucle.

Il faut lui permettre de comprendre les possibilités et aussi les limites que l'on donne au robot (par exemple ici en terme de socialité ou en matière de gestion des engagements). Il y a des codes à créer et à intégrer. C'est pourquoi nous pensons que ce concept de reconnaissance mutuelle est clé et qu'un défi sera de la mettre en oeuvre conjointement du côté du robot et aussi de la personne avec laquelle il interagit.

6.4 Contributions Associées

[Fernandez Castro 2019] Victor Fernandez Castro, Aurélie Clodic, Rachid Alami, Elisabeth Pacherie. Commitments in Human-Robot Interaction. 2019.
`https://jeannicod.ccsd.cnrs.fr/ijn_03084105/file/Fernandez-Castro-Clodic-Alami-Pacherie-Commitments-in-HRI-AI-HRI%202019.pdf`

[Clodic 2019] Aurélie Clodic, Rachid Alami, Elisabeth Pacherie, Víctor Fernández Castro. Commitments in Human-Robot Interaction. 8th Joint Action Meeting, Jul 2019, Genova, Italy. 2019.
`https://hal.archives-ouvertes.fr/LAAS-IDEA/ijn_03084105v1`

[Fernández Castro] (A venir) Víctor Fernández Castro, Amandine Mayima, Kathleen Belhassein, Aurélie Clodic (forthcoming) "The Role of Commitments in Socially Appropriate Robotics" In Gransche, M. Nerurkar, J. Bellon, S. Nähr-Wagener (eds) Technology Socialization ? Social appropriateness for computer-based systems. Springer

— Les deux premières contributions reprennent nos propositions sur l'apport que peut avoir la modélisation de la notion d'engagement (commitment) dans le cadre d'une interaction humain-robot. La dernière, une version plus élaborée, est en cours de révision pour une publication dans un ouvrage.

[Fernández Castro 2020] Víctor Fernández Castro, Raul Hakli, Aurélie Clodic. What does it take to be a social agent ?. Appeared in : Nørskov, M., Seibt J., Quick O. 2020. Culturally Sustainable Social Robotics—Proceedings of Robophilosophy 2020., Series Frontiers of AI and Its Applications, IOS Press, Amsterdam., 2020, 10.3233/FAIA200954 .
`https://hal.archives-ouvertes.fr/hal-02963044/document`

— Cet article est le résultat d'un travail commun lors de l'organisation d'un workshop pour Robo-Philosophy 2020. Le but de cet article est de présenter une liste (d'inspiration philosophique) d'exigences pour permettre la mise en place d'une stratégie d'interprétation sociale dans un cadre d'interaction humain-robot. Une telle liste ne vise pas à détailler les processus cognitifs derrière la socialité mais à fournir une caractérisation (sans mise en œuvre) des capacités et compétences associées à la socialité. Nous utilisons la notion de posture intentionnelle de Dennett comme fondement méthodologique pour étudier l'agentivité intentionnelle et l'étendre à une posture sociale qui prend en compte les caractéristiques sociales du comportement. Nous discutons des exigences de base de la socialité et des différentes manières de les comprendre, et suggérons quelques avantages potentiels à les comprendre de manière instrumentale dans le contexte de la robotique sociale.

[Belhassein 2022] Kathleen Belhassein, Víctor Fernández-Castro, Amandine Mayima, Aurélie Clodic, Elisabeth Pacherie, et al.. Addressing joint action challenges in

HRI : Insights from psychology and philosophy. Acta Psychologica, Elsevier, 2022, 222, pp.103476. 10.1016/j.actpsy.2021.103476 .
https://hal.archives-ouvertes.fr/hal-03505047

— Cet article est l'aboutissement du travail mené lors du projet JointAction4HRI. Après avoir exploré la littérature actuelle sur l'action conjointe en interaction humain-robot, aboutissant à une définition plus précise de ces enjeux, le présent article propose quelques perspectives empruntées à la psychologie et à la philosophie montrant le rôle clé de la communication dans les interactions humaines. De la reconnaissance mutuelle entre individus à l'expression de l'engagement et des attentes sociales, nous soutenons que les signaux de communication peuvent faciliter la coordination, la prédiction et la motivation dans le cadre d'une action conjointe.

Chapitre 7

Faire les choses bien

7.1 Introduction

Le domaine de la psychologie auquel on va s'intéresser maintenant en parlant des études utilisateurs, couvre un spectre très large allant de l'ergonomie à la psychologie du développement en passant par la psychologie sociale. C'est une chose que m'ont apprise mes études de psychologie et aujourd'hui je sais faire la différence. Mais il faut bien avoir conscience que selon la discipline c'est un peu comme si vous parliez d'une part à une spécialiste du contrôle d'altitude chez les drones et d'autre part à une spécialiste du contrôle de la marche bipède sur les robots humanoïdes. Cela n'a, tout simplement, rien à voir.

Mes premiers souvenirs d'études utilisateurs datent du projet Cogniron (cf section 2.6). Dans le cadre des travaux sur la planification et l'exécution de mouvements prenant en compte l'humain (*human-aware motion*) et plus particulièrement la navigation, une collaboration avait été mise en place avec l'équipe de Kerstin Dautenhan pour réaliser des études avec des utilisateurs. Cela avait amené à plusieurs publications communes, e.g. [Sisbot 2005] et [Dautenhahn 2006]. Je n'étais pas directement impliquée, mais c'était la première fois que de telles études étaient réalisées par l'équipe et c'est peu dire que la communication a été compliquée au départ. Les études utilisateurs, quand elles sont réalisées de manière sérieuse, demandent la définition d'un cadre très précis, des questions, des hypothèses, des variables qui ne souffrent pas l'approximation. C'est parfois frustrant pour nous roboticiens d'avoir l'impression de ne tester qu'une petite partie de ce qui nous intéresse.

Très franchement, je me rappelle me demander à l'époque si l'on ne mettait pas la charrue avant les bœufs. J'avais un peu l'impression qu'on cherchait à choisir la couleur des sièges passagers avant de savoir faire voler l'avion... C'est je dois dire une impression qui me poursuit encore quelquefois dans les couloirs des conférences de mon domaine, mais là n'est pas la question.

Ces études, datant donc de 2005 et 2006, sont ressorties des années plus tard sous mon radar. Nous avions un étudiant de M2 en psychologie à qui nous avions demandé d'étudier la bibliographie des études utilisateurs sur la navigation et l'interaction humain-robot. Et lors d'un point, il me montre un slide avec d'après [Dautenhahn 2006] d'une part "une grande majorité de sujets humains, lorsqu'ils sont assis, préfèrent qu'un robot s'approche du côté gauche ou droit" et d'autre part "l'approche frontale a été considérée comme inconfortable, peu pratique, dans certains cas même menaçante ou conflictuelle, et devrait donc être évitée" (*"Results from the two HRI trials indicate that a large majority of human subjects, when seated, preferred a robot to approach from either the left or right side. The frontal approach was seen as uncomfortable, impractical, in some cases even threatening or confrontational, and should thus be avoided."[Dautenhahn 2006]*). Alors bien sûr c'est notre travail d'expliquer à l'étudiant comment il faut lire et interpréter un article (et c'est ce que nous avons fait avec lui bien sûr), mais c'est aussi notre travail de faire attention à ne pas rendre des conclusions trop définitives, trop générales qu'il sera ensuite difficile

de venir contredire ("et bien on se demande bien pourquoi vous avez fait cette étude, il y en a déjà une qui le disait déjà.."). J'ai pu analyser cela rapidement, car je connaissais ces travaux, et leurs limites notamment sur le type de plate-forme utilisé et le nombre de participants, mais j'aurais pu passer à côté sinon.

Dans ce chapitre, je vais dans un premier temps, expliquer pourquoi l'évaluation avec des utilisateurs dans le domaine de la robotique est compliquée en me basant sur l'analyse de James E. Young et al. [Young 2010]. Ensuite, en prenant appui sur les travaux d'Astrid Weiss [Weiss 2011], je vais expliquer les différentes perspectives qui sont en concurrence lorsque l'on travaille sur une étude utilisateur et pourquoi il est important d'avoir un dialogue informé avec l'ensemble des disciplines en présence. Ensuite, en me basant sur un travail effectué avec Kathleen Belhassein et Guilhem Buisan [Belhassein 2019] je vais expliquer quels sont les travers à éviter lors de la réalisation d'études utilisateurs en robotique et les points auxquels il faut être attentif. Je conclurais en introduisant une tâche benchmark, nécessaire pour permettre la réplicabilité, la tâche du Directeur [Sarthou 2021b] et en faisant quelques remarques.

7.2 Pourquoi ce domaine de l'interaction personne-robot est unique ?

C'est quelque chose dont nous avons tous conscience, mais qui est très bien expliqué par James E. Young et al. [Young 2010]. Les gens ont naturellement tendance à traiter les robots de la même manière qu'ils peuvent traiter des créatures vivantes. Les gens attribuent facilement une intentionnalité aux actions des robots, quelles que soient leurs capacités réelles ou leurs conceptions explicites. Ainsi pour beaucoup de gens, interagir avec un robot ressemble plus à interagir avec un animal ou une autre personne qu'avec une technologie - le robot est perçu comme un acteur physique et social actif dans notre monde quotidien. Pour cette raison, les gens sont prêts à percevoir les robots comme prenant des décisions autonomes et intelligentes. Cela contribue aussi au développement d'attentes (souvent irrationnelles) concernant les capacités des robots (non lié à leurs capacités réelles). Cette expérience que les gens vivent (ou croient vivre) lorsqu'ils sont confrontés à nos robots est problématique lorsque l'on cherche à évaluer une interaction. Cependant, on doit s'attacher à réduire le problème en donnant aux gens les clés pour appréhender nos robots et leurs limites.

7.3 Qu'est ce qu'on cherche à évaluer finalement ?

Tout article étiqueté interaction humain-robot ou presque doit aujourd'hui s'accompagner d'une étude utilisateur sous peine de se faire rejeter sur cet unique fait. À mon sens, ce n'est pas équitable, il y a souvent beaucoup de travail dans la conception et l'implémentation d'un système en soi, surtout s'il est ambitieux et devoir y adjoindre une étude utilisateur, outre que cela prend du temps (et de la place dans un article), n'a pas toujours de sens de prime abord. In fine, plus un système est simple, plus il sera facile à expliquer et à évaluer, et l'on donne la prime à des études souvent très morcelées. Aujourd'hui, des conférences comme HRI (*Human-Robot Interaction Conference*) mettent en avant différents types de soumissions pour éviter ces travers (`https://humanrobotinteraction.org/2022/guides-to-submission-types/` renvoyant vers `https://chi2021.acm.org/`

`for-authors/presenting/papers/guide-to-a-successful-submission`), mais il est à mon sens toujours latent. On pourrait également imaginer évaluer d'une part la qualité technique de la contribution (est-ce que c'est une idée qui fait avancer le domaine) et d'autre part son évaluation par les utilisateurs et que ces deux types d'évaluation soit faites d'une part par les roboticiens pour la partie technique et d'autre part par les psychologues pour la partie évaluation.

Cependant, il faut aussi nous remettre en question et se poser la question de ce qu'on cherche à évaluer. Il y a une grande hétérogénéité des plates-formes, peut-on comparer des recherches effectuées sur Pepper et des recherches effectuées sur PR2 ? Veut-on évaluer le système robotique lui-même, en interaction avec des utilisateurs dans un cadre donné ou est-ce qu'on souhaite évaluer un algorithme particulier (e.g. de calcul de trajectoire pour la navigation d'un robot de manière autonome en environnement humain ; de planification de tâches pour savoir si les décisions prises par le robot sont adéquates ; de gestion d'un dialogue lors d'une collaboration humain-robot ; etc..) ?

Le plus important c'est aussi de comprendre que les psychologues et les philosophes peuvent avoir des lignes de recherche différentes que celles des roboticiens. Ils n'ont pas forcément les mêmes attentes et les mêmes vues sur ce qu'ils souhaitent évaluer.

Pendant ma thèse j'avais été frappé par les travaux qu'Astrid Weiss notamment sur le robot HRP2 et la définition du cadre USUS pour l'évaluation de l'interaction humain-robot (*The USUS Evaluation Framework for Human-Robot*[Weiss 2011]). Le cadre d'évaluation proposé est basé sur un ensemble d'indicateurs ciblant les facteurs d'utilisabilité, d'acceptation sociale, d'expérience utilisateur et d'impact sociétal comme objectifs d'évaluation. Les facteurs sont sélectionnés pour identifier des scénarios de travail collaboratif socialement acceptables où des robots humanoïdes peuvent être déployés de manière bénéfique pour convaincre la société de soutenir positivement l'intégration de robots humanoïdes dans un environnement de travail humain. La motivation principale pour choisir ces facteurs est de soutenir les approches d'évaluation centrées sur l'utilisateur en HRI, allant au-delà des études d'utilisabilité pures. Ainsi, ce cadre cherchait à répondre à une question plus générale qui était de savoir si les gens sont prêts à accepter de tels robots dans la société et permettait ainsi de donner une vision holistique de l'évaluation des robots humanoïdes (*(the proposed framework) intends to be applied to answer the general question if people experience robots as a support for cooperative work and accept them as part of society and thus give an holistic view on evaluating humanoid robots [Weiss 2011]*). Je trouve que ce travail montre bien le "pas de côté" que l'on fait lorsque l'on se place dans une perspective différente de celle d'une équipe de roboticiens développant une application derrière leurs claviers. Ce sont des éléments que l'on retrouve également dans les travaux de Johanna Seibt qui a développé l'approche Robotique Sociale Integrative (*Integrative Social Robotics (ISR)*) [Seibt 2016, Seibt 2018, Fischer 2020].

7.4 Quels moyens pour évaluer ?

Nous ne parlerons ici que des questionnaires.

À côté de questionnaires maisons réalisées de manière ad hoc pour des tâches précises, il existe déjà des questionnaires permettant d'évaluer une interaction humain-robot.

Ainsi, le questionnaire Godspeed [Bartneck 2008] permet de mesurer la perception du robot par l'humain avec des questions relatives à l'anthropomorphisme ou à l'intelligence perçue. Ce questionnaire est centré sur l'évaluation de la perception du robot et ne traite

pas de l'évaluation de l'interaction et de l'utilisabilité du système.

Le questionnaire ROSAS (pour RObotic Social Attributes Scale) [Carpinella 2017] s'inspire du questionnaire Godspeed et de la littérature de la psychologie sociale et mesure le jugement et la perception des personnes par rapport aux attributs sociaux du robot, comme la chaleur ou la compétence.

Même si ces questionnaires sont intéressants, ils sont peu utiles lorsque l'on cherche à évaluer les décisions de haut niveau prises par le robot. Par exemple, si l'on considère les décisions pour répondre à ces questions dans le cadre d'une collaboration : quelle action effectuer pour faire avancer la collaboration, qui la réalise et quand ? L'objectif ici va être d'évaluer la capacité du robot à prendre les bonnes décisions au bon moment (et on l'espère obtenir alors une évaluation positive de la personne avec laquelle il interagit). C'est dans ce cadre que Camille Vrignaud, Sandra Devin et al. ont proposé dans [Devin 2018] le questionnaire PeRDITA (*(Pertinence of Robot Decisions In joinT Action*) visant à évaluer la pertinence des décisions du robot lors d'une action jointe. On cherche ici à évaluer si le robot prend les bonnes décisions, pas s'il est smart, mignon, sympa, ou autre, c'est une différence de perspective importante.

À côté de ces questionnaires relatifs au domaine de l'interaction humain-robot, les études utilisateurs ne se privent pas d'utiliser des questionnaires ou tests psychologiques existants. Cela n'est pas sans poser problème.

Par exemple, cela peut paraître une bonne idée d'utiliser un questionnaire de personnalité pour évaluer les utilisateurs dans une interaction humain-robot, mais il faut surtout y réfléchir à deux fois. Chacun des questionnaires existants et validés en psychologie le sont dans un cadre très précis pour une utilisation elle aussi très précise et pour donner des indications sur des aspects particuliers dans un cadre défini. La validité d'un questionnaire ou d'un test permet de s'assurer qu'on va mesurer effectivement ce qu'on est censé mesurer, mais il faut prendre en compte les dimensions et les hypothèses dans lesquelles cette validité est assurée.

Il existe bien entendu des tests et des questionnaires utilisés dans d'autres disciplines que l'on peut utiliser, mais il faut le faire à bon escient. Ces tests mesurent quelque chose de précis et il faut les utiliser dans ce cadre.

De même, si un questionnaire est validé dans une langue, cela ne suffit pas que quelqu'un réalise une traduction, il y a toute une procédure à suivre pour valider un questionnaire dans une nouvelle langue. C'est malheureusement l'un des travers d'un des tests très utilisé en HRI aujourd'hui. Les traductions ont été réalisées par des chercheurs, sans contrôle, et nous nous trouvons aujourd'hui avec des problèmes de traduction insolubles.

7.5 Comment faire les choses bien ? (ou pour le mieux)

Avec Kathleen Belhassein et Guilhem Buisan, nous avons proposé un certain nombre d'éléments qu'il est important d'avoir en tête lorsque l'on réalise une étude utilisateur dans le domaine de l'interaction humain-robot [Belhassein 2019].

1. Plus il y a d'utilisateurs, mieux c'est. Cela améliore la validité statistique et cela peut effacer certains biais.
2. Élargir le recrutement. Vos collègues en savent long sur la technologie : recrutez au hasard des personnes dans votre boulangerie, dans le supermarché, à l'aide de flyers...

3. Soyez rigoureux avec votre protocole. Il y a tellement de variables incontrôlées indésirables. N'en soyez pas une !

4. Laisser l'utilisateur s'habituer au robot et à son comportement avant de commencer l'expérience. Afin de diminuer l'effet "wouaw" et rapprocher les mesures d'une utilisation à plus long terme.

5. Assurez-vous que votre expérience est physiquement et psychologiquement sûre. En plus de blesser un utilisateur, vous risquez de biaiser vos mesures si l'expérimentateur est stressé à propos de quelque chose qui ne va pas.

6. Mesurez objectivement si votre robot est utile. Un robot peut être très satisfaisant, mais il sera vite oublié s'il ne fait rien.

7. Utilisez les bons outils pour vos mesures. Des outils utilisés largement et standardisés donneront de la crédibilité à votre étude. Les questionnaires ne sont pas une fin en soi.

8. Utilisez (ou réalisez) des outils et validez les théoriquement vis-à-vis du domaine de l'interaction humain-robot (et publiez-les s'ils n'existent pas). Cela profitera à toute la communauté.

9. Donnez le plus de détails possible sur votre étude. Donnez le code source, les schémas matériels ou les références et la description de l'environnement afin de faire en sorte que d'autres personnes puissent reproduire votre expérience.

10. Lorsque vous faites un Magicien d'Oz, soyez rigoureux. Émulez uniquement un petit composant non évalué de votre robot. Écrivez les règles que vous suivez, respectez-les et publiez-les avec votre étude.

À cette liste, ils ajoutent plusieurs remarques concernant les conclusions que l'on peut donner à une étude utilisateur en indiquant qu'il faut savoir rester humble et éviter les phrases chocs pour décrire un résultat. Une étude n'est généralement réalisée que dans un cadre précis, avec un robot particulier et des utilisateurs qui le sont tout autant, il ne fait pas sens de tirer des conclusions générales de ce genre d'étude. Ainsi on évitera les phrases trop générales du style : "lorsqu'un robot navigue parmi les humains il doit forcément faire ce comportement plutôt que celui-là, car.." . Dites plutôt, "avec ce robot particulier, dans ce contexte particulier, dans cette tâche particulière, dans cet environnement particulier, alors nous avons observé que..". Donner des conclusions générales peut avoir des conséquences désastreuses parce que si votre article est publié, il pourra être pris comme référence et opposé à d'autres chercheurs. C'est ainsi que parfois des conclusions générales issues d'une étude réalisée dans un contexte très restreint deviennent des généralités et empêchent ensuite certaines recherches de se développer (pourquoi continuer à faire des recherches sur le sujet puisque tel article établit que...). J'ajoute qu'en tant que reviewer de conférences ou de journaux nous avons la responsabilité de ne pas laisser passer cela.

Il existe aujourd'hui des publications comme le livre édité par Céline Jost, Human-Robot Interaction Evaluation Methods and Their Standardization [Jost 2020] ou des sites web, e.g : `https://tianleimin.github.io/HRI-Methodology-Guidelines/main` et sa version interactive `https://tianleimin.github.io/HRI-Methodology-Guidelines/` pour appréhender le sujet et découvrir les bonnes pratiques.

7.6 Et alors ?

L'un des problèmes posés par l'évaluation est la réplicabilité. Nous avons un spectre de plates-formes et de logiciels associés très larges et chacun nous définissons et redéfinissons des scénarios et des démos au fil de nos projets. Avec Guillaume Sarthou, Amandine Mayima, Guilhem Buisan et Kathleen Belhassein, nous avons proposé dans The Director Task : a Psychology-Inspired Task to Assess Cognitive and Interactive Robot Architectures [Sarthou 2021b] d'utiliser la tâche du Directeur comme une tâche de benchmark pour l'étude de l'interaction humain-robot. Dans cet article, nous proposons une description détaillée de la tâche mise en place, une implémentation et des idées de recherches qui pourraient être menées grâce à cette tâche.

Je voudrais juste finir en disant que je considère que je ne sais toujours pas faire une étude utilisateur. Je connais les bases, je sais lire un protocole, mais je ne suis pas à l'aise avec l'analyse des données et l'interprétation. Je reste persuadée qu'on doit rester humble et savoir considérer là où notre expertise s'arrête. Au jour d'aujourd'hui, si je veux faire une étude utilisateur, je m'assure d'avoir les soutiens pour la faire.

7.7 Contributions associées

[Gharbi 2015] Mamoun Gharbi, Pierre-Vincent Paubel, Aurélie Clodic, Ophélie Carreras, Rachid Alami, and Jean-Marie Cellier. Toward a better understanding of thecommunication cues involved in a human-robot object transfer. In 24th Inter-national Symposium on Robot and Human Interactive Communication, IEEERO-MAN, pages 319–324, Kobe, Japan, August 2015. DOI : 10.1109/ROMAN.2015.7333626 URL : `https://hal-laas.archives-ouvertes.fr/hal-01628249/`

— Cet article est le résultat de la première étude utilisateur à laquelle j'ai participé et qui a aboutit. C'est un travail collaboratif qui a été mené à bien par Mamoun Gharbi avec des membres de l'équipe CLLE de l'université Jean Jaurès, Pierre-Vincent Paubel, Ophélie Carreras et Jean-Marie Cellier. Il présente une étude préliminaire sur le comportement de la tête d'un robot lors d'un geste de don. Nous avions mis en évidence un pattern terminant par Objet - Tête du receveur, laissant penser que lorsque le robot donne un objet à quelqu'un (ici à travers un dépot sur une table) il serait préférable à la fin du geste du bras de regarder l'objet puis le visage de la personne à qui l'on donne l'objet.

[Belhassein 2017] Kathleen Belhassein, Aurélie Clodic, Hélène Cochet, Marketta Niemelä, Päivi Heikkilä, Hanna Lammi, and Antti Tammela. Human-Human Guidance Study, December 2017 URL : `https://hal.laas.fr/hal-01719730/`

— Dans ce rapport est présenté le travail préliminaire mené sur les études humain-humain lors de l'analyse de la tâche de guidage avec les véritables guides travaillant à Ideapark.

[Devin 2018] Sandra Devin, Camille Vrignaud, Kathleen Belhassein, Aurélie Clodic, Ophélie Carreras, and Rachid Alami. Evaluating the Pertinence of Robot Decisions in a Human-Robot Joint Action Context : The PeRDITA Questionnaire. In 27th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2018), Nanjing, China, August 2018 DOI : 10.1109/RO-MAN.2018.8525785 URL : `https://hal.laas.fr/hal-01906101`

— Cet article est le résultat du travail du master de Psychologie (parcours Ergonomie Cognitive, Innovation Technologique et Facteur Humain (ECIT-FH)) de Camille Vrignaud en collaboration avec Sandra Devin (alors en thèse). Camille a proposé un questionnaire original et avec Sandra elles ont monté et réalisé une étude utilisateur pour permettre l'évaluation d'une partie du travail de thèse de Sandra.

[Belhassein 2019] Kathleen Belhassein, Guilhem Buisan, Aurélie Clodic, and Rachid Alami. Towards methodological principles for user studies in Human-Robot Interaction. In Test Methods and Metrics for Effective HRI in Collaborative Human-Robot Teams Workshop, ACM/IEEE International Conference on Human-Robot Interaction, Daegu, South Korea, March 2019. URL : `https://hal.laas.fr/hal-02282600`

— Dans cet article à deux voix, Kathleen et Guilhem ont mis en ordre l'ensemble des éléments de discussion qui sont apparus lorsque nous discutions autour des études utilisateurs envisagés ou en cours. On y trouve les travers, ce à quoi il faut faire attention et quelles sont les bases à connaître pour réaliser une étude utilisateur correctement.

[Sarthou 2021b] Guillaume Sarthou, Amandine Mayima, Guilhem Buisan, Kathleen Belhassein, and Aurélie Clodic. The Director Task : a Psychology-Inspired Task to Assess Cognitive and Interactive Robot Architectures. In 30th IEEE InternationalConference on Robot and Human Interactive Communication (RO-MAN 2021), Vancouver (Virtual), Canada, August 2021 DOI : 10.1109/RO-MAN50785.2021.9515543 URL : `https://hal.science/hal-03327222/`

— Dans cet article est présenté le travail autour de la tâche du Directeur comme proposition de tâche Benchmark pour l'étude de l'interaction humain-robot. De manière intéressante, on trouvera aussi un chapitre sur cette tâche dans chacune des thèses des auteurs. Un très joli travail d'équipe.

[Heikkilä 2019] Päivi Heikkilä, Hanna Lammi, Marketta Niemelä, Kathleen Belhassein, Guillaume Sarthou, Antti Tammela, Aurélie Clodic, and Rachid Alami. Should a robot guide like a human ? A qualitative four-phase study of a shopping mall robot. In The Eleventh International Conference on Social Robotics (ICSR2019), Madrid, Spain, November 2019. URL : `https://hal.laas.fr/hal-02285035`

— Faisant suite à [Belhassein 2017], cet article présente les résultats des études préliminaires réalisés avec nos collègues de VTT, Päivi Heikkilä, Hanna Lammi, Marketta Niemelä et Antti Tammela. L'écriture a été compliquée car nous n'étions pas toujours en accord sur la rigueur des méthodes utilisées et sur les conclusions qui étaient faites. Il faut savoir faire des compromis parfois car on ne maîtrise pas tout. Cela donne quand même une perspective intéressante sur les travaux réalisés.

Chapitre 8

Penser aux conséquences

8.1 Introduction

Il est nécessaire de penser aux conséquences de l'introduction dans nos sociétés des systèmes que nous développons et in fine il serait même mieux de les anticiper. Cependant, ce n'est pas forcément possible. Selon Johanna Seibt, la robotique sociale est confrontée à ce qu'elle appelle la « triple impasse de description, d'évaluation et de réglementation » [Seibt 2018]. Cette triple impasse, qu'elle présente comme basée sur le dilemme de Collingridge, se présente comme suit :

(i) Actuellement, nous manquons d'une compréhension théorique complète, systématiquement intégrée et détaillée de l'interaction humain-robot - une description théorique des «mécanismes» cognitifs, des expériences humaines et des processus de «création de sens» pratiques et conceptuels dans l'interaction humain-robot.

(ii) Sans une telle théorie descriptive, nous ne pouvons même pas commencer à évaluer les applications de la robotique sociale en ce qui concerne leurs impacts sur le bien-être individuel et social.

(iii) Sans évaluations fondées sur la recherche, une réglementation responsable du marché de la robotique et une élaboration responsable des politiques ne peuvent pas démarrer. Étant donné que les réglementations existantes interdisent les études à long terme nécessaires à une théorie descriptive pleinement pertinente de l'interaction homme-robot, la triple impasse est complète.

Pour répondre à ce dilemme, l'approche Robotique Sociale Integrative (*Integrative Social Robotics (ISR)*) a été développée par Johanna Seibt et al. ([Seibt 2016, Seibt 2018, Fischer 2020]). L'approche ISR propose que les applications de robotique sociale soient développées lors d'un processus de collaboration continue entre des équipes interdisciplinaires et pose cinq principes :

— le principe de processus (*the process principle*) : ce sont les interactions et pas les robots qui sont la cible du processus de conception (*Interactions, not robots as the target of the design process*)

— le principe de qualité (*quality principle*) : avoir un large champ interdisciplinaire. Il est important d'inclure des disciplines qui peuvent approfondir notre compréhension et notre conception des interactions sociales et socioculturelles ainsi que le contexte de connaissances spécifiques.

— le principe de complexité (*the complexity principle*) : se mettre en perspective à différents niveaux (*First, Second and Third person perspective*). Cela signifie par exemple se poser les questions suivantes : à qui est destinée l'application visée ? Quelles perceptions les utilisateurs ont-ils du robot ? Comment se percevront-ils dans leurs interactions avec le robot (notamment par rapport aux autres personnes) ? Comment la société dans son ensemble interprétera-t-elle cette interaction par rapport aux normes socioculturelles et éthiques ?

— le principe de contexte (*the context principle*) : quelles sont les caractéristiques contextuelles pertinentes ?
— le principe que les "valeurs priment" (*the "values first" principle*) : se posent ici les questions des valeurs à considérer et de la recherche d'une plus-value à l'utilisation d'un robot. Est définie ici également une maxime de non-remplacement (*non-remplacement maxim*) indiquant qu'il serait préférable que les actions qu'on prévoit pour le robot soient des actions qui ne puissent pas être réalisées par un humain.

Je n'ai pas personnellement pris part à la démarche, mais je trouve que c'est une piste intéressante. Ce qu'on peut en retenir dans tous les cas, c'est que c'est une question à laquelle nous devons être sensibles dès aujourd'hui. On ne peut pas attendre que les systèmes arrivent sur le marché pour se poser les questions. Il faut aider le régulateur dès aujourd'hui. Une mise en pratique intéressante de la méthode peut être trouvée dans Kerstin Fischer et al. [Fischer 2020]. Ils observent notamment qu'il y a souvent un écart entre ce que les roboticiens veulent implémenter et ce dont les gens ont réellement besoin. C'est quelque chose que j'avais déjà rencontré lors de mon expérience en entreprise (les équipes techniques avaient fait quelque chose de très performant et innovant, mais cela n'avait pas satisfait les équipes utilisatrices qui attendaient un outil couvrant mieux leurs besoins et surtout d'une manière simple pour eux). En cela l'approche ISR est intéressante, car elle met les utilisateurs finaux dans la boucle dès le début du développement. Cependant, il faut bien voir aussi que les objectifs et les horizons dans une perspective de recherche et dans une perspective applicative peuvent être différents.

Dans ce chapitre, je vais revenir sur trois contributions auxquelles j'ai participé sur le sujet. La première est liée à *Design for values*, l'idée qu'il faut intégrer dans nos systèmes la capacité de pouvoir expliquer son comportement et notamment vis-à-vis d'un système de valeurs éthiques, la seconde est liée à ma participation au workshop Robotique, IA et humanité : science, éthique et politique et reprend des éléments de sa déclaration finale et la troisième est issue d'un travail avec des juristes sur les possibles réglementations à mettre en place au vu des problèmes potentiels amenés par l'introduction des robots.

8.2 Une responsabilité éthique en tant que développeur de systèmes

Je vais ici décrire les grandes lignes d'un article qui est le fruit d'un travail d'équipe (avec Virginia Dignum, Frank Dignum, Javier Vázquez-Salceda, Manuel Gentile, Samuel Mascarenhas, et Agnese Augello) lors du montage d'un projet et des idées préliminaires de Virginia Dignum (qui l'amèneront un peu plus tard à publier son livre *Responsable AI* [Dignum 2019]). L'article s'intitule Design for Values for Social Robot Architectures [Dignum 2018].

Dans son rapport ([for Ethical Considerations in Artificial Intelligence 2017], IEEE Global Initiative on Ethics explique que "si les machines s'engagent dans les communautés humaines en tant qu'agents autonomes, alors ces agents devront suivre les normes sociales et morales de la communauté. Une étape nécessaire pour permettre aux machines de le faire, c'est d'identifier ces normes". L'intégration des robots sociaux dans nos sociétés nécessite qu'ils soient capables de prendre des décisions susceptibles d'affecter la vie des personnes qui les entourent. Afin de s'assurer que ces robots se comporteront selon

des principes éthiques partagés, un changement important dans la conception et le développement des robots sociaux est nécessaire, un changement dont l'objectif principal est d'améliorer la transparence éthique plutôt que la performance technique, et de placer les valeurs humaines au cœur de la conception de robots.

Afin de rendre les robots dignes de confiance, de prendre conscience des enjeux éthiques qu'impliquent les interactions humain-robot, et de garantir que les interactions soient sûres, éthiques et acceptables pour les humains, nous devons définir des processus de conception pour inclure le raisonnement et la validation éthiques dans la conception d'interactions sociales.

Pour cela nous proposons de suivre les principes suivants [1] :

— *Accountability* : L'*accountability* est nécessaire pour une interaction de confiance et fait référence à la nécessité de pouvoir expliquer et de justifier ses décisions et ses actions à ses utilisateurs et aux autres personnes avec lesquelles le système interagit. Pour garantir l'*accountability*, les décisions doivent pouvoir être déduites et expliquées par les algorithmes décisionnels utilisés. Cela inclut le besoin de représentation des valeurs morales et des normes sociétales en vigueur dans le contexte de l'opération, que l'agent utilise pour la délibération. L'*accountability* en IA requiert à la fois la fonction de guider l'action (en formant des croyances et en prenant des décisions) et la fonction d'explication (en plaçant les décisions dans un contexte plus large et en les classant selon les valeurs morales).

— *Responsibility* : La *responsibility* est requise pour l'autonomie. En imaginant l'autonomie comme une variable scalaire, allant de l'absence d'autonomie à l'autonomie complète, et comprenant différents niveaux d'autonomie d'action, de plan, d'objectif et de motivation, alors la *responsibility* indique le point de cette échelle où un acteur humain est en charge de la décision. Inversement, du point de vue du robot, la *responsibility* correspond aux capacités requises pour évaluer ses décisions et identifier les erreurs ou les résultats inattendus, et ainsi transmettre le "jeton" de *responsibility* à son utilisateur, fournissant suffisamment d'informations sur les raisons pour lesquelles il ne peut plus prendre de décision par lui-même. En règle générale, de nombreux acteurs sont impliqués dans le processus qui mène à une décision du robot : les développeurs, les fabricants, les utilisateurs, les décideurs politiques, etc. Au fur et à mesure que la chaîne de responsabilité se développe, des moyens sont nécessaires pour lier les décisions du robot à la juste utilisation des données et aux actions des acteurs impliqués dans la décision du robot.

— *Transparency* : La *transparency* est associée à l'adaptabilité et fait référence à la nécessité de décrire, d'inspecter et de reproduire les mécanismes par lesquels les systèmes d'IA prennent des décisions et apprennent à s'adapter à leur environnement, ainsi qu'à la gouvernance de l'utilisation des données créées. L'un des principaux problèmes des approches d'apprentissage automatique est que, malgré leurs performances élevées, elles manquent de transparence, ce qui est souvent appelé une "boîte noire". Même si l'attention portée à ce problème est croissante, la plupart des algorithmes actuels (d'apprentissage en profondeur) sont incapables de lier les décisions aux entrées, et ne peuvent donc pas expliquer leurs actes.

Ayant en tête ces principes, l'article propose les bases d'une méthode de conception pour prendre en compte des valeurs (*Design for Values*) lors du développement des ro-

1. nous gardons les noms en anglais car en français *accountability* et *responsability* se traduisent tous les deux par responsabilité

bots sociaux. La première étape consiste à identifier les valeurs morales que le robot doit respecter. Même si, à un niveau abstrait, les valeurs sont universellement partagées, les gens et les sociétés diffèrent dans la manière dont ces valeurs universelles abstraites sont interprétées. Étant donné que les robots sociaux interagiront directement avec les personnes dans des contextes sociaux, les valeurs prises en compte dans la conception du robot doivent être alignées sur leurs interprétations contextuelles. La seconde étape vise à aligner le comportement sur les valeurs. La responsabilité morale est associée à la capacité de délibération morale. En supposant qu'un ensemble approprié de valeurs pour le robot a été identifié, il est alors nécessaire de déterminer comment le robot doit se comporter par rapport à ces valeurs. En particulier, il est nécessaire de déterminer comment ces valeurs doivent être hiérarchisées et comment traiter les dilemmes moraux, c'est-à-dire les situations dans lesquelles chaque action possible violera une ou plusieurs valeurs. Les approches de la délibération morale reflètent des théories éthiques de longue date, telles que l'utilitarisme (faire de son mieux pour le plus) ou déontologique/kantien (impératif catégorique).

8.3 Une responsabilité en tant que scientifique citoyen

En 2019, j'avais été invitée à participer et présenter nos travaux au workshop Robotique, IA et humanité : science, éthique et politique (Robotics, AI, and Humanity : Science, Ethics, and Policy) organisé conjointement par l'académie des sciences et par l'académie des sciences sociales du Vatican `https://www.pas.va/en/events/2019/robotics.html`. Ce workshop dépassait le cadre de mon domaine mais cela avait été très riche de discuter avec des économistes, des politiciens, des psychologues, des informaticiens, des roboticiens, bref un panel de scientifiques divers et (très) variés.

Suite à ce workshop avait été rédigé une "Déclaration finale sur la robotique, l'IA et l'humanité : science, éthique et politique" dont on peut trouver l'intégralité en ligne `https://www.pas.va/en/events/2019/robotics/final_statement.html`. Je reprends ici une partie de la déclaration.

(11.) Au sein de la recherche sur l'apprentissage automatique, il existe une ligne de développement qui vise à identifier les justifications fondamentales de la conception d'agents cognitifs ; de telles justifications permettraient de dériver des théorèmes caractérisant les possibilités et les limites des agents intelligents. Les agents cognitifs agissent dans un environnement ouvert, partiellement ou complètement inconnu afin d'atteindre des objectifs. Les concepts clés d'un cadre fondamental pour l'IA comprennent : les agents, les environnements, les récompenses, les scores locaux, les scores globaux, le modèle exact d'interaction entre les agents et les environnements, et une spécification des ressources de calcul disponibles des agents et des environnements. Un "agent intelligent" peut être défini comme un agent qui peut atteindre des objectifs dans un large éventail d'environnements. Le sujet de la cognition avec des ressources limitées mérite une exploration plus approfondie afin d'atteindre un cadre fondamental significatif et largement intégré pour l'intelligence artificielle.

