HAL
open science

Interaction affective et expressive. Compagnon artificiel-humain

Jérémy Riviere

To cite this version:

Jérémy Riviere. Interaction affective et expressive. Compagnon artificiel-humain. Autre [cs.OH]. Université de Grenoble, 2012. Français. NNT : 2012GRENM045 . tel-00780213

HAL Id: tel-00780213

https://theses.hal.science/tel-00780213v1

Submitted on 23 Jan 2013

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.

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Informatique**

Arrêté ministériel : Août 2006

Présentée par

Jérémy Rivière

Thèse dirigée par **Sylvie Pesty**

préparée au sein **du Laboratoire d'Informatique de Grenoble (LIG)**
et de **L'Ecole Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique (EDMSTII)**

Interaction affective et expressive
Compagnon artificiel - humain

Thèse soutenue publiquement le **10 octobre 2012**,
devant le jury composé de :

M. Dominique Duhaut
Professeur des Universités, Université de Bretagne Sud, Président
M. Jean-Claude Martin
Professeur des Universités, Université de Paris Sud, Rapporteur
M. Nicolas Sabouret
Professeur des Universités, Université de Paris Sud - Supelec, Rapporteur
Mme Gaëlle Calvary
Professeur des Universités, Université de Grenoble, Examinatrice
M. Dominique Longin
Chargé de Recherche CNRS, Institut de Recherche en Informatique de Toulouse, Examinateur
Mme Frédérique Segond
Docteur, Responsable Centre de Recherche et Développement Viseo, Examinatrice
Mme Sylvie Pesty
Professeur des Universités, Université de Grenoble, Directeur de thèse

Résumé

Le but de ce travail de thèse est de rendre sincère un Agent Conversationnel Animé (ACA) pour, d'une part, améliorer sa crédibilité du point de vue de l'humain, et d'autre part contribuer à le rendre acceptable dans une relation privilégiée compagnon artificiel - humain.

La première partie de ce travail consiste en la création d'un Langage de Conversation Multimodal (LCM) pour ACA, constitué d'Actes de Conversation Multimodaux (ACM) tels que `promettre`, `s'excuser` ou `exiger` : ces ACM permettent à l'agent d'être crédible dans sa sincérité, en exprimant multimodalement ses états mentaux de façon appropriée. Les états mentaux portés par les ACM sont formalisés en logique : la volonté de représenter des états mentaux issus de raisonnements complexes (basés d'une part sur le raisonnement contrefactuel et d'autre part sur les normes et les buts de l'agent), dont l'expression se fait avant tout par le langage [Oatley 1987], a amené à mettre en place le modèle BIGRE (Beliefs, Ideals, Goals, Responsibility, Emotions). Ce modèle, basé sur une logique de type BDI (*Belief, Desire, Intention*), permet de représenter également des émotions que nous appelons complexes, telles que la réjouissance, la gratitude ou le regret. La catégorie expressive des ACM contient les actes exprimant ces émotions complexes (e.g. `remercier, regretter`) : l'expression se fait par le langage, conjointement aux autres expressions multimodales appropriées. Le LCM est implémenté dans l'ACA Greta, ce qui permet une évaluation de ce langage en termes de crédibilité et de sincérité perçues par l'humain.

La deuxième partie de ce travail porte sur les capacités de raisonnement de l'ACA : dans le but de permettre à l'agent de raisonner dans le dialogue, c'est-à-dire mettre à jour ses états mentaux et ses émotions et sélectionner son intention communicative, un moteur de raisonnement a été mis en place. Ce moteur de raisonnement est basé sur le cycle de comportement BDI - Perception, Décision, Action - et les opérateurs du modèle BIGRE, permettant ainsi la manipulation d'états mentaux issus de raisonnements complexes (dont les émotions complexes). Les ACM qui composent notre langage sont intégrés dans le moteur, et sont utilisés pour atteindre l'intention communicative de l'ACA : par exemple, si l'agent a l'intention d'exprimer sa gratitude, il construit un plan pour satisfaire son intention, formé des ACM `remercier` ou `féliciter`, selon le degré de l'émotion. Un type d'intention communicative, déclenché par des règles d'obligation du discours, participe à la régulation locale du dialogue. L'ACA étant de plus affectif, sa sincérité l'amène à exprimer toutes ses émotions. La généricité de ce moteur de raisonnement permet de l'implémenter dans l'ACA Greta (où il est en lien avec le LCM) et dans l'agent MARC. L'expression multimodale des ACM avec l'agent MARC a été rendue possible par l'intégration des checks de Scherer dans le moteur de raisonnement que nous avons adapté au contexte du dialogue. Une évaluation du moteur de raisonnement avec l'agent MARC montre que les états mentaux déduits par le moteur sont appropriés à la situation, et que leur expression (l'expression de la sincérité de l'agent) est également appropriée.

Abstract

The aim of this thesis is to make an Embodied Conversational Agent (ECA) sincere in order to, on one hand, improve its believability from the human's point of view, and on the other hand make it acceptable in an affective relationship between an artificial companion and a human.

The first part of this work consists in creating a Multimodal Conversational Language (MCL) for an ECA, made up of Multimodal Conversational Acts (MCA) such as `promise`, `apologise` or `demand`. These MCA allow the agent to appear believable in its sincerity, by multimodally expressing its mental states in an appropriate way. The mental states carried by the MCA are formalised in logics : our will to represent mental states stemming from complex forms of reasoning (based on counterfactual reasoning or on the agent's norms and goals) that are mainly expressed via language [Oatley 1987] led us to design the BIGRE model (Beliefs, Ideals, Goals, Responsibility, Emotions). This model, based on a BDI-like logic (*Belief, Desire, Intention*), allows us to also represent some particular emotions that we call *complex* emotions, such as rejoicing, gratitude or regret. The expressive category of MCA contains acts that express these complex emotions (e.g. `thank`, `regret`) : they are expressed by the language, along with the other appropriate multimodal expressions. We implemented the MCL in the ECA Greta, which enabled an evaluation of this language in terms of believability and sincerity perceived by the human.

The second part of this work is about the ECA's reasoning capabilities : in order to allow the agent to reason in the dialogue, that is to update its mental states and its emotions and select its communicative intention, we designed a reasoning engine. This reasoning engine is based on the BDI behaviour cycle of Perception - Decision - Action and on the operators from the BIGRE model, thus enabling the manipulation of mental states resulting from complex reasoning (including complex emotions). The MCA in our language are part of our reasoning engine, and are used to achieve the ECA's communicative intention : for example if the ECA intends to express its gratitude, it builds a plan to achieve this intention, that consists of the MCA `thank` or `congratulate`, depending on the intensity of the emotion. One type of communicative intention, triggered by discourse obligations rules, participates in the local regulation of dialogue. Further, since the ECA is emotional, its sincerity brings it to express all its emotions. The generic character of this reasoning engine allowed us to implement it in the ECA Greta (where it is linked with the MCL) as well as in the agent MARC. The multimodal expression of MCA by the agent MARC was made possible by integrating Scherer's checks in the reasoning engine that we adapted to the context of dialogue. An evaluation of the reasoning engine with the agent MARC shows that the mental states deduced by the engine are appropriate to the situation, and that their expression (the expression of the agent's sincerity) is also appropriate.

Remerciements

J'aimerais commencer ici ce périlleux exercice que sont les remerciements par une remarque : je ne remercierais jamais assez (et surtout pas dans les lignes qui suivent) tous ceux que j'aimerais remercier et grâce à qui j'en suis là aujourd'hui - soyez-en prévenus ! Mais puisqu'il faut le faire, allez, je m'y colle :

Merci tout d'abord aux rapporteurs d'avoir eu le courage et la patience d'évaluer ce manuscrit, et merci aux membres du jury d'avoir accepté d'évaluer ma soutenance, qui j'espère leur plaira !

Je voudrais ensuite remercier Sylvie pour m'avoir donné la chance de commencer cette thèse, pour m'avoir pour m'avoir poussé tout le long et m'avoir considéré comme un interlocuteur valable (avec des idées bonnes et - le plus souvent malheureusement ! - moins bonnes) en recherche. Je pense qu'une grande partie de ce travail de thèse vient directement de la manière dont tu m'as encadré, et je t'en suis très reconnaissant. Je remercie également Carole pour son aide précieuse dans le développement du moteur de raisonnement, dans la formalisation des actes, en anglais, entre autres ! Je voterai pour toi lorsqu'il faudra encadrer un prochain thésard ! Merci aussi à toute l'équipe toujours prête à se serrer les coudes lors des passages difficiles - merci notamment à Cyrille, c'est chouette de ne pas être tout seul à rédiger son manuscrit les longues journées d'été. Et merci à mes compagnons de resto u qui améliorent magiquement les repas du midi.

A un niveau plus personnel, je remercie mes parents d'avoir toujours cru en moi - même (et surtout) quand j'ai eu mon bac au rattrapage, et quand j'ai redoublé l'année suivante ma 1ère année à l'université ![1]. Alors un grand grand merci, quoi qu'il arrive je reste votre fils et c'est bien de se savoir aimé. Merci évidemment à l'ensemble de ma famille, Bastien, mes grands-parents, qui s'agrandit toujours - on ne touche pas à la famille (accent italien).

Merci également à tous mes copains de Toulouse (Jocelyn, Ramou, Yannick, Noélie, Fred et tous les autres) et de Béziers (Max, pour ne citer que toi) de garder le contact depuis mon départ "en exil dans la montagne" ; c'est chouette de se revoir à intervalles (ir)réguliers pour échanger des nouvelles, débattre de politique, se coucher très tard ... Et merci aux plus si nouveaux copains de Grenoble (Laure, Tof, Guillaume et tous les autres !) de me supporter dans la semaine, de me changer les idées, d'être aussi en thèse, d'en sortir ou de vouloir y entrer ...

Et enfin, un énorme merci à toi Mathilde d'être toujours là, de me connaitre aussi bien et de me supporter (dans tous les sens du terme) depuis 6 ans ! Tu sais mieux que je ne pourrais l'écrire ici à quel point tu es importante pour moi - alors j'arrête ici. Merci à ceux que j'ai (immanquablement) oublié et qui se reconnaitront !

[1] De fait, je profite ici de cette espace pour tirer la langue à tous ceux qui voudraient - comme ça a été discuté à une époque - imposer une moyenne minimale pour rentrer à l'université ; avec eux je n'aurais pas eu la chance d'aller au bout de mes études.

Table des matières

I Introduction générale **1**

1 Contexte **3**
1.1 Le personnage virtuel : une nouvelle (mais ancienne) forme d'interaction 4
1.2 Le personnage virtuel comme compagnon artificiel 5
1.3 Réalisme et crédibilité des personnages virtuels 6

2 Problématique **9**
2.1 Proposition : un agent expressif, affectif et sincère 9
2.2 Objectifs 11
2.3 Démarche et contributions 11
2.4 Evaluations 12
2.5 Le projet CECIL 12
2.6 Organisation du manuscrit 12

II Etat de l'art : Agents Conversationnels Animés, langage et émotion **15**

3 Théories psychologiques des émotions **19**
3.1 Théories évolutionnistes des émotions 19
3.1.1 La théorie des émotions discrètes d'Ekman 19
3.1.2 La théorie d'Oatley et Johnson-Laird 20
3.1.3 Les théories dimensionnelles 21
3.2 Comment se déclenche l'émotion ? Les théories de l'évaluation cognitive. 21
3.2.1 La théorie d'évaluation cognitive de Lazarus 23
3.2.2 La théorie d'évaluation cognitive de Scherer 23
3.2.3 La théorie d'évaluation cognitive OCC 24
3.2.4 Les théories sociales de l'évaluation cognitive 26

4 ACA et émotions **27**
4.1 Expressions multimodales chez les ACA 27
4.1.1 Expressions faciales et regard 27
4.1.2 Gestes 28
4.1.3 Autres expressions multimodales 29
4.2 ACA affectifs et expressifs 29
4.2.1 MARC : Multimodal Affective and Reactive Characters . 29
4.2.2 EMA : EMotion and Adaptation 31
4.2.3 CrossTalk : un exemple de "scripting" de scène 32
4.3 ACA : émotions et états mentaux BDI 33
4.3.1 Max et l'architecture WASABI 33
4.3.2 FearNot ! 34
4.3.3 Un agent rationnel empathique 35

5 ACA, états mentaux et langage **37**
5.1 Bibliothèque de comportements : traduction des états mentaux . 38
5.1.1 Le standard SAIBA 38
5.1.2 Expressions faciales performatives 40
5.1.3 Rétroactions (*Backchannels*) 40
5.2 Modèles de dialogue chez les ACA 42
5.2.1 FearNot ! 42
5.2.2 Le projet MRE 43

6 Conclusions de l'état de l'art **45**
6.1 Résumé 45
6.2 Conclusion 47

III Langage de Conversation Multimodal **49**

7 Appuis théoriques **53**
7.1 La théorie des Actes de Discours 53
7.1.1 Formes des actes illocutoires 53
7.1.2 Verbes illocutoires 54
7.1.3 Travaux connexes 55
7.2 Les états mentaux et les émotions dans la théorie des Actes de Discours 56
7.2.1 La condition de sincérité 56
7.2.2 Le mode d'accomplissement 58
7.2.3 Le degré de puissance 58

8 Formalisation logique des émotions et des actes expressifs **59**
8.1 Définition des émotions complexes 59
8.1.1 Responsabilité et raisonnement contrefactuel 59
8.1.2 Formalisation Logique : le modèle BIGRE 60
8.1.3 Axiomes liés à la formalisation 62
8.2 Formalisation des actes expressifs 62

9 Le Langage de Conversation Multimodal **65**
9.1 Introduction du degré de puissance 65
9.2 Extension de la formalisation des actes expressifs aux usages assertifs, engageants et directifs 66
9.3 Préconditions et postconditions des actes 68
9.3.1 Préconditions 68
9.3.2 Effets 68
9.4 Expression multimodale des actes 69

10 Implémentation et évaluation du LCM **73**
10.1 Greta et le standard SAIBA : implémentation du LCM 73
10.2 Evaluation du LCM 75
10.2.1 Protocole 75
10.2.2 Résultats 76

IV Moteur de raisonnement 81

11 Appuis théoriques 85
11.1 Le modèle BDI (*Belief - Desire - Intention*) 85
11.1.1 Emotions et BDI 87
11.2 Théories des plans du discours et BDI 87
11.3 Les "niveaux" de dialogue 88
11.3.1 Obligations du discours 89

12 Le moteur de raisonnement 91
12.1 Le moteur de raisonnement dans le cycle de déclenchement des émotions 91
12.2 Architecture du moteur de raisonnement 92
12.3 Perception et mise à jour des états mentaux 94
12.3.1 Déduction des effets des actes 96
12.3.2 Conflits de croyances 96
12.3.3 "Saturation" de la base 97
12.4 Evaluation : déclenchement des émotions complexes 97
12.5 Délibération : sélection de l'intention communicative 98
12.5.1 L'intention *émotionnelle* 99
12.5.2 L'intention *d'obligations* 99
12.5.3 L'intention *globale* 99
12.5.4 Exception 100
12.6 Planification pour atteindre l'intention 100
12.7 La personnalité de l'agent 101
12.8 Exemples de dialogues 102
12.9 Généricité du moteur de raisonnement 105

13 Implémentation et évaluation 107
13.1 Implémentation 107
13.1.1 Implémentation du moteur de raisonnement en Prolog . . 107
13.1.2 Architecture logicielle Java-Prolog 109
13.1.3 L'agent GRETA et le LCM 110
13.1.4 L'agent MARC et les *checks* de Scherer 112
13.2 Evaluation du moteur de raisonnement 116
13.2.1 Scénario 117
13.2.2 Protocole 120
13.2.3 Résultats 122

V Conclusion et perspectives 135

14 Résumé 137
14.1 Langage de Conversation Multimodal 137
14.2 Moteur de raisonnement 138
14.3 Conclusion générale 139

15 Perspectives 141
15.1 Perspectives à court terme 141
15.2 Perspectives à long terme 142

VI Annexes **145**

A Langage de Conversation Multimodal **147**
A.1 Version française 147
A.2 Version anglaise 147
A.3 Formalisation des préconditions et effets des ACM 148

B Module de raisonnement **155**
B.1 Guide d'utilisation 155
B.2 Exemples de scénario : organisation des vacances 155
B.2.1 Base de connaissances 155
B.2.2 Listes des ACM et contenus propositionnels 156
B.2.3 Exemple de dialogue dans le scénario 156
B.3 Protocole d'évaluation du module 157
B.4 Résultats détaillés de l'évaluation 158

C Publications **165**

Première partie

Introduction générale

1

Contexte

« Nous sommes encore loin d'avoir besoin de *Blade Runner* ! », me disait l'autre jour un ami en rigolant, devant le personnage virtuel présent sur un site de vente en ligne de meubles suédois. Il faisait bien sûr référence au célèbre film de science-fiction réalisé par Ridley Scott[1], dans lequel la police entraîne une unité spéciale (les *Blade Runner*) pour différencier les humains des *réplicants*, êtres artificiels génétiquement conçus. De nos jours, les personnages virtuels tendent néanmoins vers l'équivalent informatique des *réplicants*, malgré leur manque de crédibilité. En effet, on a pu observer ces dernières années un intérêt grandissant pour les personnages virtuels.

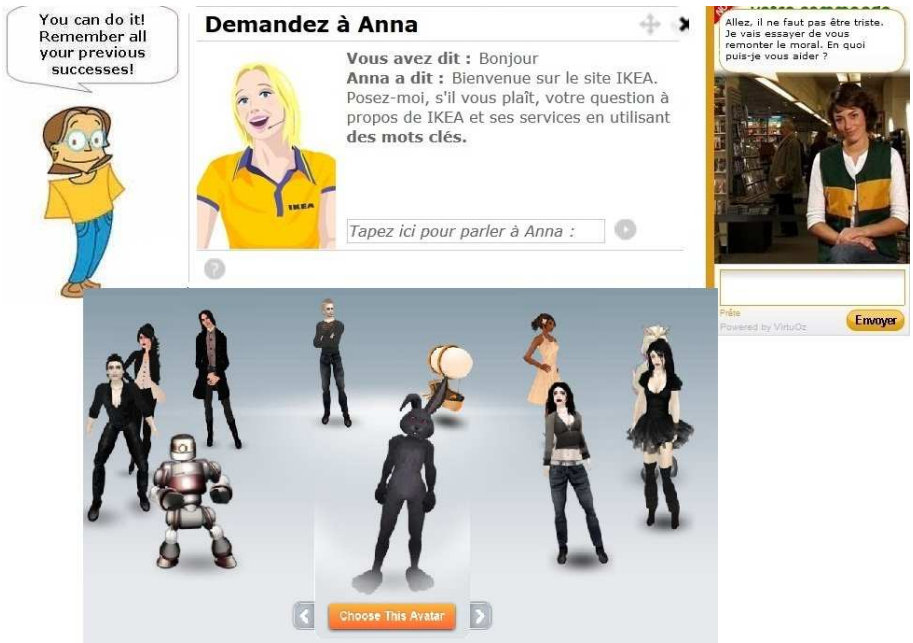

FIG. 1.1 – De gauche à droite : agent pédagogique [Jaques 2009], agents conseillers de sites (*Ikea*,*Fnac*) ; en-dessous, avatars de Second-Life (*http ://secondlife.com/*).

[1]grandement inspiré du roman *Les androïdes rêvent-ils de moutons électriques ?* de Philip K. Dick

Ils sont présents en ligne (sites de vente, communautés d'utilisateurs - Second Life...), dans les jeux vidéos, et occupent de plus en plus de place dans les systèmes interactifs consacrés à l'éducation ou à l'information.

1.1 Le personnage virtuel : une nouvelle (mais ancienne) forme d'interaction

Cet intérêt pour les personnages virtuels trouve sa source dans 2 phénomènes : d'une part la démocratisation de l'informatique, et d'autre part l'évolution de l'utilisation de l'informatique. Au 3ème semestre 2011, 73% des français possèdent un ordinateur[2], contre 34% en 2000. Ce sont donc de plus en plus d'utilisateurs différents aux compétences variées qui utilisent des dispositifs d'interaction traditionnels (souris d'ordinateur, clavier) comme des dispositifs plus récents (écran tactile, souris 3D...). Néanmoins, il existe une forme d'interaction qui nous est naturelle et que nous utilisons tous les jours : la communication humaine.
Faire évoluer les dispositifs d'interaction vers des dispositifs plus intuitifs et naturels est donc une nécessité pour des interactions plus efficaces entre machines et usagers. Cette nécessité est d'autant plus forte que les moyens d'interaction traditionnels ne sont plus optimaux pour des usages de moins en moins "bureautique" de l'informatique : jeux vidéos, réseaux sociaux, réalité augmentée, logiciels éducatifs...
Le personnage virtuel est donc né de la nécessité d'incarner un nouveau (mais ancien) moyen d'interaction, la communication humaine. Ces personnages, lorsqu'ils intègrent toutes les modalités de communication de l'interaction humaine, sont définis comme des Agents Conversationnels Animés (**ACA** ou ECA- *Embodied Conversational Agent* en anglais) [Poggi 2005, Courgeon 2008].