(12.) Un aspect central de l'apprentissage par l'expérience est la représentation et le traitement de connaissances incertaines. En l'absence d'hypothèses déterministes sur le monde, il n'y a pas de conclusion logique non triviale qui puisse être tirée du passé pour tout événement futur. En conséquence, il est d'un intérêt fondamental d'analyser la structure de l'incertitude comme une question à part entière. De

nombreuses recherches restent à mener afin de traduire des propositions préliminaires et des méthodes formelles du domaine théorique à l'ingénierie de solutions algorithmiques efficaces.

(14.) Jusqu'à récemment, les mathématiques fondamentales avaient peu (voire pas du tout) de questions éthiques à l'ordre du jour. Cependant, étant donné que les mathématiciens et les concepteurs de logiciels sont au cœur du développement de l'IA, il est essentiel qu'ils tiennent compte des implications éthiques de leur travail. À la lumière des questions qui se posent de plus en plus sur la fiabilité des systèmes autonomes, les développeurs d'IA ont la responsabilité - qui devrait idéalement devenir une obligation légale - de créer des systèmes robotiques fiables et contrôlables.

(15.) Des recherches majeures sont en cours dans des domaines qui nous définissent en tant qu'êtres humains, tels que le langage, le traitement des symboles, l'apprentissage, l'auto-évaluation, le jugement, (...). Les agents informatiques formés par l'apprentissage par renforcement et les cadres d'apprentissage en profondeur démontrent des performances exceptionnelles dans des tâches jusqu'ici considérées comme insolubles. Bien qu'il manque encore une base solide pour une théorie générale des agents cognitifs computationnels, l'avancée conceptuelle et pratique de l'intelligence artificielle a atteint un état dans lequel les questions d'éthique et de sécurité et l'impact sur la société en général deviennent des problèmes urgents. Par exemple, les inférences basées sur l'IA des sentiments des personnes dérivées des données de reconnaissance faciale sont un tel problème.

(16.) La diffusion de la robotique modifie profondément les rapports humains et sociaux dans toutes les sphères de la société, dans la famille comme dans le travail et dans la sphère publique. Ces modifications prennent le caractère de processus d'hybridation entre les caractéristiques proprement humaines des relations et celles artificielles, donc entre réalité analogique et réalité virtuelle. Par conséquent, il est nécessaire d'accroître la recherche scientifique sur les questions concernant les effets sociaux qui découlent de la délégation d'aspects pertinents de l'organisation sociale à l'IA et aux robots. Un objectif de cette recherche devrait être de comprendre comment il est possible de gouverner les processus de changement pertinents et de produire ces biens relationnels qui réalisent un épanouissement humain vertueux dans le cadre d'un développement sociétal durable et équitable.

(20.) L'IA peut servir la bonne gouvernance, y compris l'identification et la prévention des transactions illégales, par exemple l'argent provenant d'activités criminelles telles que le trafic de drogue, la traite des êtres humains ou les greffes illégales. Cependant, lorsque l'IA est entre les seules mains des entreprises, les revenus de l'IA peuvent ne pas être redistribués équitablement. Ces nouvelles technologies ne doivent pas devenir des instruments pour asservir les gens ou marginaliser davantage les pauvres.

(21.) Nous avons examiné les applications de l'IA (et des technologies émergentes connexes) dans les domaines de la médecine et des soins de santé, de la mobilité et des transports, de la fabrication et de l'agriculture. Des opportunités majeures ont été relevées et une attention considérable a été accordée aux applications robotiques/IA dans chacun de ces domaines considérés séparément. Cependant, une perspective sectorielle sur l'IA et la robotique a ses limites. Il semble nécessaire d'avoir une image plus complète des liens entre les applications et de se concentrer

sur les politiques publiques qui facilitent la justice et l'équité globales couvrant tous les aspects de l'IA et de la robotique.

(22.) À moins qu'elle ne soit canalisée dans l'intérêt du public, l'IA peut susciter des inquiétudes importantes pour l'économie et la stabilité de la société. Des emplois pourraient être perdus au profit des appareils informatisés, ce qui entraînerait une augmentation de la disparité des revenus et des lacunes dans les connaissances. Les progrès de l'automatisation et l'augmentation de l'offre de main-d'œuvre artificielle, en particulier dans les secteurs agricole et industriel, peuvent réduire considérablement l'emploi dans les économies émergentes. Grâce aux liens au sein des chaînes de valeur mondiales, les travailleurs des pays à faible revenu peuvent également être affectés par les robots dans les pays à revenu élevé, ce qui pourrait réduire le besoin d'externaliser les emplois de routine vers les anciennes régions à bas salaires. Cependant, l'utilisation de robots pourrait également augmenter la demande de main-d'œuvre en réduisant le coût de production, ce qui conduit à l'expansion industrielle. Des estimations fiables des nouveaux emplois créés dans les industries de conception et de fabrication de robots font défaut, mais sont nécessaires.

(28.) La crainte de prendre du retard dans la course au développement de nouvelles applications militaires ne doit pas devenir une excuse pour court-circuiter la recherche, les essais et une formation adéquate en matière de sécurité. Étant donné que la conception des armes tend à s'éloigner des infrastructures à grande échelle vers des systèmes autonomes, décentralisés et miniaturisés, les risques d'effets destructeurs dus à la conception des systèmes et aux défaillances potentielles seront considérablement amplifiés par rapport à la plupart des systèmes opérationnels aujourd'hui. Ce potentiel accru d'externalités négatives doit être compensé par des investissements proportionnés dans la sécurité et la formation. Les outils d'IA devraient améliorer, mais non nuire, à l'exercice d'un jugement sain et moral par le personnel militaire.

(29.) Un effort international concerté doit viser à identifier les applications spécifiques de l'IA qui présentent des risques d'escalade. Les États devraient convenir de mesures concrètes pour réduire le risque de guerres facilitées par l'IA et éventuellement intensifiées. L'attention ne doit pas seulement être portée sur les dangers que la technologie remplace les personnes dans les sphères militaires, mais aussi sur le danger que l'IA peut représenter pour l'exercice de la "réflexion stratégique" dans les situations de conflit.

(30.) En ce qui concerne les systèmes d'armes létales autonomes, compte tenu de l'état actuel des compétences techniques (et dans un avenir prévisible), aucun système ne devrait être déployé qui fonctionne en mode non supervisé. Les lignes de responsabilité humaine doivent être maintenues afin que le respect des lois de la guerre internationalement reconnues puisse être assuré et les violations sanctionnées

(31.) Les efforts de développement soutenus par les États de machines intelligentes devraient viser le bien commun. L'impact sur les biens et services publics, ainsi que sur la santé, l'éducation, le bonheur et la durabilité, doit être primordial. L'IA peut avoir des préjugés inattendus ou des conséquences inhumaines, notamment la segmentation de la société et les préjugés raciaux et sexistes, et ceux-ci doivent être traités avant qu'ils ne se produisent. Il s'agit de problèmes nationaux et mondiaux, et ces derniers nécessitent une plus grande attention de la part des Nations Unies.

(35.) Les risques d'applications manipulatrices de l'IA pour façonner l'opinion publique

et l'ingérence électorale doivent être pris en compte et des contrôles nationaux et internationaux sont nécessaires.

(36.) L'IA et la robotique offrent de grandes opportunités et comportent des risques ; par conséquent, les réglementations doivent être conçues de manière appropriée par des institutions publiques légitimes, ne pas entraver les opportunités et ne pas stimuler la prise de risques excessifs et les préjugés. Cela nécessite un cadre dans lequel le discours sociétal public inclusif est informé par les sciences de différentes disciplines, et tous les segments de la société sont en mesure de participer.

(38.) Les entreprises devraient créer des conseils d'éthique et de sécurité et se joindre à des organisations à but non lucratif qui visent à établir les meilleures pratiques et normes pour le déploiement bénéfique de l'IA et de la robotique. Des protocoles appropriés pour la sécurité des robots et des systèmes à base d'IA doivent être développés, tels que la vérification dupliquée par des équipes de conception indépendantes. La réussite de tests d'éthique et de sécurité, évaluant par exemple l'impact social ou les préjugés raciaux cachés, devrait devenir une condition préalable à la sortie de nouveaux logiciels d'IA. Des commissions civiles externes procédant à une évaluation récurrente et transparente de toutes les technologies, y compris militaires, devraient être envisagées.

(39.) Les scientifiques et les ingénieurs, en tant que concepteurs de dispositifs d'IA et de robots, ont la responsabilité d'essayer activement de garantir que leurs inventions et innovations sont sûres et peuvent être utilisées à de bonnes fins.

8.4 Une responsabilité vis-à-vis des consommateurs finaux

En 2020, dans le cadre de l'institut d'Intelligence Artificielle de Toulouse (ANITI `https://aniti.univ-toulouse.fr/`), nous avons eu l'occasion de réaliser un travail avec les juristes Claire Boine et Céline Castets-Renard dans un article intitulé In Love with a Corporation without Knowing It : An Asymetrical Relationship [Boine 2020]. Je vais reprendre ici les grandes lignes et les propositions que nous avions défini dans cet article.

8.4.1 Introduction

La plupart des domaines du droit reposent sur l'hypothèse du libre arbitre. En droit pénal, les peines sont le plus souvent justifiées par la justice rétributive, l'idée que les criminels doivent proportionnellement souffrir en retour, et par la théorie de la dissuasion, qui stipule que la perspective d'être puni dissuade les individus de commettre des crimes en premier lieu. Les deux théories supposent que les criminels auraient pu choisir de ne pas commettre leur crime mais ne l'ont pas fait. Sans libre arbitre, la justice punitive serait tout simplement cruelle et la dissuasion inefficace. En droit des contrats, les contrats sont valables s'ils ne sont pas signés sous la contrainte. En dehors de la protection des personnes vulnérables, la loi suppose que les individus prennent leurs décisions librement au quotidien. Le libéralisme économique repose également sur le libre arbitre, car l'allocation du marché repose sur le fait que les agents sont intrinsèquement rationnels et attribuent de la valeur aux biens en fonction de ce qu'ils sont prêts à payer. Néanmoins, les législateurs ont également reconnu qu'il existe un différentiel de pouvoir entre les individus et les entreprises, et que les consommateurs ne peuvent exercer leur libre arbitre s'ils n'ont pas accès à des informations complètes et loyales sur les produits. Le droit de la consomma-

tion tente de compenser ce déséquilibre. Par exemple, les fabricants de cigarettes doivent étiqueter les emballages avec des avertissements sanitaires pour s'assurer que les consommateurs sont conscients des risques néfastes du produit pour leur santé. Cependant, une nouvelle technologie récemment introduite sur le marché crée une asymétrie de pouvoir sans précédent entre les entreprises et les individus : les robots sociaux.

8.4.2 Questions éthiques et juridiques

Dans un premier temps on s'intéresse aux questions éthiques et juridiques posées par l'arrivée des robots sociaux. En particulier, les juristes s'intéressent aux préjudices potentiels découlant de l'utilisation d'un robot social et aux manières de les atténuer.

La première préoccupation concerne la vie privée. Pendant longtemps, la collecte des données a été cloisonnée entre différents domaines, la plupart des utilisateurs de technologies ne comprennent pas les conséquences de la collecte de données sur Internet sur leur vie privée [on Ethics of Autonomous 2017]. Ryan Calo [Calo 2011] identifie trois façons dont les robots peuvent porter atteinte à la vie privée des citoyens : la surveillance directe, l'introduction de nouveaux points d'accès dans des espaces historiquement protégés (par exemple, le gouvernement ou un individu peut pirater un robot social et voir l'intérieur du lieu où il se trouve), et la signification sociale (par exemple, le robot collecte des informations sur la façon dont l'utilisateur interagit avec lui, ce qui est un nouveau type de données). De plus, les robots peuvent facilement manipuler les humains et les pousser à leur donner des informations au cours des conversations, surtout si les humains développent des sentiments pour eux. Le risque de manipulation s'étend au-delà de la collecte de données. Que se passe-t-il si votre assistant robot, qui fait les courses pour vous, sélectionne des marques spécifiques dans lesquelles son entreprise détient des actions ? Que faire si votre compagnon robot vous dit qu'il est triste de ne pas avoir la dernière mise à jour, qui a un coût supplémentaire ? Et si votre robot social commençait à vous donner son avis politique ?

La seconde préoccupation concerne l'exercice du consentement et du libre arbitre. En droit des contrats, le consentement doit être libre et véritable pour que le contrat soit valable. Si le consentement a été donné sur la base d'une fausse déclaration, le contrat est nul. En droit de la consommation, les entreprises sont légalement tenues de fournir des informations sur les produits qu'ils vendent. L'information doit être précise. Il est illégal pour les entreprises d'adopter des comportements susceptibles d'induire en erreur ou de tromper les consommateurs [dir]. Lorsqu'un robot dit à un utilisateur qu'il l'aime, l'entreprise qui programme le robot induit effectivement le consommateur en erreur en lui faisant croire à la fois que le robot est capable d'éprouver des sentiments et qu'il l'aime. Il s'agit clairement d'un cas de tromperie du consommateur.

L'utilisation des robots sociaux pose la question du consentement à trois niveaux différents. Premièrement, elle survient lorsque les consommateurs achètent l'appareil. Les entreprises doivent donner des informations transparentes, des informations précises et exhaustives sur les robots et le fait que, malgré les apparences, ils n'ont pas d'émotions ou de sensibilité. Deuxièmement, la question du consentement et du libre arbitre revient une fois que les consommateurs ont grandi en développant un attachement pour les robots sociaux. Même si, à l'achat de leurs robots, les consommateurs ont reçu des informations indiquant que leurs appareils ne peuvent pas faire l'expérience de vraies émotions, pouvons-nous vraiment nous attendre à ce qu'ils s'en souviennent des mois plus tard,

après que leur assistant virtuel leur a régulièrement envoyé des SMS pour leur dire qu'ils lui avaient manqué ou leur a montré de l'affection ? De plus, une fois que le consommateur est émotionnellement attaché, il n'est pas évident qu'il pourrait se séparer du robot même si c'était dans son intérêt. En conséquence, un simple contrat de vente comprenant des informations sur le robot et ses dommages potentiels ne peut être suffisant pour y remédier étant donné que les robots ont des effets à long terme qui varient dans le temps. Troisièmement, le libre arbitre est également pertinent pour la responsabilité. Si un robot social influence quelqu'un pour enfreindre la loi, l'individu est-il entièrement responsable ? Peut-on invoquer indûment cette influence ? Si oui, comment le prouverait-on ? Quel degré d'implication serait nécessaire pour plaider la cause ?

8.4.3 Propositions

Face à ces questions, Claire Boine fait plusieurs propositions.

Les premières concernent les personnes vulnérables Comment pouvons-nous protéger les personnes particulièrement vulnérables comme les enfants ou les personnes âgées des risques liés à l'utilisation des robots sociaux ?

Dans le cas des personnes vulnérables, nous proposons que les robots ne devraient jamais pouvoir parler d'eux-mêmes. Parler de soi implique une conscience. L'utilisation du "je" marque le passage de l'objet au sujet. Les robots ne devraient pas faire semblant d'avoir des émotions ou des opinions. Des robots qui parleraient par exemple de leurs soi-disant sentiments ou qui feraient semblant de se rappeler des souvenirs d'avant que leur propriétaire ne les achète seraient trompeurs. Il est important que les entreprises développant ces robots fassent le maximum pour éviter la confusion, cela ne doit pas incomber aux enfants ou aux personnes séniles qui pourraient tout prendre pour argent comptant. Les robots n'ont pas besoin de faire semblant d'avoir une conscience ou faire preuve d'émotions afin d'accomplir leurs fonctions. Un robot social qui joue aux cartes avec une personne en EHPAD ne doit pas parler comme si elle avait des préférences, des émotions, des valeurs et des opinions. C'est ce que nous appelons le principe de non-tromperie.

Deuxièmement, nous proposons que les personnes vulnérables ne soient jamais laissées seules avec des robots sociaux. Un robot compagnon dans une maison de retraite peut être laissé dans la salle commune pour favoriser les échanges entre les résidents, sous la surveillance d'un soignant. Un robot comme Robear, qui peut soulever des personnes hors d'un lit, peut être utilisé pour aider un véritable soignant. Aucun d'entre eux ne devrait remplacer les soins humains ou l'interaction humaine. Ce principe de non-remplacement a été mis en avant par Seibt et al. comme le "principe des valeurs d'abord" : "les applications de la robotique sociale doivent respecter la maxime de non-remplacement : les robots ne peuvent faire que ce que les humains devraient, mais ne peuvent pas faire" [Seibt 2018]. La présence d'une personne non vulnérable peut atténuer le risque d'influence indue et l'asymétrie de pouvoir entre un particulier et une entreprise. De la même manière, les robots sociaux offrent une plate-forme supplémentaire pour que les enfants jouent ou apprennent, en interaction avec leurs parents ou gardiens. C'est ce que nous appelons le principe de non-substitution.

Les secondes concernent l'ensemble des utilisateurs La première proposition est de limiter chaque robot à un seul objectif. Le système de sécurité domestique ne devrait pas lire d'histoires aux enfants. L'assistant virtuel qui peut accéder aux informations de votre carte de crédit pour vous commander des repas ne devrait pas vous envoyer des messages pour vous dire qu'il vous manque. En d'autres termes, nous plaidons pour que chaque application robotique ait une portée étroite. Cela peut limiter à la fois le risque d'attachement émotionnel et le pouvoir des entreprises en cas de faute professionnelle. Par exemple, cela limite la quantité de données pouvant être collectées. Cela limite également le temps passé avec et le degré de dépendance à chaque appareil. En complément de cette première proposition, nous proposons de limiter les connaissances que les robots peuvent acquérir. Nous partons du principe que le robot n'a besoin de connaître que les informations directement pertinentes pour son objectif. Par exemple, un robot compagnon ne doit pas collecter de données sur les personnes qui vous rendent visite.

La seconde proposition est que la réglementation des robots sociaux devrait être basée sur deux catégories mutuellement exclusives : 1) les robots dont la fin n'est pas liée à l'attachement émotionnel mais qui peuvent le générer de manière accessoire ; et 2) les robots dont la fin nécessite un certain attachement émotionnel comme moyen ou dont la fin est un attachement émotionnel. Pour les robots de la première catégorie, ceux dont le but ne nécessite pas d'attachement émotionnel, nous proposons d'éviter la conception anthropomorphique. L'anthropomorphisme comprend à la fois l'apparence humaine et les caractéristiques humaines (histoires de fond, fausses émotions, opinions…). La conception du robot devrait être basée sur son objectif. Afin de pallier l'attachement potentiel aux robots sociaux de la seconde catégorie, nous proposons deux types de solutions. Le premier type est basé sur la conception. Afin de limiter le risque de confusion quant à savoir si les robots ont une individualité, les robots pourraient régulièrement changer d'apparence lorsque cela est possible. Ils pourraient également changer de voix régulièrement. Le second type de solution consiste à introduire des limitations spatiales et temporelles. Une façon d'introduire des limites de temps et d'espace serait d'avoir une relation contractuelle entre les consommateurs et les entreprises de robotique, qui indique la durée d'embauche d'un robot, avec une liste exhaustive des tâches qu'ils pourraient et devraient entreprendre.

En plus de mettre en œuvre des lois pour limiter l'attachement humain aux robots, nous suggérons d'ajouter de nouvelles lois sur la protection des consommateurs pour s'assurer que les individus comprennent avec précision la nature du produit. Compte tenu de l'effet différentiel dans le temps, nous proposons que les consommateurs soient informés régulièrement, même après avoir acheté un appareil. Comme un fumeur se voit rappeler les dangers de la dépendance chaque fois qu'il achète un nouveau paquet, les utilisateurs de robots pourraient recevoir des informations régulières qui ont été mises à jour avec de nouvelles connaissances scientifiques impartiales qui n'ont pas été parrainées par des sociétés de robotique. Les informations devraient être compilées par des agences publiques telles que la Federal Drug Administration ou la Federal Trade Commission.

8.5 Conclusion

Dans ses Lignes directrices en matière d'éthique pour une IA digne de confiance [Commission 2019], la Commission européenne affirme que "les systèmes d'IA doivent être centrés sur l'humain, en s'appuyant sur l'engagement de mettre leur utilisation au service de l'humanité et du bien commun, avec pour objectif d'améliorer le bien-être et la

liberté des êtres humains. S'ils offrent de brillantes possibilités, les systèmes d'IA soulèvent également certains risques qui doivent être traités de manière appropriée et proportionnée. Nous sommes à présent face à une occasion unique de façonner leur élaboration. Nous voulons pouvoir nous fier aux environnements sociotechniques auxquels ils sont intégrés, et nous voulons que les concepteurs de systèmes d'IA obtiennent un avantage concurrentiel en intégrant une IA digne de confiance à leurs produits et services. Cet objectif nécessite de chercher à optimiser les avantages offerts par les systèmes d'IA tout en veillant à prévenir et réduire le plus possible les risques qu'ils présentent." Dans ce même document [Commission 2019] est précisé ce que l'on entend par IA digne de confiance, "Une IA digne de confiance comporte trois éléments : 1) elle doit être licite, en assurant le respect des législations et réglementations applicables, 2) elle doit être éthique, en assurant l'adhésion à des principes et valeurs éthiques, et 3) elle doit être robuste, sur le plan tant technique que social, pour faire en sorte que, même avec de bonnes intentions, les systèmes d'IA ne causent pas de préjudices involontaires. Chaque élément est nécessaire mais ne suffit pas pour parvenir à une IA digne de confiance. L'idéal serait que ces trois éléments fonctionnent en harmonie et se chevauchent. Lorsque des tensions apparaissent, nous devrions nous efforcer de les aligner."

Nous pensons qu'il est important d'intégrer cette préoccupation dans notre réflexion scientifique. Dans ce chapitre nous avons présenté différents aspects qui doivent être envisagés pour permettre cette prise en compte.

8.6 Contributions Associées

- [Dignum 2018] Virginia Dignum, Frank Dignum, Javier Vázquez-Salceda, Aurélie Clodic, Manuel Gentile, Samuel Mascarenhas, and Agnese Augello. Design for Values for Social Robot Architectures. In Envisioning Robots in Society – Power, Politics,and Public Space, pages 43–52. 2018
  `https://hal.laas.fr/hal-01943831`
  - Cet article est le fruit d'un travail d'équipe lors du montage d'un projet et des idées préliminaires de Virginia Dignum (qui l'amèneront un peu plus tard à publier son livre Responsable AI [Dignum 2019]). L'intégration des robots sociaux dans les sociétés humaines nécessite qu'ils soient capables de prendre des décisions susceptibles d'affecter la vie des personnes qui les entourent. Afin de s'assurer que ces robots se comporteront selon des principes éthiques partagés, un changement important dans la conception et le développement des robots sociaux est nécessaire, un changement dont l'objectif principal est d'améliorer la transparence éthique plutôt que la performance technique, et de placer les valeurs humaines au cœur de la conception des robots. Dans ce résumé, nous discutons du concept de prise de décision éthique et de la manière d'instaurer la confiance selon les principes d'autonomie, de responsabilité et de transparence (ART).
- [Boine 2020] Claire Boine, Céline Castets-Renard, Aurélie Clodic, and Rachid Alami. In Love with a Corporation without Knowing It : An Asymetrical Relationship. In Nørskov, M., Seibt J., and Quick O. 2020., editors, Culturally Sustainable Social Robotics- Proceedings of Robophilosophy 2020., Frontiers of AI and Its Applications,. IOS Press, August 2020.
  `https://hal.archives-ouvertes.fr/hal-02963070`

— Dans ce papier, les juristes Claire Boine et Céline Castets-Renard ont dialogué avec nous roboticiens. La plupart des domaines du droit reposent sur l'hypothèse du libre arbitre, qui se manifeste dans l'expression du consentement. Nous examinons la nature des émotions humaines envers les personnages fictifs et les robots sociaux, et questionnons le concept de consentement dans le contexte de ces relations fictives non réciproques. Nous concluons que les politiques doivent réglementer l'utilisation des robots sociaux afin de protéger les consommateurs, et en particulier les plus vulnérables, d'une asymétrie de pouvoir entre eux et les entreprises robotiques. Nous proposons différentes solutions réglementaires et de conception en fonction de l'objectif des robots et du type d'utilisateurs.

Chapitre 9

Se confronter au monde réel

Dans ce chapitre, je vais parler du projet MuMMER (MultiModal Mall Entertainment Robot), un projet européen qui a commencé en mars 2016 pour se finir en mars 2020.

Ce cadre du projet MuMMER va nous permettre d'aborder un point important lorsqu'il s'agit de robotique et plus précisément d'interaction humain-robot : l'intégration. On ne peut pas vraiment tester nos systèmes s'il n'y a pas des personnes dans la boucle et le terrain d'expérimentation ultime c'est bien entendu la "vraie" vie.

J'aimerais dans ce chapitre expliquer l'intérêt du travail en équipe lorsque l'on veut réaliser une démonstration intégrée où l'on souhaite que le robot soit "autonome", mais aussi les difficultés que l'on peut rencontrer et l'équilibre qu'il faut trouver entre les contraintes techniques et les motivations scientifiques.

Je terminerai par mon analyse sur plusieurs aspects qui me semblent importants. Tout d'abord en prenant ma casquette de responsable technique pour indiquer les bonnes pratiques qui je pense sont à considérer lorsque l'on fait partie d'un "gros" projet comme celui-ci. Ensuite en prenant ma casquette de co-directrice de thèse pour expliquer les pour et les contre qu'implique l'engagement d'un doctorant dans ce type de projet avec un peu d'autocritique dedans.

Le travail sur le projet MuMMER a été réalisé par les doctorants du LAAS Kathleen Belhassein, Guilhem Buisan, Harmish Khambhaita, Amandine Mayima, Yoan Sallami, Guillaume Sarthou, Phani-Teja Singamaneni, Jules Waldhart encadrés par Rachid Alami et moi-même avec l'aide des ingénieurs de la plate-forme robots Christelle Ecrepont et Matthieu Herrb. Ce travail a bénéficié aussi de la collaboration de l'ensemble des partenaires du projet (University of Glasgow (UGA), Heriot-Watt University (HWU), Idiap Research Institute (IDIAP), Softbank Robotics Europe (SBR), VTT Technical Research Centre of Finland (VTT), Ideapark).

9.1 Le pitch de départ

Le projet MuMMER MultiModal Mall Entertainment Robot (`https://cordis.europa.eu/project/id/688147/fr`) proposait de faire progresser les technologies nécessaires pour développer des applications de robotique de divertissement pour le grand public. Plus précisément, le projet visait au développement d'un robot humanoïde (basé sur la plate-forme Pepper, cf figure 9.1) capable de s'engager et d'interagir de manière autonome et naturelle dans l'environnement dynamique d'un centre commercial en offrant une expérience engageante et divertissante. En utilisant des méthodes de co-conception, l'idée était de travailler dès le départ avec les parties prenantes (e.g. les clients, les employés, les commerçants) pour développer des comportements robotiques vraiment engageants (e.g. raconter des blagues, jouer à des jeux, mais aussi fournir des conseils, des informations et collecter des commentaires des clients). Le but était que le robot affiche un comportement socialement approprié, combinant une interaction basée sur la parole avec de la commu-

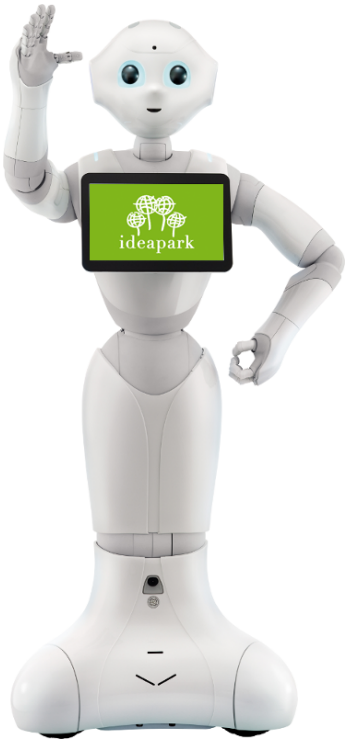


FIGURE 9.1 – Le robot Pepper de SoftBankRobotics

nication non verbale et de la navigation prenant en compte l'humain. Le lieu de déploiement du robot était un grand centre commercial public en Finlande, Idearpak Lempaala (`https://lempaala.Ideapark.fi/`) dont on peut retrouver le plan et un visuel figure 9.2. L'une des originalités du projet était l'utilisation de l'approche de co-conception pour étudier et favoriser l'acceptation des robots grand public (et ainsi influencer positivement le marché de la robotique de service).

Comme on peut le voir, le projet était très ambitieux. Sans divulgâcher la suite de ce chapitre, non, nous n'avons pas, je pense influencé les marchés de la robotique de service, mais je pense qu'il est important d'expliquer le chemin parcouru lors d'un projet comme celui-ci, car il se retrouve pour beaucoup de projets et nous aurions tort de ne pas essayer d'apprendre en cours de route.

9.2 L'organisation du projet et notre implication

Le consortium du projet MuMMER comprenait sept partenaires : University of Glasgow (UGA), Heriot-Watt University (HWU), Idiap Research Institute (IDIAP), LAAS-CNRS (LAAS), Softbank Robotics Europe (SBR), VTT Technical Research Centre of Finland (VTT), Ideapark.

Le projet était divisé en 9 workpackages, comme illustré figure 9.3. Le LAAS était responsable du (et majoritairement impliqué dans) le workpackage 5 (WP5) : Tâches interactives et contrôle du mouvement en prenant en compte l'humain (*Human-Aware interactive tasks and motion control*). Ce workpackage avait pour but de développer les capacités nécessaires à la planification des mouvements et des tâches du robot ainsi que leurs exécutions en interaction ou à proximité des humains.

FIGURE 9.2 – Carte et photo de l'environnement d'Ideapark

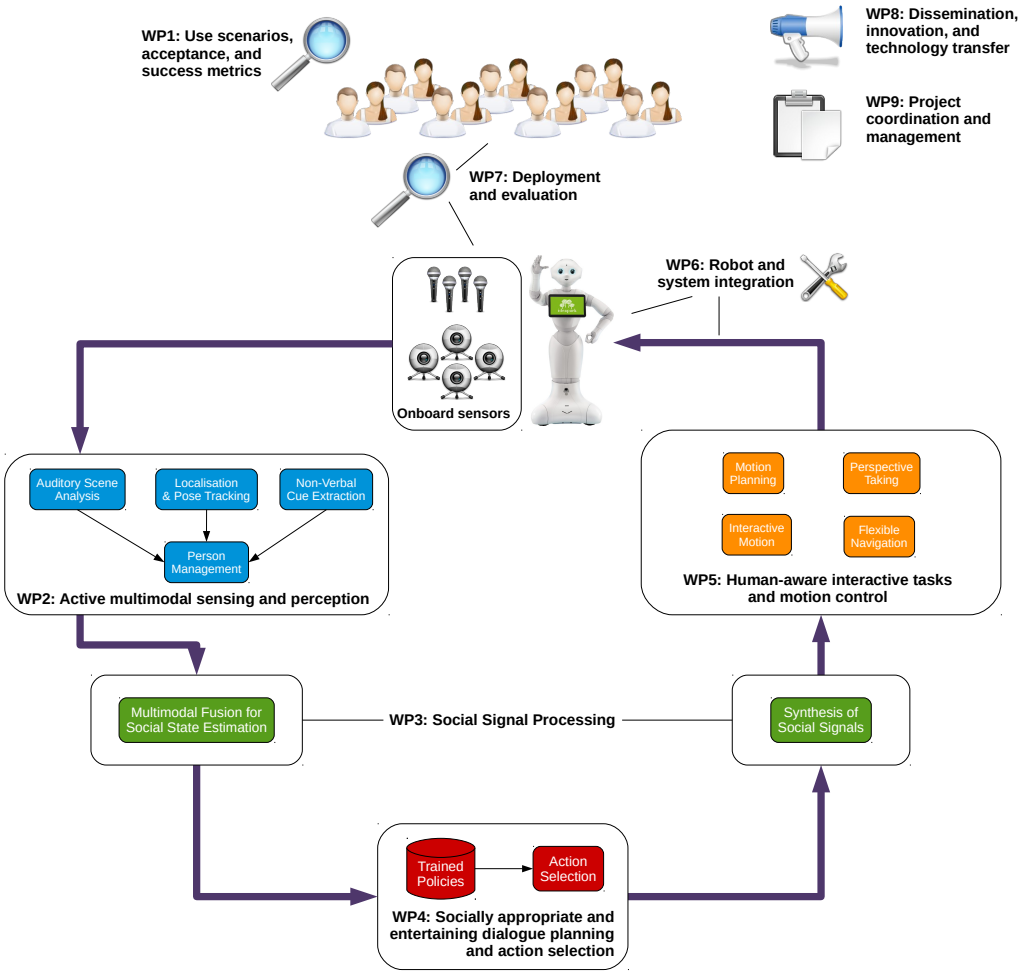


FIGURE 9.3 – Relation entre les différents workpackages du projet MuMMER

FIGURE 9.4 – Zone de déploiement du robot à Ideapark

9.3 Un peu de logistique pour commencer

9.3.1 Choix d'un scénario

Il est vite apparu que le robot avait des capacités de navigation limitées, ce qui est un souci lorsque votre rôle dans le projet est justement de développer des capacités de navigation... Après discussion avec les personnels d'Ideapark et nos collègues de VTT, nous avons choisi de nous concentrer sur un scénario où le robot interagit avec les humains et se déplace uniquement localement : une tâche dite de description d'itinéraire où le robot, sans trop bouger, est capable d'expliquer à un visiteur comment atteindre une destination donnée à partir de sa position actuelle. À côté de cette tâche de description d'itinéraire, d'autres scénarios ont été développés par les autres partenaires du projet, mais je n'en parlerai pas ici.

Une fois la décision prise de ne pas faire de navigation 'au long cours', il a été décidé que le robot serait déployé dans un environnement d'environ 4mx4m non loin de l'actuel point d'accueil du centre commercial. La version finale de la zone de déploiement est visible sur la figure 9.4. La définition de cet environnement a demandé du travail. Il a fallu équiper le sol avec un lino mat pour éviter les problèmes avec certains capteurs de Pepper sensibles aux réflexions si le sol est trop brillant. Il a fallu équiper le "toit" de la zone avec des panneaux pour éviter qu'il y ait trop de réverbérations pour le son (l'environnement dans un centre commercial comme celui-ci est très bruyant, il y a des musiques d'ambiance, des publicités, etc.). Ces panneaux ont aussi aidé pour les algorithmes de vision, dans un environnement ouvert comme celui-ci (nous étions sous un toit en verre), la lumière peut beaucoup changer entre un jour où il fait soleil (les rayons du soleil peuvent gêner la perception), un jour où il pleut ou un jour où il neige. Il a fallu prévoir un endroit où arriverait le réseau (ainsi que le câble ethernet qui serait branché au robot pendant le déploiement, les PCs du robot lui-même n'étant pas assez puissants pour faire tourner les logiciels). Cet endroit accueillait aussi le portable sur lequel était lancé l'ensemble des logiciels. Il a fallu prévoir une arrivée d'électricité pour permettre le rechargement du robot quand c'était nécessaire.

Cela peut paraître anecdotique, mais cela ne l'est pas, il est très facile de faire échouer un déploiement si l'on n'a pas pensé aux détails. Si vous prévoyez que la personne puisse parler au robot, mais qu'in fine l'environnement cible est trop bruyant pour que la reconnaissance de la parole fonctionne, cela met en défaut tout le reste. C'est d'autant plus

FIGURE 9.5 – La maquette de l'environnement avec les signes ajoutés pour indiquer les magasins

important quand votre environnement cible, comme ici le centre commercial, n'est pas accessible facilement.

9.3.2 Création d'un environnement de test accessible

L'autre souci était que le lieu d'intégration était difficilement accessible. Or le développement d'une tâche pour un robot autonome demande de pouvoir réaliser un ensemble de tests pour pouvoir d'une part assurer l'intégration des différents logiciels qui vont être utilisés et ensuite pour tester la ou les tâches dans leur ensemble. C'est pourquoi nous avons construit au LAAS un environnement complet qui émulait celui d'Ideapark. Pour ce faire, nous avons choisi plusieurs magasins présents à Ideapark avec un maximum de variété : restaurants (burger, pizzeria, thaï, etc.), magasins de vêtements (homme, femme, les deux, sportswear) ou distributeurs automatiques de billets. Ce n'est pas seulement le type de magasin qui a été émulé, mais aussi leurs positions dans l'environnement (par exemple, l'utilisation du premier étage, les portes des bâtiments et les grandes surfaces comme on peut le trouver à Ideapark). Nous avons également choisi une place pour le robot dans cet environnement, qui permet d'envisager la plupart des cas qui pourraient se produire à Ideapark, par ex. nous émulons Funpark au premier étage qui est visible depuis la position du robot et nous avons plusieurs façons d'y accéder. De la même manière, même les portes d'accès au parking ont été intégrées.

Avec cet environnement, nous avions au LAAS une maquette réaliste et accessible de l'environnement Ideapark au LAAS (cf. figure 9.5). Cette maquette a été utilisée par le projet pour des semaines d'intégration avant le déploiement sur site et par le LAAS, en collaboration avec les autres partenaires, pour développer, ajuster et tester la tâche de guidage tout au long du projet.

9.4 Développements

Un projet européen comme celui-ci est avant tout un projet collaboratif. Ici, je vais expliquer les développements que nous avons réalisés au LAAS mais aussi les différentes collaborations que nous avons menées avec les autres WP (et donc partenaires) pour mener

ce projet à bien.

9.4.1 Définition du scénario et conception des fonctions nécessaires (travaux communs WP1/WP5 et collaboration avec le projet JointAction4HRI)

Afin d'assurer la conception et la mise en œuvre des fonctions pertinentes et leur articulation, le LAAS a mené avec VTT une première étude exploratoire basée sur l'interaction humain-humain lorsqu'un guide fournit des informations et des conseils d'itinéraire à un visiteur du centre commercial Ideapark. L'étude a été réalisée au plus près de la zone où allait être déployé le robot.

Sur la base d'entretiens préliminaires avec des guides travaillant à Ideapark, une liste de magasins souvent demandés par les clients a été sélectionnée. Par la suite, des enregistrements vidéos des guides donnant des indications pour atteindre ces magasins ont été réalisés. Après la traduction finnois-anglais des dialogues par VTT, les vidéos ont été analysées au LAAS. Nous nous sommes concentrés sur la communication non verbale, et en particulier les différents types de gestes utilisés pour guider, les positions des deux protagonistes par rapport au magasin cible et à leur interlocuteur, et l'alternance des regards. Un rapport [Belhassein 2017], donnant les premières indications à prendre en compte pour que le guidage du robot soit efficace et compris par les clients, est issu de cette étude exploratoire. À titre d'exemple, cette étude nous a permis de constater une utilisation préférentielle de la main ipsilatérale (i.e. du même côté) au champ visuel de la cible. Dans la lignée de la littérature existante sur les études gestuelles, nous avons également remarqué que la main utilisée pour pointer un élément de référence est orientée verticalement dans le cas des magasins (référents verticaux) ou des actions dirigées (comme le chemin à prendre ou les virages), alors que dans le cas des référents horizontaux (par exemple les escalators), la main est orientée horizontalement (paume vers le sol).

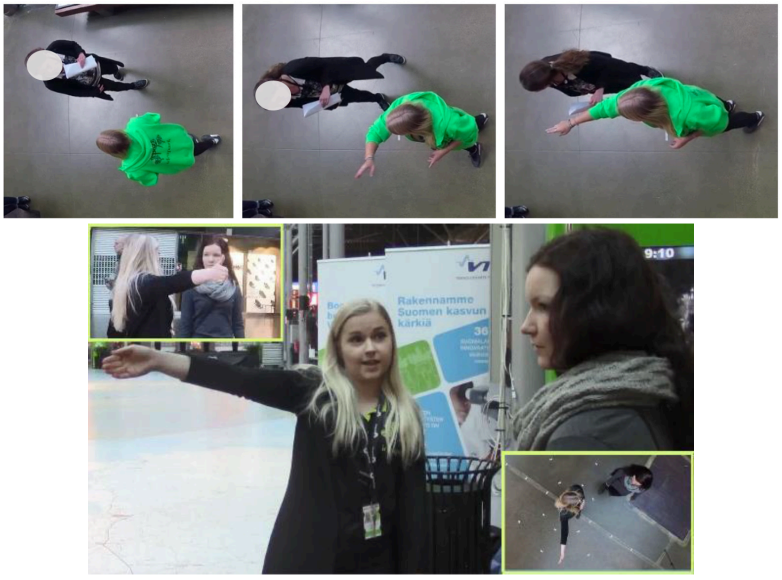


FIGURE 9.6 – Illustration des placements des protagonistes et des gestes utilisés pendant l'étude exploratoire

Une deuxième étude exploratoire a ensuite été menée, toujours en collaboration avec VTT, ajoutant des situations complexes pouvant apporter une certaine confusion aux

participants. Encore une fois, les signaux sociaux ont été analysés. Le protocole utilisé et les résultats (publié dans [Heikkilä 2019]) ont fourni des éléments de compréhension des comportements typiques de guidage. Entre autres résultats, cette étude a mis en évidence l'enchaînement de la description d'itinéraire. Ainsi, un premier geste déictique (qui désigne un objet (locuteur, lieu, date…) dont le référent dépend de la situation d'énonciation) sur un passage visible (couloir, ou si la boutique demandée est au deuxième étage, l'escalator) précède les explications sur les directions à prendre. Les résultats les plus intéressants concernaient les situations pouvant prêter à confusion ou aboutir à une incompréhension du client. Par exemple, l'utilisation d'un seul canal de transmission (ex. : le geste sans la parole), le choix de points de repère qui ne sont pas toujours adaptés, s'il y a plusieurs descriptions d'itinéraires dans une même explication ou lorsque la distance n'est pas précisée.

Contributions associées :

[Belhassein 2017] **Human-human guidance study**
Belhassein Kathleen, Clodic Aurélie, Cochet Hélène, Niemela Markettä, Heikkila Päivi, Lammi Hanna, and Tammela Antti (2017). Techical Report hal-01719730, LAAS-CNRS, CLLE, VTT.
`https://hal.laas.fr/hal-01719730/document`

[Heikkilä 2019] **Should a robot guide like a human ? A qualitative four-phase study of a shopping mall robot**
Päivi Heikkilä, Hanna Lammi, Marketta Niemelä, Kathleen Belhassein, Guillaume Sarthou, Antti Tammela, Aurélie Clodic, Rachid Alami The Eleventh International Conference on Social Robotics (ICSR 2019), Nov 2019, Madrid, Spain
`https://hal.laas.fr/hal-02285035/file/VTT_LAAS_ICSR_2019.pdf`

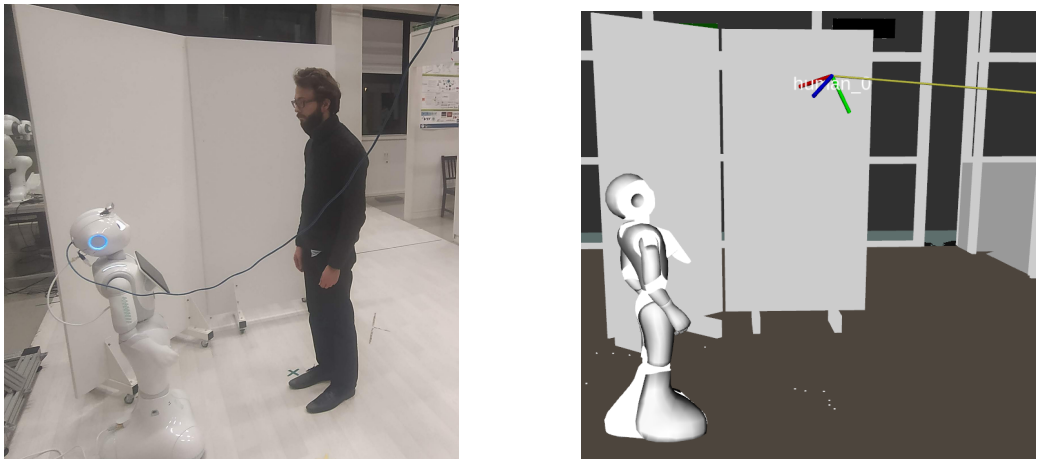
9.4.2 Planificateur de perspective visuelle partagée (SVP Planner) (WP5.1)

Pour assurer au mieux la tâche de description d'itinéraires, nous avons développé un planificateur de perspective visuelle partagée (shared visual perspective (SVP) planner). Ce composant calculait pour une direction donnée, le meilleur couple pour la position du robot (guide) d'une part et la position de l'humain (visiteur) d'autre part pour permettre une bonne visibilité de la destination comme illustré figure 9.7.

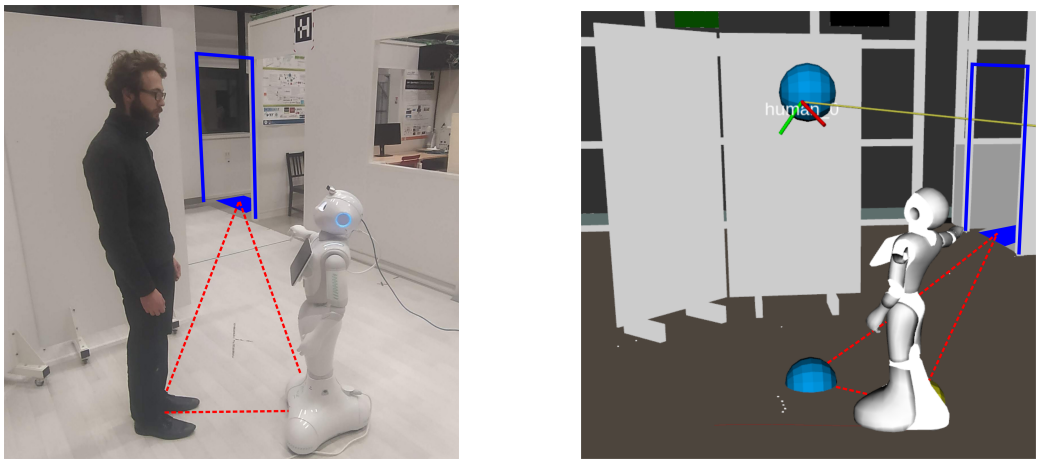
Contributions associées :

[Waldhart 2019] **Reasoning on Shared Visual Perspective to Improve Route Directions**
Jules Waldhart, Aurélie Clodic, Rachid Alami 2019 28th IEEE International Conference on Robot & Human Interactive Communication, Oct 2019, New Delhi, India
`https://hal.archives-ouvertes.fr/hal-02283904/file/ROMAN19_0173_FI.pdf`

[Waldhart 2018] **Planning Human and Robot Placements for Shared Visual Perspective**
Jules Waldhart, Aurélie Clodic, Rachid Alami Workshop on Robotic Co-workers 4.0 : Human Safety and Comfort in Human-Robot Interactive Social Environments 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct 2018, Madrid, Spain
`https://hal.archives-ouvertes.fr/hal-01972096/document`



(a) Positions initiales de l'humain et du robot. L'humain demande au robot une direction qui est derrière le robot.



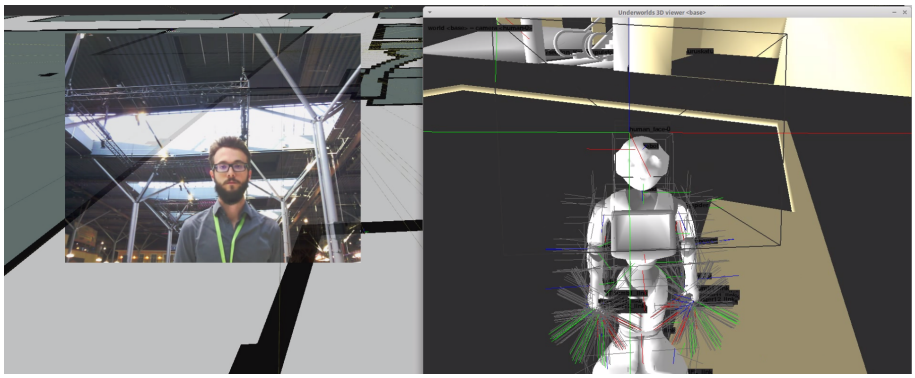
(b) Le robot et l'humain sont à leurs positions finales. Les sphères bleues sont les positions calculées pour l'humain par le robot. Le robot pointe le passage (représenté par le cadre bleu). On peut observer, en rouge, le triangle formé entre l'humain, le robot et le passage.

FIGURE 9.7 – Positions initiales et finales calculées par le planificateur de perspective visuelle partagée. À gauche les images, à droite des captures d'écran de la visualisation 3D Rviz.

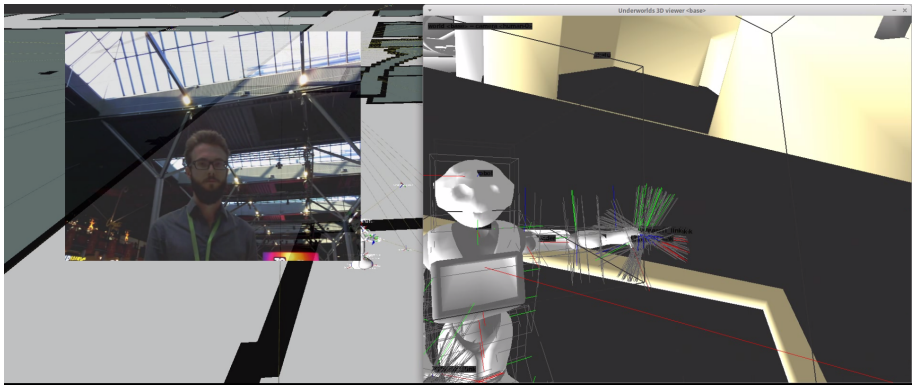
9.4.3 Prise de perspective et raisonnement géométrique : UnderWorlds (avec WP5.2 avec WP2)

Pour pouvoir alimenter le planificateur précédent, il faut avoir accès non seulement à la perspective du robot, mais également à la (simulation de) la perspective de l'humain. Pour cela, nous avons développé un composant d'évaluation de la situation : Underworlds. Underworlds permet de représenter l'environnement du robot sous forme de structures de données distribuées en temps réel, contenant à la fois des graphiques de scène (pour la représentation de géométries 3D) et des chronologies (pour la représentation d'événements temporels). Dans Underworlds, l'environnement est considéré comme un ensemble de mondes qui peuvent chacun avoir des granularités spatiales et temporelles différentes, et peuvent hériter les uns des autres. Il fournit également un ensemble de bibliothèques clientes et d'outils de haut niveau pour introspecter et manipuler les modèles d'environne-

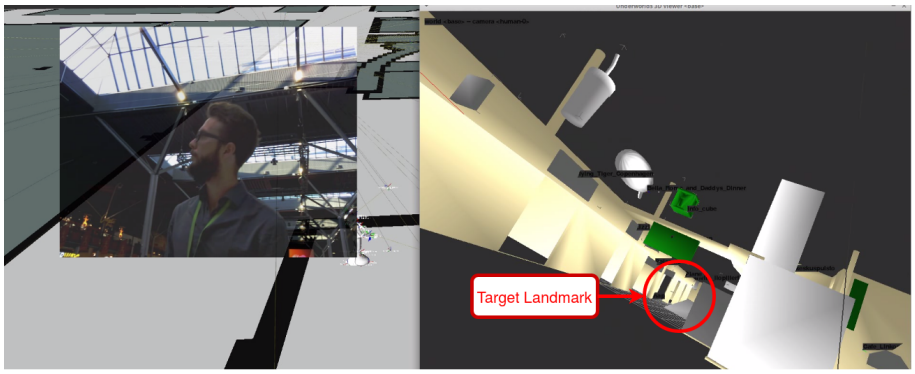
ment. La figure 9.8 montre des exemples de calcul de perspective dans le cadre de la tâche qui nous intéresse. C'est aussi ce composant qui assure le calcul de visibilité des points de repère et de chacun des agents.



(a) L'interaction commence, la personne demande d'aller au magasin *Illopileri*



(b) Le robot se déplace à la position donnée par SVP planner pour pouvoir pointer vers le premier point de repère sur la route.



(c) Ensuite, le robot vérifie que la personne a regardé le point de repère (en rouge) en utilisant une estimation de sa perspective.

FIGURE 9.8 – Exemples d'estimation et de suivi de la perspective de l'humain pendant la tâche de guide. Les images à gauche sont celles de la caméra du robot, à droite on peut voir une estimation de la perspective de la personne avec laquelle le robot interagit.

Contributions associées :

[Lemaignan 2018] **UNDERWORLDS : Cascading Situation Assessment for Robots** Séverin Lemaignan, Yoan Sallami, Christopher Wallbridge, Aurélie Clodic, Tony Belpaeme, Rachid Alami IEEE/RSJ International Conference on Intelligent Robots and Systems

(IROS 2018), Oct 2018, Madrid, Spain.
https://hal.laas.fr/hal-01943917/file/lemaignan2018underworlds.pdf

[Sallami 2018] **Implementing an implicit false belief task by cascading situation assessment with Underworlds**
Sallami Yoan, Lemaignan Séverin, Clodic Aurélie, and Alami Rachid (2018). Workshop "Toward a framework for joint action :what about theory of mind ?" at Robotics Science and System conference (RSS'18), Pittsburgh,USA

9.4.4 Génération de routes et système de gestion d'ontologies travail conjoint WP5/WP4

Dans la tâche de description d'itinéraire, le robot n'accompagne pas le client jusqu'à sa destination, mais doit expliquer le chemin à emprunter. Cette tâche, qui nous paraît si naturelle dans un contexte humain-humain, nécessite en fait un ensemble de connaissances (par exemple sur le lieu, sur les repères possibles dans ce milieu particulier) et de « bonnes pratiques » (par exemple pour donner un chemin facile à suivre si possible) qu'il faut modéliser si l'on veut l'implémenter sur un robot. En collaboration avec HWU, les travaux du LAAS dans cette tâche ont produit une représentation spatiale sémantique, un composant de génération de route ainsi qu'un système de gestion d'ontologie : "Ontologenius".

Pour déterminer les « bonnes pratiques » et identifier les références utilisées dans le cadre d'un centre commercial, une étude en quatre phases a été menée dans le centre commercial Ideapark [Heikkilä 2019]. Les résultats de cette étude sont conformes aux trois pratiques de base identifiées par Allen [Allen 2003] pour la communication des connaissances sur les itinéraires. Elles peuvent se résumer ainsi : a) respecter l'ordre spatio-temporel, b) se focaliser sur l'information sur les points d'intérêt et c) utiliser des repères facilement identifiables par l'auditeur. L'étude nous a également permis d'identifier le type de questions que le robot devrait gérer. Nous avons remarqué que les clients ne demandaient souvent pas un magasin en particulier, mais un type de magasin (par exemple un restaurant) ou des lieux où certains articles sont vendus (par exemple un magasin où trouver des chaussures). De plus, nous avons remarqué qu'un magasin peut être nommé de différentes manières, ce qui signifie que le robot doit connaître toutes ces variantes (ou du moins les plus courantes).

Afin de mettre en œuvre les pratiques identifiées et de doter le robot de toutes les connaissances nécessaires sur le centre commercial, le LAAS a mis en place une première contribution sur la représentation spatiale. La représentation spatiale sémantique [Sarthou 2019a] (*Semantic Spatial Representation SSR*) vise à fusionner une représentation topologique et une représentation sémantique. Elle nous a permis de développer les deux fonctionnalités présentées dans Mallot et Basten [Mallot 2009] pour la tâche de description d'itinéraire. Elles consistent d'une part, à sélectionner une séquence de lieux menant à l'objectif et d'autre part, à gérer les connaissances déclaratives pour choisir la bonne action à expliquer à chaque point de la séquence. Avec le SSR, nous représentons l'ensemble des connaissances dans une ontologie avec au moins les trois classes identifiées par Kuipers [Kuipers 2000] dans sa hiérarchie sémantique spatiale : région, chemin et lieu. De plus, nous étendons la notion de lieu avec les concepts d'intersection et d'interface de chemin et la notion de chemin avec les concepts de couloir et d'espace ouvert. En plus de ces classes, nous avons défini douze propriétés pour décrire la topologie dans la représentation sémantique. En plus du SSR, nous avons développé trois algorithmes pour calculer

des itinéraires d'un point à un autre, sélectionner le "meilleur" itinéraire en tenant compte des préférences du client et générer une explication.

Puisque la représentation spatiale sémantique est décrite dans une ontologie et que tous les algorithmes qui vont déterminer le chemin à suivre pour sa verbalisation sont basés sur les connaissances représentées dans cette ontologie, nous avons besoin d'un logiciel efficace pour la gérer. Ontologenius [Sarthou 2019b] est un logiciel open source, léger, compatible ROS qui permet de stocker des connaissances sémantiques, de les explorer efficacement et d'effectuer des raisonnements sur ces connaissances. Ontologenius permet de partager cette connaissance sémantique entre tous les composants d'une architecture robotique permettant ainsi d'avoir un référentiel unique de connaissances. De plus, Ontologenius permet de traiter plusieurs bases de connaissances distinctes, correspondant aux connaissances du robot et aux connaissances estimées de ses partenaires humains, tout en prenant en compte la notion de connaissances communes de base. Il est également important d'observer que, grâce à ce logiciel, nous sommes en mesure de stocker les noms de concepts dans différentes langues et de définir des synonymes pour ces noms. De plus, ce logiciel prend en charge les ambiguïtés provenant de la reconnaissance de la parole. Par exemple, de cette manière, si le système de reconnaissance de la parole affiche `[dinsko]`, le logiciel est capable de le corriger dans `[Dinsko]`, d'autres exemples du déploiement : `[ti ti nalle / Ti-ti Nallen Koti]`, `[jewelsport / Juvesport]`, `[manga / Mango]`, `[mimi / Minimani]`, `[lord / Nordea]`.

Contributions associées :

[Sarthou 2019b] **Ontologenius : A long-term semantic memory for robotic agents**
Guillaume Sarthou, Aurélie Clodic, Rachid Alami IEEE International Conference on Robot & Human Interactive Communication (IEEE Ro-MAN 2019), Oct 2019, New Delhi, India
`https://hal.archives-ouvertes.fr/hal-02283342/file/ontologenius_final.pdf`

[Sarthou 2019a] **Semantic Spatial Representation : a unique representation of an environment based on an ontology for robotic applications**
Guillaume Sarthou, Rachid Alami, Aurélie Clodic Combined Workshop on Spatial Language Understanding (SpLU) and Grounded Communication for Robotics (RoboNLP), Jun 2019, Minneapolis, United States
`https://hal.archives-ouvertes.fr/hal-02177140/file/semantic_spatial_representation.pdf`

9.4.5 Supervision d'une tâche interactive Humain-Robot (WP5.3 avec WP2 et WP4)

Cette tâche a conduit au développement d'un composant de supervision par le LAAS qui est spécialement dédié à la mise en œuvre de l'activité de robot de guide. Ce composant gère les décisions du robot concernant le moment d'agir, ce qu'il faut dire et ce qu'il faut faire pendant la tâche. Il communique via ROS avec plusieurs composants du système : la perception, le raisonnement géométrique, l'ontologie, le planificateur SVP, la navigation et le dialogue. Le LAAS a travaillé en étroite collaboration avec WP2 (Idiap) et WP4 (HWU) pour les tests et l'intégration. La figure 9.9 montre le flux de tâches de guidage, son processus et ses interactions avec d'autres composants.

Ce composant a été initialement codé comme une machine à états basée sur la bibliothèque Smach (`http://wiki.ros.org/smach`). Cependant, à mesure que la tâche devenait plus sophistiquée, les machines à états étaient de moins en moins appropriées. Il n'était pas possible de concevoir certains comportements souhaités en raison des limitations de la

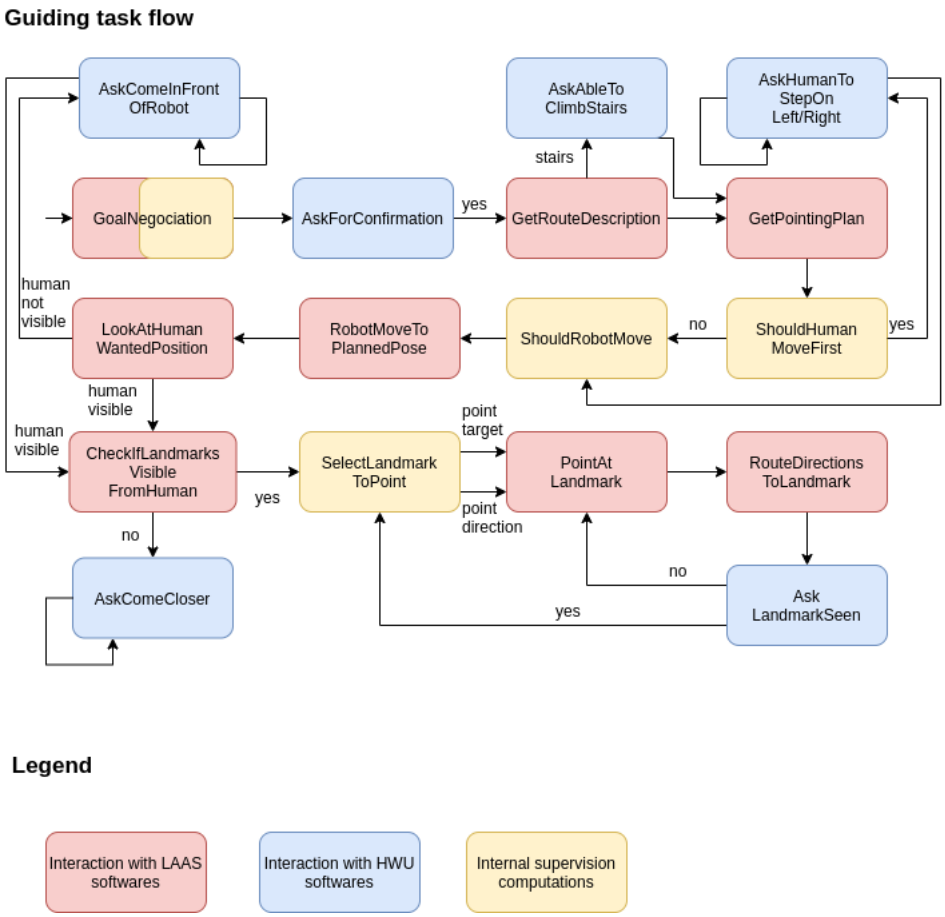


FIGURE 9.9 – WP5.3 Supervision : Représentation de la tâche de guidage

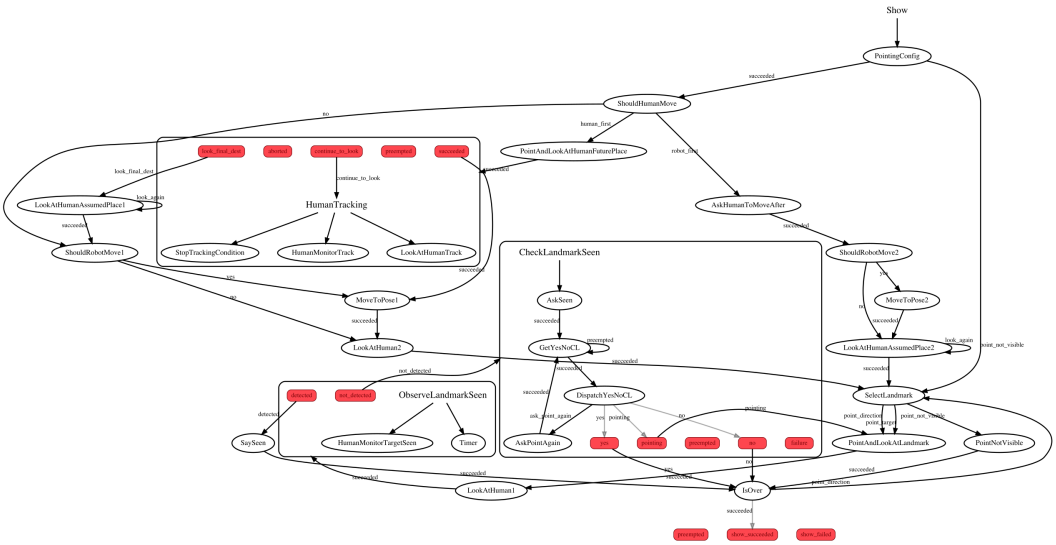


FIGURE 9.10 – WP5.3 Supervision : Augmentation de la complexité de la machine à états au fur et à mesure du développement de la tâche

bibliothèque et le graphe devenait complexe et difficile à appréhender et à modifier comme on peut le voir sur la figure 9.10. Ce composant a ensuite été entièrement réimplémenté à l'aide de Jason, un système basé sur le paradigme Belief Desire Intention et compatible ROS (`http://jason.sourceforge.net/wp/`) permettant plus de flexibilité et une conception modulaire. Dans cette dernière version, un composant pour gérer les sessions d'interaction a été ajouté. Il choisit quand commencer l'interaction lorsqu'une personne semble vouloir interagir avec le robot. Ensuite, il maintient l'attention du robot sur cette personne jusqu'à la fin de l'interaction, c'est-à-dire lorsque la personne part ou dit au revoir. Une session d'interaction comprend des tâches de chatbot et des tâches de guidage avec une seule personne. Enfin, un travail sur la mesure de la qualité de l'interaction a été initié et mis en œuvre. L'idée est de trouver quelles mesures pourraient être utilisées pour déterminer si l'interaction se passe bien (ou pas). Ces mesures ont été utilisées lors de l'étude utilisateur finale.

Contributions associées :

[Mayima 2019] **Evaluation of the Quality of Interaction from the robot point of view in Human-Robot Interactions**
Amandine Mayima, Aurélie Clodic, Rachid Alami 1st Edition of Quality of Interaction in Socially Assistive Robots (QISAR) Workshop, Nov 2019, Madrid, Spain
`https://hal.laas.fr/hal-02403081/file/QoI_paper.pdf`

9.4.6 Planification des mouvements, des placements et des postures (WP5.1)

Comme indiqué précédemment, le planificateur de perspective partagée (SVP Planner) fournit des positions pour le robot ainsi que la personne avec laquelle il interagit pour permettre un pointage efficace. Ainsi, le robot doit pouvoir naviguer de sa position actuelle à la position de pointage proposée par le planificateur SVP tout en (1) respectant ses contraintes cinématiques (2) en évitant les obstacles statiques et mobiles et (3) en esquivant les humains et en assurant leur sécurité physique et psychologique (par exemple, ne pas les surprendre ou les effrayer). Pour gérer les déplacements de la base du robot, un logiciel de planification de trajectoire basé sur une *Human Aware Time Elastic Band (HATEB)* a été mis en place [Khambhaita 2017b]. Cet algorithme est capable de planifier et d'exécuter des trajectoires tout en assurant la satisfaction des contraintes cinématiques du robot, en évitant les obstacles non humains statiques et mobiles et en planifiant des solutions de co-navigation respectant les contraintes sociales avec les humains. En effet, en exécutant l'optimisation sur les trajectoires planifiées du robot et de l'humain, nous sommes en mesure de trouver une solution de co-navigation pour le robot et l'humain et d'ajuster cette solution en fonction de nombreuses contraintes (par exemple, le robot ou l'humain doit-il fournir le plus d'efforts au cas où survient un conflit de trajectoires). Cependant, HATEB a besoin d'un calcul précis et en temps réel des positions et des mouvements humains. C'est quelque chose que nous n'avons pas pu obtenir avec précision à Ideapark (par exemple, lorsque le robot bouge, il perd la position de la personne car le seul capteur utilisable est la vision). Pour cette raison, ce planificateur a été adapté [Singamaneni 2020a].

9.5 Résultats

9.5.1 Intégration et déploiement (WP7 avec WP5)

Plusieurs semaines d'intégration ont eu lieu à Ideapark (pour le LAAS : 1 semaine en septembre 2018, juin 2019 et septembre 2019) pour permettre le déploiement final. Pendant ces semaines de travail intensif, il s'agit avec les autres partenaires du projet, de vérifier que tout ce qui a été testé fonctionne "sur le terrain", e.g.

> *"Je suis allé à la semaine de déploiement en juin 2019 pour affiner l'intégration avec les partenaires et essayer d'éliminer le plus de bugs possible, en passant 55 heures dans le centre commercial et j'ai refait la même chose pour le déploiement final de septembre 2019. Pour ce dernier déploiement en équipe, nous devions finaliser les installations des logiciels sur l'ordinateur qui fonctionnera tout seul pendant les 3 mois de déploiement en monde réel. Nous avons également continué à tester nos logiciels et à les rendre autonomes, c'est-à-dire faciles à lancer et faciles à redémarrer en cas de problème. Nous avons eu des difficultés avec l'installation et l'utilisation de la localisation. Nous y sommes parvenus finalement mais avec la contrainte que le robot ne puisse pas interagir avec les gens pendant une minute à la fin d'une session d'interaction pour qu'il puisse chercher un marqueur à relocaliser. En effet, après quelques tâches de navigation, le robot dériverait et se perdrait puis pointerait dans la mauvaise direction." Amandine*

Le déploiement à long terme a eu lieu de septembre 2019 à décembre 2019 pour un total de 42 jours. Durant cette période, au moins une personne de l'équipe technique a été impliquée dans le lancement et le suivi à distance de la démo. Cela signifie que tout a été préparé par tous les partenaires du projet avant de permettre un lancement de démonstration à distance et "aussi simple que possible". L'équipe d'Ideapark a apporté son soutien avec la présence de l'un d'entre eux pendant le temps de travail du robot. C'était pour des raisons de sécurité. Pour des raisons de sécurité également, le déplacement du robot n'a pas été utilisé lors de cette phase de déploiement.

9.5.2 Evaluation (WP7 et WP5)

Après le déploiement final du système robotique, en coopération avec VTT et le personnel d'Ideapark, le LAAS a mené une étude utilisateur auprès des clients d'Ideapark afin d'évaluer comment le comportement du robot influence l'efficacité de la description de l'itinéraire et la qualité perçue de l'interaction par les clients. Plus précisément, nous avons comparé trois versions du système :

- Condition 1 : dialogue : le robot explique seulement le parcours
- Condition 2 : dialogue + pointage : le robot explique le parcours et pointe sur le premier élément du parcours
- Condition 3 : dialogue + pointage + prise de perspective (système complet) : le robot explique l'itinéraire et se déplace pour pointer le premier élément de l'itinéraire de manière à ce que le client voie l'élément pointé

Nous avons choisi trois boutiques avec des critères différents selon leur visibilité et leur éloignement de la zone centrale où le robot a été déployé. Chaque participant était confronté aux trois conditions. Les magasins ont été contrebalancés dans les conditions expérimentales. L'ordre des conditions était également contrebalancé pour chaque participant. Les interactions entre les participants et le robot étaient enregistrées sur vidéo, et

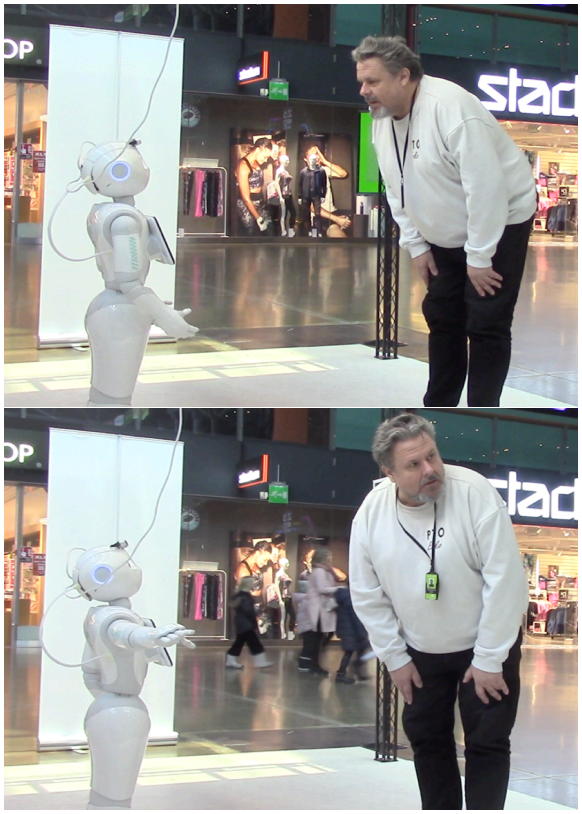


FIGURE 9.11 – Illustration d'un participant en interaction avec le robot. Sur la seconde image, le participant regarde la direction pointée par le robot.

les participants avaient, entre chaque condition, deux questionnaires (questionnaire PeRDITA [Devin 2018] + un supplémentaire sur la qualité de l'interaction) à remplir et un test informatique (logiciel PsychoPy) dans lequel le participant devait trouver son chemin selon les indications données précédemment par Pepper, grâce aux photos du centre commercial. Un entretien à la fin de l'expérience a permis aux chercheurs de recueillir les impressions des participants. 30 participants ont pris part à l'étude. La figure 9.11 montre un exemple d'interaction. Les résultats sont en cours d'analyse.

Le recrutement des participants étant toujours difficile, surtout ici à l'étranger, nous avions construit un site internet (`https://sites.google.com/view/pepper-exam/home`) associé à un questionnaire de recrutement (`https://docs.google.com/forms/d/e/1FAIpQLSc3NHiffXpc0WLJ-rYDBu-IxrBzVnxn4ZHizkF5zaGgcyLYWw/viewform`). Tout cela nous a permis d'avoir suffisamment de participants pour notre étude (30 au final). C'était assez difficile, car tout dans l'étude des utilisateurs était en finnois, c'est pourquoi nous avons été aidés par nos collègues de VTT et le personnel d'Ideapark.

9.6 Conclusions

Notre travail dans ce projet visait à développer les capacités nécessaires dans la planification des mouvements, les tâches et la coordination des mouvements et enfin l'exécution, pour permettre à un robot autonome d'agir et d'interagir avec les humains et à proximité physique des humains. Nos objectifs ont été atteints. La tâche de guidage, ou plus précisément de description de route, est une partie de la démonstration globale du projet

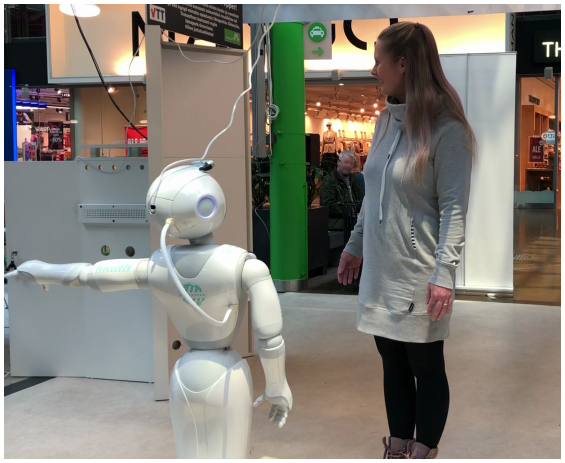
(a) Le client écoute Pepper au début de la tâche.

(b) Le client regarde dans la direction pointée par Pepper.

(c) Le client écoute Pepper au début de la tâche.

(d) Le client écoute Pepper tandis qu'il indique la direction à suivre

(e) Le client écoute Pepper au début de la tâche.



(f) Le client regarde dans la direction pointée par Pepper.

FIGURE 9.12 – Des exemples de déroulement de tâches de description réalisés lors du déploiement à Ideapark.