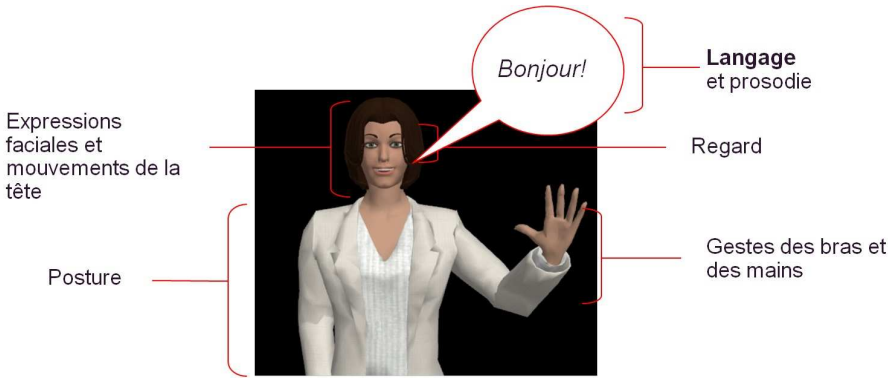

Fig. 1.2 – L'Agent Conversationnel Animé (ACA) : un modèle de la communication humaine, verbale et non-verbale

[2]Enquête GfK / Médiamétrie - Référence des Equipements Multimédias

1.2 Le personnage virtuel comme compagnon artificiel

Le concept de *compagnon artificiel* est apparu récemment : il peut être défini [Pesty 2012] comme un système interactif intelligent destiné à entretenir une relation à long terme et privilégiée avec l'usager, par exemple en dialoguant, en lui offrant divers services ou encore en jouant. La capacité de tels systèmes à établir une relation avec l'usager est de plus en plus considérée comme un facteur clé pour une amélioration significative des interactions, en termes d'intuitivité, de naturel et d'efficacité. Pour parvenir à mettre en place une telle relation affective entre un utilisateur et le personnage virtuel, ce dernier doit d'abord satisfaire aux exigences des Interfaces Homme-Machines (IHM) traditionnelles, telles que l'ergonomie, la rapidité et la fiabilité.

Il doit de plus répondre aux attentes importantes que son apparence humaine engendre chez les utilisateurs. En effet, les personnages virtuels sont de plus en plus **réalistes**, en termes de rendu de leur apparence et d'animation, du fait notamment des progrès en informatique graphique (cf. figure 1.3) ; ce même phénomène se produit également avec les robots humanoïdes.

FIG. 1.3 – De gauche à droite, les ACA Steve [Rickel 1997], Greta [Poggi 2005] et MARC [Courgeon 2008, Courgeon n C].

Ce réalisme engendre des attentes importantes chez les utilisateurs : ils prêtent à ces personnages de grandes capacités de raisonnement (intentions, états mentaux...) et les extrapolent pour prévoir leur **comportement** [Lee 2005]. Lorsque ces attentes ne correspondent pas au comportement réel du personnage (ce qu'on peut qualifier de **dissonance**), il se produit chez les utilisateurs un phénomène de rejet ou de répulsion à l'égard de ces personnages : ce phénomène perceptif est connu sous le nom d'*uncanny valley* [Mori 1970]. L'*uncanny valley* peut également intervenir au sein du réalisme du personnage ou entre les différentes modalités d'expression : si le rendu de son apparence est supérieur en qualité à son animation (dynamique des gestes, expressions faciales etc.), cela résulte en un sentiment négatif chez les utilisateurs. Un utilisateur ne pourra donc pas construire de relation affective durable avec le compagnon artificiel tant qu'il y aura cette dissonance entre ses attentes, issues du réalisme de l'agent, et le comportement de cet agent. En effet, d'après [Dautenhahn 1997], « *interactions [...] with humans create a need for all these [agents] to behave "naturally", i.e. in a way which is acceptable and comfortable to humans, so that the human user or collaborator can accept artificial agents as companions or "interaction partners".* »[3].

[3]Littéralement : les interactions avec les humains créent une nécessité pour ces agents

1.3 Réalisme et crédibilité des personnages virtuels

Il est important de définir ici ce que nous appelons **réalisme** et **crédibilité** d'un personnage virtuel. Ces deux notions englobent la plupart des travaux dans le domaine des personnages virtuels : le réalisme concerne ce que j'appelle le *comportement esthétique* de l'agent virtuel et la crédibilité son *comportement cognitif*.

Le domaine de l'informatique graphique contribue à améliorer le **réalisme** du personnage virtuel, i.e. son *comportement esthétique* : non seulement le rendu de son apparence physique (e.g. le rendu de la peau, de la chevelure, des rides) mais aussi l'animation de son *comportement cognitif* (ses gestes, ses expressions faciales etc.). Du point de vue de l'humain, la perception du réalisme du personnage influence sa crédibilité, car ces deux notions sont deux facettes du comportement perçu du personnage. Dans la mesure où ma thèse est un travail se situant en intelligence artificielle, nous n'aborderons pas cet aspect ici pour nous concentrer sur la crédibilité de l'agent.

La notion de **crédibilité** est liée au *comportement cognitif* du personnage virtuel. Ce comportement est vu, selon les travaux, comme un comportement émotionnel [Bates 1994], un comportement social [Prendinger 2001], ou comme l'ensemble des actes du personnage [de Rosis 2003], et peut également prendre en compte sa personnalité [McRorie 2009]. Malgré cette diversité, on peut identifier des caractéristiques communes dans la définition de la crédibilité [Demeure 2011] ; un personnage virtuel est *crédible* lorsqu'il montre :

1. un comportement adéquat (plausible et approprié étant donnée la situation) ;
2. la capacité d'exprimer multimodalement ses émotions ;
3. et la capacité de prendre en compte son environnement social (raisonnement, prise de décision...).

Les sciences sociales (la psychologie, la psychologie sociale) et l'intelligence artificielle sont les domaines de recherche qui concernent l'aspect cognitif du comportement des agents, en intégrant les émotions, la communication (verbale et non-verbale) et le raisonnement. Dans le but d'améliorer la crédibilité des personnages virtuels, et donc de réduire la distance entre leur *comportement cognitif* et les attentes des utilisateurs, la majorité des travaux sur les personnages virtuels se sont intéressés principalement au rôle de l'émotion dans le comportement.

Le rôle central de l'émotion dans la communication, la cognition et plus généralement dans les relations sociales a été montré dans de nombreux travaux tant en psychologie qu'en informatique [Bates 1994, Piaget 1989]. Picard a souligné plus précisément la présence constante des émotions chez l'utilisateur lors d'interactions avec la machine [Picard 1997]. De ce fait, un grand nombre de personnages virtuels sont dotés de modèles cognitifs des émotions leur permettant de prendre en compte des variables de leur environnement pour en déduire

d'avoir un comportement "naturel", i.e. qui soit acceptable et confortable pour l'humain ; de cette manière, l'utilisateur humain peut accepter les agents artificiels comme compagnons ou "partenaires d'interaction"

leurs propres émotions et ensuite les exprimer ([Courgeon 2008, Paiva 2006, Kopp 2003], cf. chapitre 4). Les interactions entre l'humain et le personnage virtuel s'en trouvent améliorées, tant en termes de crédibilité que de performance, c'est-à-dire de succès de la communication. D'autres recherches se sont orientées vers la prise en compte des émotions dans le raisonnement des personnages, et s'intéressent à leur influence sur leurs croyances et leur comportement ([Adam 2009b, Pereira 2008], cf. chapitre 11).

Rendre **explicite** les émotions des personnages virtuels (de manière réaliste, i.e. avec les expressions appropriées) améliore donc leur crédibilité. Ce résultat est cohérent avec d'autres travaux [Dautenhahn 1997, Breazeal 1999], qui affirment que l'interaction sociale entre un humain et un personnage virtuel (ou un robot humanoïde) n'est possible que si le personnage exprime **ses états mentaux** : il doit faire *croire* à l'humain qu'il a des croyances, des désirs et des intentions, et cela passe par l'expression de ses émotions. Cependant, exprimer seulement les émotions du personnage virtuel suffit-il pour réduire l'*uncanny valley*, et mettre en place une relation affective et durable avec l'humain ?

2
Problématique

La problématique de ce travail de recherche est la suivante : le phénomène perceptif d'*uncanny valley* - une dissonance entre les attentes des utilisateurs, issues du **réalisme** du personnage, et son comportement (sa **crédibilité**) - peut provoquer un sentiment de déception et de rejet chez l'humain ; ce sentiment rend difficile la mise en place d'une relation affective et durable entre un compagnon artificiel et lui. Dans le but de réduire cet effet, certains travaux ont choisi de baisser le réalisme du personnage, réduisant ainsi les attentes des utilisateurs (comme c'est la cas dans certains films d'animation ou jeux vidéos par exemple) ; d'autres travaux (la majorité) ont choisi de s'inspirer des théories des émotions en psychologie, et rendre ainsi les personnages virtuels capables d'exprimer leurs émotions. Lorsque ces personnages expriment leurs émotions (essentiellement de manière non verbale, comme nous le verrons dans le chapitre 4), ces émotions deviennent **explicites** pour l'utilisateur, et leur comportement est de ce fait jugé plus crédible par l'humain.

Cependant, les émotions ne sont pas les seuls états mentaux qu'un individu exprime lorsqu'il communique avec d'autres individus [Searle 1983] : on peut vouloir communiquer ses croyances, son intention de faire quelque chose etc. Dans le but d'augmenter la crédibilité des ACA, et ainsi réduire encore la dissonance entre attentes et comportement, il leur faut donc rendre **explicite** l'ensemble de leurs états mentaux, et pas seulement leurs émotions. La proposition que nous faisons consiste à rendre **sincère** l'ACA, et à ce qu'il soit crédible dans l'expression de sa sincérité : un ACA **expressif, affectif et sincère**, capable de raisonner dans un environnement social, exprime donc l'ensemble de ses états mentaux de manière crédible.

2.1 Proposition : un agent expressif, affectif et sincère

La *sincérité* d'un ACA est liée aux états mentaux de l'agent : le dictionnaire Atilf[1] définit la sincérité comme étant « *la qualité d'une personne sincère, qui exprime des sentiments réellement éprouvés, qui ne cache pas ses pensées* ». Un ACA expressif sincère exprime donc l'ensemble de ses états mentaux ; lorsqu'il est également affectif, i.e. capable de "ressentir" des émotions, il exprime également ses émotions "ressenties".

[1] Analyse et Traitement Informatique de la Langue Française, http ://atilf.atilf.fr/

Nous faisons l'hypothèse que l'apport de la sincérité dans le raisonnement et dans l'expression des ACA est double :

1. L'expression de l'ensemble des états mentaux de l'ACA sincère, de manière crédible, rend son comportement explicite et transparent pour l'humain, et permet de réduire ainsi l'effet de dissonance entre son comportement et les attentes de l'humain ;
2. La sincérité étant un élément important dans le jugement affectif [Fiske 2007, Rosenberg 1968], un ACA sincère est plus crédible dans un rôle de compagnon artificiel qu'un ACA insincère, et plus à même de mettre en place une relation affective et à long terme.

Ce deuxième point est en accord avec plusieurs travaux en psychologie [Fiske 2007, Rosenberg 1968], qui affirment qu'un individu perçoit socialement un autre individu suivant deux dimensions distinctes (cf. figure 2.1) : la dimension sociale (également appelée cordialité, *warmth* en anglais) et la dimension de compétence (*competence* en anglais). La dimension sociale contient des caractéristiques qui sont liées à la perception de l'intention, dont la gentillesse (*friendliness*), l'amabilité (*helpfulness*), la **sincérité**, la fiabilité (*trustworthiness*) et la moralité ; la dimension de compétence contient pour sa part des caractéristiques qui sont liées à la perception des capacités, dont l'intelligence, la compétence (*skill*), la créativité et l'efficacité.

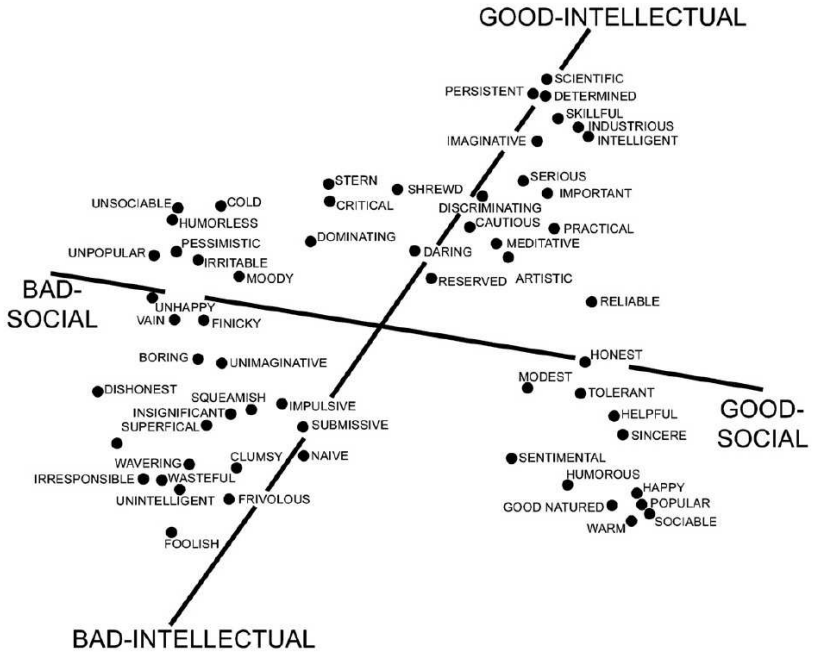

Fig. 2.1 – Les deux dimensions de compétence (*good-intellectual / bad-intellectual*) et sociale (*good-social / bad-social*) introduites par [Rosenberg 1968]. La **sincérité** est une des caractéristiques positives de la dimension sociale.

Récemment, il a été montré que les ACA étaient également perçus socialement selon ces deux dimensions par les utilisateurs [Niewiadomski 2010, Demeure 2011] ;

il a de plus été montré que ces deux dimensions étaient liées à la crédibilité de l'agent, sans que soient précisées dans quelles mesures les différentes caractéristiques qui composent ces dimensions influencent ce jugement.

2.2 Objectifs

Rendre un ACA sincère, c'est-à-dire qui exprime tous ses états mentaux, nous amène à poursuivre deux objectifs principaux.

Premièrement, il faut que l'ACA **soit crédible dans l'expression de sa sincérité** du point de vue de l'humain : si, par exemple, l'ACA regrette sincèrement quelque chose et qu'il s'excuse avec un grand sourire, il n'apparaitra pas très crédible dans sa sincérité pour l'humain. L'ACA doit donc pouvoir exprimer l'ensemble de ses états mentaux de manière crédible. Comme on le verra dans le chapitre 3, un certain nombre d'émotions (e.g. joie, tristesse, dégoût) s'expriment principalement de manière non-verbale (e.g. expressions faciales) ; cependant, lorsqu'il s'agit d'émotions plus complexes (e.g. regret, culpabilité, gratitude) ou d'autres états mentaux (e.g. une intention, un désir, des croyances), la modalité principale utilisée pour leur expression reste le langage [Searle 1983, Oatley 1987], lié aux autres modalités. Pour que l'ACA puisse être crédible dans sa sincérité, un des objectifs de ma thèse consiste donc à examiner les liens qui existent entre le langage, les états mentaux et les autres modalités.

Deuxièmement, il faut que l'ACA **ait des états mentaux** et qu'il puisse les déduire de ses interactions avec son environnement social : il faut donc lui permettre de raisonner dans cet environnement, mettre à jour ses états mentaux (**dont ses émotions**) et décider de son *intention communicative*, i.e. les états mentaux qu'il va vouloir exprimer dans le contexte de la situation.
De notre proposition et de nos objectifs découle un certain nombre de verrous scientifiques auxquels il est primordial d'apporter des solutions. La façon de définir et de modéliser les émotions (sur quelles théories psychologiques se base-t-on ?), la nature de la relation qui existe entre langage et émotions et l'influence de ces émotions, liées à la sincérité de l'agent, sur son comportement cognitif, verbal et non-verbal sont les principaux points auxquels j'ai tenté de répondre.

2.3 Démarche et contributions

Pour mener à bien ces deux objectifs principaux, nous proposons de 1/ lier fortement les états mentaux (dont les émotions) de l'agent et sa façon de communiquer **dans son ensemble** et 2/ donner les moyens à l'agent de raisonner (pour le dialogue) à partir de ce lien de sincérité.

Premièrement, il s'agit d'identifier les liens qui existent entre langage et états mentaux dans la littérature (théories philosophiques du langage, théories psychologiques des émotions). Cette recherche a permis de définir une certaine catégorie d'émotions fortement liées au langage : je qualifierai ici ces émotions de *complexes*. Ces liens ont été exploités dans le but de spécifier et d'implémenter un Langage de Conversation générique pour ACA (à la manière des ACL[2]) liant langage, états mentaux et expression multimodale. Ainsi, dans le cadre d'un

[2]Langage de Communication pour Agents dans le domaine des systèmes multi-agents, le plus connu étant ACL-FIPA (http ://www.fipa.org/index.html)

dialogue ACA - humain, ce **Langage de Conversation Multimodal** (LCM) permet à un agent d'être *crédible* dans sa *sincérité*, c'est-à-dire de s'excuser, par exemple, en exprimant multimodalement une émotion de regret.

Deuxièmement, il s'agit de s'intéresser au comportement cognitif de l'agent, et plus particulièrement à sa façon de raisonner dans le dialogue. J'ai donc construit un **moteur de raisonnement** de l'agent, basé sur ses croyances et sur ses émotions, pour déduire son intention communicative en utilisant ce Langage de Conversation Multimodal.

2.4 Evaluations

Il ne s'agit pas ici d'évaluer le réalisme de l'animation et du rendu visuel de l'agent. En effet, mon travail se situe dans le champ de l'intelligence artificielle et n'entend pas améliorer ces points, appartenant à un champ de recherche différent (l'informatique graphique). Il s'agit donc de mesurer comparativement ce qu'apportent, en termes de crédibilité et de sincérité ressenties, le Langage de Conversation Multimodal (LCM) et le moteur de raisonnement à des personnages virtuels existants.
Deux évaluations distinctes sont nécessaires pour mesurer d'un côté l'apport du LCM en termes de crédibilité et de sincérité du point de vue de l'usager, et mesurer de l'autre la crédibilité de l'expressivité du dialogue déduit par le moteur de raisonnement. Il sera également important d'évaluer si l'ensemble du système améliore, selon notre ojectif de départ, la nature affective de la relation entre l'agent et l'utilisateur.

2.5 Le projet CECIL

Ma thèse fait partie du projet ANR CECIL[3], regroupant 3 partenaires : l'équipe LiLaC de l'IRIT, l'équipe MAGMA du LIG et le LTCI. Une partie du travail sur les émotions complexes que je vais présenter ici a été menée par l'équipe LiLaC ; lorsque c'est le cas, ce sera précisé dans le manuscrit. Les partenaires du LTCI ont fourni quant à eux les expressions multimodales des émotions *complexes*, définies et évaluées par des sujets.

2.6 Organisation du manuscrit

Il est nécessaire avant tout de situer mon travail dans un contexte global de recherche. La partie suivante (partie II) fait un état des lieux de la recherche sur les agents virtuels et en particulier sur la place qu'occupent l'expression des états mentaux chez ces agents. Quel est l'état de l'art actuel des ACA ? Quelle est la place des émotions dans les différents travaux ? Comment se décident et s'expriment leurs états mentaux ?

La suite du mémoire, de la même façon que le travail effectué au cours de ma thèse, se divise en deux parties : une partie sur le LCM (partie III) et une partie sur le moteur de raisonnement (partie IV). Chacune de ces parties rappelle tout d'abord les fondements théoriques de la proposition, puis expose la proposition

[3] *Complex Emotion in Communication, Intercation and Language*, http ://www.irit.fr/CECIL/

et détaille son implémentation. Les résultats des évaluations de ces propositions sont présentées en conclusion de ces parties.
Je ferai dans une dernière partie (partie V) un bilan de ce travail. Il sera également intéressant de se pencher sur les perspectives qu'apporte l'ensemble de ce travail, à court et long termes.

Deuxième partie

Etat de l'art : Agents Conversationnels Animés, langage et émotion

Introduction

Les ACA sont des personnages virtuels qui intègrent toutes les modalités de la communication humaine. L'amélioration de la crédibilité du comportement des ACA étant un enjeu majeur, deux questions se posent lors de leur développement : « comment l'ACA décide-t-il de ce qu'il va exprimer ? » (son **intention communicative**), puis « comment l'ACA exprime-t-il ce qu'il a décidé ? ». L'importance des émotions dans le raisonnement et la communication a été montrée dans de nombreux travaux en psychologie comme en informatique [Bates 1994, Piaget 1989] ; les humains considérant les personnages virtuels comme des partenaires sociaux et interagissant avec eux socialement [Picard 1997, Breazeal 1999], c'est naturellement que la majorité des travaux ont doté les ACA de modèles des émotions et de capacités d'expressions de ces émotions. Ces modèles se basent sur des théories des émotions en psychologie pour le déclenchement [Scherer 2001] ou l'expression [Ekman 1992] des émotions des agents. Ces théories sont présentées dans un premier chapitre (chapitre 3) ; les modalités d'expression des émotions chez les ACA et plusieurs travaux majeurs, intégrant des modèles des émotions aux ACA, sont présentés dans un second chapitre (chapitre 4).

D'autres travaux s'intéressent à l'expression d'états mentaux différents des émotions dans le dialogue [Aylett 2004], par des expressions faciales [Poggi 2000] ou des rétroactions [Schroder 2012]. Le standard SAIBA [Vilhjalmsson 2007] notamment, propose une bibliothèque de comportements, traduisant les intentions communicatives des ACA, au sein d'une architecture logicielle commune pour la génération de ces comportements. L'ensemble de ces travaux sont présentés dans le chapitre 5.

L'ensemble des travaux sur les ACA présentés dans cet état de l'art sont résumés dans le chapitre 6, qui conclut cette partie.

« C'est le mérite de la science, d'être exempte d'émotion. »
Oscar Wilde

3

Théories psychologiques des émotions

Comme nous avons pu le voir dans l'introduction, l'émotion joue un rôle très important dans la communication humaine et dans la cognition. De ce fait, un grand nombre d'ACA (cf. chapitre 4) privilégient l'expression de "leurs" émotions à partir de modèles s'appuyant sur des théories psychologiques des émotions. Je présente dans ce chapitre un historique rapide des théories des émotions en psychologie et celles qui ont été retenues dans les travaux de recherche sur les ACA.