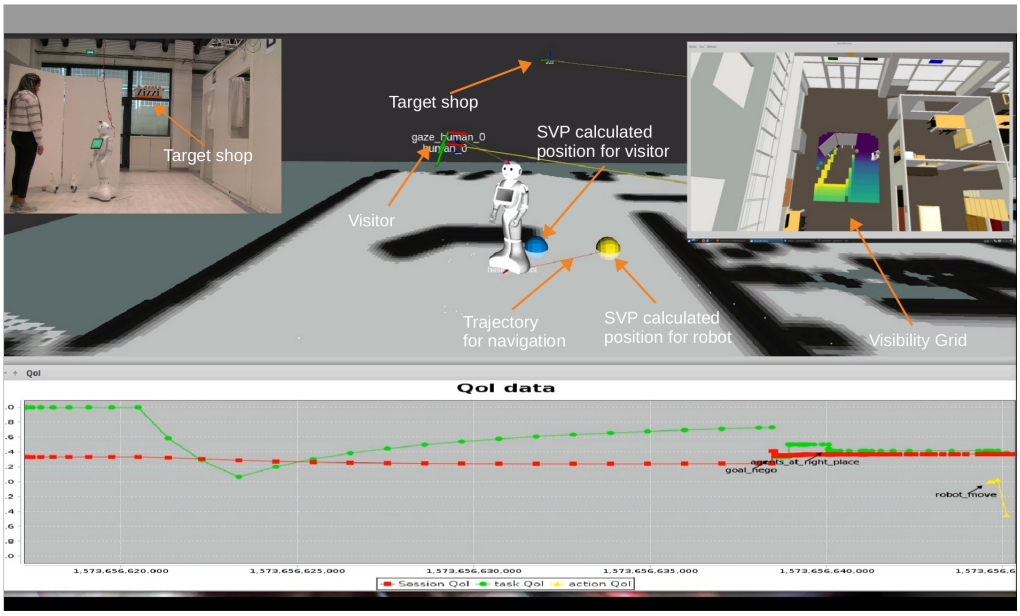


FIGURE 9.13 – Illustration dans le contexte de la maquette du LAAS

MuMMER qui a été illustrée dans le centre commercial. Une vidéo du scénario, illustrée à la figure 9.13, est disponible [Singamaneni 2020b] : `https://youtu.be/KfOs23wFzIQ`. Nous avons relevé le défi de déployer le robot dans le centre commercial, dans un environnement réel, avec de "vraies" personnes. Nous pouvons être assez satisfaits du résultat puisque le système a fonctionné pendant trois mois à Ideapark. Cependant, cela nous a amenés à poser des questions concernant la sécurité. En effet, pour pouvoir faire fonctionner la navigation avec Pepper, nous avons dû désactiver une sécurité. Nous avons donc fait le choix que le robot ne bougerait pas si une personne de l'équipe du projet n'était pas présente. De plus, une personne du personnel d'Ideapark était toujours présente auprès du robot pendant le déploiement. Oui, le robot ne semble pas être "dangereux" mais dans ce genre d'environnement, où vous avez des enfants par exemple, vous ne pouvez pas laisser place au doute en termes de sécurité. À l'avenir, nous devrons réfléchir à la manière de gérer un tel défi.

9.7 Introspection

J'aimerais revenir ici sur cette expérience du projet MuMMER, de l'expérience du travail au sein d'un "gros" projet européen et du travail d'intégration d'une démonstration en environnement réel. Tout ce que je vais dire ici, on le sait tous déjà et ce sont des évidences, mais çà n'empêche qu'il est bon de le rappeler.

9.7.1 Avec ma casquette de responsable technique

Dans un projet européen, il y a beaucoup de partenaires, il faut s'assurer dès le départ d'avoir un consortium avec des gens que vous connaissez ou avec qui vous avez envie de travailler. Ensuite, vous ne travaillerez peut-être pas avec tout le monde, ce n'est pas grave. L'important c'est de trouver des points d'accroche avec les partenaires qui vous intéressent en priorité. Il y a des choses pour lesquelles vous arriverez au bout et d'autres que vous

abandonnerez en cours de route.

Au fil du projet, vont se former plusieurs lignes directrices. Normalement, c'est au coordinateur de suivre et d'assurer la cohérence de l'ensemble. Si ce n'est pas le cas, assumez ce que vous voulez faire et assurez-vous du soutien des partenaires avec qui vous souhaitez travailler.

Lorsqu'il y a dans le projet une démonstration intégrée avec des logiciels provenant de divers partenaires, il faut assurer assez tôt dans le projet que chacun des partenaires partage la même vision. Il faut réaliser une conception où chacun trouve sa place. Il faut penser le plus tôt possible aux interfaces et au moment où il sera possible de faire l'intégration. Prévoyez DES semaines d'intégration où tout le monde va travailler de concert pendant plusieurs jours. Prévoyez des versions 0 de chacun des logiciels même avec des versions ultra-simplifiées, mais qui permettront aux autres composants utilisant vos résultats de commencer à travailler (bien sûr, assurez-vous que cette version 0 simplifiée n'est pas buggée). De plus un logiciel qui fonctionne, ce n'est pas un logiciel qui marche une fois dans un cas particulier (mais encore faut-il que l'on n'ait pas présenté ce cas particulier comme l'objectif...). Dans tous les cas, quand il est question de déploiement, il vaut mieux une version simplifiée qui fonctionne, qu'une version plus avancée, mais pas encore mature. Votre version plus avancée pourra être démontrée ailleurs (ou plus tard), privilégiez la simplicité et la robustesse quand il s'agit d'intégration. L'intégration ce n'est pas débugger UN logiciel, c'est débugger le fonctionnement de plusieurs logiciels ENSEMBLE. Bref, pensez à la supervision qui va avoir besoin d'utiliser tous les logiciels.

Il ne faut pas négliger les détails du terrain d'expérimentation et les envisager dès que possible. De plus, si vous souhaitez évaluer votre système, réfléchissez y dès que possible. Normalement cela devra être en ligne avec ce que vous développerez.

Au niveau de votre propre équipe aussi, il faut que quelqu'un ait la "vision" projet de l'ensemble des développements qui doivent être réalisés (généralement c'est la personne qui a participé au montage du projet) mais plus important il faut que cette vision soit partagée avec toutes les personnes qui travailleront sur le projet dans le laboratoire et surtout bien montrer comment l'ensemble des contributions vont venir se nourrir les unes des autres. Quels sont les défis scientifiques des uns et des autres (et, s'il y a lieu, le défi scientifique que représente l'intégration de l'ensemble). Si vous avez un engagement important dans un projet européen (je ne parle même pas d'une coordination), embauchez un post-doctorant pour la gestion interne du projet, la gestion des doctorants et des recherches associées. C'est un travail très important et il faut que quelqu'un puisse y consacrer son temps et son énergie à plein temps. Bien entendu, il faut aider et accompagner son travail, peu de personnes à part les permanents seront là du début à la fin du projet, et encore moins si l'on compte la période de montage du projet.

9.7.2 Avec ma casquette de co-directrice de thèse

Il peut être très intéressant pour un doctorant d'être partie prenante d'un projet de ce type. Déjà, car cela peut les financer, on est alors dans un cas particulier où normalement ce que doit faire le doctorant pendant sa thèse fait partie des attendus du projet. Cela peut être aussi intéressant de manière plus générale, car cela peut donner de la visibilité à son travail et lui donner l'opportunité de travailler avec des laboratoires en pointe sur tel ou tel aspect qui n'est pas étudié dans son labo. Cependant, un projet européen ce sont aussi des livrables et des échéances pour lesquels il faut assurer un travail souvent

conséquent. Il est donc important de définir clairement les buts et les objectifs de chacun d'une part pour sa thèse et d'autre part pour son travail pour le projet s'il doit en faire partie. Oui, un travail de thèse se construit et évolue au cours du temps, mais lorsqu'on parle d'intégration on parle aussi de besoin de synchronisation à certains moments et pour cela il faut un minimum de stabilité. C'est d'autant plus important qu'une thèse dure trois ans alors qu'un projet peut durer parfois plus longtemps.

Deuxième partie

Comment on va essayer de continuer

Chapitre 1

Perspectives de recherche

1.1 Introduction

Dans les médias, certaines personnes expriment leur crainte que les robots ne prennent le dessus sur les humains avant la fin du siècle. Cependant, les personnes qui ont travaillé avec de vrais robots savent mieux quels pas de géant doivent encore être franchis, même pour que les robots soient robustes dans les interactions quotidiennes avec les humains. La plupart des ressources du robot sont utilisées pour des fonctionnalités de base telles que la navigation dans des environnements partiellement connus, la reconnaissance d'objets, d'obstacles et de personnes. Avec tous les développements matériels et logiciels, ces fonctionnalités s'améliorent et, en outre, deviennent plus robustes. À mesure que la technologie évolue, un nombre croissant de chercheurs se sont concentrés sur le développement de robots capables d'impliquer et d'assister les utilisateurs dans différents domaines.

D'obstacles à éviter dans l'environnement pour les premiers robots autonomes, on souhaite aujourd'hui que les personnes puissent devenir des partenaires que le robot peut non seulement aider, mais avec lesquels il peut coopérer.

Pour atteindre cet objectif, les roboticiens ont développé un certain nombre de composants pour permettre la prise en compte de l'humain, par exemple au LAAS :

— lors de l'évaluation de la situation, e.g. en évaluant la situation non seulement du point de vue du robot, mais également du point de vue de la personne avec laquelle il interagit, et aussi et surtout en rendant cette évaluation accessible (e.g. calcul de isOn via du raisonnement géométrique) [Milliez 2014] [Lemaignan 2018] [Sallami 2019]

— lors de la planification de tâches, e.g. en calculant un plan prenant en compte les contributions possibles de l'ensemble des agents (et pas seulement du robot lui même) [Montreuil 2007] [Alami 2006] [Milliez 2016b] [Silva 2015] [Lallement 2014] [Buisan 2021]

— lors de l'exécution de tâches, e.g. en assurant l'exécution des actions par le robot mais également le monitoring des actions de la personne avec laquelle il interagit [Fiore 2014] [Devin 2017b] [Mayima 2021b]

— lors de l'exécution de fonctions tel que la navigation ; e.g. en assurant la prise en compte de tous les agents présents (e.g. human-aware navigation et manipulation) [Sisbot 2012] [Kruse 2014] [Khambhaita 2017b] [Khambhaita 2017a]

Des recherches ont été réalisées sur la façon dont les robots devraient se comporter dans une interaction sociale avec des agents humains, comme par exemple [Breazeal 2003],[Gockley 2007], [Satake 2009], [Leite 2013], [de Greeff 2015]. Cependant, même si ces premiers exemples sont des étapes importantes sur le terrain, il reste vrai que la plupart des aspects sociaux manipulés par ces robots sont encodés dans la tâche elle-même et ils ne présentent pas de cadre générique/général pour gérer une diversité de tâches dans une diversité de contextes (sociaux). Si nous voulons que les robots fonctionnent dans notre environnement quotidien, ils doivent également s'intégrer dans le monde social.

Dans leurs expériences, M. Veloso et al. montrent comment les gens sont prêts à aider un robot si il leur demande (par exemple, d'appuyer sur des boutons dans un ascenseur pour les amener au bon étage) et introduisent le concept d'autonomie symbiotique, dans laquelle les robots incluent explicitement des actions pour demander de l'aide aux humains dans leurs politiques de comportement [Veloso 2018]. Cependant, il n'est pas clair quelles choses peuvent être demandées et lesquelles ne devraient pas être demandées ou quoi faire si les gens ne se conforment pas à la demande. Il est clair que les robots doivent commencer à délibérer sur leur contexte social lorsqu'ils interagissent avec les gens. Cela pose d'emblée la question de la complexité. Il est nécessaire de donner aux robots des connaissances sociales et des capacités de raisonnement d'une manière qui limite le nombre de délibérations nécessaires dans chaque situation tout en permettant une flexibilité suffisante pour réagir de manière robuste à toute circonstance inattendue.

Un travail récent et intéressant allant dans ce sens est le travail de [Shih 2021]. Ils considèrent que dans le contexte d'une collaboration multiagent ou humaine-IA, les agents d'IA devraient être capables de s'adapter rapidement aux nouveaux partenaires en étant flexibles pour ajuster leur comportement dépendant de la convention spécifique au partenaire tout en réutilisant le même comportement dépendant des règles. Ils proposent un cadre qui sépare explicitement les représentations dépendantes des conventions et les représentations dépendantes des règles à travers des interactions répétées avec de multiples partenaires. Si nous proposons de fonder notre travail sur le cadre plus général des pratiques sociales (plutôt que sur les conventions), nous allons dans le même sens.

1.2 La robotique sociale

La robotique sociale vise à construire des agents sociaux artificiels incarnés qui peuvent assumer des rôles sociaux pour des tâches individuelles et collaboratives, et agir de manière autonome dans les espaces physiques et sociaux des interactions sociales humaines. Les robots sociaux sont testés en tant qu'enseignants, réceptionnistes, guides d'aéroport ou assistants domestiques. Des experts indépendants ont jugé que la robotique sociale était la deuxième plus importante parmi les dix premières technologies émergentes (Forum économique mondial) et l'association européenne de robotique prédit dans son programme de recherche stratégique que d'ici 2025 "la robotique influencera nos vies à tous égards aussi bien au travail et qu'à la maison". Cependant, comme l'ont souligné les chercheurs en sciences sociales, nous ne devons pas sous-estimer les complexités qui surviennent lorsque de nouveaux agents sociaux sont introduits dans une société. Si nous souhaitons réussir à amener des agents autonomes artificiels dans les espaces publics et domestiques, nous devons les doter d'un modèle sophistiqué du monde social qui reflète les normes et valeurs socioculturelles de la communauté dans laquelle le robot doit être placé, y compris des considérations générales sur l'éthique dans les systèmes autonomes ([for Ethical Considerations in Artificial Intelligence 2017], [Bryson 2017]). Mais est-ce même possible ?

Pour illustrer les défis auxquels les roboticiens sont confrontés, nous allons imaginer quelques scénarios.

Premièrement, imaginez que vous ayez un robot d'accueil et d'assistance dans une entreprise. Le robot accueille les visiteurs et peut les guider jusqu'à leur interlocuteur, collecte et distribue le courrier et les colis, arrose les plantes, apporte le café dans les salles de réunion, peut aider les travailleurs sur des tâches spécifiques, etc. Pour accomplir ces

différentes tâches, le robot doit être équipé de caméras et de microphones, mais il ne serait sûrement pas souhaitable que le robot enregistre toutes les conversations au sein de la société et filme en continu. D'un autre côté, il ne serait pas non plus efficace si un humain était nécessaire pour surveiller les actions du robot à tout moment. Au contraire, le robot devrait être un agent social autonome adapté aux normes et valeurs socioculturelles de la communauté et de la société qui l'intègrent. Le robot doit être capable de comprendre les rôles respectifs (droits, obligations, statuts, habitudes, etc.) que détient chaque personne, et quels types de comportement sont appropriés en interaction avec chacun d'eux - par exemple, qu'un visiteur ne doit pas être laissé seul ; qu'un visiteur ne peut pas demander des informations privées sur l'entreprise (ou inversement, le robot doit être sensible au fait qu'aucune information confidentielle ne doit être délivrée devant un visiteur) ; qu'une commande du patron soit classée avec la priorité la plus élevée si on le lui demande, et ainsi de suite. De plus, le robot doit être capable de reconnaître qu'il est dans une situation d'incertitude décisionnelle où différentes lignes de poursuites possibles doivent être évaluées. Les rôles et les réglementations avec lesquels le robot travaille doivent être ajustables pour des occasions spéciales (par exemple, maladie, vacances, occasions spéciales d'entreprise) et ne s'appliquer que dans certains contextes - si le robot apporte du café dans la salle de réunion et que le contenu de la réunion est confidentiel, il ne doit pas filmer ou enregistrer pour protéger la confidentialité.

Pour le second exemple, imaginez que vous possédiez un robot personnel chez vous. Il nettoie la maison, prépare les repas, aide les enfants à faire leurs devoirs, aide la grand-mère à localiser les objets manquants, arrose les plantes, etc. Pour accomplir ces différentes tâches, le robot doit être équipé de caméras et de microphones, mais il n'est pas souhaitable non plus que le robot enregistre toutes les conversations dans la maison et filme en continu. D'un autre côté, il ne serait pas non plus efficace si un humain était nécessaire pour surveiller les actions du robot à tout moment. Au contraire, le robot devrait être un agent social autonome adapté aux normes et valeurs socioculturelles de la communauté et de la société qui l'intègrent. Le robot doit être capable de comprendre les rôles respectifs (droits, obligations, statuts, habitudes, etc.) que détiennent chaque membre d'une famille, et quels types de comportement sont appropriés en interaction avec chacun d'eux - par exemple, que les enfants, mais pas les parents doivent être envoyés au lit à 21 heures ; que le conseil de la grand-mère pour les enfants n'est prioritaire que si les parents ne sont pas là, et ainsi de suite. De plus, le robot doit être capable de reconnaître qu'il est dans une situation d'incertitude décisionnelle où différentes lignes de poursuites possibles doivent être évaluées. Les rôles et les réglementations avec lesquels le robot travaille doivent être ajustables pour des occasions spéciales (par exemple, anniversaires, maladie, vacances) et s'appliquer uniquement dans certains contextes - si votre robot vous accompagne lorsque vous faites vos courses, il ne doit pas filmer ou enregistrer d'autres personnes afin que leur vie privée reste protégée.

Afin d'anticiper les difficultés potentielles liées à l'introduction des robots sociaux, nous devons agir de manière proactive et développer des outils théoriques et des modèles logiciels pour l'acquisition de compétences sociales et l'apprentissage éthique, de manière autonome ou interactive, chez les robots. Dans ce cadre, les objectifs globaux associés aux projets de recherche qui sont présentés dans la section suivante et sur lesquels je souhaite travailler sont les suivants :

Objectif 1 : Proposer un modèle pour définir et gérer le contexte social des robots. Ce modèle devra être capable de gérer plusieurs niveaux de contexte (éventuellement entre-

lacés). Il doit être capable de considérer que quelque chose a une signification particulière dans un contexte donné (ex : dans un jeu avec des enfants, une feuille de papier peut tenir lieu de billet de banque, mais pas à l'épicerie…). Il devra pouvoir inclure l'apprentissage de l'interaction quotidienne et évoluer. Il devra être capable de déduire que quelque chose est inconnu ou qu'une connaissance concernant la situation est incorrecte.
Objectif 2 : Concevoir et implémenter une architecture de contrôle pour exécuter ce modèle sur une variété de scénarios. Cette architecture de contrôle devra pouvoir tirer parti du modèle social pour rendre le robot entièrement sensible au contexte. Cela signifie, conscients du ou des rôles joués par le robot et les autres agents, capables de reconnaître les protocoles (sociaux) et de les utiliser de manière appropriée. De plus, l'architecture de contrôle sera conçue pour prendre en compte les valeurs et fournir des explications de comportement basées sur le contexte social.
Objectif 3 : Analyser des scénarios et des logiciels existant au regard du modèle ainsi défini. L'idée est de rétro-concevoir ce qui existe déjà pour trouver chaque paramètre qui peut être réglé (par exemple, avoir la possibilité de régler l'algorithme de vision pour suivre un seul visage à la fois (celui du propriétaire du robot) et pas les autres pour respecter la vie privée). Dans ce travail, nous nous intéresserons également à trouver des biais potentiels existants dans ces algorithmes (par exemple un algorithme de vision qui ne fonctionne pas pour les personnes portant des lunettes).

1.3 La notion de pratiques sociales

Les pratiques sociales (*social practices (SP)*) sont des façons acceptées de faire les choses, contextuellement et matériellement médiatisées, qui sont partagées entre les acteurs et routinisées au fil du temps [Reckwitz 2002] [Schatzki 1996]. Notre proposition est que les pratiques sociales pourraient servir de base à la modélisation de l'interaction sociale [Clodic 2018a]. Bien entendu, même si nous proposons de commencer par l'utilisation du modèle des pratiques sociales, d'autres possibilités doivent être explorées. Par exemple, les apports de la micro sociologie (e.g. l'inattention civile *civil inattention* ou les règles de conduite [Goffman 1963]) ou de la psychologie cognitive (e.g. les lisseurs de coordination *coordination smoothers* [Vesper 2017]).

En sciences sociales, les pratiques sociales (SP) sont définies à partir des notions suivantes [Holtz 2014].

— **Matériel** : couvre tous les aspects physiques de l'exécution d'une pratique, y compris le corps humain (concerne les aspects physiques d'une situation).
— **Signification** : fait référence aux problèmes qui sont considérés comme pertinents par rapport à ce matériel, c'est-à-dire les compréhensions, les croyances et les émotions (relatif aux aspects sociaux d'une situation)
— **Compétence** : désigne les compétences et les connaissances qui sont requises pour effectuer la pratique (relative à la notion de délibération sur une situation).

Les individus et les sociétés développent généralement un ensemble de pratiques au fil du temps qui peuvent être appliquées dans différentes situations. De plus, selon la situation, la personnalité et les compétences d'un individu, la réalisation d'une pratique sera un processus plus ou moins automatique ou délibéré. Par exemple, une salutation peut être une simple poignée de main rapide et automatique dans un environnement familier, tandis que la salutation doit être plus élaborée au début d'une réunion de négociation internationale. À chaque utilisation, des éléments de la pratique, notamment des savoir-

faire, des significations et des finalités, sont reconfigurés et adaptés [Shove 2012]. Par conséquent, l'utilisation de pratiques sociales comprend un apprentissage constant des individus utilisant la pratique sociale dans des contextes en constante évolution. De cette façon, les pratiques sociales guident le processus d'apprentissage des agents de manière naturelle.

Nous pensons que ce concept de pratiques sociales (SP) tel qu'il a été étudié en sciences sociales peut permettre de gérer la complexité des interactions cognitives, sociales et physiques de manière à ce que les robots puissent agir de manière robuste, flexible et efficace dans le cadre d'une interaction humain-robot. Nous proposons d'utiliser ce concept de pratiques sociales pour gérer les interactions car les humains eux-mêmes utilisent les pratiques sociales exactement dans le même but, c'est-à-dire simplifier les interactions sociales et les standardiser autant que possible. Nous ne voyons pas les pratiques sociales comme des scripts ou des protocoles rigides pour l'interaction, mais plutôt comme un ensemble de procédures généralisables.

1.4 Compétences sociales et éthiques pour un robot autonome

Notre objectif est d'intégrer la notion et les modèles proposés de pratiques sociales dans une architecture robotique. Le but est de permettre aux robots d'avoir un modèle du contexte social dans lequel ils évoluent, de reconnaître le ou les rôles qu'y tiennent chacun des acteurs et ainsi reconnaître les protocoles sociaux et les utiliser de manière appropriée. Pour cela, nous proposons de créer un modèle opérationnel des pratiques sociales qui permettra d'encadrer les interactions sociales auxquelles le robot participe. Il s'agit pour l'instant d'une architecture conceptuelle qui n'a pas encore été implémentée.

1.4.1 Pratiques sociales appliquées

Nous allons utiliser dans ce qui suit la description réalisée par Frank Dignum pour décrire un modèle des pratiques sociales à destination des informaticiens [Dignum 2015] et que nous avions aussi utilisé dans [Clodic 2018a].

Contexte (*Context*) La partie Contexte décrit les objets physiques qui sont utilisés ou nécessaires dans la pratique, les rôles et les acteurs impliqués ainsi que le lieu et le moment de la pratique. Cela ne décrit pas un contexte complet, mais seulement les parties nécessaires à la pratique.
Par exemple : dans un contexte robotique, il fournira des informations permettant une détection active pour trouver les objets et les personnes dans l'environnement.

Activités (*Activities*) La partie Activités décrit une sorte d'ontologie d'actions et de capacités qui sont couramment utilisées ou assumées dans la pratique.
Par exemple : dans un contexte robotique, cette partie de la pratique sociale peut aider à la planification des tâches car elle donne les actions et les exigences nécessaires pour les exécuter et ainsi le robot peut démarrer à partir de cet ensemble limité d'actions.

Significations (*Meanings*) La partie Significations facilite l'interprétation de la situation en termes de situation sociale abstraite et conduit à des manières alternatives de planifier pour atteindre un objectif. Les valeurs indiquent quel type d'objectifs

généraux sont servis par les actions. Les valeurs peuvent être utilisées pour évaluer les considérations morales et éthiques qui jouent un rôle dans la pratique sociale.

Attentes (*Expectations*) La partie Attentes au sein de la pratique définit l'ordre des actions, les paires déclencheur-action, les normes que l'on s'attend à suivre.
Par exemple, dans un contexte robotique, cette partie sera utilisée par la planification du robot (les attentes conduisent à une concentration de plans et d'actions possibles) et la perception (le robot recherchera activement les événements auxquels il s'attend à ce moment dans la pratique sociale).

1.4.2 Le *Social Practice Kernel*

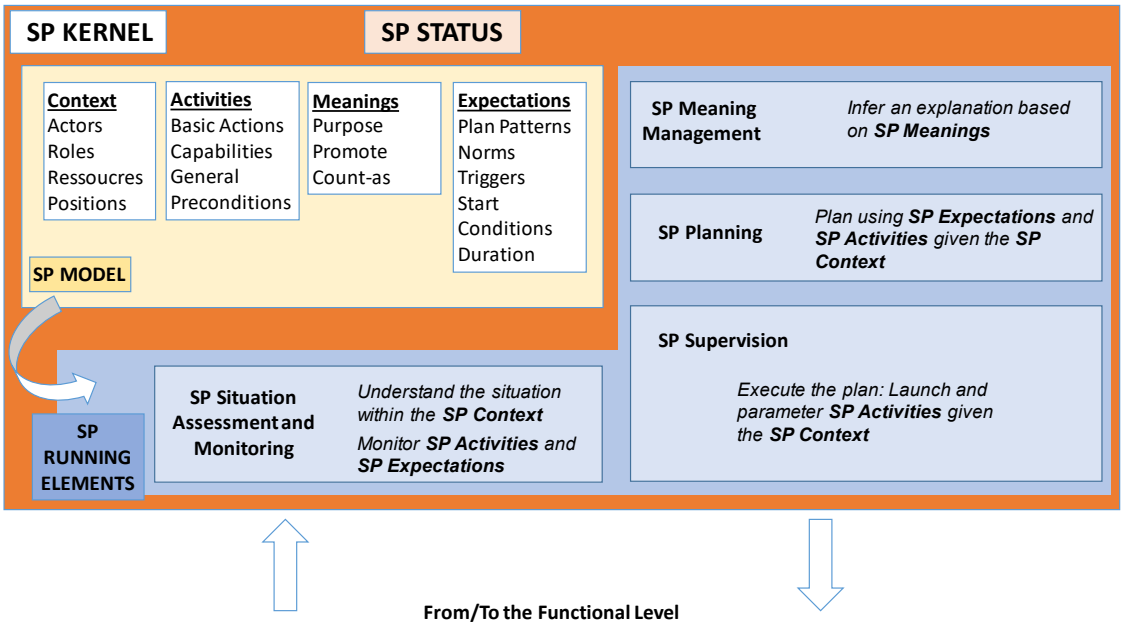


FIGURE 1.1 – Esquisse du noyau de pratique sociale (SP-Kernel)

Nous proposons de modéliser une pratique sociale sous la forme d'un kernel qu'on nommera *Social Practice Kernel*. Ce kernel contiendra non seulement la modélisation, mais aussi les éléments nécessaires à la pratique sociale, cela concerne notamment la compréhension de l'environnement, la planification, l'instanciation et la réalisation des actions. Un *Social Practice Kernel* représente une instanciation d'une pratique sociale. Nous proposons une première esquisse avec la figure 1.1. En plus de ce *Social Practice Kernel*, nous définissons un ensemble d'éléments de fonctionnement pour gérer la réalisation de la pratique sociale :

— *SP Situation Assessment and Monitoring* dont le rôle sera de focaliser l'attention sur les aspects pertinents de la pratique sociale considérée et de fournir une compréhension de la situation adaptée à la pratique sociale.
— *SP Planning* qui fournira un plan pour la réalisation de la pratique sociale (dans sa manière la plus simple, cela pourrait être un comportement) qui sera géré par le composant *SP Supervision*.
— *SP Meaning Management* devra pouvoir fournir une explication concernant le comportement du robot dans la pratique sociale.

1.4.3 L'architecture

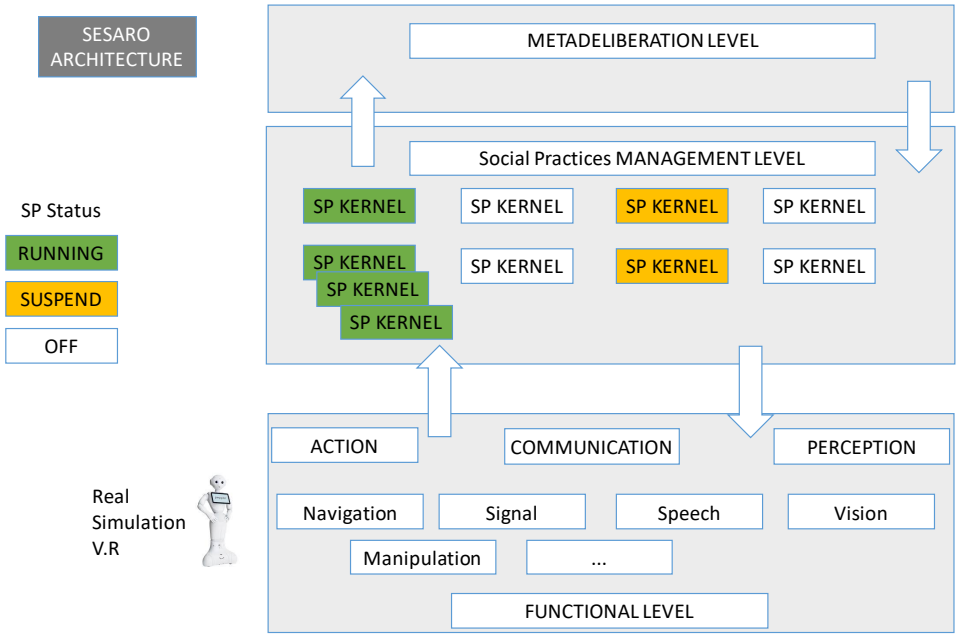


FIGURE 1.2 – Esquisse de l'architecture proposée

À partir de cet élément clé de Social Practice Kernel, la figure 1.2 représente une esquisse préliminaire de l'architecture que nous souhaitons développer. Elle est nommée SESARO pour Social and Ethical Skills for Autonomous RObot. Elle comporte trois niveaux : un *metadeliberation level*, un *social practices management level*, un *functional level* que nous allons chacun décrire plus en détail.

1.4.4 Le *metadeliberation level*

Le *metadeliberation level* représente le modèle social cognitif du système. Il est représenté figure 1.3. Notre objectif est de rechercher une solution générale afin d'être "agnostique" du type de robot ou de la tâche.

Il sera équipé d'un *Planner* qui décidera, à un rythme lent, quel(s) objectif(s) poursuivre compte tenu de l'état du monde et d'un *Agenda*. Nous parlons de rythme lent parce que, comme vous le faites avec votre propre agenda, il n'est pas prévu que le robot change souvent son ou ses objectifs généraux.

Par ailleurs, un ensemble de modèles, d'ontologies et de mémoires seront définis, tels que :

Social States History pour rassembler les données d'une interaction précédente ;

Agent/Actors Database si nous envisageons une interaction quotidienne avec la même personne, nous aurons besoin de recueillir des informations à leur sujet [1] ;

Social Practices Ontology connaissances stockées sur les pratiques sociales disponibles ;

Knowledge Base Ontology connaissances communes sur le monde.

Lorsqu'un objectif (ou un ensemble d'objectifs) devient actif, un *Goal(s) Planner* calculera un plan pour réaliser le ou les objectifs. À partir de ce plan, qui sera représenté

1. Nous nous occuperons des problèmes de confidentialité que ce type de bases de données pourrait entraîner et ajuster les connaissances que le robot est capable de gérer en conséquence

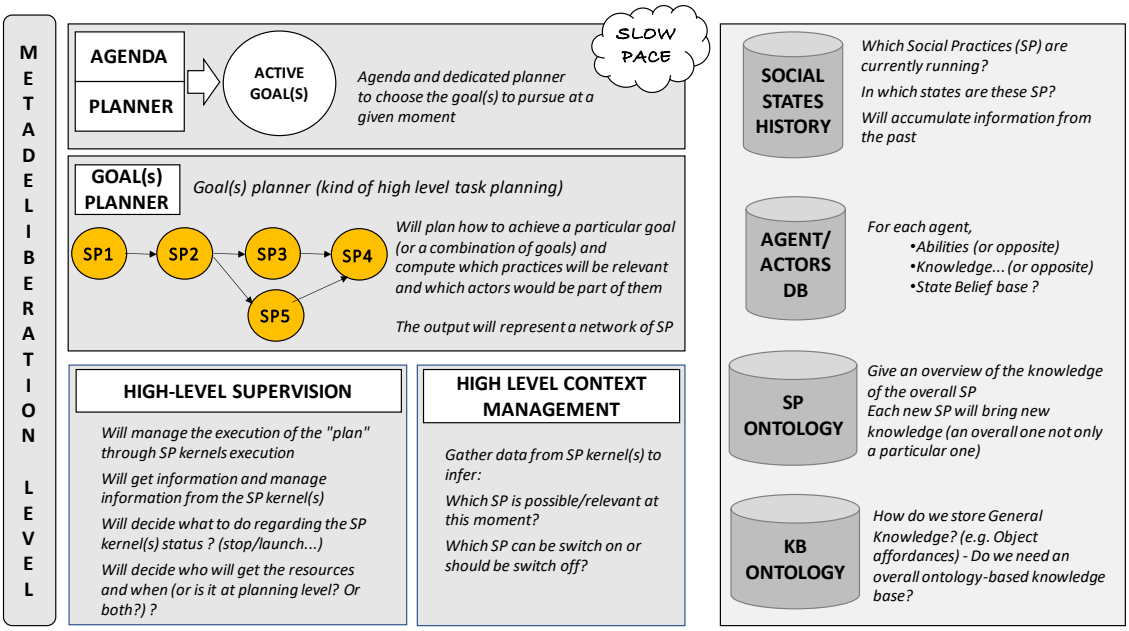


FIGURE 1.3 – Le *metadeliberation level*

comme un réseau de Pratiques Sociales (SP), il sera possible de : décider quel(s) noyau(s) SP doit être activé et quand, décider quel(s) noyau(s) SP doit avoir le jeton de ressources (si par exemple deux SP ont besoin de faire bouger le robot à un même moment), contrôler l'état de chaque noyau SP et superviser leur exécution. Cela sera fait par un composant *High-Level Supervision* et un composant *High-Level Context Management*.

1.4.5 Le *social practices management level*

L'élément clé de ce niveau est le *Social Practice Kernel*. Le *social practices management level* sera piloté par le *metadeliberation level*. Le système défini doit être capable de prendre en compte :

— des SP fonctionnant en parallèle (par exemple, SP "Apporter le courrier aux collaborateurs" pourrait fonctionner en parallèle avec SP "Dire bonjour lorsque vous rencontrez un collaborateur")
— des SP imbriquées (par exemple SP "Se comporter dans une salle de réunion" imbriquée dans SP "Se comporter en entreprise")

1.4.6 Le *functional level*

Il existe aujourd'hui un certains nombres de fonctions disponibles pour les robots (qu'il soit hardware, en simulation ou en VR). Chaque robot, équipé d'un ensemble de capteurs et d'actionneurs et de logiciels associés, dispose d'un ensemble de compétences. Des compétences de base tel que : goto, pick, place, say, look_at, face detection/recognition ; et des compétences plus interactives : regarder/identifier/reconnaître quelqu'un dans la scène, parler à quelqu'un, naviguer vers quelqu'un, reconnaître et suivre les ordres de base, remettre un objet… Si nous voulons bénéficier des modèles de pratiques sociales, nous devons réfléchir à la manière de redéfinir certaines de ces fonctions pour en extraire un éventuel réglage (via des paramètres ou des coûts). Par exemple :

— au niveau de la perception, nous devrions pouvoir régler l'algorithme de reconnaissance faciale pour ne regarder que celui impliqué dans la ou les pratiques sociales. De plus, au niveau fonctionnel, il devrait être possible de reconnaître des signaux (par exemple, agiter les mains) mais l'interprétation (par exemple, Bonjour ou Viens) se ferait au niveau de l'évaluation de la situation concernant les pratiques sociales en cours.
— au niveau de l'action, nous devrions être en mesure d'ajuster le volume et la vitesse de la parole du robot au moment précis de la pratique sociale que nous envisageons, par ex. dans le scénario bureau, entrant dans une salle de réunion, le robot adaptera son volume si la réunion est déjà en cours ou si les participants sont en pause.

Nous pensons qu'il sera nécessaire d'analyser les composants existants et peut être de les modifier pour permettre au robot de prendre en compte le contexte social et les exigences éthiques.

Il est encore trop courant aujourd'hui que ces paramètres ne soient pas rendus disponibles ou pire que des décisions soient prises au niveau fonctionnel alors qu'elles devraient pouvoir être pilotées par les aspects décisionnels. Ainsi un algorithme de planification et d'exécution de trajectoires qui s'occuperait de prendre toutes les décisions à son niveau empêchent toute prise en compte du contexte au niveau décisionnel et c'est bien dommage. Il est peut-être nécessaire que le robot roule lentement même si l'environnement semble dégagé parce qu'il y a des personnes âgées ou des enfants. Cela le contexte social peut nous le dire, mais il faut pouvoir ensuite le paramétrer au niveau de la navigation.

1.4.7 Gestion des connaissances et des données

Dans l'architecture que nous avons représentée dans la figure 1.2, au *metadeliberation level* et pour chaque *social practive kernel*, il existe un certain nombre d'endroits où les connaissances et les informations doivent être stockées, mises à jour et récupérées.

Nous n'avons pas encore une idée définitive de toutes les connaissances qu'il faudrait stocker ni quelles seraient les structures nécessaires. Cependant, nous avons déjà mis en place des architectures (e.g. [Lemaignan 2017][Sarthou 2021b][Sarthou 2021a]) qui ont bénéficié d'une gestion des connaissances cohérente et à partir desquelles nous pouvons définir plusieurs exigences :

— stockage des connaissances de différents points de vue (eg stockage de la connaissance du robot concernant la SP mais aussi celle de la personne avec qui il va interagir (eg un visiteur n'aura pas la même connaissance de la SP de l'entreprise qu'un collaborateur)) [Alami 2011, Clodic 2007b] ;
— utilisation autant que possible d'un système de gestion des connaissances basé sur une ou des ontologies pour bénéficier des capacités de raisonnement associées [Lemaignan 2010, Sarthou 2019b] ;
— être capable de stocker des connaissances dans le temps et de gérer une mémoire pour le système [Sarthou 2019b], mémoriser des événements et en tirer parti (ex : j'ai été informé qu'un des collaborateurs est absent une semaine donc je m'occuperai différemment de son courrier pendant cette période) ;
— être capable de gérer la génération d'expressions de références ([Buisan 2020b]) pour partager cette connaissance soit avec l'humain avec lequel le robot interagit, soit avec d'autres composants du système [Buisan 2020a].