3.1 Théories évolutionnistes des émotions

C'est Darwin [Darwin 1872] qui propose la première grande théorie psychologique des émotions. Il présente les émotions d'un point de vue évolutionniste : pour lui, les émotions sont des réponses adaptées à l'environnement, en nombre limité, issues de l'évolution et de ce fait, universelles. Il s'intéresse particulièrement aux expressions de ces émotions et à leur **fonction communicative** et adaptative au sein de l'espèce humaine. Un certaine nombre de théories des émotions s'inscrivent dans la lignée de Darwin, parmis lesquelles on peut citer la théorie d'Ekman [Ekman 1992], celle d'Oatley et Johnson-Laird [Oatley 1987] et les théories dimensionnelles [Russell 1980, Russell 1977].

3.1.1 La théorie des émotions discrètes d'Ekman

La théorie des émotions discrètes d'Ekman [Ekman 1992] est l'une des théories les plus connues. Elle définit six émotions selon leurs expressions faciales, appelées émotions "primaires" (cf. Figure 3.1) : la peur, la colère, le dégoût, la joie, la tristesse et la surprise. La théorie repose sur l'universalité des expressions de ces six émotions : elles sont universelles dans le sens où n'importe quel individu, quelles que soient sa culture et son expérience personnelle, peut les reconnaitre exprimées sur le visage d'un autre. Cette universalité s'explique, selon Ekman, par leur fonction de communication : informer les autres de son propre état émotionnel et de la situation qui a déclenché l'émotion. Le fait que les émotions "primaires" soient définies par leur expression faciale fait que cette théorie a été largement utilisée dans le domaine des ACA [Courgeon 2008, de Rosis 2003].

FIG. 3.1 – Les six émotions "primaires" d'Ekman définies selon leurs expressions faciales. De gauche à droite, un exemple d'expression de la colère, de la peur, du dégoût, de la surprise, de la joie et de la tristesse.

3.1.2 La théorie d'Oatley et Johnson-Laird

La théorie d'Oatley et Johnson-Laird [Oatley 1987] se situe également dans la lignée des théories évolutionnistes, en mettant en avant le rôle communicatif des émotions et leur aspect universel ; elle admet cependant un processus cognitif dans le déclenchement de certaines émotions (qui se rapproche, comme on va le voir, des théories de l'évaluation cognitive). Les auteurs distinguent deux types d'émotions, déclenchées par deux processus différents : les émotions "de base", déclenchées à un niveau inconscient, et les émotions "complexes", déclenchées à un niveau plus cognitif et concient. Ces deux niveaux ont un rôle avant tout communicatif :

- le niveau inconscient sert pour la communication émotionnelle ou "non-propositionnelle" : les émotions "de base" sont définies comme étant des primitives sémantiques, sans contenu propositionnel ; on trouve principalement les émotions "primaires" d'Ekman ;
- le niveau conscient sert pour la communication "propositionnelle" : les émotions "complexes" sont des émotions "de base" modulées par un contenu propositionnel (i.e. elles nécessitent le recours à des représentations de soi, de la situation, de l'action [Tcherkassof 2008]) ; on peut trouver par exemple le remords, causé par la conscience (représentations morales) d'avoir mal agi (représentation de l'action) lors d'un événement passé (représentation de la situation).

Les auteurs proposent des règles sémantiques pour différencier les émotions "de base" des émotions "complexes", en appliquant la règle du "je sais pourquoi". Selon eux, il est sémantiquement correct de dire, par exemple, « je suis triste, mais je ne sais pas pourquoi » : ceci est vrai pour toutes les émotions "de base", celles-ci ne nécessitant pas de représentation particulière et donc étant privées de contenu propositionnel. A l'inverse, lorsqu'il y a une cause à l'émotion (et qu'elle nécessite le recours à une représentation), l'on a à faire à une émotion complexe : il est donc sémantiquement incorrect de dire « j'ai des remords, mais je ne sais pas pourquoi ».

De plus, les auteurs affirment qu'une émotion "complexe" implique toujours une émotion "de base" (celle-ci constituant, en plus d'un contenu propositionnel, l'émotion "complexe"). L'aspect langagier de cette définition des émotions

est très intéressante : on peut distinguer des émotions qui s'expriment principalement *via* des expressions faciales (les émotions "de base" ou primaires) et des émotions plus complexes, qui ont besoin du langage pour être communiquées et reconnues par les autres.

3.1.3 Les théories dimensionnelles

Enfin, dans la lignée de Darwin, les théories dimensionnelles ([Russell 1980, Russell 1977]) situent les expressions faciales des émotions "primaires" (et les termes émotionnels - appelés labels - qui les définissent) dans un espace à deux ou trois dimensions élémentaires. Russel utilise les dimensions d'activation (ou d'implication émotionnelle) et de valence (degré de plaisir ou déplaisir), puis, avec Mehrabian, ajoute la dominance comme troisième dimension, formant ainsi le modèle PAD (Pleasure, Arousal, Dominance). L'espace que forment les deux (ou trois) dimensions contient donc les labels émotionnels des six émotions "primaires", mais permet également (par jugements subjectifs de sujets) de "remplir" les espaces entre ces labels émotionnels (cf. Figure 3.2).

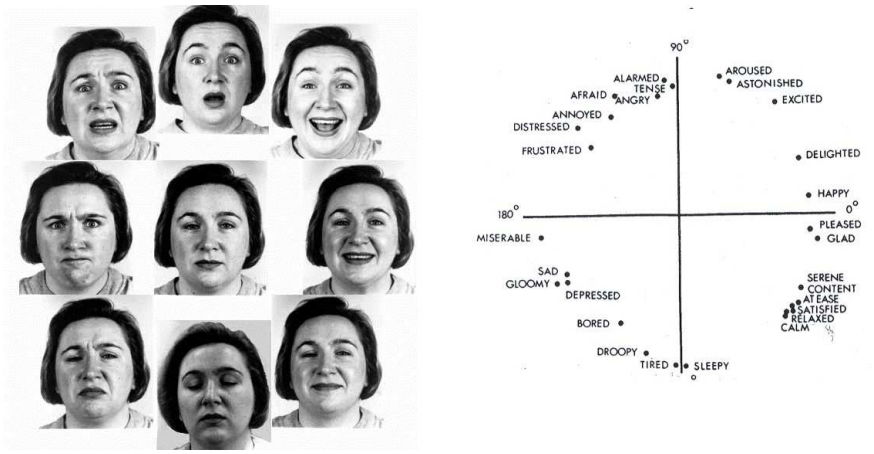

FIG. 3.2 – Le modèle à deux dimensions de Russel [Russell 1980] : la dimension horizontale (valence) va de déplaisir à plaisir (de gauche à droite) et la dimension verticale (arousal) de basse activation vers haute activation (de bas en haut).

L'ensemble de ces théories dans la lignée de Darwin s'intéressent donc aux émotions comme des réponses automatiques adaptatives, et principalement à leur expression et à leur **fonction communicative** ; un autre courant théorique, initié par James [James 1884] puis Cannon [Cannon 1927], s'intéresse au processus de **déclenchement des émotions**, comme nous allons le voir.

3.2 Comment se déclenche l'émotion ? Les théories de l'évaluation cognitive.

A l'inverse de Darwin, James [James 1884] s'est interessé plus particulièrement au processus de **déclenchement de l'émotion**, au travers d'un célèbre exemple, illustré par la figure 3.3.

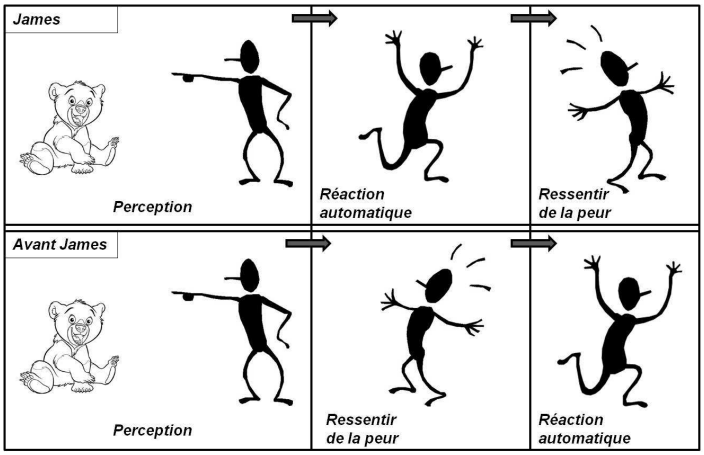

FIG. 3.3 – Comment se déclenche l'émotion de peur ? En haut, je vois l'ours, je cours (suite à une réaction automatique), *donc* j'ai peur. En bas, je vois l'ours, j'ai peur, *donc* je cours.

James affirme que les réactions automatiques du corps, en réponse à l'environnement, déclenchent les changements physiologiques, et la prise de conscience de ces changements déclenche les émotions. Lorsque je vois un ours, mon corps réagit automatiquement en prenant la fuite, et la prise de conscience des changements induits par la fuite (e.g. l'accélération du rythme cardiaque) déclenche l'émotion de peur.

Le problème de cette théorie, soulevé par Cannon [Cannon 1927], est qu'il n'y a pas de couple unique réaction physiologique - émotion : une émotion peut être induite par plusieurs réactions physiologiques, et *vice-versa*. Ce problème a amené les psychologues (d'abord Schachter [Schachter 1959], puis Arnold [Arnold 1960]) à introduire une évaluation cognitive dans le déclenchement des émotions. Selon Arnold, la perception de l'environnement, qui initie les changements physiologiques et les émotions, résulte d'un processus d'évaluation (*appraisal* en anglais) d'un certain nombre de critères, dont les croyances et les attentes d'un individu : ce processus représente le coeur des théories de l'évaluation cognitive (cf. figure 3.4).

Chaque individu, pour une même situation, peut avoir une évaluation (consciente ou non) différente de cette situation, et donc une réaction émotionnelle différente : dans l'exemple de l'ours, un dompteur (ou un berger, un chasseur...) n'aura pas forcément la même évaluation de la situation et ne ressentira pas la même émotion.

L'émotion en elle-même, résultant du processus d'évaluation, se décline en plusieurs composants : une réaction physiologique, une expression motrice (la communication de l'émotion par des expressions), une motivation à l'action (comment se traduit l'émotion en termes d'actions, e.g. la peur peut se traduire par la fuite) et un ressenti subjectif (*subjective feeling* en anglais, ce qui fait que l'on nomme l'état émotionnel que l'on ressent) La plupart des psychologues actuels admettent le rôle majeur de la cognition dans le déclenchement des émotions. Les théories de l'évaluation cognitive permettent un nombre infini d'émotions : l'émotion en elle-même est vue comme une combinaison de

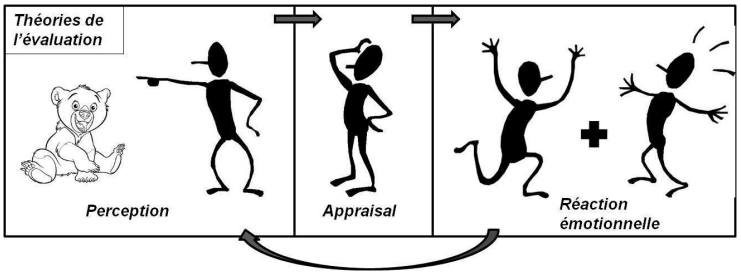

FIG. 3.4 – Cycle de perception, appraisal et réaction émotionnelle et physiologique. Je vois l'ours - j'évalue la situation (significativité de l'événement, conséquences sur mes buts...), consciemment ou non - je réagis en prenant la fuite et en ayant peur.

critères d'évaluation, là où les théories dans la lignée de Darwin n'acceptent qu'un nombre fini d'émotions (chacune liée à une expression faciale).

Un certain nombre de travaux majeurs [Scherer 1984, Ortony 1988, Lazarus 1991] héritent directement de la théorie de l'évaluation cognitive d'Arnold. Les critères d'évaluation diffèrent plus ou moins selon les travaux, bien que l'on puisse trouver un certain nombre de similarités. On s'intéresse ici tout particulièrement aux théories de l'évaluation cognitive qui servent de modèles aux ACA que je présente dans le chapitre 4.

3.2.1 La théorie d'évaluation cognitive de Lazarus

Lazarus [Lazarus 1991] affirme qu'il existe deux niveaux d'évaluation cognitive, et donc deux types de critères d'évaluation : ceux se rapportant à la pertinence de l'événement par rapport aux buts de l'individu (l'évaluation "primaire"), et ceux se rapportant aux capacités d'adaptation de l'individu face aux conséquences de l'événement (l'évaluation "secondaire"). D'après lui, ces deux évaluations ne se font pas en séquence mais sont complémentaires, pouvant être suivies dans n'importe quel ordre. Le tableau 3.1 présente ces six critères selon les deux niveaux d'évaluation.

Le modèle EMA [Gratch 2004], que je présente dans le chapitre suivant (chapitre 4) s'appuie sur la théorie de Lazarus pour déclencher les émotions des ACA. Cette théorie est néanmoins peu adaptée à une implémentation informatique, notamment à cause de la subjectivité de certains de ses critères [Adam 2007], tel que l'implication de l'individu dans la situation, et à cause du fait qu'il n'y a pas de correspondance claire entre les *patterns* que forment les critères et les émotions.

3.2.2 La théorie d'évaluation cognitive de Scherer

La théorie de l'évaluation cognitive de Scherer [Scherer 1984, Scherer 2001] décrit le processus d'évaluation cognitive (appelé le "Componential Process Model" - CPM) comme une évaluation en séquence de plusieurs critères, les *checks* ("Stimulus Evaluation Checks"). Ces *checks* sont divisés en 4 "objectifs d'évaluation", comme le montre le tableau 3.2.

Evaluation	Critère	Définition [Lazarus 1991]
Evaluation "primaire"	Pertinence aux buts	L'événement est pertinent s'il concerne particulièrement les propres buts de l'individu.
	Rapport aux buts	Déterminer si l'événement (et ses conséquences) vont contre/dans le sens des buts.
	Implication de l'individu (*type of ego-involvement*)	Ce critère peut se traduire par l'implication personnelle de l'individu dans la situation
Evaluation "secondaire"	Attribution de la responsabilité	Déterminer qui est responsable de l'événement.
	Potentiel d'adaptation	Evaluation des choix qui s'offrent à l'individu, en réponse à l'événement, et sélection de ceux qui sont les meilleurs pour l'individu.
	Attentes futures	Les conséquences de l'événement si l'individu ne fait rien

TAB. 3.1 – Les critères d'évaluation de Lazarus par niveau, évalués de façon complémentaire.

On peut remarquer que certains des *checks* de Scherer rejoignent les critères de Lazarus détaillés ci-dessus. Les principales différences se situent au niveau du critère d'implication de l'individu dans la situation, que Scherer traduit par les normes de l'individu, et au niveau de la pertinence d'un événement, à laquelle Scherer ajoute la nouveauté et le plaisir intrinsèque. Surtout, le processus d'évaluation de Scherer consiste en une séquence de critères évalués l'un après l'autre, tandis que Lazarus distingue deux niveaux complémentaires.

Scherer s'est de plus intéressé à l'expression motrice des *checks* (et des émotions qu'ils représentent) sur le visage. Ellgring et lui [Scherer 2007] ont montré que chaque *check* pouvait être associé à un mouvement musculaire du visage limité dans le temps appelé Unité d'Action (*Action Units* - AU - en anglais, introduites par Ekman et Friesen [Ekman 1978], que je détaille dans le chapitre suivant). Lors de l'évaluation en séquence des *checks*, les AU sont activées sur le visage, chaque AU précédemment activée se combinant avec la suivante, pour former une expression faciale dynamique. Courgeon [Courgeon n C] illustre bien ce processus en implémentant dans l'agent virtuel MARC un modèle d'évaluation adapté de la théorie de Scherer, associé à l'expression des *checks* sur le visage (cf. chapitre 4) ; je présente également une implémentation de ces *checks*, adaptés au dialogue compagnon artificiel - humain, dans le chapitre 13.

3.2.3 La théorie d'évaluation cognitive OCC

La théorie OCC [Ortony 1988] est une théorie d'évaluation cognitive particulièrement adaptée à l'implémentation, car pensée en ce sens [Adam 2007] : par exemple, l'intensité d'une émotion est définie par un ensemble de variables associées aux critères d'évaluation. Elle est à l'origine de nombreux modèles émotionnels pour les agents virtuels [Paiva 2006, Jaques 2009, Klesen 2003]. La

Objectif d'évaluation	*Check*	Définition [Scherer 2001]
Pertinence	Nouveauté	Probabilité et prédictibilité de l'événement.
	Plaisir intrinsèque	Nature de l'événement (e.g. bon ou mauvais), indépendamment des buts de l'individu.
	Pertinence aux buts	L'événement est pertinent si il concerne particulièrement les propres buts de l'individu.
Implications	Attribution de la responsabilité	Déterminer qui est responsable de l'événement.
	Probabilité des conséquences	Déterminer les conséquences probables de l'événement.
	Rapport aux buts	Déterminer si l'événement (et ses conséquences) vont contre/dans le sens des buts.
Potentiel d'adaptation	Contrôle et pouvoir	Evaluation des choix qui s'offrent à l'individu, en réponse à l'événement, et sélection de ceux qui sont les meilleurs pour l'individu.
Rapport aux normes	Normes internes et externes	Déterminer si l'événement (et ses conséquences) vont contre/dans le sens des normes sociales (externes), morales et personnelles (internes).

TAB. 3.2 – Les *checks* de Scherer par objectif d'évaluation, évalués en séquence.

théorie OCC différencie trois types d'évaluation : une évaluation de l'événement lui-même (en termes de conséquences pour l'individu ou pour autrui), une évaluation des actions (en termes de responsabilité de l'événement) et une évaluation de l'objet sur lequel porte l'événement. La figure 3.5 montre ces trois évaluations, les différents critères qui les composent, et quelques émotions déclenchées en fonction de ces critères.

Ces évaluations ne se font pas en séquence : c'est en fonction du "focus" de l'individu (ce sur quoi il se concentre) lors du stimulus que se décide le type d'évaluation qu'il va effectuer. L'exemple d'un agent virtuel, d'un utilisateur et d'un régime de bananes, proposé par Bartneck [Bartneck 2002], illustre bien les différentes évaluations du même stimulus : un agent virtuel aime les bananes (virtuelles) et l'utilisateur lui en donne un régime entier. Si pendant l'événement, l'agent se concentre sur ses conséquences, il évalue les conséquences pour l'utilisateur et pour lui : pour l'utilisateur, il a de la *compassion* (car l'utilisateur a des bananes en moins), et concernant les conséquences pour lui, il a de la *satisfaction*. S'il se concentre plutôt sur l'action, cela va déclencher une émotion d'*admiration*. Enfin, s'il se concentre sur l'objet de l'événement, cela va déclencher une émotion d'*attirance*.

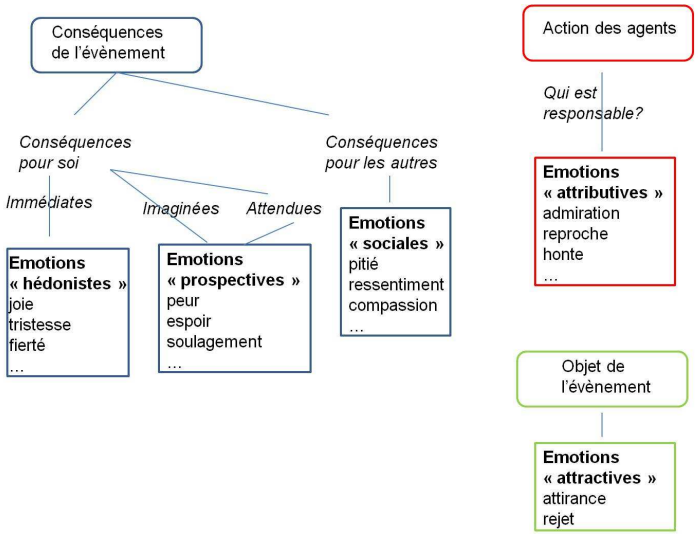

FIG. 3.5 – Les trois types d'évaluation du modèle OCC.

3.2.4 Les théories sociales de l'évaluation cognitive

Pour terminer, certains travaux ont mis en avant l'influence de la fonction sociale des émotions (montrée par [Averill 1985, Rimé 1992, Parkinson 1996]) dans le processus d'évaluation cognitive d'une situation. Ces travaux définissent ainsi l'émotion comme une "construction sociale" [Gross 2011]. Manstead et Fischer [Manstead 2001] proposent de prendre en compte, dans l'évaluation cognitive d'un individu et dans l'expression de ses émotions, les émotions des autres personnes proches de l'individu : ils définissent ainsi l'évaluation cognitive sociale (*social appraisal* en anglais). Un exemple "intuitif" illustre bien le concept d'évaluation cognitive sociale : imaginons deux personnes qui reçoivent une même nouvelle (une nomination de l'un d'entre eux par exemple) ; cette nouvelle est très bonne pour l'un et mauvaise pour l'autre ; dans ce cas, il est *socialement admis* que celui nominé va réguler l'expression de son émotion de joie par rapport à l'autre personne, non selectionnée.

« Un bon acteur sait mettre de l'émotion dans l'action et de l'action dans l'émotion. »

Charlie Chaplin

4 ACA et émotions

4.1 Expressions multimodales chez les ACA

La capacité de s'exprimer multimodalement permet aux ACA de s'approcher de plus en plus de la communication humaine : on connait en effet depuis longtemps l'importance des expressions faciales [Ekman 1992], des gestes [Ekman 1972] ou encore du regard [Argyle 1976] dans les interactions humaines, qui viennent compléter ou se substituer à la modalité langagière. Le développement d'ACA plus crédibles, capables d'exprimer multimodalement leurs émotions, est donc devenu un enjeu majeur pour l'amélioration des interactions homme-machine, en termes de succès de la communication mais aussi de crédibilité et d'acceptabilité des ACA pour les humains. L'expression des états mentaux (dont les émotions) des ACA s'effectue *via* plusieurs modalités, que je liste ci-dessous.