1.4.8 Évaluation de la situation, évaluation du contexte et interprétation

En robotique, l'évaluation de la situation est nécessaire pour interpréter les relations entre les robots, les personnes et l'environnement, et fournit des données pour prédire les capacités, les connaissances et les comportements des humains qui interagissent avec le robot. Dans l'architecture que nous proposons, cela est nécessaire à deux niveaux. Au *metadeliberation level*, le *High-Level Context Management* fournira une compréhension du contexte par rapport au(x) objectif(s) du robot. Au *social practices management level*, dans chaque *social practices kernel*, un composant *SP Situation Assessment and Monitoring* fournira une compréhension du contexte concernant la ou les pratiques sociales.

1.4.9 Planification et monitoring

Dans l'architecture, nous considérons deux niveaux de planification. Au *metadeliberation level* serait décidé quels sont les *Active Goal(s)* et comment cela se traduit dans un réseau de pratiques sociales avec le *Goal(s) Planner*. Au *social practice management level*, pour chaque *social practice kernel* (éventuellement intriqué ou imbriqué), *SP Planning* planifiera en utilisant *SP Model* (*SP Expectations* et *SP Activities* étant donné le *SP Context*).

Malgré un objectif de tâche partagé et connu, un humain et un robot peuvent avoir plusieurs façons d'atteindre leurs objectifs grâce à l'interaction. Nous considérons que plusieurs SP pourraient être en compétition et donc plusieurs plans. Nous développerons un composant de gestion et de surveillance des attentes qui permettra au robot de comprendre sa situation par rapport à ces plans ou plans d'action possibles.

1.4.10 Supervision

Comme suggéré par l'architecture, l'exécution et la supervision seront nécessaires à différents niveaux.

Au *metadeliberation level* représenté figure 1.3, un composant *High-Level Supervision* sera consacré à la gestion des objectifs de haut niveau du robot. Il devra lancer/suspendre/arrêter/... le(s) *social practice kernel* pertinent(s) à partir du *social practice management level* chaque fois que nécessaire et surveillera leur exécution.

Ensuite, pour chaque *social practice kernel*, nous avons également représenté un composant *Social Practice Supervision* (cf. figure 1.1). Ce composant *SP Supervision* exécutera le(s) plan(s)/action(s) donné(s) par le composant *SP Planning*. Le plan donné doit être adapté pour l'exécution en fonction des contraintes de la/des pratique(s) sociale(s) en cours, par exemple je ne dois pas interrompre quelqu'un, je ne parle pas trop fort dans le couloir, je n'interromps pas une discussion en cours. Le composant *SP Supervision* devrait être capable d'ajuster les actions ou les compétences en fonction d'une ou plusieurs pratiques sociales particulières, par ex. pour remettre un courrier à un salarié, je prendrai soin de le placer à un endroit visible sur son bureau (ce qui pourrait être la spécialisation du "put"). La composante *SP Supervision* sera également en charge du suivi de la tâche (ex : est-ce que l'autre agent suit le plan) et par conséquent pourra renseigner si la ou les pratiques sociales choisies sont bien adaptées (ou non) au contexte actuel.

Pour concevoir et implémenter ces composants, nous pourrons profiter de nos travaux précédents [Devin 2017b], [Devin 2016b], [Fiore 2014], [Clodic 2009].

L'un des aspects intéressants apportés par la pratique sociale est qu'elle nous permettra

d'avoir un accès direct à un ensemble d'attentes et de signaux correspondants (quelque chose que nous avons esquissé [Castro 2019]). Par exemple, la secrétaire attend du robot qu'il livre un colis à un employé dans un délai donné. S'il arrive que le robot soit en retard pour une raison ou une autre, il doit envoyer un signal à la secrétaire pour l'avertir, non pas parce que la tâche a échoué, mais parce que l'attente de la secrétaire quant à sa réalisation possible doit être confortée. Une autre chose intéressante qui est souvent laissée de côté est le fait qu'il existe une hiérarchie, ce que nous appelons des pratiques imbriquées, et garder cette hiérarchie apporte des possibilités intéressantes pour envisager la réalisation de la tâche à différents niveaux (un peu comme nous l'avons fait [Clodic 2009]). Par exemple, considérons que la secrétaire demande au robot de livrer un colis à un employé - le robot doit être capable de gérer la tâche au niveau de la secrétaire (par exemple être capable de reconnaître que le travail a été fait ou est impossible pour quelle qu'en soit la raison) et au niveau de l'employé (par exemple, être en mesure de livrer le colis comme prévu). Si le robot rencontre la secrétaire avant de livrer le colis, il doit pouvoir expliquer pourquoi cela n'a pas encore été fait et quel est le plan. Pour ce faire, nous pourrons profiter du composant *SP Meaning and Management*.

1.4.11 Apprendre des pratiques au fil des actions

La planification de l'interaction avec les pratiques sociales nécessite qu'un concepteur spécifie l'ensemble des pratiques sociales avant une interaction. Mais le concepteur n'aura probablement qu'une connaissance incomplète des rôles, des normes, des conventions et des pratiques sociales que nous, en tant qu'êtres humains, définissons explicitement ou implicitement pour gérer notre vie quotidienne ensemble. D'un autre côté, une approche qui vise à apprendre des modèles et des représentations des états latents humains, tels que les pratiques sociales, à partir des seules observations du comportement humain peut échouer à récupérer le vrai modèle, puisque plusieurs modèles peuvent expliquer le comportement observé. Au lieu de cela, nous souhaitons développer la capacité d'un concepteur à spécifier des modèles partiels de SP, puis permettre au concepteur, ainsi qu'aux observateurs et collaborateurs humains, d'affiner de manière interactive les modèles de pratique sociale du robot et d'améliorer ainsi les interactions sociales au fil du temps. Nous pourrons par exemple envisager d'étendre le travail commencé pour lier l'apprentissage et la planification [Renaudo 2015].

L'objectif sera par exemple d'affiner les modèles de SP, via les entrées des observateurs humains quant à l'acceptabilité (ou l'inacceptabilité) des interactions, par exemple le robot a effectué telle action dans un contexte donné et on lui dit que dans ce contexte particulier il aurait fallu faire autrement.

1.5 En travaux...

J'ai proposé ici la définition d'une architecture robotique basée sur les pratiques sociales. Ces pratiques sociales seront utilisées comme connaissance à priori concernant les attendues et les possibilités offertes par un contexte social particulier. Ces connaissances permettront d'orienter les systèmes de perception du robot vers des régions d'intérêts. Je pense que cela donnera aux robots la capacité de mieux interpréter les intentions des humains et des autres robots, d'anticiper les changements dans l'environnement, d'agir conformément aux attentes des autres agents (humain ou non), et de raisonner sur le plan

d'action approprié pour répondre aux exigences de multiples tâches (éventuellement imbriquées et/ou intriquées). La modélisation des pratiques sociales permettra de capturer les attentes de l'interaction entre le robot et les autres agents (données par les rôles des agents définis par le contexte social) ce qui permettra de guider la prise de décision.

Les projets en cours vont me permettre de travailler au développement de cette architecture et des éléments qui la composent.

L'un des premiers aspects qui doit être étudié et travaillé selon moi est la notion de *social practice kernel* qui formera l'élément de base de l'architecture que je souhaite développer. Au niveau d'une *social practice* s'effectue une focalisation sur certains éléments de connaissances, une polarisation sur certains moyens de faire ou d'appréhender la situation. La vue globale est ensuite amenée par le *metadeliberation level*. Il permettra d'enchaîner plusieurs *social practices* mais aussi de gérer leur possible imbrication ou intrication. C'est à ce niveau aussi que seront gérées les connaissances et les possibles mémoires qu'il faudra mettre en œuvre. La compréhension de la situation prendra également en compte le contexte global pour gérer le pilotage de l'ensemble. Le niveau fonctionnel n'est pas à négliger. Il faut je pense s'équiper d'un ensemble d'actions de base pour commencer mais ensuite vraiment répandre l'idée qu'il faut donner des moyens aux systèmes décisionnels pour pouvoir adapter le comportement d'une fonction au contexte courant (i.e. plus fin que seulement démarrer/arrêter une fonctionnalité).

Le projet ai4hri permet de travailler sur les aspects planification et supervision dans un contexte social en prenant en compte plusieurs pratiques (e.g. être dans un magasin, interagir avec un client, interagir avec un employé, etc) et une possible interaction multipartites avec des horizons différents (selon que l'on interagisse avec un employé (plusieurs fois par jour ou même sur plus long terme) ou un client (sur du court terme)). Le projet DISCUTER permettra d'analyser comment la communication peut être intégré au modèle d'interaction sociale et donc au modèle de *social practice kernel* et comment le faire évoluer ainsi que la connaissance associée. Le projet ELSA permettra d'analyser comment prendre en compte les affordances sociales et leur apprentissage. Le projet EPIIC permettra d'apporter des moyens d'évaluation de l'interaction et peut être des moyens d'analyse sur les moyens d'interaction les plus pertinents.

Surtout chacun de ces projets propose la mise en place de scénarios d'interaction qui nous permettront de travailler au niveau architecture en utilisant des éléments que nous aurons développé mais aussi des éléments développés dans d'autres laboratoires.

Pour finir, je voudrais dire que j'ai longtemps voulu attendre pour écrire cette HDR d'avoir ma contribution claire nette et précise à présenter. Lorsque j'écris ces quelques lignes j'ai toujours cette impression que j'ai surtout donné des pistes et des opportunités, insufflé des idées, expliqué des concepts, favorisé des rencontres et qu'aujourd'hui et bien... le chemin ne fait que commencer... y a plus qu'à !

Chapitre 2

Projets en cours

Je suis actuellement impliquée dans quatre projets ANR :

— le projet ANR tri-latéral franco-germano-japonais **ai4hri**, *Artificial Intelligence for Human-Robot Interaction (ai4hri)* - Intelligence Artificielle pour l'Interaction Human-Robot

— le projet ANR JCJC **EPIIC** (porté par Raphaëlle Roy), *ElectroPhysiological Involuntary Inputs for Collaborative robotics enhancement (EPIIC)* - Utilisation de données issues de capteurs physiologiques pour l'amélioration de l'interaction humain-robot

— le projet ANR PRCI **ELSA**, *Effective Learning of Social Affordances for human-robot interaction (ELSA)* - Apprentissage d'affordances sociales en interaction humain-robot

— le projet ANR ASTRID **DISCUTER**, *Dialogue Interactif Structuré, Consolidé et Unifié pour la réalisationde Tâches En Robotique (DISCUTER)*

Dans cette partie, je vais donner une description de chacun de ces projets et expliquer mon implication.

2.1 Artificial Intelligence for Human-Robot Interaction (ai4hri)

Le projet **ai4hri** est un projet ANR issu d'un appel à projet trilatéral France-Allemagne-Japon en Intelligence Artificiele (IA). Le projet de 36 mois a démarré le 1/03/2021. Il regroupe :

— pour la France, l'équipe Robotique et InteractionS du LAAS-CNRS avec moi-même (PI), Rachid Alami (DR CNRS) et Ely Repiso Polo (Post-Doc recrutée en 09/2021 pour la durée du projet) (financement obtenu : 296 579 euros)

— pour le Japon, le HRI lab de l'Université de Kyoto, dirigé par le professeur Takayuki Kanda (PI) `https://www.robot.soc.i.kyoto-u.ac.jp/en/` avec Dražen BRŠČIĆ et Malcolm Doering

— pour l'Allemagne, l'équipe de l'Institut d'Intelligence Artificielle de l'Université de Brême, dirigée par le professeur Michael Beetz `https://ai.uni-bremen.de/start` avec Mona Abdel-Keream

2.1.1 Description

Si nous envisageons que les humains partagent leur espace de travail avec des robots de service, les robots doivent apprendre à interagir avec les humains d'une manière socialement appropriée. Le problème est que l'interaction sociale est très complexe car elle dépend d'une multitude de facteurs psychologiques, sociologiques et autres.

L'exemple principal que nous utilisons dans le projet ai4hri est celui un robot-employé dans un magasin de détail (figure 2.1). Nous envisageons que le robot travaille pendant les heures d'ouverture dans le magasin avec un autre employé pour assurer le fonctionnement du magasin, mais aussi qu'il puisse aider les clients de diverses manières. Par exemple, le robot doit : a) connaître le magasin et les articles qu'il contient, de sorte qu'il soit par ex. capable de détecter les changements comme les articles manquants et de réapprovisionner les étagères ; b) être capable d'interagir socialement avec les clients d'une manière qui leur semble naturelle, y compris les saluer, donner des informations sur le magasin, les guider vers les articles souhaités, etc. ; c) être capable de coordonner ses actions avec les autres employés, par exemple pour s'occuper des clients lorsque d'autres employés sont occupés. Nous montrerons qu'en appliquant l'architecture ai4hri, il sera possible de créer un robot compétent et capable d'interagir pour les magasins de détail. Bien sûr, il s'agit d'un cas d'utilisation. Notre objectif est de concevoir l'architecture et ses composants de manière à

FIGURE 2.1 – Dessin représentant un exemple visé dans le projet ai4hri, un robot employé dans un magasin ayant des connaissances sur le magasin et étant capable d'interagir aussi bien avec d'autres employés ou avec un client

ce qu'elle puisse être utilisée dans d'autres contextes (par exemple, maison, atelier, etc.).

Au cours de ce projet, les capacités que nous souhaitons développer sont les suivantes :

1. Modéliser les connaissances et développer des capacités de raisonnement à la fois sur l'environnement et sur les autres agents qui s'y trouvent
2. Travailler conjointement et/ou interagir avec d'autres agents en fonction de leurs rôles
3. Apprendre une interaction sociale
4. Prendre en charge une interaction dans sa globalité (quelle soit perçue comme une interaction court-terme ou une interaction long-terme)

Avec le projet ai4hri, grâce à une collaboration étroite des partenaires, nous avons l'ambition de concevoir, réaliser et étudier un cadre commun pour l'interaction humain-robot. L'architecture que nous construisons a pour but de fournir un système de gestion des connaissances associées à des capacités de raisonnement et d'apprentissage. Le système sera conçu pour gérer une interaction dans sa globalité. L'architecture et les composants associés seront utilisés par (et bénéficieront à) chaque partenaire. Ils seront également mis à la disposition de la communauté (comme le fait déjà UniHB : e.g. `knowrob.org/home`, `cram-system.org`).

2.1.2 Rôle

Je suis Coordinatrice et Responsable scientifique du projet pour la France.

Le projet est organisé en 6 workpackages :

WP1 Architecture
WP2 Représentation des connaissances et raisonnements associés
WP3 Apprentissage d'interactions sociales
WP4 Planification et exécution d'actions conjointes
WP5 Intégration, collecte de données, études utilisateurs
WP6 Dissemination et Management

Les interconnexions entre les workpackages sont illustrées par la figure 2.2.

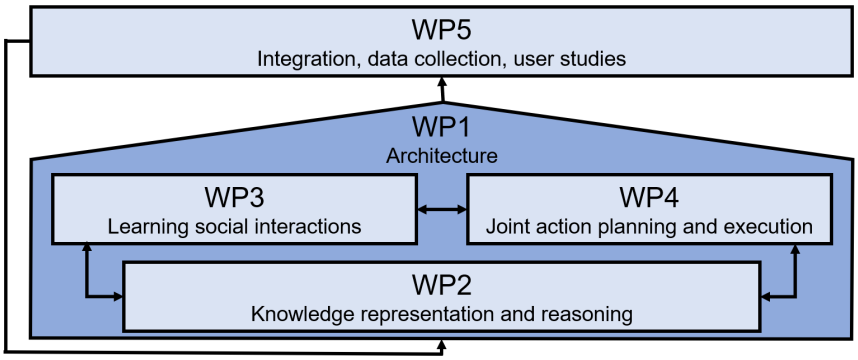


FIGURE 2.2 – Interconnexion entre les différents workpackages du projet

2.1.3 Statut

Le travail du LAAS se situe principalement au niveau du workpackage 4 Planification et exécution d'actions conjointes. Nous sommes en train de développer des capacités de planification et de supervision de tâches.

Planification Dans une première tâche de planification, l'objectif est de mettre en œuvre un composant qui permet au robot de planifier à la fois pour le robot et pour l'humain ([Alami 2006, Lallement 2018]).

De plus, si l'on considère la définition des représentations de tâches partagées, on dit qu'elles "ne se contentent pas de spécifier à l'avance quelles sont les tâches respectives de chacun des co-agents, elles fournissent également des structures de contrôle qui permettent aux agents de surveiller et de prédire ce que leurs partenaires font, permettant ainsi une coordination interpersonnelle en temps réel" [Knoblich 2011]. Nous avons développé un système de surveillance qui permet au robot de déduire l'état du plan et l'état des actions à la fois de son point de vue et du point de vue de l'humain dans [Devin 2016a]. Nous voulons maintenant ajouter ces informations au niveau de la planification. Pour ce faire, nous étudions le moyen d'utiliser la modélisation des engagements comme moyen de gérer les attentes et les signaux lors de l'interaction [Castro 2019].

Nous examinons également la capacité du robot à prendre en compte le contexte social au niveau de la planification, par ex. considérer l'interaction à court terme avec un client comme quelque chose de différent de l'interaction à long terme avec un employé. Ce travail s'inspire de travaux récents sur les pratiques sociales [Clodic 2018b]. De plus, pour certains modèles d'interaction spécifiques, nous travaillons avec **WP2** et **WP3** pour tirer parti de l'apprentissage de l'interaction sociale.

Supervision et Execution Dans cette tâche, nous développons un système capable d'exécuter le plan qui sera fourni dans la tâche précédente. Ce système de supervision doit être capable de gérer l'exécution du plan conjoint (exécution de l'action du robot tout en surveillant l'action humaine si nécessaire) ainsi que les imprévus.

Il est divisé en deux parties. La première est liée au processus de vérification du plan. La question à laquelle il faut répondre ici est la suivante : le plan suivi actuellement est-il le meilleur pour l'instant ? En effet, s'il n'est pas bon au niveau de l'interaction que le robot change de plan trop fréquemment nous allons mesurer la qualité de l'interaction et la qualité du plan pour décider quoi faire. Nous examinons les options où le changement de plan devrait être discuté avec l'autre agent, cette communication sera liée au plan.

La deuxième partie est consacrée à l'exécution du plan. Au niveau de l'exécution, le robot doit exécuter sa propre action tout en surveillant l'action de l'autre (chaque fois que nécessaire ou possible). Il doit être capable de gérer les imprévus ainsi que la communication liée à la gestion des objectifs ainsi que les engagements/attentes associés à cet objectif (par exemple, signaler qu'il attend quelque chose de l'humain). Parfois, le système doit délibérer pour choisir qui fera une action ou poser une question pour décider.

L'idée est d'être capable au final d'avoir un système d'exécution capable de prendre en charge une interaction avec plusieurs personnes. Le système devra également être capable de s'adapter aussi bien au contexte global (e.g. beaucoup vs peu de personnes dans le magasin, nous adaptons le volume des haut-parleurs du robot) qu'au contexte social (par exemple, interaction à court terme avec un client, interaction à long terme avec un autre employé).

2.2 EPIIC

Le projet EPIIC (ElectroPhysiological Involuntary Inputs for Collaborative robotics enhancement - Utilisation de données électrophysiologiques pour améliorer la collaboration humain-robot) est le projet ANR JCJC porté et obtenu par Raphaelle Roy en 2021. Le projet démarre en janvier 2022 pour une durée de 48 mois.

2.2.1 Description

En interaction humain-robot, la mesure de la qualité de l'interaction passe généralement par des questionnaires et des mesures de performance qui s'effectuent généralement a posteriori (après l'interaction). Il n'existe pas ou peu de techniques d'évaluation continue et/ou en ligne de l'état cognitif de l'opérateur humain. Ces dernières années, le domaine de la neuro-ergonomie a vu le développement de méthodes d'acquisition en ligne de données physiologiques. C'est le cas par exemple des interfaces cerveau-ordinateur qui permettent l'utilisation de ces données pour adapter ou modifier le comportement d'un système. A notre connaissance, cette technologie n'a pas encore été très étudiée dans le domaine de l'interaction humain-robot et plus spécifiquement dans le cadre de la robotique collaborative. Or, le fait que ces capteurs soient aujourd'hui portables et mobiles permet d'envisager cette possibilité.

Le projet EPIIC souhaite fournir une première évaluation de l'utilisabilité de données électrophysiologiques à partir de capteurs portables pour évaluer la qualité de l'interaction (QoI (Quality of Interaction)) dans le cadre d'une collaboration humain-robot.

Les principaux objectifs du projet sont les suivants :

— caractériser l'état cognitif des utilisateurs (e.g. l'effort cognitif et l'"automation surprise") lors d'une interaction humain-robot en utilisant des données électrophysiologiques involontaires sur deux tâches standards en interaction humain-robot (une tâche de navigation conjointe et une tâche de manipulation interactive) (WP1)

— créer un index QoI enrichi qui prend en entrée ces données traitées via un pipeline d'estimation de l'état mental (WP2)

— utiliser cet index QoI pour améliorer l'interaction (WP3)

— promouvoir les bonnes pratiques scientifiques, notamment via le partage de données (WP4)

Pour atteindre ces objectifs, il reste plusieurs barrières technologiques et scientifiques à franchir. Premièrement, tester la faisabilité de l'utilisation des capteurs (electro)-physiologiques lors de tâche de robotique collaborative. Il y a notamment une incertitude sur l'impact des tâches elles-mêmes, en terme d'activités physiques et d'artefacts associés, sur les signaux acquis. Deuxièmement, la définition même de la notion de qualité d'interaction (QoI) et la signification des données acquises pour cette mesure. Troisièmement, comment prendre en compte la mesure de qualité d'interaction (QoI) pour améliorer l'interaction elle-même en adaptant le comportement du robot. Quatrièmement, la recherche en interaction humain-robot (HRI) et en interface cerveau-machine (BCI (brain computer interface)) souffrent d'un manque de réplicabilité et de reproductibilité.

2.2.2 Rôle

Le projet a été écrit et est porté par Raphaëlle Roy, associate professor à ISAE-SUPAERO (`https://personnel.isae-supaero.fr/raphaelle-n-roy-211/?lang=fr`).

Raphaëlle Roy mène une recherche interdisciplinaire à la croisée des sciences cognitives, des neurosciences, de l'apprentissage automatique et de l'interaction humain-machine (IHM). Son objectif principal est d'étudier comment mieux caractériser l'état mental des opérateurs pour améliorer l'IHM ainsi que la sécurité et les performances. Pour cela, elle développe des méthodes pour extraire et classer les caractéristiques à partir de données physiologiques. Dans ce projet ANR JCJC, autour d'elle et de son équipe (notamment le prof. Frédéric Dehais de l'ISAE SUPAERO), la partie robotique du projet repose sur l'expertise de l'équipe Robotique et InteractionS du LAAS avec moi et Rachid Alami.

Bien entendu, nous ne partons pas d'une page vide. Pour la partie de mesure de la qualité d'interaction côté robotique, nous pourrons nous appuyer sur les travaux de thèse d'Amandine Mayima [Mayima 2021b]. De son côté Raphaëlle Roy et son équipe ont déjà commencé à étudier l'intérêt de l'utilisation des mesures physiologiques dans le domaine de l'interaction humain-robot [Roy 2020]. On peut noter que les deux équipes ont déjà publié ensemble il y a quelques années [Dehais 2011].

2.2.3 Statut

Au printemps 2022, deux stagiaires de master 2 en co-encadrement entre Raphaelle Roy et moi vont commencer à travailler sur le projet. L'un des stagiaires sera chargé de mettre en place une tâche de manipulation interactive avec le robot PR2 du LAAS. L'autre stagiaire sera chargé de la définition du protocole et de la mise en place du recueil des données physiologiques. Les deux stagiaires devront travailler de concert pour assurer la cohérence des données récoltées.

Avec ce premier setup, nous souhaitons réaliser un recueil de données. Ce travail sera poursuivi par un travail de thèse qui débutera en septembre 2022. Cette thèse sera en co-encadrement entre Raphaelle Roy et moi. Le déroulé du projet est représenté par la figure 2.3.

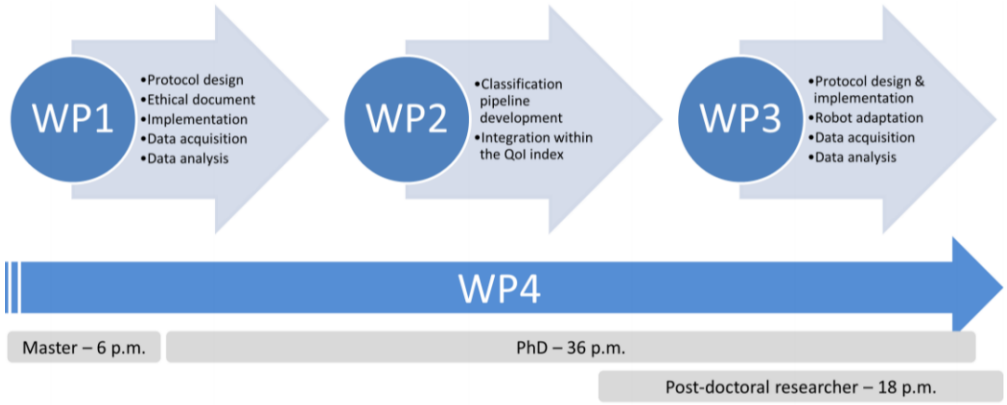


FIGURE 2.3 – Diagramme représentant l'évolution temporelle du projet EPIIC

2.3 ELSA

Le projet ANR Effective Learning of Social Affordances for human-robot interaction (Apprentissage d'affordances sociales pour l'interaction humain-robot) est un projet ANR PRCI issu de l'appel AAPG2021. C'est une collaboration entre la France et l'Autriche dont les partenaires sont :

— pour la France :
  - l'ISIR (Institut des Systèmes Intelligents et de Robotique) à Paris avec Mehdi Khamassi (Coordinateur pour la France - DR CNRS) et Benoit Girard (DR CNRS). Mehdi et Benoît ont tous deux une grande expérience des modèles cognitifs de prise de décision et d'apprentissage par renforcement pour les neurosciences et la psychologie et leur application à la robotique
  - le LAAS (laboratoire d'Analyse et d'Architecture des Systèmes) à Toulouse avec moi (Responsable Scientifique) et le support d'un ingénieur de la plate-forme robotique (financement obtenu : 200 000 euros)

— pour l'Autriche :
  - IIS le groupe de recherche sur les systèmes intelligents et interactifs (Intelligent and Interactive Systems research group IIS (`https://iis.uibk.ac.at`)) du departement d'informatique de l'université d'Innsbruck avec Justus Piater (Coordinateur pour l'Autriche, Professeur d'informatique, ses recherches portent sur l'apprentissage et l'inférence dans les systèmes sensorimoteurs) et Erwan Renaudo (Chercheur post-doctorant)
  - DiSC le centre des sciences numériques (Digital Science Center (DiSC)) de l'université d'Innsbruck avec Matthias Schurz (Assistant professor en psychologie et sciences cognitives, spécialiste de la cognition sociale humain)

Le projet a démarré en février 2022 pour une durée de 48 mois.

2.3.1 Description

L'objectif de ce projet est de doter les robots de la capacité d'apprendre les affordances sociales pour parvenir à une interaction naturelle et efficace avec les humains. Les affordances sociales signifient ici les représentations internes de l'effet attendu de l'exécution d'actions sociales spécifiques, et quand et comment cette exécution est pertinente. Par exemple, le robot doit apprendre qu'agiter la main en disant bonjour est une bonne chose à faire (e.g. c'est ce que font la plupart des humains), tandis qu'agiter la main lorsque l'humain demande au robot de saisir quelque chose est inapproprié.

Plus généralement, le concept d'affordances formalise l'idée de *ce que l'environnement offre à l'animal, en bien ou en mal*[1] [Gibson 1979]. Il a d'abord été principalement étudié dans le domaine de la psychologie, et a progressivement inspiré les recherches en robotique et en intelligence artificielle comme moyen de donner plus d'autonomie aux agents artificiels en leur permettant d'apprendre par eux-mêmes quelles commandes envoyées à leurs moteurs produisent des effets sur l'environnement, et lesquels n'en produisent pas [Sahin 2007, Chatila 2018]. Néanmoins, le concept d'affordances sociales reste largement inexploré en robotique et plus particulièrement en interaction humain-robot.

Ce que nous souhaitons montrer avec ce projet, c'est :

(1) que les robots peuvent apprendre les affordances sociales de la même manière

1. what the environment offers to the animal, either for good or ill

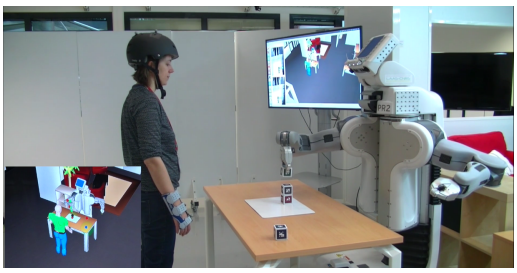


Figure 2.4 – **Scénario de manipulation interactive :** Les participants peuvent coopérer avec le robot pour atteindre un objectif particulier (*par exemple,* construire une tour avec des cubes). Le robot doit apprendre à dissocier d'une part la compétence des participants dans des tâches spécifiques et d'autre part les propriétés spécifiques des objets (*par exemple, aucun participant ne peut placer une balle au sommet de la tour cubique non pas parce qu'il est incompétent, mais plutôt parce que les balles tombent des tours*) .

qu'ils doivent apprendre les affordances des actions réalisées dans des situations non sociales ;

(2) que les robots qui pourront reconnaître de manière autonome les affordances sociales lorsque l'humain initie une action interactive répondront plus efficacement et de manière appropriée à l'humain, facilitant ainsi la coordination et la coopération humain-robot.

À cette fin, nous chercherons à doter les robots de la capacité d'explorer quelles réponses pourraient déclencher les différentes actions, mouvements ou expressions faciales qu'ils effectuent en présence d'une personne. L'idée est d'apprendre des modèles internes d'affordances sociale "générique" (i.e. qui produisent systématiquement le même effet chez les humains (par exemple, les humains sourient généralement en retour lorsque le robot sourit)), et possiblement "spécifique" à un contexte social particulier.

Pour illustrer ces processus, l'un des scénarios centraux dans lesquels nous testerons nos algorithmes d'apprentissage est un scénario de manipulation interactive (Fig. 2.4), où le robot doit engager un ou plusieurs participants à coopérer afin d'atteindre un objectif particulier (*par exemple,* construire une tour avec des cubes). Le robot doit apprendre de manière autonome que certains humains sont plus compétents que d'autres pour les aider dans des tâches spécifiques ; ou qu'aucun humain ne peut effectivement saisir un cube situé dans des zones de la table qui lui sont inaccessibles. De plus, le robot doit également reconnaître quelles actions effectuées par l'humain sont sociales et ainsi l'obliger à agir en réponse.

Alors que les affordances physiques sont bien étudiées en tant qu'approche moderne des boucles perception-action et de l'analyse de l'environnement, les affordances sociales ont été largement ignorées. Seuls quelques articles ont abordé cette question, parfois seulement partiellement. Dans ce projet, nous étudierons les problèmes suivants liés aux affordances sociales en robotique :

— Comment les robots peuvent-ils mieux prendre en compte la façon dont les personnes perçoivent et utilisent les affordances sociales des autres agents ? Quelle information signale la disponibilité d'une affordance ? Quel comportement permet d'utiliser les affordances sociales pour produire des interactions efficaces ? **(WP1)**

— Comment intégrer les affordances sociales dans une architecture modulaire de prise de décision et d'apprentissage par renforcement pour les robots ? Comment concevoir un modèle du monde de manière à englober la variabilité des réactions humaines aux sollicitations des robots ? **(WP2)**

— Comment un robot doit-il se comporter pour utiliser efficacement les affordances

sociales offertes par d'autres agents ? Quelles informations sont importantes pour représenter avec précision les effets de l'interaction afin que le robot puisse augmenter progressivement son répertoire d'actions dans des contextes sociaux ? **(WP3)**

— Comment prendre en compte et tirer parti des affordances sociales (y compris leur apprentissage) dans une architecture pour l'interaction humain-robot ? Que faut-il ajouter/modifier à la prise de décision et à la planification des tâches ? **(WP4)**

2.3.2 Rôle

Pour ce projet, en plus de mon implication et de l'implication d'un ingénieur de la plate-forme robotique, nous disposons du budget pour une thèse et de 8 mois pour un ingénieur.

Le projet ELSA répond à un manque de connaissances de base sur la façon dont les affordances sociales sont apprises et utilisées dans le contexte d'une interaction. Il permettra de développer une approche originale de la représentation des actions et des connaissances associées lors d'une interaction sociale dans un contexte robotique :

i) Aborder un sous-domaine relativement inexploré (affordances sociales en robotique) qui s'appuie sur des domaines bien étudiés (affordances en robotique mais aussi en psychologie)

ii) S'inscrire dans la continuité des travaux antérieurs du consortium, y compris des publications conjointes (projets RoboErgoSum et IMAGINE)

iii) Être hautement interdisciplinaire et impliquant des expertises complémentaires.

Le projet est divisé en 4 workpackages :

WP1 : Théorie et Pratique des affordances sociales (lead DiSC)
WP2 : Affordances sociales et prise de décision (lead ISIR)
WP3 : Affordances sociales et exécution des actions (lead IFI)
WP4 : Architecture, scénario, integration (lead LAAS)

Le travail du LAAS se situera en premier lieu au niveau du WP2/WP3 pour étudier le lien entre les affordances sociales et les actions au niveau planification et exécution. Ainsi, nous étudierons comment prendre en compte les affordances sociales au niveau de la planification des tâches en nous basant sur nos précédents travaux e.g. HATP Human-Aware Task Planner ([Montreuil 2007, Alili 2009a, Lallement 2018]). HATP est un planificateur qui tire parti d'un HTN (Hierarchical Task Network) laissant la possibilité de considérer le plan à différents niveaux de précision et qui permet également de calculer un plan pour plusieurs agents (plus précisément un plan pour chacun des agents). La gestion des aspects sociaux apportera de nouveaux défis à la planification. Parmi eux, il faudrait permettre la prise de décision lors de l'exécution (ex qui fera une action [Devin 2017b], mais aussi comment une action serait faite en fonction du contexte social (ex s'il doit y avoir un échange d'objet entre deux agents, la décision de donner à la main ou de poser à côté de la personne pourrait être considérée au moment de l'exécution)). De plus, nous étudierons comment modéliser les attentes et ajouter ces informations dans le plan (par exemple si la personne attend une action du robot, le robot devrait donner des indices s'il est engagé ou non pour le faire (on pourrait peut-être envisager des travaux sur la gestion des engagements [Fernandez Castro 2019]). Nous prévoyions également de continuer les travaux du LAAS et de l'ISIR sur l'intégration de la planification et de l'apprentissage [Renaudo 2015], en réfléchissant à la manière de mettre à jour le système de planification avec les résultats de l'apprentissage. La gestion des affordances sociales apportera également de nouveaux

défis à l'exécution. Notre système devra être capable d'exécuter le plan et de prendre en compte les imprévus et la surveillance nécessaire à la réalisation du plan. En outre, comme le propose la définition des affordances sociales, le système d'exécution doit pouvoir être à la fois socialement sensible et socialement réactif. Pour ce faire, il doit être capable de prendre en compte les attentes des robots concernant le comportement humain d'un côté et les attentes humaines concernant le comportement des robots de l'autre côté. Pour mesurer l'efficacité et la pertinence des systèmes de décision et de supervision qui en résultent, nous pourrons profiter de la mesure de qualité d'interaction que nous avons déjà développée [Mayima 2021b].

Le travail du LAAS se situera également au niveau de WP4, Architecture, scénario et intégration. Deux aspects importants mis en évidence dans la définition des affordances sociales [Carvalho 2020] (que nous avons également considérées comme des exigences pour les agents sociaux [Fernández Castro 2020]) sont d'une part le besoin de réactivité mutuelle *mutual responsiveness* et d'autre part la sensibilité aux autres agents *sensitivity to other agents*. La capacité de mettre en musique ces deux aspects sera une caractéristique clé de l'architecture robotique qui sera développée au cours du projet. Nous ne partons pas d'une page blanche car chaque laboratoire a déjà des contributions qui pourront être utilisées dans le projet. Par exemple, le LAAS et l'ISIR ont proposé une architecture prenant en compte l'intégration de l'apprentissage dans le cadre du projet Robo-ErgoSum (ANR-12-CORD-0030 `https://anr.fr/Projet-ANR-12-CORD-0030`) [Chatila 2018], et le LAAS a également déjà implémenté une architecture dédiée à l'interaction humain-robot [Lemaignan 2017].

Le projet ELSA nous amène à considérer de nouveaux défis au niveau architecture :

gestion des affordances sociales : WP1 donnera des informations sur la façon d'interpréter les affordances sociales ainsi que sur la façon de les produire. Les WP2 et WP3 profiteront de ces travaux pour proposer des moyens de les mettre en œuvre et de les gérer.

intégration de l'apprentissage : nous considérerons l'évolution de la représentation de l'action et des capacités de l'agent à travers l'apprentissage. Les affordances sociales dépendent fortement du contexte et de l'agent avec lequel le robot va interagir. Les modèles d'action, d'exécution et de planification doivent pouvoir en tenir compte ainsi que l'évolution de ces modèles dans le temps et les contextes. Cela se fera par apprentissage dans les WP2 et WP3.

gestion de l'interaction : Bien sûr, nous considérerons l'interaction dans son ensemble et notre architecture sera définie de manière à permettre au robot de réaliser une tâche en coopération avec un humain.