4.1.1 Expressions faciales et regard

Le système des FACS[1] de Ekman et Friesen [Ekman 1978] est un standard d'annotation pour les expressions du visage. Il a été initialement développé pour des psychologues et des spécialistes du comportement. Ce système définit les mouvements musculaires du visage, décomposés en 66 Unités d'Actions (AU), auxquelles on peut associer des valeurs et que l'on peut combiner pour former des expressions faciales (cf. figure 4.1).

Le système des FACS est de plus en plus utilisé par les ACA pour définir les expressions faciales, et permettre de contrôler l'animation indépendamment de la géométrie du visage. Par exemple, "jouer" l'AU 20 revient, selon que la valeur est positive ou négative, à serrer ou desserrer les lèvres. Les émotions primaires de [Ekman 1992] telles que la joie, la peur ou la tristesse peuvent ainsi être exprimées, en combinant les AU pour former ces expressions faciales. D'autres combinaisons d'AU peuvent former des émotions plus complexes, dont les expressions faciales ne sont pas spécifiées dans la littérature : ces expressions faciales peuvent être définies à partir de corpus vidéos dans lesquelles les émotions sont annotées par des "juges" puis validées par des tests de reconnaissance auprès d'autres sujets (e.g. corpus MindReading [Golan 2006]).

Spécifiquement pour l'animation du visage, le standard ISO MPEG-4 [Pandzic 2003] a été mis en place pour définir des paramètres d'animation communs. Ce standard se base également sur les muscles du visage pour définir de "nouvelles" AU,

[1] Facial Action Coding System

FIG. 4.1 – Le système des FACS : les différentes Unités d'Actions (AU) représentant les mouvements musculaires du visage (personnage Mary de MARC [Courgeon 2008]).

les FAPs (Facial Animation Parameters), utilisées pour l'animation d'un visage virtuel. Ce standard est notamment utilisé pour la reproduction de mouvements issus de captures d'expressions faciales.

Le regard est également très important dans la communication humaine, notamment pour réguler la conversation (distribuer la parole à tour de rôle), pour signaler de l'intérêt et pour exprimer des émotions [Argyle 1976]. On retrouve dans les AU décrites ci-dessus douze AU pouvant être utilisées pour diriger le regard et le rendre expressif, en jouant sur l'ouverture des paupières, le plissement des yeux etc. De nombreux ACA développent des modèles de regard pour contribuer à l'expression multimodale et à la crédibilité de l'ACA [Rickel 1998].

4.1.2 Gestes

La plupart des ACA utilisent un lexique de gestes, notamment ceux des bras et des mains, construits à partir des "illustrateurs" (*illustrators* en anglais) d'Ekman et Friesen [Ekman 1972]. Ces gestes peuvent être liés :

- à la sémantique de l'énoncé exprimé par l'agent (gestes déictiques, spatiaux et pictographiques) ;
- au sens, comme les "emblèmes", gestes culturels qui remplacent la parole (e.g. « bonjour », « au revoir », « il est fou », cf. figure 4.2) ;
- à la prosodie de cet énoncé (les "bâtons", marquant le rythme de la phrase).

Les gestes peuvent être influencés par l'état émotionnel de l'agent : par exemple, lorsque l'ACA est en colère, ses gestes sont plus brusques et leur amplitude plus grande [Wallbott 1998]. D'autres travaux ont mis en évidence la faculté des gestes à exprimer clairement les émotions primaires telles que la joie, la peur etc. [McNeill 1996, Poggi 2003].

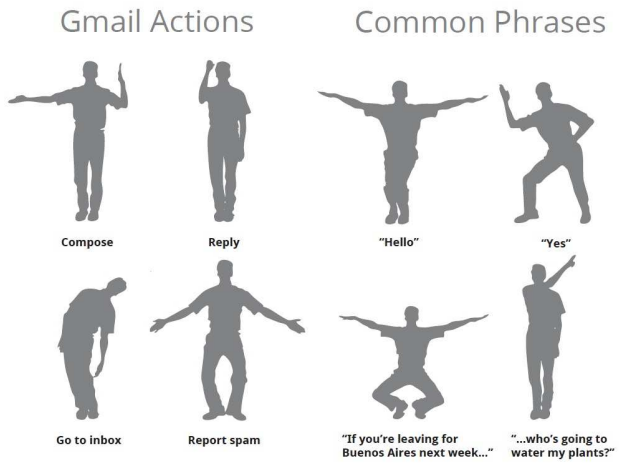

FIG. 4.2 – Quelques-unes des actions et phrases "communes" disponibles dans Gmail Motion (*poisson d'avril* 2011 de Google). Ces types de gestes peuvent être considérés comme de "faux emblèmes", en ce sens où ils veulent apparaitre comme intuitifs, innés et partagés culturellement par tous.

4.1.3 Autres expressions multimodales

La prosodie du discours [Klesen 2003] et la posture du corps [Tan 2009] interviennent également dans l'expression multimodale des émotions de certains ACA.

4.2 ACA affectifs et expressifs

Je présente dans cette section quelques travaux sur les ACA proposant des modèles de déclenchement des émotions s'appuyant sur les théories psychologiques présentées dans le chapitre précédent : le but ici n'est pas d'être exhaustif mais de mettre en évidence, à l'aide de ces travaux, les différentes méthodes de déclenchement des émotions, leurs effets sur le comportement de l'agent et leur expression multimodale.

4.2.1 MARC : Multimodal Affective and Reactive Characters

MARC [Courgeon 2008, Courgeon n C] est un ACA développé au LIMSI par Matthieu Courgeon au cours de sa thèse. Quatre théories psychologiques des émotions distinctes ont été adaptées dans ses travaux ; sur ces quatre théories, deux ont été implémentées de telle sorte que les émotions de MARC soient déclenchées de manière automatique : ce sont les théories de l'évaluation cognitive de Scherer [Scherer 2000] et de l'évaluation cognitive sociale ([Manstead 2001], *social appraisal theory* en anglais), qui nous intéressent ici.

Evaluation cognitive de Scherer

Certain des *checks* de la théorie de Scherer ont été implémentés pour déduire les émotions de MARC, dans le contexte particulier d'une partie avec un utilisateur du jeu Reversi (variante de Othello[2]). Dans ce contexte, un nouvel événement est un nouveau coup joué par son adversaire ou par l'agent lui-même. Un ensemble de règles propres au jeu Reversi a été défini pour attribuer des valeurs aux *checks*. Les émotions de MARC sont déduites selon les différentes valeurs des *checks* : la configuration particulière de ces *checks* est comparée avec la litterature [Scherer 2001] qui définit les profils des émotions primaires de joie, peur, colère, tristesse et fierté. Pour l'expression faciale, les liens entre *check* et Action Unit (AU), spécifiés par Scherer (comme je le montre dans le chapitre 3), sont utilisés.

Une expression faciale dynamique des émotions de MARC (cf. figure 4.3) est construite et est constituée de l'expression faciale de chaque *check* évalué (i.e. de certains états mentaux de l'agent) ; l'expression finale de MARC est l'expression faciale, définie par Ekman, de l'émotion primaire que forment les *checks* (sauf la fierté, qui a été définie à partir du corpus MindReading [Golan 2006]).

(0) (1) (2) (3) (4) (5)

FIG. 4.3 – MARC (à l'état neutre (0)) exprime, de gauche à droite, un événement inattendu (1), désagréable (2), qui entrave ses buts (3) et sur lequel MARC a peu de contrôle (4). L'expression finale (5) est l'émotion de tristesse telle que définie par Ekman [Ekman 1992].

Il est montré que cette expression dynamique améliore la perception qu'ont les utilisateurs des capacités cognitives de l'agent, en lui attribuant plus d'états mentaux que lors d'une expression statique. Il serait intéressant ici de prendre en compte des émotions plus complexes dans l'évaluation des événements (comme le regret par rapport à un coup précédent etc.), et d'ajouter la modalité langagière lors de l'expression des états mentaux.

Evaluation cognitive sociale

Le modèle précédent d'évaluation cognitive a été enrichi en ajoutant une dimension *sociale* à l'évaluation et l'expression des émotions de MARC. Le nouveau modèle se base notamment sur les travaux de Manstead et Fischer [Manstead 2001] en incluant un niveau de "reappraisal" (ré-évaluation) dans l'architecture cognitive de MARC : accompagné d'un autre agent virtuel émotionnel, MARC évalue un événement et exprime l'émotion déduite de son évaluation ; puis il observe la réaction émotionnelle de l'autre agent, ré-évalue l'événement une deuxième fois en prenant en compte cette réaction et exprime par le visage le résultat de cette ré-évaluation.

[2] http ://fr.wikipedia.org/wiki/Othello_(jeu)

Le phénomène de “reappraisal” est modélisé dans MARC par un certain nombre de règles : par exemple, si MARC évalue un événement comme plaisant, mais que l’autre agent exprime qu’il croit ne pas avoir de contrôle sur l’événement, alors l’aspect plaisant de l’événement pour MARC est diminué lors de la nouvelle évaluation.

Les résultats de l’évaluation sont positifs et prometteurs, dans le sens où les utilisateurs reconnaissent et comprennent les comportements sociaux des agents. Cette expérimentation ne propose cependant pas, à l’inverse de l’implémentation dans Othello, de système d’évaluation concret des événements (dans le dialogue par exemple) ; l’expérimentation se base en effet sur l’évaluation de vidéos.

4.2.2 EMA : EMotion and Adaptation

Gratch et Marsella ont proposé le modèle affectif EMA [Gratch 2004], basé sur la théorie de l’évaluation cognitive de Lazarus [Lazarus 1991] (cf. chapitre 3) : cette architecture est donc composée de deux niveaux, un processus d’évaluation et un processus d’adaptation, chacun évaluant un ensemble de critères spécifiques (attribution de la responsabilité, probabilité, rapport aux buts - désirabilité etc.). Le processus d’évaluation déclenche des émotions à partir des critères d’évaluation ; ces émotions sont décrites dans le tableau 4.1[3].

Evaluation de l’événement E	Emotion
Evénement inattendu	Surprise
Evénement désirable et probabilité qu’il survienne <1	Espoir
Evénement désirable et probabilité qu’il survienne =1	Joie
Evénement indésirable et probabilité qu’il survienne <1	Peur
Evénement indésirable et probabilité qu’il survienne =1	Tristesse
Evénement indésirable, quelqu’un d’autre est responsable et potentiel d’adaptation faible	Colère
Evénement indésirable pour quelqu’un d’autre et l’agent est responsable	Culpabilité

TAB. 4.1 – Les émotions déclenchées dans EMA par le processus d’évaluation.

EMA maintient également les croyances de l’agent sur l’état du monde, représenté par une conjonction de propositions (il fait beau $\wedge$ j’ai un vélo $\wedge$ etc.). L’agent a des buts et des intentions relativement à l’état du monde ; ces opérateurs ne sont cependant pas formalisés en logique.

Le processus d’adaptation permet de choisir, parmi un certain nombre de stratégies, celles qui contribuent à maintenir un état désirable ou au contraire, celles à adopter pour faire face à des états indésirables. Ces stratégies peuvent modifier les croyances, les buts, les intentions ou les plans des agents : face à un état indésirable par exemple, les stratégies peuvent être de prendre de la distance avec cet état (croire qu’il a peut d’importance), ou essayer de le corriger en adoptant une nouvelle intention etc. EMA est un des seuls modèles affectifs pour ACA intégrant ce processus d’adaptation.

[3]On notera ici que l’agent ne “ressent” de la culpabilité que par rapport à un événement indésirable pour l’autre, contrairement à la définition donnée par d’autres travaux [Tangney 2003]

EMA est actuellement utilisé dans le projet MRE (Mission Rehearsal Exercise, [Hill 2001]), un environnement pour l'entrainement virtuel dans le domaine militaire : des agents virtuels font partie de cet environnement et peuvent interagir avec le sujet humain, en communiquant par la voix et les gestes et en exprimant leurs émotions *via* des expressions faciales. L'entrainement consiste à jouer un scénario qui met en scène un lieutenant (le sujet humain), un sergent (l'agent virtuel), un conducteur de véhicule militaire, et une civile et son fils (figurants) avec qui le conducteur vient d'avoir un accident : le garçon est gravement blessé et la mère est très agitée. Le rôle du sergent est de rendre compte de la scène en interagissant avec le lieutenant pour l'informer et attendre ses ordres.

4.2.3 CrossTalk : un exemple de "scripting" de scène

Dans le domaine des ACA, le "scripting" consiste à annoter préalablement une scène qui va être jouée par des ACA : on attribue aux divers éléments de la scène (un événement, un objet etc.) une valeur ou un "poids" qui va influencer le comportement de l'ACA. Les effets de tous les éléments d'une scène sur les ACA sont donc décidés à l'avance : par exemple, une porte qui s'ouvre soudainement déclenche une émotion de surprise chez un ACA et de peur chez un autre.

Le système CrossTalk [Klesen 2003] consiste en une plateforme virtuelle de trois agents virtuels : deux agents "acteurs" (qui jouent une scène de vente de voiture, cf. figure 4.4) et un agent hôte (dont le rôle est d'accueillir l'utilisateur et de lui introduire la scène). Gebhard et al. [Gebhard 2003] ont présenté une extension de ce système en intégrant une dimension émotionnelle, s'appuyant sur la théorie OCC [Ortony 1988]. Les auteurs définissent pour chaque type d'évaluation (événement, action, objet, cf. chapitre 3) un facteur de déclenchement de l'émotion ("EEC", *Emotion Eliciting Condition* en anglais) : la désirabilité et la probabilité de l'événement, l'admirabilité d'une action et l'attractivité de l'objet. Ces "EEC" sont ensuite traduits en dix émotions du modèle OCC : joie, tristesse, espoir, peur, quatre émotions "attributives" (fierté, honte, reproche et admiration) et les deux émotions "attractives", attirance et rejet.

Les auteurs proposent d'annoter préalablement la scène qui va être jouée par les deux agents "acteurs" ("scripting") en termes d'"appraisal tags" et de "dialogue act tags" que le système peut traduire en "EEC". Les "appraisal tag" indiquent les valeurs d'évaluation des actions, objets et événements : par exemple, un événement peut être bon, mauvais, futur, présent ou probable, tandis qu'une action peut être effectuée par soi-même ou quelqu'un d'autre et être bonne ou mauvaise. Les "dialogue act tags" annotent, en termes d'actes de dialogue, les énoncés des agents : chaque acte est lié (indépendamment du contexte) à une valeur dans chaque "EEC", e.g. si Ritchie *attaque* (verbalement) Tina, celle-ci évalue l'événement comme étant peu désirable et répréhensible, tandis que Ritchie l'évalue comme modérement désirable. La bibliothèque d'actes de dialogue ainsi annotés est constituée de vingt-trois actes (e.g. calmer, féliciter, s'excuser, faire une mauvaise blague). L'utilisateur peut intervenir dans la scène *via* deux actes : il peut *féliciter* ou *critiquer* un des agents.

Les émotions, ainsi déduites des "EEC" traduits de l'annotation, sont exprimées par la prosodie (hauteur, vitesse et volume de la voix), par des expressions faciales "simples" (positive *vs.* négative) et le regard. Les actes de dialogue sont eux associés à des gestes "de conversation", liés à la sémantique et au sens

FIG. 4.4 – La scène de vente jouées par les agents "acteurs" (Ritchie et Tina) dans CrossTalk.

de l'énoncé exprimé par l'agent.

Considérer les actes de discours comme des "événements" à évaluer dans les théories de l'évaluation cognitive semble très intéressant. On peut cependant regretter que ces actes n'aient pas de lien formel avec les émotions et que la majorité de ces actes ne soient pas disponibles pour le dialogue agent - humain. On pourrait également imaginer que les agents évaluent eux-mêmes les événements (dont les actes des autres agents), sans passer par l'annotation des scènes. On peut également remarquer ici que l'expression des émotions et l'expression des actes de discours se font de manière indépendante l'une de l'autre.

4.3 ACA : émotions et états mentaux BDI

Trois travaux qui ont proposé de formaliser, à l'aide d'opérateurs logiques de type BDI (Belief, Desire, Intention, [Rao 1991]), les états mentaux des ACA et leurs émotions, sont ici présentés. Ces ACA perçoivent les événements, mettent à jour leurs états mentaux et évaluent ces événements en termes d'"impulsions" [Kopp 2003] ou *via* des variables d'évaluation [Paiva 2006, Ochs 2007] provenant des théories cognitives des émotions (cf. chapitre 3). Cette évaluation déclenche ou influence les émotions de ces ACA.

4.3.1 Max et l'architecture WASABI

Les travaux de Becker et Kopp [Becker 2004, Kopp 2003] ont permis de mettre en place un agent virtuel "guide de musée" appelé *Max*, capable de raisonner et d'interagir avec les visiteurs. L'architecture de *Max* consiste en un niveau cognitif, basé sur une architecture BDI, et un niveau émotionnel, dans lequel les auteurs distinguent les émotions de *Max* de ses humeurs (*mood* en anglais). Les deux niveaux s'influencent l'un l'autre : le niveau cognitif interprète ce que dit l'utilisateur, évalue cette interprétation en termes d'impulsions (positive

ou négative), et transmet au niveau émotionnel ces impulsions. Ces impulsions influent sur les émotions de *Max* et sur ses humeurs de manière linéaire et cumulative dans l'espace à 3 dimensions PAD (*Pleasure, Arousal, Dominance*, cf. chapitre 3). Les émotions *primaires* de *Max* sont situées dans cet espace, de même que d'autres états émotionnels (ennuyé, déprimé) et non-émotionnels (concentré, amical) : la figure 4.5 présente ces différents états dans l'espace PAD. Ainsi, une impulsion positive (ou négative) du niveau cognitif va influencer la position de son état émotionnel dans cet espace.

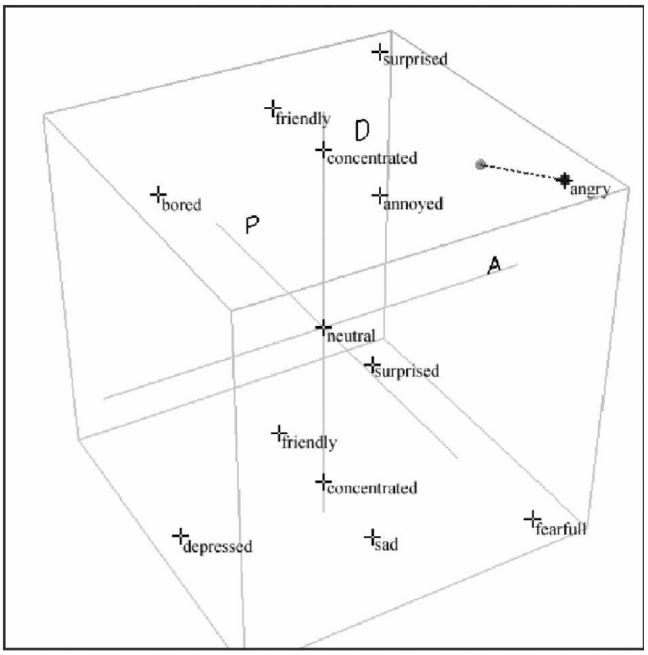

FIG. 4.5 – Les états émotionnels et non-émotionnels de *Max* situés dans l'espace PAD.

Le niveau cognitif est également influencé par le niveau émotionnel en intégrant les émotions de *Max* dans son raisonnement pour sélectionner ses plans : ainsi, lorsque *Max* est en colère par exemple, il peut choisir de quitter la scène (plan nécessitant comme précondition la colère). L'humeur de *Max* et les valeurs du triplet PAD influencent également l'animation du personnage, i.e. son souffle, son débit de parole et la fréquence de clignement des yeux.

Enfin, le modèle WASABI intègre dans *Max* des émotions *secondaires*, déclenchées au niveau cognitif, et présentes dans OCC : l'espoir, la peur confirmée (*fears-confirmed* en anglais) et le soulagement apparaissent en comparant les attentes de *Max* (issues de ses plans et de ses intentions) et l'état du monde. Ces émotions sont ensuite exprimées verbalement (par un énoncé) par *Max*.

4.3.2 FearNot !

Le système FearNot ! [Paiva 2006] consiste en un monde virtuel regroupant des personnages virtuels : ces personnages sont impliqués dans des scènes courtes

dans lesquelles le "héros", John, est un petit garçon molesté par un ou plusieurs autres personnages. Le but ici est à caractère éducatif : il amène l'utilisateur (un enfant) à regarder ces scènes et à jouer le rôle d'un "ami invisible" qui donne des conseils à John. Dans ce système, les agents mettent à jour un modèle du monde (leurs croyances), leurs buts, leurs intentions et leurs plans, à la manière des modèles BDI. FearNot ! se base sur OCC [Ortony 1988] pour modéliser les émotions des agents, et considère deux types d'émotions : les émotions "réactives" (peur, colère, joie...), déclenchées lorsque certains des critères d'évaluation d'OCC sont activés, et des émotions plus délibératives, portées vers le futur (espoir et peur sur les buts - "*InterGoal-Fear*"), déclenchées par la probabilité du plan suivi par l'agent de réussir ou d'échouer. L'agent se concentrant sur ses émotions les plus fortes, celles-ci influencent la sélection de ses actions et de ses plans. Ces émotions sont exprimées par les agents *via* des expressions faciales. FearNot ! propose également une taxonomie des actes de discours particulière au système [Aylett 2004], utilisée pour la communication entre agents et entre l'humain et l'agent, que je décris au chapitre suivant (chapitre 5).

4.3.3 Un agent rationnel empathique

Dans sa thèse, Magalie Ochs [Ochs 2007] propose de modéliser pour l'ACA Greta [Poggi 2005] certaines des émotions correspondant, dans la théorie OCC (cf. chapitre 3), aux *conséquences pour soi* et *pour les autres* d'un événement : pour soi, selon que l'événement est désirable ou pas, la joie, la peur, la colère ou la tristesse sont déclenchées ; et si cet événement est inattendu, la surprise est déclenchée. Les émotions qui sont déduites de l'évaluation des conséquences d'un événement pour les autres sont les émotions sociales ou "empathiques" : l'ACA peut être content (*happy-for*) ou triste (*sorry-for*) pour l'autre suivant la désirabilité de l'événement. Cette désirabilité est exprimée en termes d'opérateurs représentant des états mentaux, basés sur la théorie de l'interaction rationnelle [Sadek 1991] : croyance, incertitude et choix (qui représente l'intention). Par exemple, un événement est désirable pour l'agent s'il a l'intention qu'il survienne et qu'il est incertain sur sa probabilité de survenir. Les émotions ainsi déclenchées sont exprimées par des expressions faciales, que Greta peut adapter à la situation en masquant une émotion par une autre (en exprimant simultanément de la tristesse et de la colère par exemple).

5

ACA, états mentaux et langage

Les émotions ne sont pas les seuls états mentaux présents dans la communication humaine [Searle 1983] : on peut vouloir par exemple exprimer ses buts, ses croyances, ou ses désirs. Ces différentes intentions communicatives s'expriment en général par le langage, en lien avec d'autres modalités non-verbales. En effet, lorsque l'on parle de communication, on a tendance à penser directement au langage : le langage est la modalité la plus utilisée et la plus naturelle dans la communication humaine (on parle d'ailleurs des autres modalités comme étant "extra-langagières" ou "co-verbales"). Cependant, le langage reste une modalité parmi d'autres au sein de la communication humaine ; elle est fortement liée à ces autres modalités dans l'expression des états mentaux, comme on va le voir dans ce chapitre, l'exemple le plus "parlant" étant lorsque l'on observe quelqu'un parler au téléphone.

Comme je vais le montrer dans le chapitre 7, la théorie des Actes de Discours [Searle 1969, Searle 1985] reste la référence parmi les théories philosophiques du langage. Cette théorie introduit une typologie des actes de discours (assertifs, directifs, déclaratifs, engageants, expressifs) sur laquelle s'appuie un certain nombre de travaux sur les agents virtuels [Poggi 2005, Berger 2005, Heylen 2008] pour exprimer les états mentaux de l'agent ; d'autre travaux s'appuient sur des typologies différentes ([Samtani 2008] sur la typologie de Traum et Hinkelman [Traum 1992], ou [Paiva 2006] sur le vocabulaire de la maltraitance, qui est présenté ci-dessous).

Cependant, la plupart de ces travaux proposent de traduire les actes de discours en un ensemble de modalités (dont le langage), sans que l'ACA décide de son intention communicative. Par exemple, les travaux de Poggi et Pelachaud [Poggi 2000] définissent au sein du standard SAIBA (présenté ci-dessous) un lexique d'actes (tels que suggérer, ordonner...) dont les signaux correspondent à l'expression faciale des "unités cognitives" formalisant ces actes. Malheureusement, sur l'ensemble de ces travaux, peu d'actes sont en général définis, et les liens formels qui ont pu être établis ne suffisent pas à modéliser de façon correcte l'ensemble des états mentaux (dont les émotions) d'un agent.

Plusieurs travaux en linguistique [Perrault 1980, Cohen 1979] ou dans le domaine des systèmes multi-agents ([FIPA-ACL 2006], KQML) se sont intéressés au processus de sélection d'un acte de discours comme intention communicative. En revanche, peu de travaux chez les ACA intègrent ce type de processus, à l'exception des systèmes FearNot ! et du projet MRE présentés précédemment, qui sont présentés plus en détails dans ce chapitre.

5.1 Bibliothèque de comportements : traduction des états mentaux

5.1.1 Le standard SAIBA

SAIBA (Situation, Agent, Intention, Behavior, Animation, [Vilhjalmsson 2007]) est une initiative de recherche internationale dont le but est de définir une architecture logicielle commune pour la génération du comportement d'agents virtuels. L'architecture SAIBA est composée de trois modules, présentés figure 5.1.

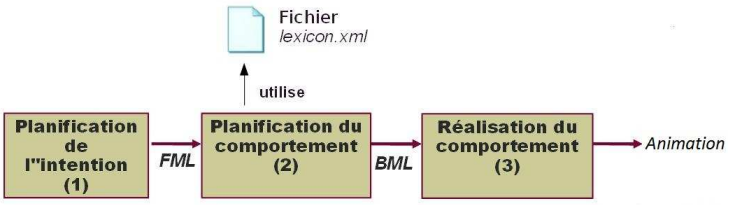

FIG. 5.1 – Le standard SAIBA

Le premier module de *Planification d'intention* décide de l'intention communicative de l'agent. Les différents ACA qui suivent le standard SAIBA implémentent de façon différente ce module, comme nous le verrons par la suite : ainsi, l'intention communicative peut être un acte communicatif[1] de plusieurs sortes, tels que rétroactions (*backchannels* en anglais), émotions, actes de langage, actions et certains gestes [Heylen 2008]. Cet acte communicatif est encodé en langage FML (Function Markup Language, [Vilhjalmsson 2005]), langage de balisage basé sur XML.

Le langage FML permet d'associer plusieurs comportements, de les combiner ou de les jouer en séquence, grâce à l'existence des différentes balises. Les balises initiales introduites par les auteurs de [Vilhjalmsson 2005] permettent de spécifier, notamment, les émotions de l'ACA, les participants et le sujet de la conversation, et l'acte de discours (appelé *performative*) que va faire l'agent. De nombreux travaux [Adam 2009a, Mancini 2008, Samtani 2008] s'intéressent à l'amélioration du langage FML en introduisant de nouvelles balises ou en définissant de nouveaux actes. Par exemple, Pelachaud et Mancini [Mancini 2008] ont introduit le langage FML-APML utilisé par l'ACA Greta, et ajouté plusieurs balises (e.g. le degré d'intensité des émotions, si elles sont ressenties ou feintes) et un certain nombre d'actes de discours [Searle 1969] dans une bibliothèque de comportements (e.g. ordonner, proposer, supplier). La figure 5.2 montre un exemple d'utilisation du FML-APML : l'intention communicative de l'ACA est de nier quelque chose (représenté par l'acte de discours *deny*, spécifié par la balise "performative") associé à un énoncé (*« This is not my fault ! »* - Ce n'est pas ma faute !).

L'intention communicative ainsi encodée en FML est envoyée au module de *Planification du comportement* pour être interprétée. Ce deuxième module

[1] on distingue l'intention communicative qui est ce qui est planifié (décidé), et l'acte communicatif qui le représente. Par exemple, l'intention de communiquer son accord peut se représenter par l'acte *accepter*.

```
<fml-apml>
  - <bml>
      + <speech language="english" text="This is not my fault!"
            voice="openmary" start="0.0" id="s1">
    </bml>
  - <fml>
        <performative start="s1:tm1" id="p1" type="deny" />
    </fml>
</fml-apml>
```

FIG. 5.2 – Encodage de l'acte nier en FML-APML.

gère un certain nombre de modalités de communication dans le but d'exprimer l'intention communicative de l'agent : la parole, l'expression du visage et les gestes sont ainsi calculés grâce à un *fichier de lexique* XML. Ce *fichier de lexique* (le *lexicon.xml* dans la figure 5.1) est la bibliothèque de comportements de SAIBA. Il contient la liste des expressions multimodales que peut jouer l'ACA (émotions, gestes déictiques...) pour traduire l'intention communicative : il est donc directement lié au langage FML, dans le sens où le module de *Planification du comportement* cherche les correspondances des balises FML avec des signaux multimodaux. Ces expressions sont construites différemment selon les différents travaux : corpus vidéos, validation par expérimentations, corpus théoriques (expressions faciales émotionnelles d'Ekman [Ekman 1992]) etc. Malgré cette diversité, le format de l'expression dans le fichier de lexique reste néanmoins le même ; par exemple, la figure 5.3 montre comment se traduit par des signaux multimodaux l'acte de discours *deny* du code FML précédent :

```
- <behaviorset name="performative-deny">
    - <signals>
        - <signal name="head=head_shake" modality="head" id="1">
              <alternative name="head=head_up" probability="0.5"/>
          </signal>
          <signal name="certainty=uncertain" modality="gesture" id="2"/>
          <signal name="eyes=close" modality="gaze" id="3"/>
          <signal name="eyes=frown" modality="face" id="4"/>
      </signals>
    + <constraints>
  </behaviorset>
```

FIG. 5.3 – La traduction de l'acte nier (*deny*) en signaux multimodaux dans le fichier lexique.

On pourrait enfin ajouter par exemple une émotion de colère (si elle est définie dans le fichier lexique) à l'expression de l'acte nier, en utilisant la balise "emotion", en-dessous de la balise "performative" :
<emotion id="e1" type="anger" start="s1 :tm1" regulation="felt" />

Le second module de *Planification du comportement* traduit l'intention communicative de l'ACA en un ensemble de signaux multimodaux, grâce au *fichier de lexique*, comme détaillé ci-dessus. Le langage BML (Behavior Markup Language, [Kopp 2006]) est utilisé pour représenter de manière détaillée (modalité, temps de début, de fin, latence etc.) ces signaux, et les transmettre au dernier module de *Réalisation de comportement*, qui vérifie les contraintes d'animation et les conflits potentiels entre ces signaux. Comme ce langage s'inscrit dans le processus concret d'animation du personnage virtuel, il n'est pas plus détaillé ici.

5.1.2 Expressions faciales performatives

Un lexique d'expressions faciales associées au regard, et basé sur l'expression d'"unités cognitives", a été construit par Poggi et Pelachaud [Poggi 1998, Poggi 2000] pour l'ACA Greta [Poggi 2005] qui suit l'architecture SAIBA : le but ce ce lexique est d'exprimer par une expression faciale les "unités cognitives" d'un acte performatif de la théorie des Actes de Discours [Searle 1969, Searle 1985], tels que suggérer, ordonner ou encore implorer. Ces "unités cognitives" sont les états mentaux de l'agent formalisés en termes d'intentions, de croyances et d'actions de l'agent ; par exemple, la formalisation de l'acte *suggérer* est de la forme :

- S (*Sender*) a le but que A (*Addressee*) fasse une action a ;
- S croit que l'action a est utile à un but de A ;
- S croit la proposition précédente à un degré faible de certitude.

Ces "unités cognitives" sont traduites par une expression faciale grâce aux AU introduites par Ekman et Friesen ([Ekman 1972], cf. chapitre précédent). La figure 5.4 montre l'expression faciale de l'acte suggérer.

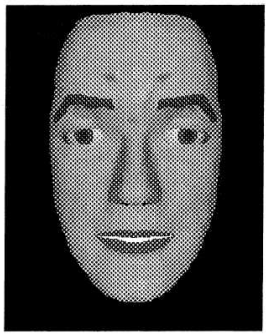

FIG. 5.4 – Expression faciale de l'acte performatif suggérer. Adaptée de [Poggi 1998].

Cette formalisation ne permet cependant pas de représenter des états mentaux complexes (reposant sur des raisonnements complexes) ou des émotions complexes ; de plus, les états mentaux sont seulement utilisés dans l'expression de ces actes performatifs, sans que l'ACA raisonne à partir d'eux.

5.1.3 Rétroactions (*Backchannels*)

Dans la communication, certains gestes sont des indicateurs de l'état cognitif de l'ACA : ils expriment des états mentaux tels que l'accord, la compréhension, ou encore l'intérêt pendant l'écoute dans le dialogue [Allwood 1991]. Ces rétroactions (*backchannels* en anglais) peuvent être des modalités non-verbales (hochements de tête, le regard et le sourire) ou verbales (énoncés courts, tels que « oui » pour signifier l'accord par exemple).

Un certain nombre de travaux ont intégré des systèmes de rétroactions chez les ACA [Gratch 2007, Maatman 2005, Morency 2010]. Les travaux réalisés dans le cadre du projet SAL[2] [Bevacqua 2008, Schroder 2012] ont intégré dans l'ACA

[2] *Sensitive Artificial Listenener*, qui fait partie du projet européen SEMAINE,

Greta [Poggi 2005] un comportement d'écoute dans le dialogue. L'utilisateur parle à l'ACA *via* un micro et une caméra : le système est capable d'analyser la prosodie du discours et les gestes de l'utilisateur, à partir desquels il va construire une représentation de son niveau d'intérêt. Ce niveau d'intérêt est utilisé pour adapter la fréquence des rétroactions : si le niveau baisse, c'est que l'utilisateur peut vouloir stopper la conversation, et donc la fréquence des rétroactions est baissée, et inversement.

Les auteurs se sont intéressés notamment à l'influence de la personnalité dans ces rétroactions. Ils proposent pour cela quatre personnalités différentes de l'ACA Greta, incarnées par quatre personnages virtuels ([McRorie 2009], cf. figure 5.5) :

- Spike est agressif et pinailleur ;
- Poppy est extravertie et optimiste ;
- Obadiah est pessimiste et dépressif ;
- et Prudence est pragmatique et pratique.

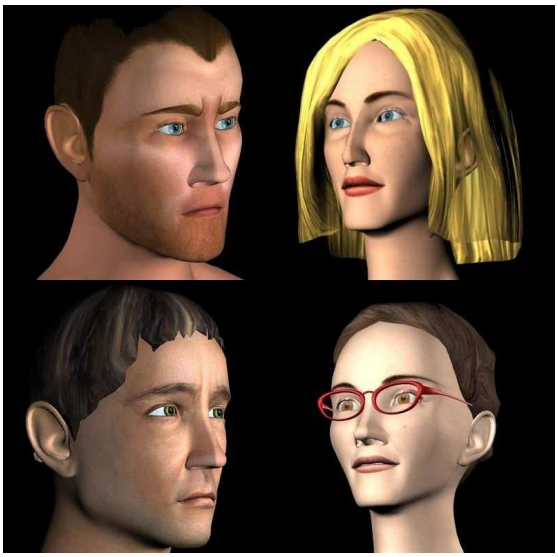

Fig. 5.5 – Les quatre personnalités de l'ACA Greta. De gauche à droite : Spike, Poppy, Obadiah et Prudence.

Chaque personnalité est décrite en fonction de ses préférences de modalités (hochements de tête, regard, gestes...) et de son expressivité, traduite dans ses mouvements (fréquence, vitesse, volume spatial, énergie, fluidité et répétition).

Douze intentions communicatives, présentes dans l'écoute dans le dialogue, ont été implémentées d'après la taxonomie d'Allwood [Allwood 1991] : elles expriment l'accord, le désaccord, l'acceptation, le refus, la croyance, l'incrédulité, l'intérêt, le désintérêt, la compréhension et l'incompréhension, et si l'ACA aime ou pas. Chacun des personnages a une propension à utiliser des rétroactions plutôt positives (Poppy et Prudence) ou négatives (Obadiah et Spike). Ces intentions communicatives se traduisent *via* une bibliothèque de comportements (fichier lexique) par les modalités citées ci-dessus : non-verbales (hochements,

http ://www.semaine-project.eu/

regard, sourire) et verbales (énoncés courts).

Le système déclenche les rétroactions de trois manières différentes :

- de manière réactive : ce déclenchement se base sur des règles probabilistes (lorsque l'humain fait tel geste, l'ACA répond par l'expression de l'intention la plus probable dans la communication) ;
- de manière cognitive : ce déclenchement se base sur les états mentaux de l'agent, définis dans un fichier XML à partir de la personnalité ;
- par imitation : ce déclenchement détermine quand et quels signaux de l'humain imiter (principalement hochement de tête et sourire).

Ces travaux montrent l'importance de l'expression de certains états mentaux de l'interlocuteur (celui qui écoute) dans le dialogue. Cependant, le but de ce système n'est pas de construire un dialogue (pas de reconnaissance des actes de l'humain) mais d'entretenir l'interaction avec l'humain grâce à ces rétroactions. Les états mentaux de l'agent sont de plus définis statiquement suivant sa personnalité, et ne sont pas mis à jour au cours du dialogue.

5.2 Modèles de dialogue chez les ACA

Cette section présente deux modèles de dialogue qui viennent enrichir les ACA de FearNot ! et du projet MRE présentés précédemment.

5.2.1 FearNot !

FearNot ! propose une taxonomie des actes de discours particulière au système [Aylett 2004]. Les auteurs se placent en effet dans le contexte de la maltraitance et construisent, à partir de l'ontologie de ce contexte, un langage entre agents et un langage agent - humain. Dans le langage entre agents, trois catégories d'actes sont définies : la catégorie "Help" (demander de l'aide, offrir de l'aide, offrir sa protection...), la catégorie "Confrontation" (ordonner, interdire, se méfier...) et la catégorie "Socialising" (saluer, introduire un sujet, inviter...). Ces actes font partie des plans et des possibilités d'action des agents et leur sélection est donc influencée par les émotions des agents et leur personnalité (agresseur *vs.* agressé). Des règles de langage ont également été mises en place pour spécifier les actes à effectuer après chaque acte reçu : après un "ordre" ou une "interdiction" par exemple, un agent peut "nier", "obéir", "demander pourquoi" ou "supplier". Ces actes s'expriment par des énoncés oraux, sans lien avec d'autres expressions multimodales.

Le dialogue entre l'agent John et l'utilisateur se fait *via* une interface textuelle (cf. figure 5.6). Deux catégories d'actes sont définies ici : "Agent" (demander un conseil, remercier l'utilisateur, confirmer...) et "User" (conseiller, ignorer, justifier...). Le système reconnait les actes de l'utilisateur grâce à de la reconnaissance de mots-clés ; l'agent peut également guider le dialogue en suggérant des phrases à l'utilisateur.

Plusieurs aspects de ce travail le rendent très intéressant, notamment la mise à jour des états mentaux des agents et le déclenchement et l'influence des émotions sur les actions des agents. Cependant, le manque de généricité des langages mis en place (liés au contexte de la maltraitance) et les émotions considérées dans le système (pas d'émotions complexes mis à part l'espoir)

FIG. 5.6 – Interaction entre l'agent John et l'utilisateur.

rendent ce système difficilement applicable au dialogue agent compagnon - humain.

5.2.2 Le projet MRE

Traum et al. [Traum 2002] ont introduit un modèle de dialogue dans le projet MRE, basé sur la taxonomie des actes de discours de Traum [Traum 1992]. Cette taxonomie reprend certains des actes de discours assertifs et directifs de la théorie des Actes de Discours [Searle 1969, Searle 1985] : ce sont des actes "orientés-tâches", dans le sens où ils contribuent à l'échange / la demande d'information, etc. ; on trouve les actes affirmer, demander une information (*info-request*), ordonner, suggérer et faire une requête. D'autres actes concernent, dans le dialogue, le tour de parole (*take-turn*, *request-turn*...), la mise en place d'un nouveau sujet de conversation (*start-topic*,*end-topic*), la prise de contact ou encore l'attention de l'agent (*give-attention*, *withdraw-attention*...).

Ce modèle de dialogue est relié au modèle affectif EMA, présenté au chapitre 4 : ces deux modèles s'influencent l'un l'autre. Le modèle de dialogue influence le modèle affectif en indiquant à celui-ci le sujet de la conversation, i.e. l'événement qui va être évalué par EMA.

Le modèle affectif influence également le modèle de dialogue : par exemple [Traum 2004], si on demande à l'agent « que se passe-t-il ici ? », le modèle affectif joue sur la réponse de l'agent en indiquant ce sur quoi porte son émotion la plus forte. EMA influence également la génération en langue naturelle des actes sélectionnés par le module de dialogue [Fleischman 2002]. La méthode utilisée ici est la mise en place d'un lexique du domaine du scénario joué ("the driver", "collided with", "was hit", etc.) et l'affectation d'un poids émotionnel à chaque entrée de ce lexique. Le problème est donc ramené à une tâche de minimisation des distances : suivant ses émotions, l'agent va choisir les mots du lexique qui minimisent la distance avec ces états. Ainsi, si l'agent souhaite présenter l'information de façon très détachée, il dira « il y a eu un accident » ; à l'inverse, s'il se sent coupable de l'accident et qu'il veut minimiser sa responsabilité, il dira

par exemple « ils nous sont rentrés dedans ».

L'expression de ces actes en langue naturelle s'accompagne d'expressions multimodales générées grâce au système BEAT (*Behavior Expression Animation Toolkit*, [Breitfuss 2008]). Ce système définit un ensemble de règles qui associent des gestes à un énoncé et dirigent le regard de l'agent. Ces règles sont des règles probabilistes (e.g. les premiers mots d'un énoncé sont associés à un regard vers l'interlocuteur 80% du temps) ; la figure 5.7, tirée de [Breitfuss 2008], montre un exemple d'énoncé annoté.

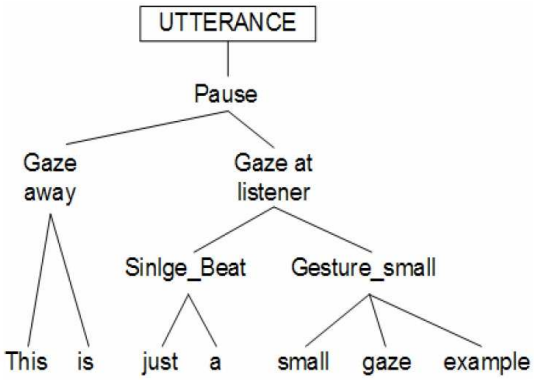

FIG. 5.7 – Annotation automatique de l'énoncé par le système BEAT.