2.4 DISCUTER

Le projet Dialogue Interactif Structuré, Consolidé et Unifié pour la réalisation de Tâches En Robotique (DISCUTER) est un projet ANR ASTRID (Accompagnement Spécifique des Travaux de Recherches et d'Innovation Défense) issu de l'appel ASTRID IA 2021 (`https://anr.fr/fr/detail/call/accompagnement-specifique-des-travaux-de-recherches-et-dinnovation-defense-appel-a-proje`

Les partenaires du projet sont :

— Linagora, une entreprise spécialisée dans les assistants virtuels et le traitement du langage (`https://www.linagora.com/fr/`) avec Julie Hunter (PI), Kate Thompson, et Jean-Pierre Lorré

— l'IRIT et l'équipe MEthodes et ingénierie des Langues, des Ontologues et du DIscours (MELODI) spécialisée dans le traitement du langage avec Nicholas Asher, Emiliano Lorini et Chloé Braut

— le LAAS-CNRS et l'équipe Robotique et InteractionS spécialisé dans l'interaction humain-robot avec moi-même et le support de le plate-forme robotique du LAAS (financement obtenu : 100 000 euros).

Le projet a démarré le 01/01/2022 pour une durée de 36 mois.

2.4.1 Description

Des robots capables d'apprendre de nouveaux concepts et de recevoir des instructions pour exécuter de nouvelles tâches par le biais d'une simple conversation - par opposition à une démonstration complète ou à un codage "en dur" par des experts - seront utiles dans un large éventail de contextes à l'avenir, qu'il s'agisse d'exécuter des tâches de construction dans des usines ou de s'engager dans des situations de conflit. Ces robots devront être capables de réagir sur-le-champ à de nouvelles informations acquises soit par l'observation du contexte non-linguistique, soit par la conversation. Le développement d'assistants véritablement conversationnels de ce type nécessitera de modéliser comment le sens de ce qui est dit dans une conversation, et ce qui se passe dans le contexte non-linguistique dans lequel elle se déroule, interagit dynamiquement avec la base de connaissances du robot et la représentation de ses propres croyances et objectifs et de ceux de son interlocuteur.

La principale contribution du projet DISCUTER (Dialogue Interactif Structuré, Consolidé et Unifié pour la réalisation de Tâches En Robotique) est d'étudier et de faire progresser les modèles et les processus nécessaires pour intégrer efficacement le dialogue en langage naturel dans les processus de prise de décision impliqués dans l'exécution de tâches conjointes humain-robot et, plus généralement, humain-système. L'objectif à long terme est d'atteindre la capacité d'utiliser le dialogue en langage naturel pour collaborer efficacement avec les humains et apprendre en interagissant avec eux. L'objectif immédiat est de produire un module de dialogue capable de gérer une conversation en utilisant des éléments de contexte non-linguistiques et des représentations d'états de croyance (qui peuvent évoluer dynamiquement).

Les travaux actuels en robotique qui exploitent les modèles cognitifs pour guider les actions d'un robot ne laissent que peu ou pas de place au rôle que peut jouer le langage dans la mise à jour de la représentation qu'a un agent de son propre état mental ou de celui de son interlocuteur. Les assistants vocaux et les chatbots actuellement sur le marché peuvent engager des conversations plus ou moins complexes, mais ils sont rigides et fragiles. Ils utilisent soit des bases de connaissances limitées et statiques, soit des modèles de langage

déterminés par la probabilité des prochains mouvements linguistiques sans aucune notion des contraintes que ces mouvements exercent sur la conversation et les mouvements futurs, ou des états cognitifs et des objectifs discursifs des participants à la conversation. Ils sont encore moins capables de saisir la relation entre ces états et la conversation en cours.

La clé du développement d'assistants plus flexibles et réactifs, qu'ils soient ou non incarnés par des robots, sera de leur donner les outils nécessaires pour reconnaître la différence entre un acte de dialogue ou un événement non linguistique inattendu ou inconnu d'une part, et un événement incohérent avec sa base de connaissances et sa représentation de ses propres états mentaux et de ceux de son interlocuteur d'autre part. Pour cela, réunissant des spécialistes du langage (LINAGORA et IRIT) et de l'interaction humain-robot (LAAS), le projet DISCUTER exploitera les travaux sur la structure et l'interprétation du discours, qui nous donneront les outils pour modéliser la cohérence entre le discours et les événements non-linguistiques et le contenu des bases de connaissances et des états cognitifs modélisés. Le projet DISCUTER utilisera ces outils pour développer un module de conversation interactive et située qui s'appliquera à tout système multimodal, qu'il s'agisse d'un assistant robotique entièrement incarné ou d'un assistant vocal. Les domaines d'application potentiels d'un tel module sont nombreux, et existent partout où la communication en langage naturel pourrait améliorer l'automatisation intelligente.

2.4.2 Rôle

Si il existe aujourd'hui des bots capables de mener des conversations pour vous faire parler de votre film préféré ou du temps qu'il fait aujourd'hui, il n'existe pas encore de système permettant de coupler la connaissance telle qu'elle est gérée au sein d'un système robotique avec un système de gestion du discours sophistiqué avec des capacités comme celles qu'on vient de décrire. Ce que nous proposons est que ces éléments soient couplés dans un système de gestion de l'interaction permettant de gérer non seulement l'avancement dans un dialogue mais aussi l'avancement dans l'acquisition des connaissances, dans le plan suivi par les participants et dans les actions associées.

Pour pouvoir nous focaliser sur cet objectif, nous devons nous abstraire des difficultés considérables concernant la perception et la réalisation des actions pour un robot en environnement réel, qui sont des thèmes importants de recherche en robotique. Nous faisons le choix dès le départ du projet de nous placer dans un environnement "calibré" et simplifié qui sera réalisé en réalité virtuelle.

Nous proposons de considérer deux types d'interaction.

— La première mettra en jeu un robot spectateur d'une interaction entre deux humains. Le système développé par le projet DISCUTER devra permettre au robot de pouvoir répondre à des questions sur la situation et proposer éventuellement des solutions pour la mener à bien. Pour être plus concret, nous proposons de nous restreindre sur certains types de relations discursives, e.g. des explications, des questions-réponses et des acquiescements.

— La seconde mettra en jeu un robot interagissant avec un humain. Le système développé par le projet DISCUTER devra permettre au robot de pouvoir réaliser une interaction complète intégrant ses propres actions à l'interaction. Encore une fois nos interactions langagières seront restreintes à des explications, des questions-réponses et des acquiescements.

Les objectifs du projet DISCUTER sont :

— Construire un corpus d'interactions multimodales humain-robot qui serviront de base de notre étude
— Etudier les interactions entre les modèles de dialogue et les modèles utilisés en interactions humain-robot, élaborer l'interface entre structure discursive et les diverses connaissances requises pour accomplir des tâches collectives
— Apprendre des structures discursives sur notre corpus via des méthodes faiblement supervisées et les intégrer dans notre modèle d'interaction
— Évaluer notre modèle

Le projet DISCUTER se découpe en 4 tâches :

Tâche 1 : Création d'un corpus de conversation situé annoté

Tâche 2 : Modélisation des éléments de l'interaction

Tâche 3 : Apprentissage et production à partir du modèle de structuration et d'interprétation de l'interaction

Tâche 4 : Evaluation

Au sein du projet DISCUTER, le LAAS dispose du budget pour 13 mois de chercheur post-doctorat et de deux stagiaires pour 6 mois.

Troisième partie

CV

Aurélie Clodic

Je suis ingénieure de recherche au LAAS-CNRS (Toulouse, France) depuis 2012 (permanente depuis 2015). J'ai obtenu mon doctorat en robotique en 2007 pour lequel j'ai élaboré, conçu et implémenté des composants pour l'interaction humain-robot dans plusieurs contextes (ex : robot guide dans un musée, assistant robotique dans le cadre du projet COGNIRON). En 2018, j'ai obtenu une licence de psychologie. Mes intérêts de recherche incluent la supervision de tâches collaboratives humain-robot ainsi que la conception et l'implémentation d'architectures robotiques dédiée à l'interaction humain-robot. Je m'intéresse aux apports de la philosophie et de la psychologie à la compréhension de l'interaction en générale et à ses composantes sociales en particulier.

Après ma thèse, j'ai passé 4 ans dans une société de services en informatique : ATOS. J'ai d'abord travaillé comme ingénieur logiciel embarqué pour le projet ATC (Air Traffic Control pour le compte d'Airbus. Dans ce contexte, je suis intervenue à tous les niveaux du cycle en V dans un contexte avionique (norme DO178-B)). J'ai ensuite intégré l'équipe ingénierie des modèles. J'y ai travaillé comme consultant d'une part, sur des projets en lien avec l'ingénierie des exigences et les lignes de produits logiciels dans le contexte automobile (Continental). J'étais également chef de projets toujours pour différents grands comptes (CNES, Astrium, Airbus).

Fin 2012, j'ai réintégré le LAAS et j'ai commencé à travailler pour le projet FP7 SAPHARI Safe and Autonomous Physical Human-Aware Robot Interaction (`http://www.saphari.eu/`) ainsi que pour le projet ANT MarDi Man-Robot Dialogue (`https://anr.fr/Project-ANR-12-CORD-0021`) principalement sur la partie supervision et la partie évaluation de la situation. Ensuite, j'ai été impliqué dans le projet ANR RoboErgoSum (`http://roboergosum.isir.upmc.fr/`, `https://anr.fr/Projet-ANR-12-CORD-0030`). J'ai travaillé sur une preuve de concept pour mélanger l'approche traditionnelle de planification des tâches avec l'apprentissage. L'idée était, étant donné qu'il est impossible pour les algorithmes d'apprentissage par renforcement d'apprendre à partir de zéro dans un contexte de collaboration humain-robot, d'utiliser un planificateur "traditionnel" (dans notre cas HATP, Human-Aware Task Planner) comme amorce pour les algorithmes d'apprentissage par renforcement. J'ai également travaillé sur une manière de gérer l'interaction de manière fluide dans un contexte d'action conjointe.

J'ai ensuite travaillé sur le projet MuMMER (H2020-688147 `http://mummer-project.eu/`) sur la conception et le développement d'un robot de divertissement dans un centre commercial en Finlande. Nous avons travaillé principalement sur une tâche de description d'itinéraires permettant aux clients d'interagir avec le robot pour se guider dans le centre commercial. J'ai participé à l'encadrement de deux thèses au sein de ce projet. Celle de Guillaume Sarthou (Knowledge representation and exploitation for interactive and cognitive robots [Sarthou 2021a]) et celle d'Amandine Mayima (Endowing the Robot with the Abilities to Control and Evaluate its Contribution to a Human-Robot Joint Action [Mayima 2021a]), toutes les deux en co-encadrement avec Rachid Alami.

Ces dernières années, j'ai travaillé sur la construction de relations avec d'autres disciplines telles que la philosophie et la psychologie.

Dans ce cadre, j'ai initié et participé au projet ANR JointAction4HRI (`https://jointaction4hri.laas.fr/`, `https://anr.fr/en/funded-projects-and-impact/funded-projects/project/funded/project/`

b2d9d3668f92a3b9fbbf7866072501ef-c67d03a23f/) où avec des psychologues du développement et des philosophes nous avons travailler à la conception d'une architecture pour l'action conjointe humain-robot. Cela a aboutit à une thèse en psychologie, soutenue en décembre 2021, en co-encadrement entre l'équipe Robotique et InteractionS et le laboratoire CLLE Cognition, Langues, Langage et Ergonomie de l'université Jean Jaurès à Toulouse, Kathleen Belhassein Propositions de stratégies communicatives pour une action jointe Humain-Robot efficace, fluide et durable [Belhassein].

J'ai aussi organisé une série de workshops « Vers un cadre d'action conjointe » (toward a framework for Joint Action fja.sciencesconf.org). En effet, il est intéressant d'observer, du point de vue d'un roboticien, que l'action conjointe humain-humain est un sujet de recherche intense en psychologie cognitive et en philosophie. Ce constat nous a conduit à mettre en place une initiative pluridisciplinaire pour créer une opportunité unique d'échanges scientifiques à travers une série d'ateliers intitulés "vers un Cadre d'Action Conjointe" (fja.sciencesconf.org). Les psychologues et les philosophes peuvent présenter les développements récents de la recherche-action conjointe, tandis que les roboticiens peuvent discuter des défis auxquels ils sont confrontés en ce qui concerne l'interaction homme-robot et plus précisément l'activité conjointe homme-robot.

Je suis aujourd'hui coordinateur et responsable scientifique du projet ANR franco-germano-japonais ai4hri artificial intelligence for human-robot interaction. Je suis responsable scientifique pour le LAAS pour les projets ANR ELSA Effective Learning of Social Affordances et ANR ASTRID IA DISCUTER Dialogue Interactif Structuré, Consolidé et Unifié pour la réalisation de Tâches en Robotique.

Je suis également impliquée dans le projet ANR JCJC EPIIC ElectroPhysiological Involuntary Inputs for Collaborative robotics enhancement.

Formation

- Doctorat, LAAS-CNRS, Toulouse (France) - 2002-2007
  - Supervision pour un robot interactif : Action et Interaction pour un robot autonome en environnement humain
  - directeurs de thèse : Rachid Alami, Raja Chatila, LAAS CNRS, Toulouse, France
- DEA Imagerie Vision Robotique, INPG, Grenoble (France) - 2001-2002
- Licence de Psychologie, Université Paris 8, 2016-1018

Expériences professionnelles

- Ingénieure de recherche en robotique (spec. Interaction Humain-Robot) — LAAS-CNRS, Toulouse (FRANCE) — 2012-PRESENT
- Responsable de projets — AtoS, Toulouse (France) - 2011-2012
  - Gestion d'équipe. Supervision de l'avancement des projets. Elaboration des réponses. Gestion de la relation client. Gestion du reporting et du planning.
  - Multi-projects team management. Projects supervision and follow-up . Elaboration of proposals. Handling customer/supplier relationship. Carried out reporting to the executive committee. Responsible for contract, planning, invoicing, delivery, risks and costs.

  - Ingénierie des exigences dans le domaine de l'automobile (Continental) et de l'espace (CNES).
- Ingénieure logiciel embarqué - ATOS, Toulouse (FRance) — 2008-2011
  - projet Air Traffic Control (ATC), programmes A350, A380, A400M (Airbus)
  - Travail selon le norme DO-178B, Software Considerations in Airborne Systems and Equipment Certification (Level C)

Chapitre 1

Activités

1.1 Activités d'encadrement

1.1.1 Thèse

Ecole doctorale de Rattachement : Ecole Doctorale 475 Mathématiques, Informatique, Télécommunications de Toulouse

2018-2021 - **Guillaume SARTHOU** - Knowledge representation and exploitation for interactive and cognitive robots [Sarthou 2021a] (co-encadrement avec Rachid Alami)

Résumé : L'arrivée des robots dans notre vie quotidienne fait émerger le besoin pour ces systèmes d'avoir accès à une représentation poussée des connaissances et des capacités de raisonnements associées. Ainsi, les robots doivent pouvoir comprendre les éléments qui composent l'environnement dans lequel ils évoluent. De plus, la présence d'humains dans ces environnements et donc la nécessité d'interagir avec eux amènent des exigences supplémentaires. Ainsi, les connaissances ne sont plus utilisées par le robot dans le seul but d'agir physiquement sur son environnement mais aussi dans un but de communication et de partage d'information avec les humains. La connaissance ne doit plus être uniquement compréhensible par le robot lui-même mais doit aussi pouvoir être exprimée et partagée. Dans la première partie de cette thèse, nous présentons Ontologenius, le logiciel développé durant ces trois ans de thèse. C'est un logiciel permettant de maintenir des bases de connaissances sous forme d'ontologies, de raisonner dessus et de les gérer dynamiquement. Nous commençons par expliquer en quoi ce logiciel est adapté aux applications d'interaction humain-robot (HRI), notamment avec la possibilité de représenter la base de connaissances du robot mais aussi une estimation des bases de connaissances des partenaires humains ce qui permet d'implémenter les mécanismes de théorie de l'esprit, telle que la prise de perspective. Nous poursuivons avec une présentation de ses interfaces. Cette partie se termine par une analyse des performances du système ainsi développé. Dans une seconde partie, cette thèse présente notre contribution à deux problèmes d'exploration des connaissances : l'un ayant trait au référencement spatial et l'autre à l'utilisation de connaissances sémantiques. Nous commençons par une tâche de description d'itinéraires pour laquelle nous proposons une ontologie permettant de décrire la topologie d'environnements intérieurs et deux algorithmes de recherche d'itinéraires. Nous poursuivons avec une tâche de génération d'expression de référence. Cette tâche vise à sélectionner l'ensemble optimal d'informations à communiquer afin de permettre à un auditeur d'identifier l'entité référencée dans un contexte donné. Ce dernier algorithme est ensuite affiné pour y ajouter les informations sur les activités passées provenant d'une

action conjointe entre un robot et un humain, afin de générer des expressions encore plus pertinentes. Il est également intégré à un planificateur de tâches symbolique pour estimer la faisabilité et le coût des futures communications. Cette thèse se termine par la présentation de deux architectures cognitives, la première utilisant notre contribution concernant la description d'itinéraire et la seconde utilisant nos contributions autour de la Génération d'Expression de Référence. Les deux utilisent Ontologenius pour gérer la base de connaissances sémantique. À travers ces deux architectures, nous présentons comment nos travaux ont amené la base de connaissances a prendre un rôle central, fournissant des connaissances à tous les composants du système.

2017-2021 - **Amandine MAYIMA** - Endowing the Robot with the Abilities to Control and Evaluate its Contribution to a Human-Robot Joint Action [Mayima 2021a] (co-encadrement avec Rachid Alami)

Résumé : Un nombre important d'études se concentrent sur des fonctionnalités rendant un robot plus utile et adaptatif tels que la planification, la perception, la gestion des connaissances, la navigation, la reconnaissance d'actions, le dialogue... Que l'on améliore ces fonctionnalités est essentiel. Cependant ce ne sont pas elles qui font collaborer un robot et un humain, mais la supervision. En effet, ce composant, tel un marionnettiste, contrôle les autres composants de l'architecture qui implémentent les fonctionnalités sus-mentionnées. En s'appuyant sur eux, il prend les décisions sur comment et quand le robot doit agir, dans une tâche de collaboration avec un humain, il décide ce que le robot doit dire, en réagissant à l'environnement, au comportement et aux communications de l'humain. Dans cette thèse, nous proposons un composant de supervision dédié à l'interaction humain-robot. Il dote le robot d'un certain nombre de capacités dans le but d'en faire le meilleur partenaire pour l'humain, telles que la modélisation de ses états mentaux ou l'adaptation à ses décisions. Par ailleurs, nous introduisons un nouveau concept : l'évaluation par le robot en temps réel de la qualité de son interaction avec un humain. Il s'agit d'une première étape vers la gestion des contingences, car plus tard, en dehors du cadre de cette thèse, cette méthode pourrait s'intégrer à la supervision, l'aidant à améliorer sa décision et ses réactions.

1.1.2 Master

- 2013 Jules Waldhart - M2 INSA (co-encadrement avec Rachid Alami) -
- 2014 Sandra Devin - M2 ENSEEIHT (co-encadrement avec Rachid Alami) - Supervision pour l'interaction humain-robot
- 2015 Gilles Lacroix - M1 UPSITECH (co-encadrement avec Grégoire Milliez) - Take part in the development of an inference framework for a PR2 robot, Development of a ROS plugin for 3D visualization, Implementation of a database manager
- 2016 Dorian Kurzaj - M1 INSA (co-encadrement avec Christelle Ecrepont) - Implementation of a communication bridge between the IoT platform OM2M and the robotic framework ROS
- 2017 Camille Vrignaud - M2 Psychologie Ergonomie (co-encadrement avec Sandra Devin, Rachid Alami, Ophélie Carreras et Loic Caroux) - Evaluation d'une interaction human-robot via la réalisation d'une étude utilisateur

— 2017 Julia Cohen - M1 INSA (co-encadrement avec Christelle Ecrepont) - Liaison entre une architecture robotique et des objets connectés
— 2017 Guillaume Sarthou - M1 INSA (co-encadrement avec Christelle Ecrepont) - Liaison entre une architecture robotique et des objets connectés
— 2018 Guillaume Sarthou - M2 INSA (co-encadrement avec Rachid Alami) - Base de connaissances et ontologies pour l'interaction humain-robot
— 2018 Nathan Compan - M2 Psychologie Ergonomie (co-encadrement avec Guilhem Buisan, Rachid Alami, Ophélie Carreras et Loic Caroux) - Etude expérimentale de l'interaction humain-robot dans le cadre de la navigation dans un espace physique partagé.
— 2018 Jason Gaudinet - M2 INSA (co-encadrement avec Christelle Ecrepont)

1.2 Activités d'enseignement

— introduction à l'interaction humain-robot - 2 à 3 h (Master IHM (depuis 2017), ISAE Supaéro (depuis 2016))

1.3 Activités d'évaluation

Je suis reviewer pour la plupart des conférences de mon domaine (HRI, RO-MAN, IROS, ICRA). Je participe également à diverses activités détaillées ci-dessous :

1.3.1 Membre de comité de programme

— membre du comité de programme (PC Member) HRI Human Robot Interaction Conference (HRI 2020)
— membre du comité de programme (PC Member) International Conference on Principles of Knowledge Representation and Reasoning (KR2021)
— membre du comité de programme (PC Member) Robophilosophy 2022 (RP2022)

1.3.2 Editeur associé

— éditeur associé (Associate Editor) IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2021)
— éditeur associé (Associate Editor) IEEE Robotics and Automation Letters IEEE RA-L since 09/2021
— éditeur associé (Associate Editor) IEEE International Conference on Robot & Human Interactive Communication (ROMAN 2022)

1.3.3 Jury de thèse

J'ai participé aux jurys des thèses suivantes :
— Agentivité conjointe lors des interactions homme-machine : comment concevoir des agents plus coopératifs ? par Aïsha Sahaï - Juillet 2019 `http://theses.fr/2019PSLEE025`
— Decisional issues during human-robot joint action par Sandra Devin - Novembre 2017 `http://theses.fr/fr/2017INPT0092`

— Architecture cognitive generique pour la coordination de strategies d'apprentissage en robotique par Remi Dromnelle - Juillet 2021 `http://www.theses.fr/s266650`
— Modèles inspirés de la psychologie du développement pour le raisonnement physique et social dans le cadre de l'action jointe humain-robot par Yoan Sallami - Janvier 2021 `http://thesesups.ups-tlse.fr/5062/`
— Desormeaux, Kevin. Temporal models of motions and forces for Human-Robot Interactive manipulation par Kevin Desormeaux - Octobre 2019 `http://thesesups.ups-tlse.fr/4546/`

1.3.4 Jury de concours

J'ai participé aux jurys de concours ITA suivants :
— concours ITA IR CNRS BAPC Ingénieur en conception et développement en expérimentation (Concours IT CNRS 29 - 2016)
— concours ITA IE INRAE Ingénieur(e) d'étude expert en techniques expérimentales (Concours IT INRAE IE20-NUMM-1 - 202)

1.4 Activités de diffusion de l'information scientifique

1.4.1 Invited talk/speaker

— conférencier invité (invited keynote) Artificial Intelligence and Human-Robot Interaction AAAI Fall Symposium 2015
— conférencier invité (invited speaker) Trust in Robots Lecture Series, Vienna, 21 December 2018
— conférencier invité (invited speaker) HUMAINT Winter school on AI and its ethical, legal, social and economic (ELSE) impact, 4-8 February 2019
— conférencier invité (invited speaker) Workshop on Robotics, AI, and Humanity : Science, Ethics, and Policy" organized jointly by the Pontifical Academy of Sciences (PAS) and the Pontifical Academy of Social Sciences (PASS), Casina Pio IV, Vatican City, 16-17 May 2019
— conférencier invité (invited speaker) Workshop MIT AeroAstro Virtual Workshop - "Humans Interacting with Autonomy", 13 January 2021

1.4.2 Dissemination grand public

— participation à un numéro des Curionautes sur les robots (`https://www.curionautes.com/sommaires/octobre-2019-a-quoi-servent-les-robots/`)
— participation à un article sur le département robotique du LAAS dans un numéro spécial de Sciences et Vie Junior sur les robots (Bienvenue dans un monde de robots SVJ Janvier 2021)
— relecture d'un manga sur les robots (édition Bayard Presse `https://www.bayard-editions.com/jeunesse/documentaire/des-9-ans-7014/les-robots`)

1.5 Organisation de workshops

Je suis l'organisatrice d'une série de workshops sur l'action jointe humain-robot.

1.5.1 Description

Depuis plus d'une décennie, le domaine de l'interaction humain-robot a généré de nombreuses contributions intéressantes pour la communauté robotique dans son ensemble. Le champ est vaste, allant de la perception (e.g., tactile ou visuelle) à l'action (e.g., manipulation, navigation) et à la décision (e.g., interaction, planification consciente de l'humain). Cependant, lorsqu'il s'agit de développer de futurs assistants robots ou coéquipiers robotiques dans des équipes mixtes humain-robot, il est nécessaire d'approfondir la compréhension de l'action conjointe humain-robot qui pourrait fournir un cadre pour les différentes contributions et études.

Il est intéressant d'observer, d'un point de vue robotique, que la recherche sur l'action conjointe humaine est un sujet de recherche intense en psychologie cognitive et en philosophie. Dans cet atelier, nous aimerions analyser les hypothèses fondamentales ainsi que les résultats empiriques détaillés de ces disciplines et les relier aux diverses activités de recherche en cours en robotique, de la conception d'architectures de contrôle à l'interaction humain-robot. Plus précisément, notre objectif est de rassembler divers aspects des travaux existants et d'examiner comment ils peuvent nous aider à définir le type de cadre intégratif nécessaire à la conception d'un robot autonome pouvant s'engager dans une interaction à long terme avec un partenaire humain. Ce cadre devrait pouvoir servir deux objectifs complémentaires. D'une part, cela doit nous aider à définir avec précision ce qu'un robot doit comprendre de l'humain avec lequel il interagit pour que l'interaction soit réussie et donc de quelles capacités le robot doit être équipé pour s'assurer qu'il peut construire cette compréhension. D'autre part, le robot doit également être compris par son partenaire humain et ce cadre devrait nous aider à clarifier comment cette compréhension fonctionne et ce qui est nécessaire pour permettre au robot de se comporter de manière appropriée et d'une manière qui manifeste ce qu'il fait au partenaire humain.

L'objectif de ces workshops est de réunir des chercheurs en robotique, des psychologues et des philosophes. Cela créera une occasion unique d'échanges scientifiques entre ces disciplines. En particulier, des psychologues et des philosophes auront l'occasion de présenter les développements récents de la recherche-action conjointe, tandis que les roboticiens pourront discuter des défis auxquels ils sont confrontés en matière d'interaction humain-robot et plus précisément d'activité conjointe humain-robot. Nous imaginons et espérons que l'atelier sera une première réunion d'une série d'ateliers dans les années à venir qui créeront un contexte dans lequel les chercheurs intéressés pourront se réunir et discuter de l'élaboration d'un cadre intégratif pour la recherche sur l'action conjointe humain-robot.

Ces workshops ont réuni de 30 à 50 personnes lors de chacune des éditions.

1.5.2 Site web

fja.sciencesconf.org

1.5.3 Liste des workshops

— Towards a Framework for Joint Action (FJA@ROMAN2014)
— Towards a Framework for Joint Action (FJA@HRI2015)
— Towards a Framework for Joint Action (FJA@ICSR2015)

- From Human-Human Joint Action to Human-Robot Joint Action (and vice-versa !) (HRJA2016)
- Towards a Framework for Joint Action - What about Common Ground ? (FJA@ROMAN2016)
- Towards a Framework for Joint Action - What about Commitments ? (FJA@ROBOPHILOSOPHY2016)
- Towards a Framework for Joint Action - RobotJAM !!! (FJA@JAM2017)
- Towards a Framework for Joint Action - What about Theory of Mind ? (FJA@RSS2018)
- Human-Robot Joint Action - Refining the understanding of joint action through an interdisciplinary perspective (HRJA2018)
- Towards a Framework for Joint Action - The Communication Challenges in Joint Action for Human-Robot Interaction (FJA@ICSR2019)
- Towards a Framework for Joint Action - Social Models for Social Robotics (FJA@ROBOPHILOSOPHY2020)

Chapitre 2

Projets passés

Projets sur lesquels j'ai collaboré et mes publications associées.

2.1 JointAction4HRI (French ANR project) 2017-2021

2.1.1 Description

Dans ce projet, nous avons étudié les composants et mécanismes clés qui sous-tendent l'action conjointe, leur contenu informationnel et la façon dont ils sont liés entre eux, temporairement, hiérarchiquement et de manière significative. Notre ambition est de contribuer, au niveau le plus fondamental, à l'identification et à la caractérisation des ingrédients nécessaires à la réalisation réussie d'une tâche collaborative dans un contexte humain-humain et humain-robot.

2.1.2 Site web

`https://jointaction4hri.laas.fr/`

2.1.3 Publications

- Kathleen Belhassein, Víctor Fernández-Castro, Amandine Mayima, Aurélie Clodic, Elisabeth Pacherie, et al.. Addressing joint action challenges in HRI : Insights from psychology and philosophy. Acta Psychologica, Elsevier, 2022, 222, pp.103476. 10.1016/j.actpsy.2021.103476 . hal-03505047
- Amandine Mayima, Aurélie Clodic, Rachid Alami. Towards Robots able to Measure in Real-time the Quality of Interaction in HRI Contexts. International Journal of Social Robotics, Springer, In press, 10.1007/s12369-021-00814-5 . hal-03326954
- Guillaume Sarthou, Amandine Mayima, Guilhem Buisan, Kathleen Belhassein, Aurélie Clodic. The Director Task : a Psychology-Inspired Task to Assess Cognitive and Interactive Robot Architectures. 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2021), Aug 2021, Vancouver (Virtual), Canada. 10.1109/RO-MAN50785.2021.9515543 . hal-03327222
- Aurélie Clodic, Rachid Alami. What Is It to Implement a Human-Robot Joint Action ?. Robotics, AI, and Humanity, Springer International Publishing, pp.229-238, 2021, 978-3-030-54172-9. 10.1007/978-3-030-54173-6_19 . hal-03184727
- Amandine Mayima, Aurélie Clodic, Rachid Alami. Toward a Robot Computing an Online Estimation of the Quality of its Interaction with its Human Partner. 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN), Aug 2020, (Virtual Conference), Italy. 10.1109/RO-MAN47096.2020.9223464 . hal-02910899

— Víctor Fernández Castro, Raul Hakli, Aurélie Clodic. What does it take to be a social agent ?. Appeared in : Nørskov, M., Seibt J., Quick O. 2020. Culturally Sustainable Social Robotics—Proceedings of Robophilosophy 2020., Series Frontiers of AI and Its Applications, IOS Press, Amsterdam., 2020, 10.3233/FAIA200954 . hal-02963044

— Kathleen Belhassein, Guilhem Buisan, Aurélie Clodic, Rachid Alami. Towards methodological principles for user studies in Human-Robot Interaction. Test Methods and Metrics for Effective HRI in Collaborative Human-Robot Teams Workshop, ACM/IEEE International Conference on Human-Robot Interaction, Mar 2019, Daegu, South Korea. hal-02282600

— Sandra Devin, Camille Vrignaud, Kathleen Belhassein, Aurélie Clodic, Ophélie Carreras, et al.. Evaluating the Pertinence of Robot Decisions in a Human-Robot Joint Action Context : The PeRDITA Questionnaire. 27th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2018), Aug 2018, Nanjing, China. 10.1109/ROMAN.2018.8525785 . hal-01906101

— Aurélie Clodic, Javier Vázquez-Salceda, Frank Dignum, Samuel Mascarenhas, Virginia Dignum, et al.. On the Pertinence of Social Practices for Social Robotics. IOS Press. Envisioning Robots in Society – Power, Politics, and Public Space, pp.36-74, 2018, 978-1-61499-931-7. 10.3233/978-1-61499-931-7-63 . hal-01943774

— Sandra Devin, Aurélie Clodic, Rachid Alami. About Decisions During Human-Robot Shared Plan Achievement : Who Should Act and How ?. Social Robotics,9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings, 10652, springer, pp.453-463, 2017, Lecture Notes in Computer Science book series, 978-3-319-70022. hal-01943959

— Kathleen Belhassein, Aurélie Clodic, Hélène Cochet, Marketta Niemelä, Päivi Heikkilä, et al.. Human-Human Guidance Study. Rapport LAAS n° 17596. 2017. hal-01719730

2.2 MuMMER- MUltiModal Mall Entertainment Robot (H2020) 2016-2020

2.2.1 Description :

Le projet MuMMER ("MultiModal Mall Entertainment Robot") proposait d'aborder le marché important et en croissance de la robotique de divertissement grand public en faisant progresser les technologies nécessaires pour prendre en charge ce domaine de la robotique. Plus précisément, le projet visait au développement d'un robot capable de s'engager et d'interagir de manière autonome et naturelle dans les environnements dynamiques d'un centre commercial public, offrant une expérience engageante et divertissante au grand public.

2.2.2 Site web :

`http://mummer-project.eu/`

2.2.3 Publications :

- Amandine Mayima, Aurélie Clodic, Rachid Alami. Towards Robots able to Measure in Real-time the Quality of Interaction in HRI Contexts. International Journal of Social Robotics, Springer, In press, 10.1007/s12369-021-00814-5 . hal-03326954
- Amandine Mayima, Aurélie Clodic, Rachid Alami. Evaluation of the Quality of Interaction from the robot point of view in Human-Robot Interactions. 1st Edition of Quality of Interaction in Socially Assistive Robots (QISAR) Workshop, The 11th International Conference on Social Robotics (ICSR 2019), Nov 2019, Madrid, Spain. hal-02403081
- Päivi Heikkilä, Hanna Lammi, Marketta Niemelä, Kathleen Belhassein, Guillaume Sarthou, et al.. Should a robot guide like a human ? A qualitative four-phase study of a shopping mall robot. The Eleventh International Conference on Social Robotics (ICSR 2019), Nov 2019, Madrid, Spain. hal-02285035
- Mary Ellen Foster, Bart Craenen, Amol Deshmukh, Oliver Lemon, Emanuele Bastianelli, et al.. MuMMER : Socially Intelligent Human-Robot Interaction in Public Spaces. Artificial Intelligence for Human-Robot Interaction Symposium (AI-HRI), AAAI Fall Symposium Series 2019, Nov 2019, Arlington, VA, United States. hal-02291690
- Yoan Sallami, Séverin Lemaignan, Aurélie Clodic, Rachid Alami. Simulation-based physics reasoning for consistent scene estimation in an HRI context. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Nov 2019, Macau, China. pp.7834-7841, 10.1109/IROS40897.2019.8968106 . hal-02925945
- Jules Waldhart, Aurélie Clodic, Rachid Alami. Reasoning on Shared Visual Perspective to Improve Route Directions. 2019 28th IEEE International Conference on Robot & Human Interactive Communication, Oct 2019, New Delhi, India. hal-02283904
- Séverin Lemaignan, Yoan Sallami, Christopher Wallbridge, Aurélie Clodic, Tony Belpaeme, et al.. UNDERWORLDS : Cascading Situation Assessment for Robots. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS

2018), Oct 2018, Madrid, Spain. 10.1109/IROS.2018.8594094 . hal-01943917
— Kathleen Belhassein, Aurélie Clodic, Hélène Cochet, Marketta Niemelä, Päivi Heikkilä, et al.. Human-Human Guidance Study. Rapport LAAS n° 17596. 2017. hal-01719730

2.3 RoboErgoSum - Self Aware Robot (French ANR project) - 2013-2017

2.3.1 Description

L'un des plus grands défis auxquels la recherche en robotique est aujourd'hui confrontée est la conception de robots capables d'agir de manière autonome au niveau cognitif et d'interagir avec les gens de manière naturelle. Le projet RoboErgoSum visait à étudier comment développer la conscience de soi d'un robot et sa capacité à effectuer un raisonnement au niveau cognitif.

2.3.2 Site Web

http ://roboergosum.isir.upmc.fr/

2.3.3 Publications

— Séverin Lemaignan, Matthieu Warnier, Emrah Akin Sisbot, Aurélie Clodic, Rachid Alami. Artificial Cognition for Social Human-Robot Interaction : An Implementation. Artificial Intelligence, Elsevier, 2017, 247, pp.45-69. 10.1016/j.artint.2016.07.002 . hal-01857498

— Aurélie Clodic, Elisabeth Pacherie, Rachid Alami, Raja Chatila. Key Elements for Human Robot Joint Action. Sociality and Normativity for Robots Philosophical Inquiries into Human-Robot Interactions, Springer, pp.159-177, 2017, Studies in the Philosophy of Sociality

— Mehdi Khamassi, Benoît Girard, Aurélie Clodic, Sandra Devin, Erwan Renaudo, et al.. Integration of Action, Joint Action and Learning in Robot Cognitive Architectures. Intellectica - La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo), Association pour la Recherche sur la Cognition, 2016, 2016/1 (65), pp.169-203. 10.3406/intel.2016.1794 . hal-01375666

— Erwan Renaudo, Sandra Devin, Benoît Girard, Raja Chatila, Rachid Alami, et al.. Learning to interact with humans using goal-directed and habitual behaviors. Ro-Man 2015, Workshop on Learning for Human-Robot Collaboration, Aug 2015, Kobe, Japan. hal-01944380

2.4 Saphari - Safe and Autonomous Physical Human-Aware Robot Interaction (FP7) - 2011-2015

2.4.1 Description

Les progrès récents de l'interaction physique humain-robot ont montré que le partage actif et sûr de l'espace de travail devient en principe possible. SAPHARI a effectué un changement de paradigme fondamental dans le développement de robots en ce sens que l'humain était au centre de toute la conception. Le projet a étudié la manière de suivre, comprendre et prédire les mouvements humains dans un environnement dynamique faiblement structuré en temps réel. Le projet s'est concentré sur des cas d'utilisation industriels nécessitant des contacts et des échanges de force dans le cadre d'un travail collaboratif humain-robot, ainsi que sur des scénarios de services professionnels dans les hôpitaux, dans lesquels un personnel médical et un robot assistant interagissent étroitement au cours du travail quotidien.