On peut néanmoins regretter que cette annotation ne prenne pas en compte les états mentaux de l'agent en s'appuyant principalement sur la syntaxe de l'énoncé.

Certains aspects de ces travaux sont intéressants, notamment la façon dont les modèles de dialogue et affectif s'influencent. Ce modèle de dialogue "orienté-tâche" est adapté à un tel scénario d'entrainement mais ne convient pas à un dialogue compagnon artificiel - humain. De plus, les émotions prises en compte s'expriment non-verbalement, et les états mentaux des agents ne sont pas formalisés.

6
Conclusions de l'état de l'art

Ce chapitre fait un résumé des différents travaux dans le domaine des ACA présentés dans cette partie, et conclut en mettant en évidence le manque de liens entre la modalité langagière des ACA et les autres modalités dans l'expression d'états mentaux complexes, tels que les buts, croyances et intentions des agents.

6.1 Résumé

Une majorité de travaux dans le domaine des ACA accordent une place importante aux émotions, à leur influence sur le comportement des agents et à leur expression. Ces travaux se basent sur des théories psychologiques des émotions pour 1/ déclencher les émotions des agents et 2/ exprimer multimodalement les émotions. J'ai présenté dans le chapitre 3 quelques théories évolutionnistes des émotions, d'une part, qui s'intéressent principalement à l'expression des émotions et à leur fonction communicative, et d'autre part quelques théories de l'évaluation cognitive, qui se concentrent sur le processus de déclenchement des émotions.

Les différentes modalités d'expression des émotions, tels que le regard, les expressions faciales et les gestes, ont été présentées dans le chapitre 4. Dans la plupart des travaux sur les ACA, une ou plusieurs modalités sont utilisées pour exprimer les émotions des agents. Ce même chapitre présente également quelques exemples d'ACA affectifs, qui ont des modèles des émotions s'appuyant sur des théories psychologiques : parmi ces ACA, certains se basent sur des logiques de type BDI (Belief, Desire, Intention, [Rao 1991]) pour mettre à jour d'autres états mentaux qui influencent le déclenchement de leurs émotions. Les émotions des agents restent néanmoins les seuls états mentaux exprimés. Un exemple de "scripting" de scène est également présenté.

Un troisième chapitre (chapitre 5) a présenté un certain nombre d'ACA qui prennent en compte et expriment d'autres types d'états mentaux, tels que les croyances, les buts et les intentions. Au sein de l'architecture SAIBA, plusieurs travaux ont proposé des lexiques de comportements pour traduire ces intentions communicatives en signaux multimodaux. D'autres travaux présentés dans ce même chapitre se sont concentrés sur le langage des agents pour exprimer leurs états mentaux, en introduisant des taxonomies d'actes de discours, parfois peu adaptées à un dialogue ACA - humain.

Le tableau 6.1 récapitule l'ensemble de ces différents travaux sur les ACA, les états mentaux qu'ils prennent en compte et leurs moyens d'expression de ces états mentaux.

ACA / Modèles	**Etats mentaux pris en compte**	**Expression multimodale des états mentaux**
MARC (section 4.2.1)	**Emotions** : théorie cognitive de Scherer	Expressions faciales
MARC v2	**Emotions** : Scherer et théorie sociale de Manstead	Expressions faciales
MAX (WASABI) (section 4.3.1)	**Emotions** : théorie dimensionnelle PAD	Expressions faciales
	Etats mentaux BDI	Enoncés
FearNot ! (sections 4.3.2 et 5.2.1)	**Emotions** : théorie d'OCC	Expressions faciales
	Etats mentaux BDI	Enoncés
Un agent empathique rationnel (section 4.3.3)	**Emotions** : théorie d'OCC	Expressions faciales : emotions "complexes", masquées et composées
	Etats mentaux de type BDI : théorie de Sadek	$\emptyset$
MRE (EMA + modèle de dialogue) (sections 4.2.2 et 5.2.2)	**Emotions** : théorie cognitive de Lazarus	Expressions faciales
	Etats mentaux BDI	Actes de discours de Traum
SEMAINE (section 5.1.3)	Rétroactions	Regard, sourire, mouvements de la tête et énoncés courts
CrossTalk (section 4.2.3)	**Emotions** : théorie d'OCC	Prosodie, regard, expressions faciales
	$\emptyset$	Gestes de conversation liés aux actes de discours
Expressions faciales performatives (section 5.1.2)	$\emptyset$	Expressions faciales d'"unités cognitives"

TAB. 6.1 – Les états mentaux et leur expression chez les ACA

6.2 Conclusion

A la vue de ce récapitulatif, on peut remarquer que la majorité des travaux s'intéresse à un état mental particulier, l'émotion, et à son déclenchement et son expression multimodale. Les émotions qui sont déclenchées sont souvent de "bas-niveau" (par opposition à des émotions plus complexes, issues de raisonnements cognitifs), et sans lien avec la modalité langagière dans leur expression. En revanche, certains travaux explicitent et manipulent des états mentaux de type BDI des ACA, pour déduire les émotions des agents ; toutefois, ces états mentaux ne sont eux-mêmes que rarement exprimés.

Selon la théorie des Actes de Discours, ces états mentaux de type BDI sont portés par des actes de discours, la plupart du temps non-expressifs, qui s'expriment par le langage **en lien** avec les autres modalités. Ce lien fait défaut chez la plupart des ACA exprimant ces états mentaux, comme on a pu le voir dans le chapitre 5. Ce manque de liens entre le langage, les autres modalités et les émotions est préjudiciable pour la crédibilité de l'ACA, notamment pour les émotions plus cognitives (e.g. les émotions "complexes" d'Oatley et Johnson-Laird telles que le remords ou la culpabilité), issues de raisonnements complexes, qui ne sont pas facilement reconnaissables exprimées sur un visage (en opposition avec les émotions "primaires"). On a besoin, dans le cadre d'un dialogue par exemple, du contexte de ce dialogue pour déterminer ce type d'émotions chez l'interlocuteur, et donc d'un lien entre langage, multimodalité et états mentaux pour les exprimer. Un langage de conversation pour ACA basé sur ce lien, que je présente dans la partie suivante, répond à la question « Comment exprimer l'intention communicative (les états mentaux) de l'ACA ? ».

Décider quels états mentaux exprimer dans un certain contexte reste un problème complexe, auquel la plupart des travaux tente de répondre en privilégiant les émotions de l'ACA. Dans le cadre d'un dialogue entre un ACA et un humain, peu de travaux s'intéressent à la décision de l'intention communicative. Sous l'hypothèse d'un agent sincère, je présente pour ma part dans la partie IV un moteur de raisonnement qui met à jour les états mentaux de l'agent, déclenche ses émotions et sélectionne son intention communicative dans une interaction avec l'humain.

Troisième partie

Langage de Conversation Multimodal

Introduction

Cette troisième partie présente un premier travail effectué au niveau de l'expression des états mentaux de l'agent : on s'intéresse ici à la **crédibilité** de l'expression de la sincérité de l'ACA (« comment l'ACA exprime-t-il ce qu'il a décidé ? ») ; dans la partie suivante (partie IV), on s'intéressera à la **déduction des états mentaux** de l'ACA et à la sélection de son intention communicative (« comment l'ACA décide-t-il de ce qu'il va exprimer ? »).

Comme on l'a vu dans la partie précédente, le langage reste la modalité principale pour exprimer des états mentaux qui ne sont pas des émotions "primaires", comme des émotions plus "complexes" (gratitude, admiration, regret...) ou des croyances, des buts, des intentions etc. Cependant, pour que l'ACA soit crédible dans l'expression de ses états mentaux, il doit accompagner son expression langagière d'expressions multimodales appropriées (e.g. un sourire quand il exprime de la gratitude ou l'expression de son regret quand il s'excuse). Les travaux actuels (cf. chapitre 6) ne font pas état d'un tel lien entre la modalité langagière des ACA, ses états mentaux et les autres modalités. Or, il est nécessaire de mettre en place ce lien, et ce pour deux raisons principales :

1. Rendre **sincère** un ACA affectif et expressif **en conservant sa crédibilité** implique une forte cohérence entre son comportement verbal, non-verbal et ses états mentaux ;
2. Permettre à l'ACA de **manifester des états mentaux complexes** en prenant en compte notamment une certaine catégorie d'émotions qui s'expriment principalement par le langage et qui sont issues d'un raisonnement complexe, impliquant des représentations de la situation, des causes etc. (regret, gratitude... cf. chapitre 8).

Le Langage de Conversation Multimodal (LCM) que nous proposons fait ce lien : ce langage s'appuie sur la théorie philosophique des Actes de Discours [Searle 1969, Searle 1985] et sur la définition d'émotions que nous appelons *complexes*, fortement liées au langage. Le LCM se présente donc sous la forme d'une bibliothèque d'actes liés à des états mentaux et à une expression multimodale de ces états mentaux : par exemple, l'acte *Remercier* est lié à l'émotion *complexe* de la gratitude qui s'exprime multimodalement d'une façon particulière.

Le premier chapitre présente la théorie des Actes de Discours et le lien entre actes et états mentaux (chapitre 7). Le chapitre suivant définit ce que sont les émotions *complexes*, leur place dans les différentes théories des émotions, et leur formalisation logique (chapitre 8). Le LCM proprement dit et les Actes de Conversation Multimodaux (ACM) qui le composent sont détaillés dans le chapitre 9. Le LCM a été développé pour être compatible avec le standard SAIBA [Vilhjalmsson 2007] ; son implémentation dans l'ACA Greta [Poggi 2005] a permis d'évaluer son influence sur la perception que les utilisateurs se font de la crédibilité et la sincérité de l'ACA. L'implémentation et les résultats de cette évaluation sont présentés dans le chapitre 10.

7
Appuis théoriques

Ce chapitre présente dans un premier temps la théorie des Actes de Discours, théorie philosophique de l'utilisation du langage humain, et dans un second temps ses liens avec les états mentaux et les émotions.

7.1 La théorie des Actes de Discours

La théorie des Actes de Discours [Searle 1969, Searle 1985] étudie l'usage du langage et sa compréhension, et cherche à identifier ce qu'elle appelle les "unités premières de signification", c'est-à-dire ce qui fait l'unité minimale dans la communication humaine. La théorie se base notamment sur les travaux d'Austin [Austin 1962], qui distingue les trois types d'actes de discours qui existent lors d'une énonciation :

1. l'acte locutoire (ou d'énonciation) : l'acte même de dire quelque chose ;
2. l'acte illocutoire : le but dans lequel l'énoncé est exprimé, la direction d'ajustement du langage au monde (poser une question, affirmer, demander...) ;
3. l'acte perlocutoire : l'effet que produit l'acte sur l'auditoire.

La théorie des Actes de Discours affirme que ce sont **les actes illocutoires** qui sont les unités premières de la signification dans l'usage et la compréhension des langues naturelles. D'après Vanderveken [Vanderveken 1988] : « *Toute pensée conceptuelle (croyance, désir, intention) est en principe exprimable lors de l'accomplissement de tels actes de discours* ». Searle et Vanderveken [Searle 1985] affirment qu'il existe seulement cinq usages des actes illocutoires qui correspondent aux cinq différentes façons fondamentales de s'exprimer : l'usage assertif, directif, engageant, déclaratif et expressif. Ces cinq usages (qui sont les "classes" d'actes) sont détaillés dans le tableau 7.1.

7.1.1 Formes des actes illocutoires

Tout acte[1] est de la forme F(P), où P est le contenu propositionnel énoncé et F la *force illocutoire* qui détermine dans quel but l'énoncé est formulé [Vanderveken 1988]. Par exemple, l'énoncé « ferme la fenêtre ! » porte un acte à usage directif où F représente un ordre et P le contenu propositionnel « fermer la fenêtre ». La force

[1] Pour améliorer la compréhension, la notion d'acte ou d'acte de discours dans la suite du manuscrit fera référence seulement aux actes illocutoires.

Usages des actes illocutoires	But de l'usage	Exemples d'actes
Assertifs	Représenter un fait comme étant actuel dans le monde	Informer, Croire, Rappeler
Directifs	Essayer de faire en sorte qu'une autre personne accomplisse une action future	Demander, Exiger, Suggérer
Engageants	S'engager à accomplir une certaine action	Promettre, Accepter, Offrir
Déclaratifs	Accomplir une action simplement par l'énonciation	Déclarer, Condamner, Renier
Expressifs	Exprimer un état psychologique à propos d'un certain fait	S'excuser, Féliciter, Déplorer

TAB. 7.1 – Les cinq classes d'actes illocutoires

illocutoire F est constituée de six composantes qui expriment chacune un aspect différent de l'acte. Ces composantes sont les conditions à respecter pour que l'acte soit accompli *avec succès* :

- *le but illocutoire* de l'acte représente l'usage que l'on en fait ; comme indiqué plus haut, le but peut être assertif, engageant, directif, déclaratif ou expressif ;
- *les conditions de sincérité* : les états mentaux du locuteur (et exprimés par lui) s'il est sincère ; par exemple, « *un locuteur [sincère] qui fait une promesse exprime l'intention de faire ce qu'il promet* » [Vanderveken 1988] ;
- *les conditions préparatoires* et *les conditions sur le contenu propositionnel* représentent le contexte d'énonciation ; par exemple, une promesse doit porter sur une action future (conditions sur le contenu propositionnel) et est supposée bonne pour l'allocutaire (conditions préparatoires), à l'inverse d'une menace ;
- *le mode d'accomplissement* : la manière dont le but illocutoire doit être atteint ; il contribue, par exemple, à différencier une demande d'un ordre, car une demande doit donner une option de refus à l'allocutaire (« peux-tu me passer le sel ? ») alors qu'un ordre invoque une position d'autorité (« que ça te plaise ou non, passe-moi le sel ! ») ;
- le *degré de puissance* de l'acte : le degré avec lequel sont exprimés les états mentaux ; par exemple, ce sont les mêmes états mentaux qui sont exprimés quand le locuteur conseille ou affirme (désir que l'autre croit ce qu'on lui dit), mais le degré de puissance de ce désir est plus fort quand il affirme.

Lorsqu'un tel acte est accompli *avec succès*, on dit en philosophie du langage qu'il est "performé" par le locuteur (de l'anglais *perform*, accomplir). La théorie des Actes de Discours définit aussi la notion de satisfaction d'un acte : un acte est satisfait lorsque ses conditions sur le contenu propositionnel sont respectées.

7.1.2 Verbes illocutoires

Pour déterminer quel acte illocutoire est porté par un énoncé, il existe un certain nombre d'indices dans la syntaxe et la morphologie même de l'énoncé qui aident à identifier le but illocutoire - l'usage - (ponctuation, intonation, mode

du verbe...) et le contenu propositionnel (sujet du verbe, conjugaison...). Dans un énoncé, le verbe tient une place importante car il peut également expliciter la force illocutoire F : s'excuser, refuser, jurer, promettre... Ces types de verbes sont appelés verbes performatifs, en ce sens où ils nomment explicitement une force illocutoire : par exemple, "je m'excuse de..." et "je refuse que..." nomment clairement l'acte expressif *S'excuser* et l'acte directif *Refuser*. A l'inverse, d'autres verbes représentant l'acte n'ont pas d'usage performatif, et nomment donc implicitement la force illocutoire : on ne se vante pas, par exemple, en disant "je me vante de...", bien que l'acte soit désigné par *Se vanter*. Ces verbes explicites et implicites forment l'ensemble des verbes illocutoires. De manière générale et pour faciliter la compréhension, on confond les notions mêmes d'acte illocutoire et de verbe en nommant l'acte par le verbe qui porte sa force illocutoire : on pourra donc avoir les actes *Affirmer*, *Offrir*, *Demander*, *Féliciter* etc.
Vanderveken [Vanderveken 1988] recense en tout 163 verbes illocutoires (ou actes) dans la langue française, et propose pour chaque usage un tableau sémantique qui relie les verbes selon leur force illocutoire et leur degré de puissance (cf. figure 7.1).

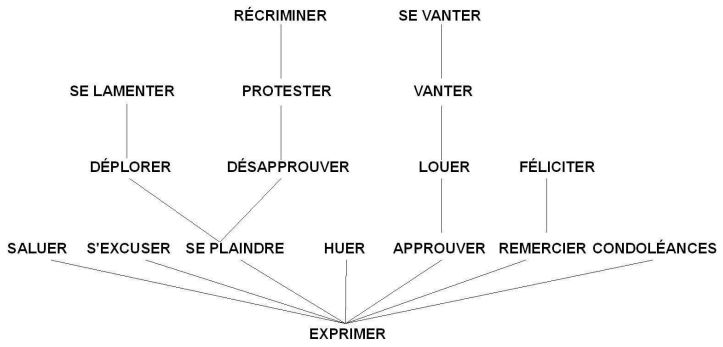

FIG. 7.1 – Tableau sémantique pour les verbes illocutoires d'usages expressifs [Vanderveken 1988]

Vanderveken et Chaib-draa [Chaib-draa 1999] ont également introduit un ensemble d'opérateurs en calcul des situations (ou logique situationnelle) afin de formaliser certains aspects de la théorie ; je n'aborderai pas ici ce point pour introduire dans le chapitre suivant (chapitre 8) une nouvelle formalisation permettant de représenter des raisonnements "complexes".

7.1.3 Travaux connexes

Un grand nombre de travaux se sont inspirés de la théorie des Actes de Discours. Le but ici n'est pas de faire un inventaire exhaustif de ces travaux, mais d'illustrer les différences d'utilisation de cette théorie.

Cohen, Perrault et Allen [Cohen 1979, Allen 1986] proposent, à la fin des années 1970, de considérer les actes de discours comme des opérateurs au sein de systèmes de planification, formalisés en termes de préconditions et d'effets à partir des croyances et des buts des agents. Ces théories à base de plans (*plan-based theories of speech act* en anglais) ont par la suite été critiquées de

par la nature de l'intention des agents. Elles introduisent cependant des idées intéressantes que je développe pour le moteur de raisonnement (cf. chapitre 11).

La théorie des Actes de Discours a servi de base à la plupart des langages de communication entre agents (ACL) dans le domaine des Systèmes Multi-Agents (SMA), tels que ACL/FIPA[2] et KQML[3]. Ces langages se composent essentiellement de quelques actes *assertifs* et *directifs* de la théorie. Ces langages ont été mis en place dans un soucis premier d'échange d'informations et d'interopérabilité entre agents logiciels, et ne sont donc pas adaptés dans le cadre d'une interaction de type conversationnel entre agents virtuels et humains [Bisognin 2004].

Des travaux précédents menés dans l'équipe [Berger 2005] se sont largement inspirés de cette théorie. Ils ont eu pour but de dépasser le simple échange d'information entre agents, et ont permis de mettre en place un *langage de conversation expressif* composé de 32 actes de conversation expressifs définis formellement en calcul des situations. Ces actes ont été sélectionnés à partir des définitions des verbes illocutoires données par [Vanderveken 1988] pour leur pertinence dans des situations comme l'assistance à l'utilisateur, la négociation, la coopération etc. Comme je le montre ci-après (cf. chapitre 9), cette pré-selection nous a permis de construire une partie du Langage de Conversation Multimodal.

L'identification des actes de discours reste une des méthodes les plus utilisées dans l'annotation et l'analyse de dialogue. Plusieurs travaux [Caelen 2003, AMIConsortium 2004] s'inspirent de la taxonomie de la théorie des Actes de Discours pour l'analyse du dialogue. Dans ces travaux, l'identification et la classification des actes dans un énoncé s'effectuent de manière unidimensionnelle : à un énoncé correspond une seule annotation, un seul acte. Ce type d'annotation a été remis en cause notamment par les auteurs de [Core 1997], qui ont proposé une taxonomie multidimensionnelle, DAMSL (*Dialog Act Markup in Several Layers*). Un énoncé peut ainsi être lié à plusieurs actes, qui appartiennent à des fonctions communicatives "forward" (c'est-à-dire à l'initiative du locuteur, tels que "Assert", "Offer", "Commit"...) ou "backward" (c'est-à-dire en réponse à l'interlocuteur, tels que "Accept", "Signal-Understanding", "Agreement"...). Un énoncé peut ainsi être lié à plusieurs actes, ce qui facilite l'annotation et permet une analyse plus fine du discours [Petukhova 2007].

7.2 Les états mentaux et les émotions dans la théorie des Actes de Discours

Certaines des composantes de la force illocutoire F détaillée ci-dessus font le lien que nous recherchons entre actes de langages et états mentaux (dont les émotions). Parmi les six composantes, nous en retenons trois que je présente ici : *la condition de sincérité*, *le mode d'accomplissement* et *le degré de puissance*.

7.2.1 La condition de sincérité

Lorsque le locuteur accomplit un acte, il manifeste également des états mentaux à propos du contenu propositionnel P. Par exemple, « *un locuteur qui*

[2] http ://www.fipa.org
[3] Knowledge Query and Manipulation Language

fait une demande exprime un désir que l'allocutaire fasse ce qu'il demande » [Vanderveken 1988], pour peut qu'il soit sincère ! La *condition de sincérité* représente donc le fait que le locuteur, d'une part, manifeste les états mentaux associés à un acte, et d'autre part possède ces états mentaux ; lorsque l'on ment (que l'on est insincère), on essaie de manifester des états mentaux (du mieux possible) que l'on ne possède pas.

On distingue dans la théorie les conditions de sincérité, d'une part, des actes assertifs, engageants, directifs et déclaratifs, et d'autre part des actes expressifs. Dans le cas des actes assertifs, engageants, directifs et déclaratifs, on peut lier la condition de sincérité au but illocutoire (l'usage que l'on fait de l'acte). Le tableau 7.2 présente les conditions de sincérité liées au but illocutoire des classes d'actes (les expressifs mis à part), telles que définies dans la théorie des Actes de Discours [Berger 2006].