2.4.2 Site Web

http ://www.saphari.eu/

2.4.3 Publications

— Mamoun Gharbi, Pierre-Vincent Paubel, Aurélie Clodic, Ophélie Carreras, Rachid Alami, et al.. Toward a better understanding of the communication cues involved in a human-robot object transfer. 24th International Symposium on Robot and Human Interactive Communication, IEEE RO-MAN, Aug 2015, Kobe, Japan. pp.319-324, 10.1109/ROMAN.2015.7333626 . hal-01628249
— Michelangelo Fiore, Aurélie Clodic, Rachid Alami. On Planning and Task achievement Modalities for Human-Robot Collaboration. International Symposium on Experimental Robotics (ISER 2014), Jun 2014, Marrakech, Morocco. 15p. hal-01149109

2.5 MarDi - Man-Robot Dialogue (French National Research Agency project) - 2012-2016

2.5.1 Description

L'objectif du projet MarDi était d'étudier l'interaction humain-machine de manière située. Les robots, dotés de capacités de perceptions, mais aussi de moyens d'agir sur l'environnement, ont permis l'intégration d'un contexte physique dans l'interaction pour la machine comme pour l'humain.

2.5.2 Publications

— Grégoire Milliez, Matthieu Warnier, Aurélie Clodic, Rachid Alami. A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. IEEE RO-MAN 2014, Aug 2014, Edinburgh, United Kingdom. pp.FrAT2.4, 10.1109/ROMAN.2014.6926399 . hal-01064546

2.6 COGNIRON (European FP6 integrated project) 2004-2008

2.6.1 Description

Les objectifs généraux de ce projet étaient d'étudier les capacités de perception, de représentation, de raisonnement et d'apprentissage des robots incarnés dans des environnements centrés sur l'humain. Le projet a développé des méthodes et des technologies pour la construction de robots cognitifs, capables d'évoluer et de développer leurs capacités en interaction étroite avec les humains de manière ouverte.

2.6.2 Site Web

`http://www.cogniron.org/final/Home.php` ; `http://www.laas.fr/robots/jido`

2.6.3 Publications

— Aurélie Clodic, Hung Cao, Samir Alili, Vincent Montreuil, Rachid Alami, et al.. SHARY : a supervision system adapted to Human-Robot Interaction. In : Khatib O., Kumar V., Pappas G.J. (eds) Experimental Robotics. Springer Tracts in Advanced Robotics, vol 54. Springer, Berlin, Heidelberg, pp.229-238, 2009. hal-01979196

— Aurélie Clodic, Vincent Montreuil, Rachid Alami, Raja Chatila. A Decisional Framework for Autonomous Robots Interacting with Humans. IEEE International Workshop on Robot and Human Interactive Communication (ROMAN), 2005, Nashville, United States. hal-01977560

— Emrah Akin Sisbot, Aurélie Clodic, Luis Felipe Marin-Urias, Mathias Fontmarty, Ludovic Brèthes, et al.. Implementing a Human-Aware Robot System. ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, Sep 2006, Hatfield, United Kingdom. hal-01977571

— Aurélie Clodic, Rachid Alami, Vincent Montreuil, Shuyin Li, Britta Wrede, et al.. A study of interaction between dialog and decision for human-robot collaborative task achievement. RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication, 2007, Jeju, South Korea. hal-01977573

— Aurélie Clodic, Maxime Ransan, Rachid Alami, Vincent Montreuil. A management of mutual belief for human-robot interaction. 2007 IEEE International Conference on Systems, Man and Cybernetics, Oct 2007, Montreal, Canada. hal-01977577

— Vincent Montreuil, Aurélie Clodic, Rachid Alami. Planning Human Centered Robot Activities. 2007 IEEE International Conference on Systems, Man and Cybernetics, Oct 2007, Montreal, Canada. hal-01977579

2.7 HR+ (French robotics program project) 2002-2005

2.7.1 Description

Explorer l'interaction humain-robot dans le cadre de la formation d'un guide de musée autonome

2.7.2 Site Web

`http://www.laas.fr/robots/rackham`

2.7.3 Publications

— Aurélie Clodic, Sara Fleury, Rachid Alami, Matthieu Herrb, Raja Chatila. Supervision and Interaction Analysis of an Autonomous Tour-guide Robot Deployment. 12th International Conference on Advanced Robotics,, Jul 2005, Seattle, United States. hal-01977554

— Aurélie Clodic, Sara Fleury, Rachid Alami, Raja Chatila, Gérard Bailly, et al.. Rackham : An Interactive Robot-Guide. IEEE International Workshop on Robots and Human Interactive Communications (ROMAN), Sep 2006, Hatfield, United Kingdom. pp.502-509. hal-00480381

Chapitre 3

Publications

3.1 Thèse

— Aurélie Clodic. Supervision pour un robot interactif : action et interaction pour un robot autonome en environnement humain. Automatique / Robotique. Université Paul Sabatier - Toulouse III, 2007. Français. tel-00196608 https://tel.archives-ouvertes.fr/tel-00196608.

3.2 Revues

1. Mehdi Khamassi, Benoît Girard, Aurélie Clodic, Sandra Devin, Erwan Renaudo, et al.. Integration of Action, Joint Action and Learning in Robot Cognitive Architectures. Intellectica - La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo), Association pour la Recherche sur la Cognition, 2016, 2016/1 (65), pp.169-203. 10.3406/intel.2016.1794 . https://hal.archives-ouvertes.fr/hal-01375666.
2. Séverin Lemaignan, Matthieu Warnier, Emrah Akin Sisbot, Aurélie Clodic, Rachid Alami. Artificial Cognition for Social Human-Robot Interaction : An Implementation. Artificial Intelligence, Elsevier, 2017, 247, pp.45-69. 10.1016/j.artint.2016.07.002 . https://hal.archives-ouvertes.fr/hal-01857498.
3. Raja Chatila, Erwan Renaudo, Mihai Andries, Omar Ricardo Chavez-Garcia, Pierre Luce-Vayrac, Raphaël Gottstein, Rachid Alami, Aurélie Clodic, Sandra Devin, Benoît Girard, Mehdi Khamassi, "Toward Self-Aware Robots". In : Frontiers in Robotics and AI (Aug. 2018), p. 88.doi :10.3389/frobt.2018.00088. https://hal.archives-ouvertes.fr/hal-01856931.
4. Aurélie Clodic, Rachid Alami. What Is It to Implement a Human-Robot Joint Action ?. Robotics, AI, and Humanity, Springer International Publishing, pp.229-238, 2021, 978-3-030-54172-9. 10.1007/978-3-030-54173-6_19 https://hal.laas.fr/hal-03184727.
5. Amandine Mayima, Aurélie Clodic, Rachid Alami. Towards Robots able to Measure in Real-time the Quality of Interaction in HRI Contexts. International Journal of Social Robotics, Springer, In press, 10.1007/s12369-021-00814-5 https://hal.laas.fr/hal-03326954.
6. Kathleen Belhassein, Víctor Fernández-Castro, Amandine Mayima, Aurélie Clodic, Elisabeth Pacherie, et al.. Addressing joint action challenges in HRI : Insights from psychology and philosophy. Acta Psychologica, Elsevier, 2022, 222, pp.103476. 10.1016/j.actpsy.2021.103476 https://hal.archives-ouvertes.fr/hal-03505047.

3.3 Conférences à comité de lecture

rq : nous avons laissé dans cette section Conférences, nos publications acceptées aux conférences qui sont ensuite éditées (c'est le cas d'ISER et IJSR dont les actes sont ensuite publiés chez Springer et de Robophilosophy dont les actes sont ensuite publiés chez Frontiers)

1. Aurélie Clodic, Vincent Montreuil, Rachid Alami, Raja Chatila. A Decisional Framework for Autonomous Robots Interacting with Humans. IEEE International Workshop on Robot and Human Interactive Communication (ROMAN), 2005, Nashville, United States `https://hal.laas.fr/hal-01977560`
2. Aurélie Clodic, Sara Fleury, Rachid Alami, Matthieu Herrb, Raja Chatila. Supervision and Interaction Analysis of an Autonomous Tour-guide Robot Deployment. 12th International Conference on Advanced Robotics„ Jul 2005, Seattle, United States. `https://hal.laas.fr/hal-01977554`.
3. Aurélie Clodic, Sara Fleury, Rachid Alami, Raja Chatila, Gérard Bailly, et al.. Rackham : An Interactive Robot-Guide. IEEE International Workshop on Robots and Human Interactive Communications (ROMAN), Sep 2006, Hatfield, United Kingdom. pp.502-509. `https://hal.archives-ouvertes.fr/hal-00480381`.
4. Emrah Akin Sisbot, Aurélie Clodic, Luis Felipe Marin-Urias, Mathias Fontmarty, Ludovic Brèthes, et al.. Implementing a Human-Aware Robot System. ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, Sep 2006, Hatfield, United Kingdom `https://hal.laas.fr/hal-01977571`.
5. Aurélie Clodic, Maxime Ransan, Rachid Alami, Vincent Montreuil. A management of mutual belief for human-robot interaction. 2007 IEEE International Conference on Systems, Man and Cybernetics, Oct 2007, Montreal, Canada. (poster) `https://hal.laas.fr/hal-01977577`.
6. Aurélie Clodic, Rachid Alami, Vincent Montreuil, Shuyin Li, Britta Wrede, et al.. A study of interaction between dialog and decision for human-robot collaborative task achievement. RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication, 2007, Jeju, South Korea `https://hal.laas.fr/hal-01977573`.
7. Vincent Montreuil, Aurélie Clodic, Rachid Alami. Planning Human Centered Robot Activities. 2007 IEEE International Conference on Systems, Man and Cybernetics, Oct 2007, Montreal, Canada `https://hal.laas.fr/hal-01977579`.
8. Emrah Akin Sisbot, Maxime Ransan, Aurélie Clodic, Rachid Alami. Supervision and Motion Planning for a Mobile Manipulator Interacting with Humans. 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), Mar 2008, Amsterdam, Netherlands `https://hal.laas.fr/hal-01979138`.
9. Aurélie Clodic, Hung Cao, Samir Alili, Vincent Montreuil, Rachid Alami, et al.. SHARY : a supervision system adapted to Human-Robot Interaction. In : Khatib O., Kumar V., Pappas G.J. (eds) Experimental Robotics. Springer Tracts in Advanced Robotics, vol 54. Springer, Berlin, Heidelberg, pp.229-238, 2009. `https://hal.laas.fr/hal-01979196`.

10. Michelangelo Fiore, Aurélie Clodic, Rachid Alami. On Planning and Task achievement Modalities for Human-Robot Collaboration. International Symposium on Experimental Robotics (ISER 2014), Jun 2014, Marrakech, Morocco. 15p. https://hal.archives-ouvertes.fr/hal-01149109.
11. Grégoire Milliez, Matthieu Warnier, Aurélie Clodic, Rachid Alami. A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management. IEEE RO-MAN 2014, Aug 2014, Edinburgh, United Kingdom. pp.FrAT2.4, 10.1109/ROMAN.2014.6926399 https://hal.archives-ouvertes.fr/hal-01064546.
12. Mamoun Gharbi, Pierre-Vincent Paubel, Aurélie Clodic, Ophélie Carreras, Rachid Alami, et al. Toward a better understanding of the communication cues involved in a human-robot object transfer. 24th International Symposium on Robot and Human Interactive Communication, IEEE RO-MAN, Aug 2015, Kobe, Japan. pp.319-324, 10.1109/ROMAN.2015.7333626 https://hal.archives-ouvertes.fr/hal-01628249.
13. Aurélie Clodic, Elisabeth Pacherie, Rachid Alami, Raja Chatila. Key Elements for Human Robot Joint Action. Sociality and Normativity for Robots Philosophical Inquiries into Human-Robot Interactions, Springer, pp.159-177, 2017, Studies in the Philosophy of Sociality, 10.1007/978-3-319-53133-5_8 . This document is an extended version of the one published in the proceedings of RoboPhilosophyconference. https://hal.laas.fr/ijn_03084126.
14. Sandra Devin, Aurélie Clodic, Rachid Alami. About Decisions During Human-Robot Shared Plan Achievement : Who Should Act and How ?. Social Robotics,9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings, 10652, springer, pp.453-463, 2017, Lecture Notes in Computer Science book series, 978-3-319-70022. https://hal.laas.fr/hal-01943959.
15. Aurélie Clodic, Javier Vázquez-Salceda, Frank Dignum, Samuel Mascarenhas, Virginia Dignum, et al.. On the Pertinence of Social Practices for Social Robotics. IOS Press. Envisioning Robots in Society – Power, Politics, and Public Space, pp.36-74, 2018, 978-1-61499-931-7. 10.3233/978-1-61499-931-7-63 https://hal.laas.fr/hal-01943774.
16. Sandra Devin, Camille Vrignaud, Kathleen Belhassein, Aurélie Clodic, Ophélie Carreras, et al.. Evaluating the Pertinence of Robot Decisions in a Human-Robot Joint Action Context : The PeRDITA Questionnaire. 27th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2018), Aug 2018, Nanjing, China. 10.1109/ROMAN.2018.8525785 .https://hal.laas.fr/hal-01906101.
17. Virginia Dignum, Frank Dignum, Javier Vázquez-Salceda, Aurélie Clodic, Manuel Gentile, et al.. Design for Values for Social Robot Architectures. Envisioning Robots in Society – Power, Politics, and Public Space, pp.43-52, 2018, 978-1-61499-931-7. 10.3233/978-1-61499-931-7-43 https://hal.laas.fr/hal-01943831.
18. Séverin Lemaignan, Yoan Sallami, Christopher Wallbridge, Aurélie Clodic, Tony Belpaeme, et al. UNDERWORLDS : Cascading Situation Assessment for Robots. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018), Oct 2018, Madrid, Spain. 10.1109/IROS.2018.8594094 . https://hal.laas.fr/hal-01943917.

19. Päivi Heikkilä, Hanna Lammi, Marketta Niemelä, Kathleen Belhassein, Guillaume Sarthou, et al.. Should a robot guide like a human ? A qualitative four-phase study of a shopping mall robot. The Eleventh International Conference on Social Robotics (ICSR 2019), Nov 2019, Madrid, Spain. https://hal.laas.fr/hal-02285035.
20. Yoan Sallami, Séverin Lemaignan, Aurélie Clodic, Rachid Alami. Simulation-based physics reasoning for consistent scene estimation in an HRI context. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Nov 2019, Macau, China. pp.7834-7841, 10.1109/IROS40897.2019.8968106 . https://hal.archives-ouvertes.fr/hal-02925945.
21. Guillaume Sarthou, Aurélie Clodic, Rachid Alami. Ontologenius : A long-term semantic memory for robotic agents. IEEE International Conference on Robot & Human Interactive Communication (IEEE Ro-MAN 2019), Oct 2019, New Delhi, India. https://hal.archives-ouvertes.fr/hal-02283342.
22. Claire Boine, Céline Castets-Renard, Aurélie Clodic, Rachid Alami. In Love with a Corporation without Knowing It : An Asymetrical Relationship. Nørskov, M., Seibt J., Quick O. 2020. Culturally Sustainable Social Robotics—Proceedings of Robophilosophy 2020., IOS Press, 2020, Frontiers of AI and Its Applications„ 10.3233/FAIA200923 https://hal.archives-ouvertes.fr/hal-02963070.
23. Guilhem Buisan, Guillaume Sarthou, Arthur Bit-Monnot, Aurélie Clodic, Rachid Alami. Efficient, Situated and Ontology based Referring Expression Generation for Human-Robot collaboration. The 29th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), Aug 2020, Naples (on line), Italy. pp.349-356, 10.1109/RO-MAN47096.2020.9223485 . https://hal.archives-ouvertes.fr/hal-02922098.
24. Víctor Fernández Castro, Raul Hakli, Aurélie Clodic. What does it take to be a social agent ?. Appeared in : Nørskov, M., Seibt J., Quick O. 2020. Culturally Sustainable Social Robotics—Proceedings of Robophilosophy 2020., Series Frontiers of AI and Its Applications, IOS Press, Amsterdam., 2020 https://hal.archives-ouvertes.fr/hal-02963044.
25. Guillaume Sarthou, Guilhem Buisan, Aurélie Clodic, Rachid Alami. Extending Referring Expression Generation through shared knowledge about past Human-Robot collaborative activity. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Sep 2021, Prague (online), Czech Republic https://hal.archives-ouvertes.fr/hal-03389504.
26. Guillaume Sarthou, Amandine Mayima, Guilhem Buisan, Kathleen Belhassein, Aurélie Clodic. The Director Task : a Psychology-Inspired Task to Assess Cognitive and Interactive Robot Architectures. 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2021), Aug 2021, Vancouver (Virtual), Canada. 10.1109/RO-MAN50785.2021.9515543 .https://hal.archives-ouvertes.fr/hal-03327222.

3.4 Workshops

1. Rachid Alami, Aurélie Clodic, Vincent Montreuil, Emrah Akin Sisbot, Raja Chatila. Toward Human-Aware Robot Task Planning. AAAI spring symposium : to

boldly go where no human-robot team has gone before, Mar 2006, Palo Alto, CA, United States. https://hal.laas.fr/hal-01977564.

2. Aurélie Clodic, Séverin Lemaignan, Amit Kumar Pandey, Lavindra de Silva, Matthieu Warnier, et al. On robot decisional abilities for human-robot joint action. Joint Action Meeting V, 2013, Berlin, Germany. 2013. (Poster) https://hal.laas.fr/hal-01979217
3. Amit Kumar Pandey, Aurélie Clodic, Rachid Alami. Vers des robots socialement intelligents. Conférence III - Intercompréhension de l'Intraspécifique à l'Interspécifique, Oct 2013, Guidel, France https://hal.laas.fr/hal-01943658
4. Amit Pandey, Aurélie Clodic, Lavindra de Silva, Séverin Lemaignan, Mathieu Warnier, et al.. Bottom up development of a robot's basic socio-cognitive abilities for joint action. 5th Joint Action Meeting (JAM), 2013, Berlin, Germany (Poster) https://hal.laas.fr/hal-01977538.
5. Erwan Renaudo, Sandra Devin, Benoît Girard, Raja Chatila, Rachid Alami, et al. Learning to interact with humans using goal-directed and habitual behaviors. Ro-Man 2015, Workshop on Learning for Human-Robot Collaboration, Aug 2015, Kobe, Japan. https://hal.laas.fr/hal-01944380.
6. Sandra Devin, Grégoire Milliez, Michelangelo Fiore, Aurélie Clodic, Rachid Alami. Some essential skills and their combination in an architecture for a cognitive and interactive robot. Workshop on the Eleventh ACM/IEEE International Conference on Human Robot Interation, Mar 2016, Christchurch, New Zealand. https://hal.archives-ouvertes.fr/hal-01330340.
7. Kathleen Belhassein, Guilhem Buisan, Aurélie Clodic, and Rachid Alami. Towards methodological principles for user studies in HumanRobot Interaction. In Test Methods and Metrics for Effective HRI in Collaborative Human-Robot Teams Workshop, ACM/IEEE International Conference on Human-Robot Interaction, Daegu, South Korea, March 2019. https://hal.laas.fr/hal-02282600.
8. Aurélie Clodic, Rachid Alami, Elisabeth Pacherie, and Víctor Fernández Castro. Commitments in Human-Robot Interaction. 8th Joint Action Meeting, July 2019. Poster.:https://hal.laas.fr/hal-02272349.
9. Mary Ellen Foster, Bart Craenen, Amol Deshmukh, Oliver Lemon, Emanuele Bastianelli, Christian Dondrup, Ioannis Papaioannou, Andrea Vanzo, Jean-Marc Odobez, Olivier Canévet, Yuanzhouhan Cao, Weipeng He, Angel Martínez-González, Petr Motlicek, Rémy Siegfried, Rachid Alami, Kathleen Belhassein, Guilhem Buisan, Aurélie Clodic, Amandine Mayima, Yoan Sallami, Guillaume Sarthou, Phani-Teja Singamaneni, Jules Waldhart, Alexandre Mazel, Maxime Caniot, Marketta Niemelä, Päivi Heikkilä, Hanna Lammi, and Antti Tammela. MuMMER : Socially Intelligent Human-Robot Interaction in Public Spaces. In Artificial Intelligence for Human-Robot Interaction Symposium (AI-HRI), AI-HRI 2019 Proceedings, Arlington, VA, United States, November 2019. AAAI Fall Symposium Series 2019. https://hal.archives-ouvertes.fr/hal-02291690.
10. Amandine Mayima, Aurélie Clodic, and Rachid Alami. Evaluation of the Quality of Interaction from the robot point of view in Human-Robot Interactions. In 1st Edition of Quality of Interaction in Socially Assistive Robots (QISAR) Workshop, Madrid, Spain, November 2019. The 11th International Conference on Social Robotics (ICSR 2019). https://hal.laas.fr/hal-02403081.

11. Guillaume Sarthou, Rachid Alami, and Aurélie Clodic. Semantic Spatial Representation : a unique representation of an environment based on an ontology for robotic applications. In Combined Workshop on Spatial Language Understanding (SpLU) and Grounded Communication for Robotics (RoboNLP), Minneapolis, United States, June 2019a. https://hal.archives-ouvertes.fr/hal-02177140.

3.5 Autres

1. Kathleen Belhassein, Aurélie Clodic, Hélène Cochet, Marketta Niemelä, Päivi Heikkilä, et al.. Human-Human Guidance Study. Rapport LAAS n° 17596. 2017 https://hal.laas.fr/hal-01719730.

Bibliographie

[Alami 1998] Rachid Alami, Raja Chatila, Sara Fleury, Malik Ghallab et Félix Ingrand. *An Architecture for Autonomy.* The International Journal of Robotics Research, vol. 17, pages pp.315–337, avril 1998. (Cité en page 7.)

[Alami 2000] R. Alami, M. Herrb, B. Morisset, R. Chatila, F. Ingrand, P. Moutarlier, S. Fleury, M. Khatib et T. Simeon. *Around the lab in 40 days [indoor robot navigation].* Dans Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065), volume 1, pages 88–94 vol.1, 2000. (Cité en page 5.)

[Alami 2006] Rachid Alami, Aurélie Clodic, Vincent Montreuil, Emrah Akin Sisbot et Raja Chatila. *Toward Human-Aware Robot Task Planning.* Dans AAAI spring symposium : to boldly go where no human-robot team has gone before, Palo Alto, CA, United States, mars 2006. (Cité en pages 24, 127 et 142.)

[Alami 2011] Rachid Alami, Mathieu Warnier, Julien Guitton, Séverin Lemaignan et Emrah Akin Sisbot. *When the robot considers the human.* Dans The 15th International Symposium on Robotics Research (ISRR), Flagstaff, United States, août 2011. (Cité en pages 19 et 135.)

[Alili 2009a] Samir Alili, Rachid Alami et Vincent Montreuil. *A Task Planner for an Autonomous Social Robot.* Dans In : Asama H., Kurokawa H., Ota J., Sekiyama K. (eds) Distributed Autonomous Robotic Systems 8. Springer, Berlin, Heidelberg, pages 335–344. Springer Berlin Heidelberg, 2009. (Cité en pages 24 et 147.)

[Alili 2009b] Samir Alili, Matthieu Warnier, Muhammad Ali et Rachid Alami. *Planning and Plan-execution for Human-Robot Cooperative Task achievement.* Dans 19th international conference on automated planning and scheduling (ICAPS), Thessaloniki, Greece, septembre 2009. (Cité en pages 51, 56 et 58.)

[Allen 2003] Gary L. Allen. *Gestures Accompanying Verbal Route Directions : Do They Point to a New Avenue for Examining Spatial Representations ?* Spatial Cognition & Computation, vol. 3, no. 4, pages 259–268, 2003. (Cité en page 115.)

[Altisen 2003] Karine Altisen, Aurélie Clodic, Florence Maraninchi et Eric Rutten. *Using Controller-Synthesis Techniques to Build Property-Enforcing Layers.* Dans Pierpaolo Degano, editeur, Programming Languages and Systems, pages 174–188, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg. (Cité en pages 63 et 67.)

[and 2014] Glenda Satne and. *What Binds Us Together.* Philosophical Topics, vol. 42, no. 1, pages 43–61, 2014. (Cité en page 72.)

[Andrews 2012] Kristin Andrews. Do apes read minds ?: Toward a new folk psychology. MIT Press, 2012. (Cité en page 72.)

[Asher 2016] Nicholas Asher, Julie Hunter, Mathieu Morey, Farah Benamara et Stergos Afantenos. *Discourse structure and dialogue acts in multiparty dialogue : The STAC corpus.* Dans Language Resources and Evaluation Conference, pages 2721–2727. 2016. (Cité en page 44.)

[Asher 2020] Nicholas Asher, Julie Hunter et Kate Thompson. *Modelling Structures for Situated Discourse.* Dialogue & Discourse, vol. 11, no. 1, pages 89–121, mars 2020. (Cité en page 44.)

[Austin 1962] John Langshaw Austin. How to do things with words. Clarendon Press, 1962. (Cité en page 74.)

[Bartneck 2008] Christoph Bartneck, Dana Kulić, Elizabeth Croft et Susana Zoghbi. *Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots.* International Journal of Social Robotics, vol. 1, no. 1, pages 71–81, novembre 2008. (Cité en page 87.)

[Belhassein] Kathleen Belhassein. *Propositions de stratégies communicatives pour une action jointe Humain-Robot efficace, fluide et durable.* PhD thesis. (Cité en page 156.)

[Belhassein 2017] Kathleen Belhassein, Aurélie Clodic, Hélène Cochet, Marketta Niemelä, Päivi Heikkilä, Hanna Lammi et Antti Tammela. *Human-Human Guidance Study*, décembre 2017. (Cité en pages 90, 91, 111 et 112.)

[Belhassein 2019] Kathleen Belhassein, Guilhem Buisan, Aurélie Clodic et Rachid Alami. *Towards methodological principles for user studies in Human-Robot Interaction.* Dans Test Methods and Metrics for Effective HRI in Collaborative Human-Robot Teams Workshop, ACM/IEEE International Conference on Human-Robot Interaction, Daegu, South Korea, mars 2019. (Cité en pages 86, 88 et 91.)

[Belhassein 2020] Kathleen Belhassein, Víctor Fernández Castro et Amandine Mayima. *A Horizontal Approach to Communication for Human-Robot Joint Action : Towards Situated and Sustainable Robotics.* Dans Culturally Sustainable Social Robotics, pages 204–214. IOS Press, 2020. (Cité en page 82.)

[Belhassein 2022] Kathleen Belhassein, Fernández-Castro Víctor, Amandine Mayima, Aurélie Clodic, Elisabeth Pacherie, Michèle Guidetti, Rachid Alami et Hélène Cochet. *Addressing joint action challenges in HRI : Insights from psychology and philosophy.* Acta Psychologica, vol. 222, page 103476, février 2022. (Cité en pages 55, 69, 72 et 83.)

[Beßler 2020] Daniel Beßler, Robert Porzel, Mihai Pomarlan, Abhijit Vyas, Sebastian Höffner, Michael Beetz, Rainer Malaka et John Bateman. *Foundations of the Sociophysical Model of Activities (SOMA) for Autonomous Robotic Agents.* arXiv preprint arXiv :2011.11972, 2020. (Cité en page 38.)

[Blodow 2011] Nico Blodow, Lucian Cosmin Goron, Zoltan-Csaba Marton, Dejan Pangercic, Thomas Rühr, Moritz Tenorth et Michael Beetz. *Autonomous Semantic Mapping for Robots Performing Everyday Manipulation Tasks in Kitchen Environments.* Dans 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), San Francisco, CA, USA, September, 25–30 2011. (Cité en page 33.)

[Boine 2020] Claire Boine, Céline Castets-Renard, Aurélie Clodic et Rachid Alami. *In Love with a Corporation without Knowing It : An Asymetrical Relationship.* Dans Nørskov, M., Seibt J. et Quick O. 2020., editeurs, Culturally Sustainable Social Robotics-Proceedings of Robophilosophy 2020., Frontiers of AI and Its Applications,. IOS Press, août 2020. (Cité en pages 99 et 103.)

[Brandom 2007] Robert B. Brandom. *The structure of desire and recognition : Self-consciousness and self-constitution.* Philosophy & Social Criticism, vol. 33, no. 1, pages 127–150, 2007. (Cité en page 72.)

[Bratman 1992] Michael E Bratman. *Shared cooperative activity.* The philosophical review, vol. 101, no. 2, pages 327–341, 1992. (Cité en page 73.)

[Bratman 1999] Michael Bratman. Intention, plans, and practical reason. Center for the Study of Language and Information, Stanford, California, 1999. (Cité en pages 8 et 9.)

[Breazeal 2003] Cynthia Breazeal. *Toward sociable robots.* Robotics and Autonomous Systems, vol. 42, no. 3, pages 167 – 175, 2003. Socially Interactive Robots. (Cité en page 127.)

[Breazeal 2004] Cynthia Breazeal, Andrew Brooks, Jesse Gray, Guy Hoffman, Cory Kidd, Hans Lee, Jeff Lieberman, Andrea Lockerd et David Chilongo. *Tutelage and Collaboration for Humanoid Robots.* International Journal of Humanoid Robotics, vol. 01, no. 02, pages 315–348, juin 2004. (Cité en page 75.)

[Brinck 2018] Ingar Brinck et Christian Balkenius. *Mutual Recognition in Human-Robot Interaction : a Deflationary Account.* Philosophy & Technology, vol. 33, no. 1, pages 53–70, décembre 2018. (Cité en pages 69 et 73.)

[Broekstra 2002] Jeen Broekstra, Arjohn Kampman et Frank Van Harmelen. *Sesame : A generic architecture for storing and querying rdf and rdf schema.* Dans International semantic web conference, pages 54–68. Springer, 2002. (Cité en pages 38 et 39.)

[Bruno 2017] Barbara Bruno, Nak Young Chong, Hiroko Kamide, Sanjeev Kanoria, Jaeryoung Lee, Yuto Lim, Amit Kumar Pandey, Chris Papadopoulos, Irena Papadopoulos, Federico Pecora*et al. The CARESSES EU-Japan project : making assistive robots culturally competent.* Dans Italian Forum of Ambient Assisted Living, pages 151–169. Springer, 2017. (Cité en page 38.)

[Bryson 2017] J. Bryson et A. Winfield. *Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems.* Computer, vol. 50, no. 5, pages 116–119, 2017. (Cité en page 128.)

[Buisan 2020a] Guilhem Buisan, Guillaume Sarthou et Rachid Alami. *Human Aware Task Planning Using Verbal Communication Feasibility and Costs.* Dans 12th International Conference on Social Robotics (ICSR 2020), volume 12 of *Social Robotics*, pages 554–565, Golden, United States, novembre 2020. Springer. (Cité en page 135.)

[Buisan 2020b] Guilhem Buisan, Guillaume Sarthou, Arthur Bit-Monnot, Aurélie Clodic et Rachid Alami. *Efficient, Situated and Ontology based Referring Expression Generation for Human-Robot collaboration.* Dans The 29th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), pages 349–356, Naples (on line), Italy, août 2020. IEEE. (Cité en pages 41, 45 et 135.)

[Buisan 2021] Guilhem Buisan et Rachid Alami. *A Human-Aware Task Planner Explicitly Reasoning About Human and Robot Decision, Action and Reaction.* Dans HRI '21 : ACM/IEEE International Conference on Human-Robot Interaction, pages 544–548, Boulder CO USA (virtual), United States, mars 2021. ACM. (Cité en page 127.)

[Bálint-Benczédi 2018] Ferenc Bálint-Benczédi et Michael Beetz. *Variations on a Theme : "It's a Poor Sort of Memory that Only Works Backwards".* Dans 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 8390–8396, 2018. (Cité en page 37.)

[Cakmak 2014] Maya Cakmak et Leila Takayama. *Teaching People How to Teach Robots : The Effect of Instructional Materials and Dialog Design.* Dans 2014 9th ACM/IEEE

International Conference on Human-Robot Interaction (HRI), pages 431–438, 2014. (Cité en page 73.)

[Calo 2011] Ryan Calo. *Robots and Privacy.* Dans Patrick Lin, George Bekey et Keith Abney, editeurs, Robot Ethics : The Ethical and Social Implications of Robotics. Cambridge : MIT Press, 2011. (Cité en page 100.)

[Carpinella 2017] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez et Steven J. Stroessner. *The Robotic Social Attributes Scale (RoSAS) : Development and Validation.* Dans 2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI, pages 254–262, 2017. (Cité en page 88.)

[Carvalho 2020] Eros Moreira Carvalho. Social affordance. Springer, Cham, 2020. (Cité en page 148.)

[Castro 2019] Victor Fernandez Castro, Aurelie Clodic, Rachid Alami et Elisabeth Pacherie. *Commitments in Human-Robot Interaction.* Dans AAAI Fall Symbosium AI-HRI, 2019. (Cité en pages 137 et 142.)

[Castro 2020] Víctor Fernández Castro et Manuel Heras-Escribano. *Social Cognition : A Normative Approach.* Acta Analytica, vol. 35, no. 1, pages 75–100, 2020. (Cité en page 72.)

[Chatila 2018] Raja Chatila, Erwan Renaudo, Mihai Andries, Omar Ricardo Chavez-Garcia, Pierre Luce-Vayrac, Raphaël Gottstein, Rachid Alami, Aurélie Clodic, Sandra Devin, Benoît Girard et Mehdi Khamassi. *Toward Self-Aware Robots.* Frontiers in Robotics and AI, vol. 5, page 88, août 2018. (Cité en pages 56, 61, 145 et 148.)

[Clark 1991] Herbert H. Clark et Susan E. Brennan. *Grounding in communication.* Dans Perspectives on socially shared cognition., pages 127–149. American Psychological Association, 1991. (Cité en page 81.)

[Clark 1996a] Herbert H Clark. Using language. Cambridge university press, 1996. (Cité en pages 33 et 48.)

[Clark 1996b] Herbert H. Clark. Using language. Cambridge University Press, mai 1996. (Cité en page 81.)

[Clark 2006] Herbert H. Clark. Social actions, social commitments. Oxford : Berg Press, 2006. (Cité en page 81.)

[Clodic 2006] Aurélie Clodic, Sara Fleury, Rachid Alami, Raja Chatila, Gérard Bailly, Ludovic Brèthes, Maxime Cottret, Patrick Danès, Xavier Dollat, Frédéric Elisei, Isabelle Ferrané, Matthieu Herrb, Guillaume Infantes, Christian Lemaire, Frédéric Lerasle, Jérôme Manhes, Patrick Marcoul, Paulo Menezes et Vincent Montreuil. *Rackham : An Interactive Robot-Guide.* Dans IEEE International Workshop on Robots and Human Interactive Communications (ROMAN), pages 502–509, Hatfield, United Kingdom, septembre 2006. (Cité en pages 6 et 75.)

[Clodic 2007a] Aurélie Clodic. *Supervision pour un robot interactif : action et interaction pour un robot autonome en environnement humain.* Theses, Université Paul Sabatier - Toulouse III, juin 2007. (Cité en pages 47, 55 et 75.)

[Clodic 2007b] Aurélie Clodic, Maxime Ransan, Rachid Alami et Vincent Montreuil. *A management of mutual belief for human-robot interaction.* Dans 2007 IEEE International Conference on Systems, Man and Cybernetics, Montreal, Canada, octobre 2007. (Cité en pages 45 et 135.)

[Clodic 2009] Aurélie Clodic, Hung Cao, Samir Alili, Vincent Montreuil, Rachid Alami et Raja Chatila. *SHARY : a supervision system adapted to Human-Robot Interaction.* Dans In : Khatib O., Kumar V., Pappas G.J. (eds) Experimental Robotics. Springer Tracts in Advanced Robotics, vol 54. Springer, Berlin, Heidelberg, pages 229–238. 2009. (Cité en pages 24, 136 et 137.)

[Clodic 2017] Aurélie Clodic, Elisabeth Pacherie, Rachid Alami et Raja Chatila. *Key Elements for Human Robot Joint Action.* Dans Sociality and Normativity for Robots-Philosophical Inquiries into Human-Robot Interactions, Studies in the Philosophy of Sociality, pages 159–177. Springer, 2017. This document is an extended version of the one published in the proceedings of RoboPhilosophy conference. (Cité en pages 13 et 28.)

[Clodic 2018a] Aurélie Clodic, Javier Vázquez-Salceda, Frank Dignum, Samuel Mascarenhas, Virginia Dignum, Agnese Augello, Manuel Gentile et Rachid Alami. *On the Pertinence of Social Practices for Social Robotics.* Dans IOS Press, editeur, Envisioning Robots in Society – Power, Politics, and Public Space, pages 36–74. 2018. (Cité en pages 130 et 131.)

[Clodic 2018b] Aurélie Clodic, Javier Vázquez-Salceda, Frank Dignum, Samuel Mascarenhas, Virginia Dignum, Agnese Augello, Manuel Gentile et Rachid Alami. *On the Pertinence of Social Practices for Social Robotics.* Dans IOS Press, editeur, Envisioning Robots in Society – Power, Politics, and Public Space, pages 36–74. 2018. (Cité en page 142.)

[Clodic 2019] Aurélie Clodic, Rachid Alami, Elisabeth Pacherie et Víctor Fernández Castro. *Commitments in Human-Robot Interaction.* 8th Joint Action Meeting, juillet 2019. Poster. (Cité en pages 73 et 83.)

[Clodic 2021] Aurélie Clodic et Rachid Alami. *What Is It to Implement a Human-Robot Joint Action ?* Dans Robotics, AI, and Humanity, pages 229–238. Springer International Publishing, février 2021. (Cité en page 28.)

[Cohen 1990] Philip R. Cohen et Hector J. Levesque. *Intention is choice with commitment.* Artificial Intelligence, vol. 42, no. 2, pages 213–261, 1990. (Cité en pages 12, 73 et 76.)

[Cohen 1991] Philip R Cohen et Hector J Levesque. *Teamwork.* Nous, vol. 25, no. 4, pages 487–512, 1991. (Cité en page 51.)

[Commission 2019] European Commission, Content Directorate-General for Communications Networks et Technology. Lignes directrices en matière d'éthique pour une ia digne de confiance. Publications Office, 2019. (Cité en pages 102 et 103.)

[Dautenhahn 2006] Kerstin Dautenhahn, Michael Walters, Sarah Woods, Kheng Lee Koay, Christopher L Nehaniv, Emrah Akin Sisbot, Rachid Alami et Thierry Simeon. *How May I Serve You ? A Robot Companion Approaching a Seated Person in a Helping Context.* Dans 1st ACM SIGCHI/SIGART conference on Human-robot interaction (HRI '06), Utah, United States, mars 2006. (Cité en page 85.)