Classes d'actes illocutoires	Conditions de sincérité
Assertifs	Lorsqu'un locuteur *affirme* le fait P, il doit croire en l'existence de P
Directifs	Lorsqu'un locuteur *demande* à l'allocutaire d'effectuer l'action représentée par P, il doit en avoir le désir
Engageants	Lorsqu'un locuteur *s'engage à* effectuer l'action représentée par P, il doit avoir l'intention d'effectuer cette action
Déclaratifs	Lorsqu'un locuteur *déclare* P, il doit avoir la croyance, l'intention et le désir d'accomplir P

TAB. 7.2 – Les conditions de sincérité selon les classes d'actes.

Dans le cas des actes expressifs, la condition de sincérité porte les émotions qu'exprime le locuteur. On considère par exemple l'acte expressif *S'excuser*, qui est l'expression d'une forme de regret : le regret d'être responsable de quelque chose de mauvais pour l'interlocuteur. D'après [Vanderveken 1988] : « *S'excuser, c'est exprimer ses regrets à un allocutaire (condition de sincérité) à propos de l'état de choses représenté par P en présupposant (condition préparatoire) que l'on est responsable de cet état de choses et qu'il est mauvais pour l'allocutaire* ». Ici, le but illocutoire de l'acte est d'exprimer du regret, cette émotion de regret étant bien portée par la condition de sincérité de l'acte. Pour ces actes expressifs, la condition de sincérité relie donc les émotions aux actes : l'expression de ces émotions est d'une part, le but même de ces actes, et d'autre part nécessaire pour que le locuteur paraisse sincère lorsqu'il *s'excuse*, *remercie* etc.

De manière plus générale, **la condition de sincérite forme un lien permanent entre les états mentaux et les différents actes** dans le sens où ces états mentaux sont ceux exprimés lors de l'acte lorsque le locuteur est sincère : l'expression de ces états mentaux lors de l'acte est donc **nécessaire** pour que le locuteur **soit crédible** dans sa sincérité.

7.2.2 Le mode d'accomplissement

Une seconde composante, *le mode d'accomplissement*, détermine comment le but illocutoire doit être accompli et les modalités d'expression des états mentaux (dont les émotions dans le cas des expressifs) portés par la condition de sincérité l'acte. Cette composante contribue à faire la différence entre une demande, par exemple, et un ordre (actes directifs) : dans le premier cas, le locuteur laisse une option de refus à l'allocutaire (par le langage, « si tu veux bien » etc., ou d'autres modalités) ; dans l'autre cas, l'expression est plus péremptoire.

Dans le cas des actes expressifs, considèrons l'exemple de l'acte *Protester* défini par [Vanderveken 1988] : « *Protester, c'est exprimer d'une façon formelle (mode d'accomplissement) sa désapprobation [...]* ». La protestation peut se traduire par le langage (tournures élaborées, rigueur, prosodie particulière...) mais également par la gestuelle, la posture ou l'expression du visage. Ainsi, les modes d'accomplissement permettent de déterminer les modalités d'expression des différentes émotions, que ce soit la modalité langagière ou d'autres modalités. On peut imaginer que cette composante pourra être par la suite influencée par exemple par la personnalité de l'ACA, le type de relation qu'il entretient avec l'utilisateur ou simplement par les modalités qu'il peut exprimer.

7.2.3 Le degré de puissance

D'après sa définition, le degré de puissance représente le degré avec lequel sont exprimés les états mentaux portés par la condition de sincérité. Il forme des liens entre différents actes de même usage, qui expriment un même état mental : par exemple, chez les directifs, *Recommander* est plus fort que *Conseiller* qui lui-même est plus fort que *Suggérer*. Les états mentaux, ici, sont les mêmes (avoir le désir que l'allocutaire fasse quelque chose), mais ne sont pas exprimés avec le même degré de puissance.

Dans le cas des expressifs, le degré de puissance représente donc le **degré d'intensité de l'émotion** portée par la condition de sincérité : exprimer de la désapprobation à un degré faible, c'est *Désapprouver*, *Protester* à un degré plus fort, et *Récriminer* à un degré encore au-dessus [Vanderveken 1988].

En général, le degré de puissance d'un acte s'exprime dans un énoncé en utilisant la ponctuation et des adverbes tels que « sincèrement » ou « franchement ». On peut également imaginer une différence dans l'expression faciale (traits plus ou moins accentués), les gestes (plus ou moins brusques) etc.

8

Formalisation logique des émotions et des actes expressifs

L'objectif que nous poursuivons ici est la mise en place d'un langage de conversation qui permet à un ACA sincère d'exprimer l'ensemble de ses états mentaux par le langage et les autres modalités. Etant donnée l'importance des émotions dans la communication et leur rôle dans la crédibilité de l'ACA, il nous faut prendre en compte ces émotions et leur expression multimodale : nous nous concentrons donc dans ce chapitre sur les actes *expressifs* de la théorie des Actes de Discours dont l'usage est l'expression des émotions.

Nous nous intéressons notamment à une catégorie particulière d'émotions que nous appelons *complexes*, qui s'expriment principalement par le langage et qui sont basées sur des raisonnements complexes comme le raisonnement contrefactuel [Kahneman 1986]. Ce chapitre présente tout d'abord la définition de ces émotions *complexes* et leur formalisation logique (le modèle BIGRE), puis détaille la formalisation des actes expressifs qui expriment ces émotions *complexes*. La définition des émotions complexes et la formalisation logique de ces émotions et des actes expressifs ont été réalisées, dans le cadre du projet ANR CECIL, par l'équipe LiLaC de Toulouse, avec une modeste contribution de notre part [Guiraud 2011].

8.1 Définition des émotions complexes

Comme nous avons pu le voir dans les théories cognitives des émotions (cf. chapitre 3), l'*émotion* au sens large du terme inclut une réaction physiologique, une expression motrice, une motivation à l'action et un ressenti subjectif, et est issue d'un processus d'évaluation cognitive. Pour notre part, nous nous plaçons dans une vue exclusivement cognitive des émotions : nous considérons l'émotion comme **un état mental particulier**, correspondant à une structure cognitive d'une certaine forme (basée sur des croyances, des buts et des raisonnements plus complexes). Cette définition est suffisante dans le cadre de nos travaux, car nous nous intéressons aux états mentaux qui sont en priorité exprimés par le langage.

8.1.1 Responsabilité et raisonnement contrefactuel

Les émotions que nous appelons *complexes* appartiennent à une catégorie bien particulière des émotions : celles basées d'une part sur le **raisonnement**

contrefactuel de l'agent qui va induire le concept de **responsabilité**, et d'autre part sur son raisonnement à propos de ses **capacités** et celles des autres agents et des **normes morales et sociales**.

Le raisonnement contrefactuel consiste à comparer la réalité avec une vue imaginaire de ce qui aurait pu se produire [Kahneman 1986]. Le concept de responsabilité est déduit du raisonnement contrefactuel : si l'on avait les capacités d'empêcher un évènement **menant à la situation actuelle**, alors on est responsable de cette situation (qui, si elle est mauvaise, induit des émotions comme le regret, ou à l'inverse, la réjouissance [Lazarus 1991, Kahneman 1986, Zeelenberg 1998]). Les normes morales et sociales (*internalisées* par l'agent) interviennent également dans le raisonnement contrefactuel pour déclencher des émotions comme, par exemple, la culpabilité ou la honte [Tangney 2003] : un agent se sent coupable de quelque chose s'il croit qu'il est responsable d'avoir violé une certaine norme (qu'il se sera appropriée), et qu'il aurait pu empêcher cet évènement d'arriver.

Le choix de ces émotions s'explique d'abord par l'importance de ces émotions dans le langage. En effet, à l'inverse des émotions primaires qui s'expriment *via* des expressions faciales prototypiques, ces émotions ont besoin du contexte (et donc du langage) pour être exprimées et reconnues par un interlocuteur [de Gelder 2011]. En neurosciences par exemple, Barret et al. [Barrett 2007] affirment que le langage participe au contexte pour la perception et la compréhension des émotions. Dans les théories psychologiques des émotions (cf. chapitre 3), l'importance du langage dans ce type d'émotions a été souligné par Oatley et Johnson-Laird [Oatley 1987] (qui parlent également d'émotions "complexes").

Ce type d'émotions tient également une place importante dans le raisonnement cognitif de l'agent, comme le montrent les théories de l'évaluation cognitive ([Scherer 2001, Ortony 1988], cf. chapitre 3). Par exemple, on retrouve parmi les checks de Scherer les notions de congruence avec les buts, rapport aux normes morales et sociales et attribution de la responsabilité. Dans un effort pour modéliser les états mentaux d'un agent, il est donc important de modéliser les émotions *complexes*.

8.1.2 Formalisation Logique : le modèle BIGRE

La formalisation proposée par l'équipe LiLaC se base sur les opérateurs de la logique **MLC** (*Modal Logic of Communication*) dans la lignée des logiques BDI [Rao 1991] : ces opérateurs permettent de représenter d'une part les états mentaux de l'agent et d'autre part l'aspect explicite et social de la communication [Guiraud 2011], comme détaillé ci-dessous. Ces opérateurs sont :

- (B) $Bel_i\varphi$: représente la croyance qu'a l'agent i en φ ;
- (I) $Ideal_i\varphi$: idéalement pour l'agent i, φ devrait être vrai ; cet opérateur correspond aux normes de l'agent, tant morales que sociales ;
- (G) $Goal_i\varphi$: l'agent i a pour but que φ soit vrai ;
- (R) $Resp_i\varphi$ est l'abréviation de $Cd_i\neg\varphi \wedge \varphi$; l'agent i pouvait (Cd_i, i.e. *could* en anglais) empêcher la situation φ, i est donc responsable de φ[1].

Certains états mentaux représentés par les BIGR sont des émotions *complexes* (E) présentées ci-dessus ; de certaines combinaisons logiques des opérateurs

[1]Par abus de langage et par soucis de lisibilité, je considère R comme un opérateur à part entière dans la suite du manuscrit.

BIGR, on peut donc déduire des émotions E (BIGR → E) : nous appelons ce modèle le **modèle BIGRE**.

Cette formalisation permet de définir en tout douze émotions de l'agent (cf. tableau 8.1), par combinaison des opérateurs B,I,G et R. Le tableau 8.1 montre la formalisation de ces émotions pour un agent i, concernant une situation φ.

$\wedge$	$Bel_i\varphi$	$Bel_i Resp_i\varphi$	$Bel_i Resp_j\varphi$
$Goal_i\varphi$	$Joie_i\varphi$	$Rejouissance_i\varphi$	$Gratitude_{i,j}\varphi$
$Goal_i\neg\varphi$	$Tristesse_i\varphi$	$Regret_i\varphi$	$Deception_{i,j}\varphi$
$Ideal_i\varphi$	$Approbation_i\varphi$	$SatisfactionMorale_i\varphi$	$Admiration_{i,j}\varphi$
$Ideal_i\neg\varphi$	$Desapprobation_i\varphi$	$Culpabilite_i\varphi$	$Reproche_{i,j}\varphi$

TAB. 8.1 – Formalisation des émotions complexes pour un agent i suivant la situation φ.

Quatre de ces émotions sont ce que nous appelons des émotions *de base*, définies sans la notion de responsabilité (première colonne dans le tableau 8.1) :

- la joie ; l'agent croit qu'un de ses buts est atteint - $Goal_i\varphi \wedge Bel_i\varphi$;
- la tristesse ; l'agent croit qu'un de ses buts a été empêché - $Goal_i\neg\varphi \wedge Bel_i\varphi$;
- l'approbation ; l'agent croit qu'un de ses idéaux est atteint - $Ideal_i\varphi \wedge Bel_i\varphi$;
- la désapprobation ; l'agent croit qu'un de ses idéaux a été violé - $Ideal_i\neg\varphi \wedge Bel_i\varphi$.

Par exemple, l'agent a pour but qu'il fasse beau (*$Goal_i$ grand soleil*) ; s'il apprend ensuite qu'il fait mauvais (*Bel_i gros nuages*), l'agent aura de la tristesse car son but qu'il fasse beau a été empêché. Dans cet exemple, personne (d'identifiable pour l'agent) n'est **responsable** du temps qu'il fait.

La prise en compte de la notion de responsabilité, issue du raisonnement contrefactuel sur ses propres actions et celles d'autrui, permet de formaliser huit émotions supplémentaires : les émotions *complexes* de regret, déception, culpabilité, reproche, satisfaction morale, admiration, réjouissance et de reconnaissance ou gratitude. Par exemple, l'émotion de culpabilité correspond à un regard porté par l'agent sur ce qu'il a fait et ce qu'il aurait pu faire à un instant donné : l'agent croit qu'un de ses idéaux a été violé ($Ideal_i\neg\varphi$) et qu'il est **responsable** de la situation ($Bel_i Resp_i\varphi$). De même, le reproche est dû à la comparaison par l'agent des capacités de l'interlocuteur et de ce qu'il a fait : l'agent croit qu'un de ses idéaux a été violé ($Ideal_i\neg\varphi$) et que l'*interlocuteur* est **responsable** de la situation ($Bel_i Resp_j\varphi$).

Ces émotions complexes, en tant qu'états mentaux, **impliquent** les émotions de base. Prenons l'exemple du regret :

$$Regret_i\varphi \stackrel{déf}{=} Bel_i Resp_i\varphi \wedge Goal_i\neg\varphi$$

L'agent i croit qu'il est responsable de la situation φ, opposée à son but.

Or, d'après la définition de la responsabilité : $Bel_i Resp_i\varphi \rightarrow \varphi$. On a donc, d'après la définition de la tristesse donnée précédemment, le théorème suivant :

$$Regret_i\varphi \rightarrow Tristesse_i\varphi$$

Ainsi, d'après notre formalisation, lorsque l'on ressent du regret (ou de la déception) pour quelque chose, on ressent également de la tristesse. De la même façon, la gratitude et la réjouissance impliquent la joie, la satisfaction morale et l'admiration impliquent l'approbation, et la culpabilité et le reproche impliquent la désapprobation.

L'introduction de ces différents opérateurs et la définition et la formalisation des émotions *complexes* est une première étape pour la formalisation des actes expressifs de la théorie des Actes de Discours [Guiraud 2011] et la mise en place de notre Langage de Conversation Multimodal.

8.1.3 Axiomes liés à la formalisation

Les opérateurs logiques Bel_i, $Goal_i$ et $Ideal_i$ introduits par la logique **MLC** satisfont un certain nombre d'axiomes associés à la logique modale [Kripke 1959, Kripke 1963]. Ces axiomes représentent des propriétés que vérifient les opérateurs. Par exemple, chacun de ces opérateurs vérifient les axiomes "d'introspections positive et négative", i.e. être conscient de ses croyances, buts et idéaux (e.g. $Bel_i\varphi \rightarrow Bel_i Bel_i\varphi$). Ces axiomes peuvent être utilisés pour énoncer des théorèmes ou déduire, dans une base de connaissances, de nouvelles connaissances.

8.2 Formalisation des actes expressifs

La définition du modèle BIGRE et des émotions *complexes* nous permet, en accord avec la théorie des Actes de Discours, de formaliser les actes expressifs dont les conditions de sincérité portent ces émotions [Guiraud 2011]. On a pu voir que le but illocutoire des actes expressifs est d'exprimer l'émotion spécifiée par les conditions de sincérité : les actes expressifs que nous souhaitons modéliser ici ont donc comme but d'exprimer les émotions *complexes*. La logique **MLC** introduit pour cela un cinquième opérateur qui représente l'accomplissement d'un acte expressif : $Exp_{i,j,H}\mu(\varphi)$.

$Exp_{i,j,H}\mu(\varphi)$

L'agent i **exprime son émotion** (notée $\mu(\varphi)$) **à propos de la situation** φ à l'agent j, devant un auditoire H ($= \emptyset$ dans le cas d'un dialogue humain-agent).

Nous définissons quatre actes expressifs *de base* à partir des actes expressifs identifiés par Vanderveken [Vanderveken 1988] et des émotions *de base*. Par exemple, exprimer de la tristesse à propos de quelque chose, c'est :

$Exp_{i,j,H}\, Tristesse_i\varphi$
$\equiv Exp_{i,j,H}(Goal_i\neg\varphi \wedge Bel_i\varphi) \stackrel{déf}{=} \mathbf{Pleurnicher}_{i,j}\varphi$

L'agent i exprime à l'agent j sa tristesse par rapport à φ ; l'acte en lui-même est $\mathbf{Pleurnicher}_{i,j}\varphi$

Le tableau 8.2 définit de la même façon les quatre actes expressifs *de base* exprimant les émotions *de base*, c'est-à-dire définies à partir des opérateurs BIG, sans la notion de responsabilité.

$Exp_{i,j,H}(...)$	Actes expressifs *de base*
$Goal_i\varphi \wedge Bel_i\varphi \equiv Joie_i\varphi$	$\stackrel{déf}{=}$ $\mathbf{Saluer}_{i,j}\varphi$
$Goal_i\neg\varphi \wedge Bel_i\varphi \equiv Tristesse_i\varphi$	$\stackrel{déf}{=}$ $\mathbf{Pleurnicher}_{i,j}\varphi$
$Ideal_i\varphi \wedge Bel_i\varphi \equiv Approbation_i\varphi$	$\stackrel{déf}{=}$ $\mathbf{Approuver}_{i,j}\varphi$
$Ideal_i\neg\varphi \wedge Bel_i\varphi \equiv Desapprobation_i\varphi$	$\stackrel{déf}{=}$ $\mathbf{Desapprouver}_{i,j}\varphi$

TAB. 8.2 – Formalisation des actes expressifs exprimant les émotions *de base*

Nous définissons également neuf actes expressifs *complexes* dont le but illocutoire est d'exprimer les émotions *complexes* (cf. tableau 8.3) : chacun de ces actes exprime une émotion *complexe* différente, excepté pour l'acte $\mathbf{S'excuser}_{i,j}\varphi$ qui est un cas particulier de l'expression du regret. La distinction des actes $\mathbf{S'excuser}_{i,j}\varphi$ et $\mathbf{Regretter}_{i,j}\varphi$ permet d'une part de respecter la définition des actes expressifs faite par Vandereken, et d'autre part de représenter deux formes de regret (le regret par rapport à ses propres buts et le regret par rapport aux buts des autres).

$Exp_{i,j,H}(...)$	Actes expressifs *complexes*
$Goal_i\varphi \wedge Bel_iResp_i\varphi \equiv Rejouissance_i\varphi$	$\stackrel{déf}{=}$ $\mathbf{SeRejouir}_{i,j}\varphi$
$Goal_i\varphi \wedge Bel_iResp_j\varphi \equiv Gratitude_{i,j}\varphi$	$\stackrel{déf}{=}$ $\mathbf{Remercier}_{i,j}\varphi$
$Goal_i\neg\varphi \wedge Bel_iResp_i\varphi \equiv Regret_i\varphi$ Cas spécifique : $Regret_i(Goal_j\neg\varphi \wedge \varphi)$	$\stackrel{déf}{=}$ $\mathbf{Regretter}_{i,j}\varphi$ $\stackrel{déf}{=}$ $\mathbf{S'excuser}_{i,j}\varphi$
$Goal_i\neg\varphi \wedge Bel_iResp_j\varphi \equiv Deception_{i,j}\varphi$	$\stackrel{déf}{=}$ $\mathbf{SePlaindre}_{i,j}\varphi$
$Ideal_i\varphi \wedge Bel_iResp_i\varphi \equiv SatisfactionMorale_i\varphi$	$\stackrel{déf}{=}$ $\mathbf{SeSatisfaire}_{i,j}\varphi$
$Ideal_i\varphi \wedge Bel_iResp_j\varphi \equiv Admiration_{i,j}\varphi$	$\stackrel{déf}{=}$ $\mathbf{Adherer}_{i,j}\varphi$
$Ideal_i\neg\varphi \wedge Bel_iResp_i\varphi \equiv Culpabilite_i\varphi$	$\stackrel{déf}{=}$ $\mathbf{Culpabiliser}_{i,j}\varphi$
$Ideal_i\neg\varphi \wedge Bel_iResp_j\varphi \equiv Reproche_{i,j}\varphi$	$\stackrel{déf}{=}$ $\mathbf{Reprocher}_{i,j}\varphi$

TAB. 8.3 – Formalisation des actes expressifs exprimant les émotions *complexes*

$\mathbf{S'excuser}_{i,j}\varphi$, en accord avec la définition de Vanderveken donnée précédemment (cf. section 7.2.1), peut se traduire de cette façon [Guiraud 2011] :

$$\mathbf{S'excuser}_{i,j}\varphi \stackrel{déf}{=} Exp_{i,j,H}(Regret_i(Goal_j\neg\varphi \wedge \varphi))$$

L'agent i exprime à l'agent j du regret par rapport au fait que j n'a pas atteint son but φ, et qu'il en est de plus responsable.

L'ensemble de ces treize actes expressifs, formalisés à l'aide des opérateurs que nous avons introduit, forment le socle du Langage de Conversation Multimodal[2]. Dans le prochain chapitre, je détaille comment, après avoir étendu notre formalisation aux actes des autres classes de la théorie des Actes de Discours, l'ajout de la multimodalité aux actes de discours fait de ces actes des **Actes de Conversation Multimodaux**.

[2]Pour plus de détails sur la formalisation et les liens logiques qui relient ces actes, le lecteur peut se référer à [Guiraud 2011].

« Et si je poussais une longue plainte déchirante pudiquement cachée sous la morsure cinglante de mon humour ravageur ? »

P. Desproges

9 Le Langage de Conversation Multimodal

Dans un premier temps, je présente tout d'abord comment l'introduction du degré de puissance (en tant que composante de la force illocutoire) enrichit l'ensemble des actes expressifs que nous avons définis pour former de nouveaux actes. Je présente ensuite la formalisation, à l'aide du modèle BIGRE, de l'ensemble des actes directifs, assertifs et engageants qui complètent nos actes expressifs (cf. section 9.2) : l'ajout d'autres classes d'actes permet à l'agent virtuel d'avoir à sa disposition un langage le plus complet possible dans le contexte d'un dialogue agent-humain.