[Davidson 1980] Donald Davidson. Essays on actions and events. Oxford University Press, septembre 1980. (Cité en page 8.)

[de Greeff 2015] Joachim de Greeff et Tony Belpaeme. *Why Robots Should Be Social : Enhancing Machine Learning through Social Human-Robot Interaction.* PLOS ONE, vol. 10, no. 9, pages 1–26, 09 2015. (Cité en page 127.)

[Dehais 2011] Frédéric Dehais, Emrah Akin Sisbot, Rachid Alami et Mickaël Causse. *Physiological and subjective evaluation of a human–robot object hand-over task.* Applied Ergonomics, vol. 42, no. 6, pages 785–791, novembre 2011. (Cité en page 144.)

[Dennett 2009] Daniel Dennett. Intentional systems theory. Oxford University Press, janvier 2009. (Cité en page 70.)

[Devin 2016a] Sandra Devin et Rachid Alami. *An Implemented Theory of Mind to Improve Human-Robot Shared Plans Execution.* Dans The Eleventh ACM/IEEE International Conference on Human Robot Interation, The Eleventh ACM/IEEE International Conference on Human Robot Interation, pages 319–326, Christchurch, New Zealand, mars 2016. IEEE Press. (Cité en pages 28 et 142.)

[Devin 2016b] Sandra Devin et Rachid Alami. *An implemented theory of mind to improve human-robot shared plans execution.* page 319–326. IEEE Press, 2016. (Cité en page 136.)

[Devin 2017a] Sandra Devin. *Decisional issues during human-robot joint action.* Theses, Université de Toulouse, 2017. (Cité en pages 47, 51, 56 et 59.)

[Devin 2017b] Sandra Devin, Aurélie Clodic et Rachid Alami. *About Decisions During Human-Robot Shared Plan Achievement : Who Should Act and How ?* Dans Social Robotics,9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings, volume 10652 of *Lecture Notes in Computer Science book series*, pages 453–463. springer, octobre 2017. (Cité en pages 28, 61, 127, 136 et 147.)

[Devin 2018] Sandra Devin, Camille Vrignaud, Kathleen Belhassein, Aurélie Clodic, Ophélie Carreras et Rachid Alami. *Evaluating the Pertinence of Robot Decisions in a Human-Robot Joint Action Context : The PeRDITA Questionnaire.* Dans 27th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2018), Nanjing, China, août 2018. (Cité en pages 61, 88, 90 et 120.)

[Dignum 2015] Virginia Dignum et Frank Dignum. *Contextualized Planning Using Social Practices.* Dans Aditya Ghose, Nir Oren, Pankaj Telang et John Thangarajah, editeurs, Coordination, Organizations, Institutions, and Norms in Agent Systems X, pages 36–52. Springer, 2015. (Cité en page 131.)

[Dignum 2018] Virginia Dignum, Frank Dignum, Javier Vázquez-Salceda, Aurélie Clodic, Manuel Gentile, Samuel Mascarenhas et Agnese Augello. *Design for Values for Social Robot Architectures.* Dans Envisioning Robots in Society – Power, Politics, and Public Space, pages 43–52. 2018. (Cité en pages 94 et 103.)

[Dignum 2019] Virginia Dignum. Responsible artificial intelligence : How to develop and use ai in a responsible way. Springer, Cham, 2019. (Cité en pages 94 et 103.)

[dir] *Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive').* (Cité en page 100.)

[Fernández Castro] Víctor Fernández Castro, Amandine Mayima, Kathleen Belhassein et Aurélie Clodic. *(Forthcoming) The Role of Commitments in Socially Appropriate Robotics.* Dans To appear in : Gransche, M. Nerurkar, J. Bellon, S. Nähr-Wagener

(eds) Technology Socialization ? Social appropriateness for computer-based systems. (Cité en pages 73 et 83.)

[Fernandez Castro 2019] V Fernandez Castro, Aurélie Clodic, Rachid Alami et Elisabeth Pacherie. Commitments in Human-Robot Interaction. working paper or preprint, septembre 2019. (Cité en pages 73, 83 et 147.)

[Fernández Castro 2020] Víctor Fernández Castro, Raul Hakli et Aurélie Clodic. *What does it take to be a social agent ?* Dans Appeared in : Nørskov, M., Seibt J., Quick O. 2020. Culturally Sustainable Social Robotics-Proceedings of Robophilosophy 2020. Series Frontiers of AI and Its Applications, IOS Press, Amsterdam., août 2020. (Cité en pages 69, 71, 83 et 148.)

[Fiore 2014] Michelangelo Fiore, Aurélie Clodic et Rachid Alami. *On Planning and Task achievement Modalities for Human-Robot Collaboration.* Dans International Symposium on Experimental Robotics (ISER 2014), page 15p., Marrakech, Morocco, juin 2014. (Cité en pages 60, 127 et 136.)

[Fischer 2020] Kerstin Fischer, Johanna Seibt, Raffaele Rodogno, Maike Kirkegård Rasmussen, Astrid Weiss, Leon Bodenhagen, William Kristian Juel et Norbert Krüger. *Integrative Social Robotics Hands-on.* Interaction Studies, 2020. (Cité en pages 87, 93 et 94.)

[Fong 2006] Terrence Fong, Clayton Kunz, Laura M. Hiatt et Magda Bugajska. *The Human-Robot Interaction Operating System.* Dans Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI '06, page 41–48, New York, NY, USA, 2006. Association for Computing Machinery. (Cité en page 75.)

[for Ethical Considerations in Artificial Intelligence 2017] The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence et Autonomous Systems. *Ethically Aligned Design : A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 2*, décembre 2017. (Cité en pages 94 et 128.)

[Foster 2019] Mary Ellen Foster, Bart Craenen, Amol Deshmukh, Oliver Lemon, Emanuele Bastianelli, Christian Dondrup, Ioannis Papaioannou, Andrea Vanzo, Jean-Marc Odobez, Olivier Canévet, Yuanzhouhan Cao, Weipeng He, Angel Martínez-González, Petr Motlicek, Rémy Siegfried, Rachid Alami, Kathleen Belhassein, Guilhem Buisan, Aurélie Clodic, Amandine Mayima, Yoan Sallami, Guillaume Sarthou, Phani-Teja Singamaneni, Jules Waldhart, Alexandre Mazel, Maxime Caniot, Marketta Niemelä, Päivi Heikkilä, Hanna Lammi et Antti Tammela. *MuMMER : Socially Intelligent Human-Robot Interaction in Public Spaces.* Dans Artificial Intelligence for Human-Robot Interaction Symposium (AI-HRI), AI-HRI 2019 Proceedings, Arlington, VA, United States, novembre 2019. AAAI Fall Symposium Series 2019. (Cité en page 34.)

[Galindo 2008] C. Galindo, J.A. Fernández-Madrigal, J. González et A. Saffiotti. *Robot task planning using semantic maps.* Robotics and Autonomous Systems, vol. 56, no. 11, pages 955–966, 2008. (Cité en page 33.)

[Gat 1992] Erann Gat. *Integrating Planning and Reacting in a Heterogeneous Asynchronous Architecture for Controlling Real-World Mobile Robots.* Dans Proceedings of the Tenth National Conference on Artificial Intelligence, AAAI'92, page 809–815. AAAI Press, 1992. (Cité en page 8.)

[Gharbi 2015] Mamoun Gharbi, Pierre-Vincent Paubel, Aurélie Clodic, Ophélie Carreras, Rachid Alami et Jean-Marie Cellier. *Toward a better understanding of the communication cues involved in a human-robot object transfer.* Dans 24th International Symposium on Robot and Human Interactive Communication, IEEE RO-MAN, pages 319–324, Kobe, Japan, août 2015. (Cité en pages 24, 55, 60 et 90.)

[Gibson 1979] James J. Gibson. The ecological approach to visual perception. Houghton Mifflin, 1979. (Cité en page 145.)

[Gilbert 2013] Margaret Gilbert. Joint commitment : How we make the social world. Oxford University Press, 2013. (Cité en pages 73 et 74.)

[Gockley 2007] Rachel Gockley, Jodi Forlizzi et Reid Simmons. *Natural person-following behavior for social robots.* Dans Proceeding of the ACM/IEEE international conference on Human-robot interaction - HRI '07. ACM Press, 2007. (Cité en page 127.)

[Goffman 1963] Erving Goffman. Behavior in public places ; notes on the social organization of gatherings. Free Press of Glencoe, New York, 1963. (Cité en page 130.)

[Greenspan 1978] P. S. Greenspan. *Behavior Control and Freedom of Action.* The Philosophical Review, vol. 87, no. 2, page 225, avril 1978. (Cité en page 74.)

[Grosz 1996] Barbara J. Grosz et Sarit Kraus. *Collaborative plans for complex group action.* Artificial Intelligence, vol. 86, no. 2, pages 269–357, 1996. (Cité en pages 12 et 80.)

[Heikkilä 2019] Päivi Heikkilä, Hanna Lammi, Marketta Niemelä, Kathleen Belhassein, Guillaume Sarthou, Antti Tammela, Aurélie Clodic et Rachid Alami. *Should a robot guide like a human ? A qualitative four-phase study of a shopping mall robot.* Dans The Eleventh International Conference on Social Robotics (ICSR 2019), Madrid, Spain, novembre 2019. (Cité en pages 91, 112 et 115.)

[Holtz 2014] G. Holtz. *Generating Social Practices.* JASSS, vol. 17, no. 1, page 17, 2014. (Cité en page 130.)

[Huang 2013] Chien-Ming Huang et Bilge Mutlu. *The Repertoire of Robot Behavior : Enabling Robots to Achieve Interaction Goals through Social Behavior.* J. Hum.-Robot Interact., vol. 2, no. 2, page 80–102, jun 2013. (Cité en page 75.)

[Hunter 2018] Julie Hunter, Nicholas Asher et Alex Lascarides. *A formal semantics for situated conversation.* Semantics and Pragmatics, vol. 11, 2018. (Cité en page 44.)

[Hunter ming] Julie Hunter et Kate Thompson. *On the Role of Relations and Structure in Discourse Interpretation.* forthcoming. (Cité en page 44.)

[Jacobsson 2008] Henrik Jacobsson, Nick Hawes, Geert-Jan Kruijff et Jeremy Wyatt. *Crossmodal Content Binding in Information-Processing Architectures.* Dans Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, pages 81–88, New York, NY, USA, 2008. ACM. (Cité en page 33.)

[Jost 2020] Céline Jost, Brigitte Le Pévédic, Tony Belpaeme, Cindy Bethel, Dimitrios Chrysostomou, Nigel Crook, Marine Grandgeorge et Nicole Mirnig, editeurs. Human-robot interaction, evaluation methods and their standardization. Springer International Publishing, 2020. (Cité en page 89.)

[Khambhaita 2016] Harmish Khambhaita, Jorge Rios-Martinez et Rachid Alami. *Head-Body Motion Coordination for Human Aware Robot Navigation.* Dans 9th International workshop on Human-Friendlly Robotics (HFR 2016), page 8p., Gênes, Italy, septembre 2016. (Cité en page 24.)

[Khambhaita 2017a] Harmish Khambhaita et Rachid Alami. *Assessing the Social Criteria for Human-Robot Collaborative Navigation : A Comparison of Human-Aware Navigation Planners.* Dans Proc. IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), page 6p., Lisbonne, Portugal, août 2017. (Cité en pages 23 et 127.)

[Khambhaita 2017b] Harmish Khambhaita et Rachid Alami. *Viewing Robot Navigation in Human Environment as a Cooperative Activity.* Dans International Symposium on Robotics Research (ISSR 2017), page 18p., Puerto Varas, Chile, décembre 2017. (Cité en pages 23, 118 et 127.)

[Knoblich 2011] Günther Knoblich, Stephen Butterfill et Natalie Sebanz. *Psychological research on joint action : theory and data.* Psychology of learning and motivation, vol. 54, pages 59–101, 2011. (Cité en pages 10, 24, 28 et 142.)

[Krahmer 2012] Emiel Krahmer et Kees Van Deemter. *Computational generation of referring expressions : A survey.* Computational Linguistics, vol. 38, no. 1, pages 173–218, 2012. (Cité en page 41.)

[Kruse 2013] Thibault Kruse, Amit Kumar Pandey, Rachid Alami et Alexandra Kirsch. *Human-Aware Robot Navigation : A Survey.* Robotics and Autonomous Systems, vol. 61, no. 12, pages pp.1726–1743, décembre 2013. (Cité en page 23.)

[Kruse 2014] Thibault Kruse, Alexandra Kirsch, Harmish Khambhaita et Rachid Alami. *Evaluating directional cost models in navigation.* Dans Proc. ACM/IEEE International Conference on Human-Robot Interaction (HRI), pages 350 – 357, Bielefeld, Germany, mars 2014. (Cité en page 127.)

[Kuipers 2000] Benjamin Kuipers. *The Spatial Semantic Hierarchy.* Artificial Intelligence, vol. 119, no. 1, pages 191–233, 2000. (Cité en page 115.)

[Lallement 2014] Raphaël Lallement, Lavindra De Silva et Rachid Alami. *HATP : An HTN Planner for Robotics.* Dans 2nd ICAPS Workshop on Planning and Robotics, Portsmouth, United States, juin 2014. (Cité en pages 24 et 127.)

[Lallement 2018] Raphaël Lallement, Lavindra De Silva et Rachid Alami. *HATP : Hierarchical Agent-based Task Planner.* Dans International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, juillet 2018. (Cité en pages 51, 142 et 147.)

[Lassila 1998] Ora Lassila, Ralph R Swick*et al.* *Resource description framework (RDF) model and syntax specification.* 1998. (Cité en page 39.)

[Leite 2013] Iolanda Leite, Carlos Martinho et Ana Paiva. *Social Robots for Long-Term Interaction : A Survey.* International Journal of Social Robotics, vol. 5, no. 2, pages 291–308, janvier 2013. (Cité en page 127.)

[Lemaignan 2010] Séverin Lemaignan, Raquel Ros, Lorenz Mösenlechner, Rachid Alami et Michael Beetz. *ORO, a knowledge management module for cognitive architectures in robotics.* Dans EEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Taipei, Taiwan, octobre 2010. (Cité en pages 37, 38 et 135.)

[Lemaignan 2017] Séverin Lemaignan, Matthieu Warnier, Emrah Akin Sisbot, Aurélie Clodic et Rachid Alami. *Artificial Cognition for Social Human-Robot Interaction : An Implementation.* Artificial Intelligence, vol. 247, pages 45–69, juin 2017. (Cité en pages 135 et 148.)

[Lemaignan 2018] Séverin Lemaignan, Yoan Sallami, Christopher Wallbridge, Aurélie Clodic, Tony Belpaeme et Rachid Alami. *UNDERWORLDS : Cascading Situation Assessment for Robots.* Dans IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018), Madrid, Spain, octobre 2018. (Cité en pages 17, 34, 35, 36, 45, 114 et 127.)

[Lenat 1989] Douglas B Lenat et Ramanathan V Guha. Building large knowledge-based systems ; representation and inference in the cyc project. Addison-Wesley Longman Publishing Co., Inc., 1989. (Cité en page 38.)

[Lörken 2008] C. Lörken et J. Hertzberg. *Grounding planning operators by affordances.* Dans International Conference on Cognitive Systems (CogSys), pages 79–84, 2008. (Cité en page 33.)

[Mallot 2009] Hanspeter A. Mallot et Kai Basten. *Embodied spatial cognition : Biological and artificial systems.* Image and Vision Computing, vol. 27, no. 11, pages 1658–1670, 2009. (Cité en page 115.)

[Masolo 2003] C Masolo, S Borgo, A Gangemi, N Guarino, A Oltramari et L Schneider. *Dolce : a descriptive ontology for linguistic and cognitive engineering.* WonderWeb Project, Deliverable D17 v2, vol. 1, pages 75–105, 2003. (Cité en page 38.)

[Mavridis 2006] N. Mavridis et D. Roy. *Grounded Situation Models for Robots : Where words and percepts meet.* Dans IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006. (Cité en page 33.)

[Mayima 2019] Amandine Mayima, Aurélie Clodic et Rachid Alami. *Evaluation of the Quality of Interaction from the robot point of view in Human-Robot Interactions.* Dans 1st Edition of Quality of Interaction in Socially Assistive Robots (QISAR) Workshop, Madrid, Spain, novembre 2019. The 11th International Conference on Social Robotics (ICSR 2019). (Cité en pages 61 et 118.)

[Mayima 2021a] Amandine Mayima. *Endowing the Robot with the Abilities to Control and Evaluate itsContribution to a Human-Robot Joint Action.* Theses, INSA, octobre 2021. (Cité en pages 32, 47, 53, 55, 59, 62, 155 et 160.)

[Mayima 2021b] Amandine Mayima, Aurélie Clodic et Rachid Alami. *Towards Robots able to Measure in Real-time the Quality of Interaction in HRI Contexts.* International Journal of Social Robotics, 2021. (Cité en pages 61, 76, 127, 144 et 148.)

[Mele 1992] Alfred Mele. Springs of action : understanding intentional behavior. Oxford University Press, New York, 1992. (Cité en page 8.)

[Michael 2014] John Michael et Elisabeth Pacherie. *On Commitments and Other Uncertainty Reduction Tools.* Journal of Social Ontology, pages 1–34, 2014. (Cité en pages 74, 76, 79 et 80.)

[Michael 2016] John Michael, Natalie Sebanz et Günther Knoblich. *The Sense of Commitment : A Minimal Approach.* Frontiers in Psychology, vol. 6, page 1968, 2016. (Cité en page 74.)

[Milliez 2014] Grégoire Milliez, Matthieu Warnier, Aurélie Clodic et Rachid Alami. *A framework for endowing an interactive robot with reasoning capabilities about perspective-taking and belief management.* Dans IEEE RO-MAN 2014, page FrAT2.4, Edinburgh, United Kingdom, août 2014. (Cité en pages 17, 19, 32, 34, 45 et 127.)

[Milliez 2016a] Grégoire Milliez. *Raisonnement sur le contexte et les croyances pour l'interaction homme-robot.* Theses, Institut National Polytechnique de Toulouse - INPT, octobre 2016. (Cité en page 34.)

[Milliez 2016b] Grégoire Milliez, Raphaël Lallement, Michelangelo Fiore et Rachid Alami. *Using human knowledge awareness to adapt collaborative plan generation, explanation and monitoring.* Dans 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2016), pages 43–50, Christchurch, New Zealand, mars 2016. (Cité en page 127.)

[Montreuil 2007] Vincent Montreuil, Aurélie Clodic et Rachid Alami. *Planning Human Centered Robot Activities.* Dans 2007 IEEE International Conference on Systems, Man and Cybernetics, Montreal, Canada, octobre 2007. (Cité en pages 24, 51, 127 et 147.)

[Morisset 2002] Benoît Morisset. *Vers un robot à comportement robuste : apprendre à combiner des modalités sensori-motrices complémentaires.* Theses, Université Paul Sabatier - Toulouse III, 2002. (Cité en page 5.)

[Muscettola 1998] Nicola Muscettola, P.Pandurang Nayak, Barney Pell et Brian C. Williams. *Remote Agent : to boldly go where no AI system has gone before.* Artificial Intelligence, vol. 103, no. 1, pages 5–47, 1998. Artificial Intelligence 40 years later. (Cité en page 8.)

[Nesnas 2003] I.A.D. Nesnas, A. Wright, M. Bajracharya, R. Simmons et T. Estlin. *CLARAty and challenges of developing interoperable robotic software.* Dans Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453), volume 3, pages 2428–2435 vol.3, 2003. (Cité en page 8.)

[Niles 2001] Ian Niles et Adam Pease. *Towards a standard upper ontology.* Dans Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001, pages 2–9, 2001. (Cité en page 38.)

[Nüchter 2008] Andreas Nüchter et Joachim Hertzberg. *Towards semantic maps for mobile robots.* Robotics and Autonomous Systems, vol. 56, no. 11, pages 915 – 926, 2008. (Cité en page 33.)

[Olivares-Alarcos 2019] Alberto Olivares-Alarcos, Daniel Beßler, Alaa Khamis, Paulo Goncalves, Maki K Habib, Julita Bermejo-Alonso, Marcos Barreto, Mohammed Diab, Jan Rosell, Joao Quintas*et al. A review and comparison of ontology-based approaches to robot autonomy.* 2019. (Cité en page 38.)

[on Ethics of Autonomous 2017] The IEEE Global Initiative on Ethics of Autonomous et Intelligent Systems. Ethically aligned design : A vision for prioritizing human well-being with autonomous and intelligent systems, version 2. IEEE, 2017. (Cité en page 100.)

[Pacherie 2007] Elisabeth Pacherie. *Is collective intentionality really primitive ?* Dans C. Penco & M. Vignolo M. Beaney, editeur, Mental processes : representing and inferring, pages 153–175. Cambridge Scholars Press, 2007. (Cité en pages 11 et 12.)

[Pacherie 2008] Elisabeth Pacherie. *The phenomenology of action : A conceptual framework.* Cognition, vol. 107, no. 1, pages 179–217, 2008. (Cité en page 9.)

[Pacherie 2011] Elisabeth Pacherie. *Framing Joint Action.* Review of Philosophy and Psychology, vol. 2, no. 2, pages 173–192, 2011. (Cité en pages 11 et 12.)

[Pacherie 2012] Elisabeth Pacherie. *The Phenomenology of Joint Action : Self-Agency vs. Joint-Agency.* Dans Axel Seemann, editeur, Joint Attention : New Developments, pages 343–389. MIT Press, 2012. (Cité en pages 9, 11, 12, 16, 17 et 19.)

[Pacherie 2021] Elisabeth Pacherie et Víctor Fernández Castro. *Robots and Resentment : Commitments, recognition and social motivation in HRI.* Dans Springer, editeur, Emotional Machines. Perspectives from Affective Computing and Emotional Human-Machine Interaction. 2021. (Cité en page 74.)

[Paprzycka 1999] Katarzyna Paprzycka. *Normative Expectations, Intentions, and Beliefs.* The Southern Journal of Philosophy, vol. 37, no. 4, pages 629–652, décembre 1999. (Cité en page 74.)

[Paulius 2018] David Paulius, Ahmad B Jelodar et Yu Sun. *Functional Object-Oriented Network : Construction & Expansion.* Dans 2018 IEEE International Conference on Robotics and Automation (ICRA), pages 5935–5941. IEEE, 2018. (Cité en page 37.)

[Prasad 2020] Pranav Krishna Prasad et Wolfgang Ertel. *Knowledge Acquisition and Reasoning Systems for Service Robots : A Short Review of the State of the Art.* Dans 2020 5th International Conference on Robotics and Automation Engineering (ICRAE), pages 36–45, 2020. (Cité en page 37.)

[Reckwitz 2002] Andreas Reckwitz. *Toward a Theory of Social Practices : A Development in Culturalist Theorizing.* European Journal of Social Theory, vol. 5, no. 2, pages 243–263, 2002. (Cité en page 130.)

[Renaudo 2014] Erwan Renaudo, Benoît Girard, Raja Chatila et Mehdi Khamassi. *Design of a Control Architecture for Habit Learning in Robots.* 07 2014. (Cité en pages 56 et 58.)

[Renaudo 2015] Erwan Renaudo, Sandra Devin, Benoît Girard, Raja Chatila, Rachid Alami, Mehdi Khamassi et Aurélie Clodic. *Learning to interact with humans using goal-directed and habitual behaviors.* Dans Ro-Man 2015, Workshop on Learning for Human-Robot Collaboration, Kobe, Japan, août 2015. (Cité en pages 56, 61, 137 et 147.)

[Renaudo 2016] Erwan Renaudo. *Des comportements flexibles aux comportements habituels : Meta-apprentissage neuro-inspiré pour la robotique autonome.* Theses, Université Pierre et Marie Curie (Paris 6), juin 2016. (Cité en pages 56 et 59.)

[Robinson 2012] Jeffrey D. Robinson. Dans Jack Sidnell et Tanya Stivers, editeurs, The handbook of conversation analysis, volume 121, chapitre Overall Structural Organization. John Wiley & Sons, 2012. (Cité en page 49.)

[Roth 2004] Abraham Sesshu Roth. *Shared Agency and Contralateral Commitments.* The Philosophical Review, vol. 113, no. 3, pages 359–410, 2004. (Cité en page 73.)

[Roy 2020] Raphaëlle N. Roy, Nicolas Drougard, Thibault Gateau, Frédéric Dehais et Caroline P. C. Chanel. *How Can Physiological Computing Benefit Human-Robot Interaction ?* Robotics, vol. 9, no. 4, 2020. (Cité en page 144.)

[Sahin 2007] Erol Sahin, Maya Cakmak, Mehmet Dogar, Emre Ugur et Göktürk Üçoluk. *To Afford or Not to Afford : A New Formalization of Affordances Toward Affordance-Based Robot Control.* Adaptive Behavior - ADAPT BEHAV, vol. 15, pages 447–472, 12 2007. (Cité en page 145.)

[Sallami 2018] Yoan Sallami, Séverin Lemaignan, Aurélie Clodic et Rachid Alami. *Implementing an implicit false belief task by cascading situation assessment with UNDERWORLDS.* Workshop "toward a Framework for Joint Action" at Robotics Science and Systems (RSS 2018), juin 2018. Poster. (Cité en page 115.)

[Sallami 2019] Yoan Sallami, Séverin Lemaignan, Aurélie Clodic et Rachid Alami. *Simulation-based physics reasoning for consistent scene estimation in an HRI context.* Dans 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 7834–7841, Macau, China, novembre 2019. IEEE. (Cité en pages 34, 45 et 127.)

[Sallami 2021] Yoan Sallami. *Developmental psychology inspired models for physical and social reasoning in human-robot joint action.* Theses, Université Paul Sabatier - Toulouse III, janvier 2021. (Cité en page 34.)

[Sarthou 2019a] Guillaume Sarthou, Rachid Alami et Aurélie Clodic. *Semantic Spatial Representation : a unique representation of an environment based on an ontology for robotic applications.* Dans Combined Workshop on Spatial Language Understanding (SpLU) and Grounded Communication for Robotics (RoboNLP), Minneapolis, United States, juin 2019. (Cité en pages 40, 45, 115 et 116.)

[Sarthou 2019b] Guillaume Sarthou, Aurélie Clodic et Rachid Alami. *Ontologenius : A long-term semantic memory for robotic agents.* Dans IEEE International Conference on Robot & Human Interactive Communication (IEEE Ro-MAN 2019), New Delhi, India, octobre 2019. (Cité en pages 45, 116 et 135.)

[Sarthou 2021a] Guillaume Sarthou. *Knowledge representation and exploitation for interactive and cognitive robots.* Theses, UPS Toulouse - Université Toulouse 3 Paul Sabatier, septembre 2021. (Cité en pages 32, 37, 135, 155 et 159.)

[Sarthou 2021b] Guillaume Sarthou, Amandine Mayima, Guilhem Buisan, Kathleen Belhassein et Aurélie Clodic. *The Director Task : a Psychology-Inspired Task to Assess Cognitive and Interactive Robot Architectures.* Dans 30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2021), Vancouver (Virtual), Canada, août 2021. (Cité en pages 45, 86, 90, 91 et 135.)

[Satake 2009] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro et N. Hagita. *How to approach humans ?-strategies for social robots to initiate interaction.* Dans 2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pages 109–116, 2009. (Cité en page 127.)

[Saxena 2014] Ashutosh Saxena, Ashesh Jain, Ozan Sener, Aditya Jami, Dipendra K Misra et Hema S Koppula. *Robobrain : Large-scale knowledge engine for robots.* arXiv preprint arXiv :1412.0691, 2014. (Cité en page 38.)

[Scanlon 2000] Thomas Scanlon. What we owe to each other. Belknap Press of Harvard University Press, Cambridge, Mass, 2000. (Cité en page 72.)

[Schank 1977] Roger C. Schank et Robert P. Abelson. Scripts, plans, goals, and understanding : an inquiry into human knowledge structures. Psychology Press, 1977. (Cité en page 72.)

[Schatzki 1996] Theodore Schatzki. Social practices : a wittgensteinian approach to human activity and the social. Cambridge University Press, New York, 1996. (Cité en page 130.)

[Schegloff 1973] Emanuel A Schegloff et Harvey Sacks. *Opening up closings.* Semiotica, vol. 8, no. 4, pages 289–327, 1973. (Cité en page 49.)

[Schlenoff 2015] Craig Schlenoff, E Prestes, PJ Sequeira Gonçalves, Mathieu Abel, Yacine Amirat, S Balakirsky, ME Barreto, JL Carbonera, A Chibani, S Rama Fiorini*et al. Ieee standard ontologies for robotics and automation.* 2015. (Cité en page 38.)

[Searle 1983] John R. Searle. Intentionality : An essay in the philosophy of mind. Cambridge University Press, 1983. (Cité en page 8.)

[Sebanz 2006] Natalie Sebanz, Harold Bekkering et Günther Knoblich. *Joint action : bodies and minds moving together.* Trends in Cognitive Sciences, vol. 10, no. 2, pages 70–76, 2006. (Cité en page 15.)

[Seibt 2014] Johanna Seibt. *Varieties of the 'As If' : Five ways to simulate an action.* Frontiers in Artificial Intelligence and Applications, vol. 273, pages 97–104, 01 2014. (Cité en pages 15 et 69.)

[Seibt 2016] Johanna Seibt. *"Integrative Social Robotics" : A New Method Paradigm to Solve the Description and the Regulation Problem ?* Dans Johanna Seibt, Marco Nørskov et Søren Schack Andersen, editeurs, What Social Robots Can and Should Do, Frontiers in Artificial Intelligence and Applications, pages 104–115. IOS Press, 2016. null ; Conference date : 17-10-2016 Through 21-10-2016. (Cité en pages 87 et 93.)

[Seibt 2018] Johanna Seibt, Malene Flensborg Damholdt et Christina Vestergaard. *Five Principles of Integrative Social Robotics.* Dans Robophilosophy/TRANSOR, 2018. (Cité en pages 87, 93 et 101.)

[Shih 2021] Andy Shih, Arjun Sawhney, Jovana Kondic, Stefano Ermon et Dorsa Sadigh. *On the Critical Role of Conventions in Adaptive Human-AI Collaboration.* Dans 9th International Conference on Learning Representations (ICLR), May 2021. (Cité en page 128.)

[Shove 2012] E. Shove, M. Pantzar et M. Watson. The dynamics of social practice. Sage, 2012. (Cité en page 131.)

[Sidner 2003] Candace L Sidner et Christopher Lee. *Engagement rules for human-robot collaborative interactions.* Dans International Conference on Systems, Man and Cybernetics (SMC), volume 4, pages 3957–3962. IEEE, 2003. (Cité en pages 50 et 75.)

[Sidner 2005] Candace L. Sidner, Christopher Lee, Cory D. Kidd, Neal Lesh et Charles Rich. *Explorations in engagement for humans and robots.* Artificial Intelligence, vol. 166, no. 1, pages 140–164, 2005. (Cité en page 75.)

[Silva 2015] Lavindra de Silva, Raphaël Lallement et Rachid Alami. *The HATP hierarchical planner : Formalisation and an initial study of its usability and practicality.* Dans IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015), Hamburg, Germany, septembre 2015. (Cité en pages 51 et 127.)

[Singamaneni 2020a] Phani-Teja Singamaneni et Rachid Alami. *HATEB-2 : Reactive Planning and Decision making in Human-Robot Co-navigation.* Dans International Conference on Robot & Human Interactive Communication, 2020, Online, Italy, août 2020. (Cité en page 118.)

[Singamaneni 2020b] Phani-Teja Singamaneni, Amandine Mayima, Guillaume Sarthou, Yoan Sallami, Guilhem Buisan, Kathleen Belhassein, Jules Waldhart et Aurélie Clodic. *Guiding Task through Route Description in the MuMMER Project.* Dans HRI '20 : ACM/IEEE International Conference on Human-Robot Interaction, pages 643–643, Cambridge, United Kingdom, mars 2020. ACM. (Cité en page 122.)

[Siposova 2018] Barbora Siposova, Michael Tomasello et Malinda Carpenter. *Communicative eye contact signals a commitment to cooperate for young children.* Cognition, vol. 179, pages 192–201, 2018. (Cité en page 74.)

[Sirin 2007] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur et Yarden Katz. *Pellet : A practical owl-dl reasoner.* Journal of Web Semantics, vol. 5, no. 2, pages 51–53, 2007. (Cité en page 38.)

[Sisbot 2005] Emrah Akin Sisbot, Rachid Alami, Thierry Simeon, Kerstin Dautenhahn, Michael Walters et Sarah Woods. *Navigation in the presence of humans.* Dans 5th IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan, décembre 2005. (Cité en page 85.)

[Sisbot 2007] Emrah Akin Sisbot, Luis Félipé Marin-Urias, Rachid Alami et Thierry Simeon. *A Human Aware Mobile Robot Motion Planner.* IEEE Transactions on Robotics, vol. 23, no. 5, pages 874 – 883, octobre 2007. (Cité en pages 23 et 24.)

[Sisbot 2010] Emrah Akin Sisbot, Luis Marin-Urias, Xavier Broquère, Daniel Sidobre et Rachid Alami. *Synthesizing Robot Motions Adapted to Human Presence.* International Journal of Social Robotics, vol. 2, no. 3, pages 329 – 343, septembre 2010. (Cité en page 24.)

[Sisbot 2012] Emrah Akin Sisbot et Rachid Alami. *A Human-Aware Manipulation Planner.* IEEE Transactions on Robotics, vol. 28, no. 5, pages 1045 – 1057, octobre 2012. (Cité en pages 24 et 127.)

[Stenmark 2013] Maj Stenmark et Jacek Malec. *Knowledge-Based Industrial Robotics.* Dans SCAI, pages 265–274, 2013. (Cité en page 38.)

[Sutton 2018] Richard S. Sutton et Andrew G. Barto. Reinforcement learning : An introduction. The MIT Press, second édition, 2018. (Cité en page 56.)

[Tambe 1997] M. Tambe. *Towards Flexible Teamwork.* Journal of Artificial Intelligence Research, vol. 7, pages 83–124, septembre 1997. (Cité en page 8.)

[Tenorth 2013] Moritz Tenorth et Michael Beetz. *KnowRob : A knowledge processing infrastructure for cognition-enabled robots.* The International Journal of Robotics Research, vol. 32, no. 5, pages 566–590, 2013. (Cité en page 38.)

[Tenorth 2017] Moritz Tenorth et Michael Beetz. *Representations for robot knowledge in the KnowRob framework.* Artificial Intelligence, vol. 247, pages 151–169, 2017. Special Issue on AI and Robotics. (Cité en page 37.)

[Tomasello 2005] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne et Henrike Moll. *Understanding and sharing intentions : The origins of cultural cognition.* Behavioral and Brain Sciences, vol. 28, no. 5, page 675–691, 2005. (Cité en pages 11, 13 et 23.)

[Varadarajan 2011] K.M. Varadarajan et M. Vincze. *Ontological Knowledge Management Framework for Grasping and Manipulation.* Dans IROS Workshop : Knowledge Representation for Autonomous Robots, 2011. (Cité en page 33.)

[Veloso 2018] Manuela M. Veloso. *The Increasingly Fascinating Opportunity for Human-Robot Interaction : The CoBot Mobile Service Robots.* ACM Transactions on Human-Robot Interaction, May 2018. (Cité en page 128.)

[Vesper 2010] Cordula Vesper, Stephen Butterfill, Günther Knoblich et Natalie Sebanz. *A minimal architecture for joint action.* Neural Networks, vol. 23, no. 8, pages 998–1003, 2010. Social Cognition : From Babies to Robots. (Cité en page 24.)

[Vesper 2017] Cordula Vesper, Ekaterina Abramova, Judith Bütepage, Francesca Ciardo, Benjamin Crossey, Alfred Effenberg, Dayana Hristova, April Karlinsky, Luke McEllin, Sari R. R. Nijssen, Laura Schmitz et Basil Wahn. *Joint Action : Mental Representations, Shared Information and General Mechanisms for Coordinating with Others.* Frontiers in Psychology, vol. 07, janvier 2017. (Cité en page 130.)

[Waldhart 2018] Jules Waldhart, Aurélie Clodic et Rachid Alami. *Planning Human and Robot Placements for Shared Visual Perspective.* Dans Workshop on Robotic Coworkers 4.0 : Human Safety and Comfort in Human-Robot Interactive Social Environments 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, Madrid, Spain, octobre 2018. (Cité en page 112.)

[Waldhart 2019] Jules Waldhart, Aurélie Clodic et Rachid Alami. *Reasoning on Shared Visual Perspective to Improve Route Directions.* Dans 2019 28th IEEE International Conference on Robot & Human Interactive Communication, New Delhi, India, octobre 2019. (Cité en page 112.)

[Wallace 1998] R Wallace. Responsibility and the moral sentiments. Harvard University Press, Cambridge, Mass, 1998. (Cité en page 74.)

[Warnier 2012] Mathieu Warnier, Julien Guitton, Séverin Lemaignan et Rachid Alami. *When the robot puts itself in your shoes. Managing and exploiting human and robot beliefs.* Dans IEEE RO-MAN : The 21st IEEE International Symposium on Robot and Human Interactive Communication, page 7p., Paris, France, septembre 2012. (Cité en page 19.)

[Weiss 2011] Astrid Weiss, Regina Bernhaupt et Manfred Tscheligi. *The USUS evaluation framework for user-centered HRI.* Dans New Frontiers in Human–Robot Interaction, pages 89–110. John Benjamins, 2011. (Cité en pages 86 et 87.)

[Wielemaker 2003] Jan Wielemaker, A Th Schreiber et BJ Wielinga. *Prolog-based infrastructure for RDF : performance and scalability.* 2003. (Cité en page 38.)

[Young 2010] James E. Young, JaYoung Sung, Amy Voida, Ehud Sharlin, Takeo Igarashi, Henrik I. Christensen et Rebecca E. Grinter. *Evaluating Human-Robot Interaction.* International Journal of Social Robotics, vol. 3, no. 1, pages 53–67, octobre 2010. (Cité en pages 69 et 86.)

[Zawidzki 2013] Tadeusz Zawidzki. Mindshaping : a new framework for understanding human social cognition. MIT Press, Cambridge, MA, 2013. (Cité en page 72.)