L'ajout des expressions multimodales appropriées à ces actes font d'eux des Actes de Conversation Multimodaux (ACM), qui autorisent un agent virtuel à exprimer l'ensemble des états mentaux liés à un acte. Pour les Actes de Conversation Multimodaux expressifs, dont le but illocutoire est d'exprimer une émotion *complexe*, l'expression multimodale sera *a priori* l'expression de l'émotion. Je détaille ci-après ce qu'il en est, de même que le cas des Actes de Conversation Multimodaux assertifs, engageants et directifs qui composent le Langage de Conversation Multimodal (cf. section 9.4).

9.1 Introduction du degré de puissance

L'ajout du degré de puissance, tel que défini dans la théorie des Actes de Discours (cf. section 7.2.3), dans notre définition des actes expressifs, permet de créer de nouveaux actes : on peut imaginer par exemple qu'une émotion *complexe* de satisfaction morale avec une forte intensité amènera l'agent à utiliser l'acte expressif *Se vanter* plutôt que *Se satisfaire*.

Pour limiter le nombre d'actes tout en permettant une certaine granularité, nous avons introduit deux niveaux dans les actes expressifs (en accord avec l'ensemble des verbes recensés par Vanderveken [Vanderveken 1988]) : les actes de degré de puissance neutre et ceux de degré supérieur. Le tableau 9.1 montre l'ensemble des actes expressifs que nous avons définis, selon les émotions qu'ils expriment et les degrés de puissance de ces expressions. Nous considérons les actes expressifs exprimant les émotions *de base* comme étant de degré neutre, et ne possédant pas de degré supérieur de part la nature de ces émotions : celles-ci sont incluses dans les autres émotions complexes, comme je l'ai montré précédemment.

Du point de vue de la formalisation, le degré de l'émotion *complexe* dépend du degré des buts ou des idéaux concernés. Cela revient à mettre en place un opérateur qui représente les buts (resp. idéaux) "faibles" de l'agent, un opérateur

Emotion complexe	Acte expressif de **degré neutre**	Acte expressif de **degré supérieur**
$Rejouissance_i\varphi$	$\mathbf{SeRejouir}_{i,j}\varphi$	$\mathbf{SeFeliciter}_{i,j}\varphi$
$Gratitude_{i,j}\varphi$	$\mathbf{Remercier}_{i,j}\varphi$	$\mathbf{Feliciter}_{i,j}\varphi$
$Regret_i\varphi$	$\mathbf{Regretter}_{i,j}\varphi$	$\mathbf{DemanderPardon}_{i,j}\varphi$
	$\mathbf{-S'excuser}_{i,j}\varphi$	
$Deception_{i,j}\varphi$	$\mathbf{SePlaindre}_{i,j}\varphi$	$\mathbf{SeLamenter}_{i,j}\varphi$
$SatisfactionMorale_i\varphi$	$\mathbf{SeSatisfaire}_{i,j}\varphi$	$\mathbf{SeVanter}_{i,j}\varphi$
$Admiration_{i,j}\varphi$	$\mathbf{Adherer}_{i,j}\varphi$	$\mathbf{Complimenter}_{i,j}\varphi$
$Culpabilite_i\varphi$	$\mathbf{Culpabiliser}_{i,j}\varphi$	$\mathbf{SeRepentir}_{i,j}\varphi$
$Reproche_{i,j}\varphi$	$\mathbf{Reprocher}_{i,j}\varphi$	$\mathbf{Protester}_{i,j}\varphi$

TAB. 9.1 – Actes expressifs suivant leur différent degré de puissance

qui représente les buts (resp. idéaux) “neutres” et un opérateur qui représente les buts (resp. idéaux) “forts”. Cette distinction est cohérente avec plusieurs travaux sur la logique BDI [Casali 2011, Blee 2009]. Cependant, pour des raisons de facilité de lecture et de simplification, je conserve ici un seul opérateur auquel est attribué neutre (0 ou sans degré) ou fort (1). Ainsi, entre $\mathbf{SeRejouir}_{i,j}\varphi$ et $\mathbf{SeFeliciter}_{i,j}\varphi$ par exemple, la différence de degré se situe dans le but initial de l'agent inclus dans la définition de l'émotion de réjouissance : on a ainsi $Goal_i\varphi \wedge Bel_i Resp_i\varphi$ (degré neutre) pour *se réjouir* et $Goal_i(1)\varphi \wedge Bel_i Resp_i\varphi$ (degré fort) pour *féliciter*.

9.2 Extension de la formalisation des actes expressifs aux usages assertifs, engageants et directifs

Dans le but d'obtenir un langage générique et le plus complet possible, il est nécessaire d'ajouter aux actes expressifs déjà définis de nouveaux actes, appartenant aux autres usages de la théorie des Actes de Discours : des actes engageants, directifs, assertifs et déclaratifs.

Les actes expressifs, comme on l'a vu, ont pour but d'exprimer leur condition de sincérité, l'émotion *complexe* intrinsèque à l'acte. Dans les autres classes, les conditions de sincérité restent les états mentaux que doit avoir l'agent lors de l'acte, mais ne représentent pas nécessairement une émotion ; surtout, exprimer ces états mentaux n'est pas le but de ces actes (cf. tableau 7.2). Par exemple, si l'on *affirme* que *Blade Runner* est un excellent film (acte assertif), le but de l'acte est de présenter ce fait comme étant actuel dans le monde ; de plus, pour être sincère, on doit en avoir la croyance (être convaincu que *Blade Runner* est un excellent film). Néanmoins, pour représenter ce lien de sincérité, la formalisation de tels actes passent par la formalisation des états mentaux portés par la condition de sincérité et exprimés lors de l'acte, à l'aide de l'opérateur $Exp_{i,j,H}$.

Les travaux de [Berger 2006] ont permis de mettre en place un *langage de conversation expressif* composé de 32 actes de conversation sélectionnés parmi les verbes illocutoires français pour leur pertinence dans des situations comme l'assistance à l'utilisateur, la négociation, la coopération etc. Cette pré-selection

m'a permis, dans le contexte d'une interaction affective et expressive compagnon artificiel - humain, de conserver 17 de ces actes appartenant aux différentes classes[1].

Le tableau 9.2 présente la formalisation des états mentaux portés par la condition de sincérité de ces actes. Les états mentaux sont ceux définis selon les usages des actes dans la théorie des Actes de Discours (tableau 7.2) : croire à ce que l'on dit pour les assertifs, avoir l'intention de faire ce à quoi on s'engage, pour les engageants, et avoir le désir (le but) que l'autre fasse ce qu'on lui demande pour les directifs.

Actes	Formalisation des états mentaux
Assertifs	*i croit à ce qu'il asserte (φ) et a le but que j sache φ*
Dire	$Bel_i\varphi \wedge Goal_i Bel_j\varphi$
Informer	$Bel_i\varphi \wedge Goal_i Bel_j\varphi$
Affirmer	$Bel_i\varphi \wedge Goal_i(1)Bel_j\varphi$
Nier	$Bel_i\neg\varphi \wedge Goal_i Bel_j\neg\varphi$
Croire	$Bel_i\varphi \wedge Goal_i(-1)Bel_j\varphi$
Contredire	$Bel_i\varphi \wedge Goal_i Bel_j\varphi$
Rappeler	$Bel_i\varphi \wedge Goal_i Bel_j\varphi$
Engageants	*i a l'intention de faire ce à quoi il s'engage (φ) et a le but de le faire savoir à j*
Promettre	$Goal_i Resp_i\varphi \wedge Goal_i Bel_j Goal_i Resp_i\varphi$
Assurer	$Goal_i Resp_i\varphi \wedge Goal_i(1)Bel_j Goal_i Resp_i\varphi$
Accepter	$Goal_i Resp_k\varphi \wedge Goal_i Bel_j Goal_i Resp_k\varphi$
Refuser	$\neg Goal_i Resp_k\varphi \wedge Goal_i Bel_j\neg Goal_i Resp_k\varphi$
Offrir	$Goal_i Resp_i\varphi \wedge Goal_i Bel_j Goal_i Resp_i\varphi\wedge$
Renoncer	$\neg Goal_i Resp_i\varphi \wedge Goal_i Bel_j\neg Goal_i Resp_i\varphi$
Directifs	*i a le but que j fasse ce qu'il lui demande (φ) et que j le sache*
Demander	$Goal_i Resp_j\varphi \wedge Goal_i Bel_j Goal_i Resp_j\varphi$
Exiger	$Goal_i Resp_j\varphi \wedge Goal_i(1)Bel_j Goal_i Resp_j\varphi$
Conseiller	$Goal_i Resp_j\varphi \wedge Goal_i(-1)Bel_j Goal_i Resp_j\varphi$
Suggérer	$Goal_i Resp_j\varphi \wedge Goal_i(-2)Bel_j Goal_i Resp_j\varphi$

TAB. 9.2 – Formalisation des conditions de sincérité des actes assertifs, engageants et directifs.

Certains de ces actes (i.e. *Promettre* et *Assurer*) expriment les mêmes états mentaux, mais avec des degrés de puissance différents. Ces degrés sont formalisés dans les degrés des différents buts illocutoires :

- assertifs : le but que l'autre croit ce qu'on lui dit ;
- directifs : le but que l'autre sache qu'on veut lui faire faire quelque chose ;
- engageants : le but que l'autre croit en l'engagement.

Le degré de puissance permettra par la suite d'avoir, comme pour les actes expressifs, une plus grande granularité dans le choix des actes et dans l'expression

[1] Les actes appartenant aux déclaratifs sont absents, car, d'après Vanderveken, ces actes « *nécessitent en général une position d'autorité du locuteur dans une institution extra-linguistique (tribunal, religion...)* » [Vanderveken 1988] ; je fais ici l'hypothèse qu'il n'y a pas de rapport d'autorité entre l'ACA et l'utilisateur.

de l'agent. Pour se conformer aux tableaux sémantiques de ces actes proposés par Vanderveken [Vanderveken 1988], deux niveaux de degré ont été ajouté, qui complètent les degrés neutre et supèrieur : un degré très faible (-2) et faible (-1).

D'autres actes expriment les mêmes états mentaux avec les mêmes degrés de puissance : ces actes se différencient de part leur contexte d'énonciation (les conditions préparatoires et les conditions sur le contenu propositionnel). Par exemple, si l'on prend les actes *Informer* et *Dire*, on ne peut informer quelqu'un de P que si l'on croit que l'autre ne sait pas P ; en revanche, l'acte *Dire* représente la "force illocutoire primitive" (le degré 0) d'assertion, et n'impose pas de contraintes à son utilisation.

Enfin, certains de ces actes n'expriment des états mentaux qu'à un seul degré ; celà s'explique par l'absence d'autres verbes faisant cette distinction dans la langue française (e.g. l'acte *Nier*).

9.3 Préconditions et postconditions des actes

Dans le but de permettre à un agent d'utiliser les actes que nous avons définis formellement, il est nécessaire de spécifier pour chacun d'eux ses préconditions et ses postconditions (qu'on appelle les effets), à la manière, par exemple, des approches par plan du discours [Cohen 1979] ou des Langages de Communication pour Agents (e.g. ACL-FIPA). La formalisation logique de ces préconditions et effets des actes est utilisée par la suite dans un moteur de raisonnement pour ACA (cf. chapitre 12). On identifie, dans les effets des actes, les effets d'émission d'un acte (c'est l'agent qui a accompli l'acte) et les effets de réception (c'est l'humain qui a accompli l'acte). Le tableau 9.3 présente la formalisation logique des préconditions et effets de quelques actes (cf. Annexe A.3 pour le tableau complet).

9.3.1 Préconditions

La formalisation des préconditions est composée :

- des conditions de sincérité de l'acte : notre hypothèse principale étant que l'agent doit être sincère, il doit respecter les conditions de sincérité de l'acte (définies formellement en section 9.2) pour l'accomplir ;
- des conditions préparatoires et des conditions sur le contenu propositionnel, telles que définies par Vanderveken [Vanderveken 1988] : ces conditions permettent de représenter le contexte d'énonciation propre à chaque acte.

Par exemple, pour l'acte *Informer*, les conditions de sincérité sont celles propres aux actes assertifs : l'agent i croit φ et a le but que l'humain j croit φ. Les autres conditions sont propres à l'acte *Informer* : l'agent i ne croit pas que j croit déjà φ.

9.3.2 Effets

Les effets lors de l'accomplissement de l'acte sont la croyance de l'agent de ses effets *perlocutoires*, c'est-à-dire les conséquences de l'acte sur l'interlocuteur

Actes	Formalisation des préconditions/effets
Informer$_{i,j}\varphi$	
Préconditions	$Bel_i\varphi \wedge Goal_i Bel_j\varphi \wedge \neg Bel_i Bel_j\varphi$
Effets d'émission	$Bel_i Bel_j\varphi$
Effets de réception	$Bel_i Bel_j\varphi \wedge Bel_i Goal_j Bel_i\varphi$
Promettre$_{i,j}\varphi$	
Préconditions	$Goal_i Resp_i\varphi \wedge Goal_i Bel_j Goal_i Resp_i\varphi \wedge Bel_i Goal_j\varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_i\varphi \wedge Committed_i\varphi$
Effets de réception	$Bel_i Goal_j Resp_j\varphi \wedge Bel_i Bel_j Goal_i\varphi$
Assurer$_{i,j}\varphi$	
Préconditions	$Goal_i Resp_i\varphi \wedge Goal_i(1) Bel_j Goal_i Resp_i\varphi \wedge Bel_i Goal_j\varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_i\varphi \wedge Committed_i\varphi$
Effets de réception	$Bel_i Goal_j Resp_j\varphi \wedge Bel_i Bel_j Goal_i\varphi$
Demander$_{i,j}\varphi$	
Préconditions	$Goal_i Resp_j\varphi \wedge Goal_i Bel_j Goal_i Resp_j\varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_j\varphi$
Effets de réception	$Bel_i Goal_j Resp_j\varphi$
SeRejouir$_{i,j}\varphi$	
Préconditions	$Goal_i\varphi \wedge Bel_i Resp_i\varphi \stackrel{déf}{=} Rejouissance_i\varphi$
Effets d'émission	$Bel_i Bel_j Rejouissance_i\varphi$
Effets de réception	$Bel_i Goal_j\varphi \wedge Bel_i Bel_j Resp_j\varphi$
Remercier$_{i,j}\varphi$	
Préconditions	$Goal_i\varphi \wedge Bel_i Resp_j\varphi \stackrel{déf}{=} Gratitude_{i,j}\varphi$
Effets d'émission	$Bel_i Bel_j Gratitude_{i,j}\varphi$
Effets de réception	$Bel_i Goal_j\varphi \wedge Bel_i Bel_j Resp_i\varphi$

TAB. 9.3 – Préconditions et effets (lorsque l'agent accomplit l'acte, ou lorsque l'acte est accompli par l'humain) de quelques actes.

(l'humain). Par exemple, après avoir *informer* j de φ, l'agent i croit que j croit φ.

A l'inverse, la réception d'un acte amène l'agent à croire les états mentaux portés par les conditions de sincérité de l'acte, et exprimés par l'humain (dont les émotions exprimées par les actes expressifs et supposées ressenties par l'humain). Ainsi, lorsque l'humain j *informe* l'agent i de φ, i croit que j croit φ et que j a le but que i croit φ. L'agent croit donc (naïvement peut-être) que l'humain est lui aussi sincère.

Pour la catégorie des actes engageants, le fait que l'agent i s'engage publiquement à faire l'action énoncée est représenté par l'attribut $Committed_i$ (e.g. dans l'acte *Promettre*). Cette attribut peut contribuer par la suite, comme on le verra (cf. chapitre 12), à la sélection de certaines intentions de l'agent.

9.4 Expression multimodale des actes

Les dix-sept actes assertifs, engageants et directifs, ajoutés aux vingt-et-un actes expressifs exprimant les émotions *complexes* et *de base*, forment l'ensemble des Actes de Conversation (cf. tableau 9.4) qui vont être manipulés par les

agents virtuels lors d'interactions avec l'utilisateur. Dans le but de former un Langage de Conversation Multimodal liant langage, états mentaux et multimodalité, et ainsi améliorer notamment la **crédibilité** de l'agent lors de l'acte, il est nécessaire d'attacher à ces actes l'expression multimodale appropriée.

Assertifs	Engageants	Directifs	Expressifs
`Dire`	`Promettre`	`Demander`	`Se réjouir`
`Informer`	`Assurer`	`Exiger`	`Se féliciter`
`Affirmer`	`Accepter`	`Suggérer`	`Remercier`
`Nier`	`Refuser`	`Conseiller`	`Féliciter`
`Croire`	`Offrir`		`Regretter`
`Contredire`	`Renoncer`		`S'excuser`
`Rappeler`			`Demander pardon`
			`Se plaindre`
			`Se lamenter`
			`Se satisfaire`
			`Se vanter`
			`Adhérer`
			`Complimenter`
			`Culpabiliser`
			`Se repentir`
			`Reprocher`
			`Protester`
			`Saluer`
			`Pleurnicher`
			`Approuver`
			`Désapprouver`

TAB. 9.4 – Bibliothèque complète d'Actes de Conversation.

Sous l'hypothèse d'un agent sincère et expressif, l'expression multimodale appropriée à un Acte de Conversation est celle des états mentaux exprimés lors de l'acte et portés par les conditions de sincérité. Comme nous l'avons déjà vu, concernant les actes expressifs, l'expression multimodale appropriée est celle de l'émotion *complexe* intrinsèque à l'acte. Par exemple, lorsque l'agent `Remercie` l'utilisateur, il exprime multimodalement l'émotion *complexe* de gratitude.

Pour les autres classes d'actes, exprimant des états mentaux non-émotionnels, on peut se baser par exemple sur les travaux de Poggi (expression d'"unités cognitives", [Poggi 2000]) ou de Scherer (expression de l'évaluation des checks, [Scherer 2007]), détaillés dans aux chapitres 5 et 3.

Pour les actes expressifs comme pour les autres classes d'actes, l'expression multimodale d'états mentaux identiques à des degrés différents est modulée par le degré. Par exemple, lorsque l'agent `Félicite` l'utilisateur au lieu de le `Remercier`, il exprime également de la gratitude mais à un degré plus élevé : l'expression multimodale associée reste celle de la gratitude mais les gestes, la posture et l'expression faciale qui la composent sont exagérés (joués avec une plus grande amplitude, latence, durée etc., cf. chapitre 4).

Il est important ici de préciser que la manière de définir ces expressions

multimodales, que ce soit l'expression d'émotions ou d'autres états mentaux, dépend fortement de plusieurs facteurs :

- le contexte : les références culturelles de l'humain et la nature de sa relation avec l'agent ;
- le système de l'ACA lui-même : la façon de décrire les expressions faciales, les gestes etc... (langage de programmation, FACs vs. Action Units...).

La problématique du contexte dans l'expression des états mentaux est très complexe. On trouve par exemple des différences dans l'expression entre deux cultures : les japonais n'expriment pas la réjouissance comme les anglais, ou les français comme les américains. De la même façon, au sein d'une même culture, deux individus n'expriment pas de la même façon leurs états mentaux. Et le même individu, suivant le contexte d'énonciation et le type de relation qu'il entretient avec son interlocuteur, ne s'excuse pas de la même manière qu'avec un autre interlocuteur, par exemple moins proche et sur le lieu de travail. Il est donc nécessaire de faire dans un premier temps un certain nombre d'hypothèses simplificatrices, que l'on pourra par la suite ne plus prendre en compte pour intégrer ces nouvelles problématiques. Ces hypothèses sont les suivantes :

> (H1) Les expressions multimodales des états mentaux, telles que définies dans ce travail, concernent un ACA féminin de culture française.
> (H2) Ces expressions sont celles d'un individu particulier et ne prétendent pas être celles de tous les individus de culture française. Par contre, elles sont significativement reconnues par eux grâce à des études de perception menées sur des sujets [Niewiadomski 2009].
> (H3) Le contexte d'énonciation consiste en un dialogue informel entre un humain de culture française et son ACA compagnon.

La définition des expressions multimodales dépend également du système dans lequel va être implémenté ce langage : on ne définira pas de la même manière l'émotion *complexe* de culpabilité par exemple pour les ACA de FearNot ! [Paiva 2006] et ceux de MRE [Hill 2001]. Le but ici de proposer un tel langage est bien de fournir un outil pour que les agents virtuels puissent s'exprimer d'une manière qui apparaisse plus sincère et crédible aux yeux de l'humain, en liant les états mentaux à exprimer lors d'un acte. Dans le cadre de cette thèse, et plus précisément du projet CECIL, le Langage de Conversation Multimodal a pour vocation à être implémenté dans l'ACA Greta [Poggi 2005] et plus largement dans le standard SAIBA. Le LTCI, partenaire du projet, a fourni l'ensemble des expressions multimodales définies pour Greta. Le prochain chapitre présente comment s'est faite cette implémentation et les résultats de l'évaluation d'une partie de ce langage.

10
Implémentation et évaluation du LCM

Ce chapitre présente l'ACA Greta et l'implémentation du Langage de Conversation Multimodal dans Greta, et ainsi dans le standard SAIBA. Cette implémentation nous a permis de mener une évaluation du LCM, dont je détaille également le protocole et les résultats.

10.1 Greta et le standard SAIBA : implémentation du LCM

Greta [Poggi 2005] est un ACA développé au LTCI[1], capable d'exprimer multimodalement ses états mentaux, *via* le regard, les gestes, les expressions faciales et la posture. Greta est conforme au standard SAIBA [Vilhjalmsson 2007], architecture logicielle commune pour la génération du comportement d'agents virtuels, présentée dans le chapitre 5.

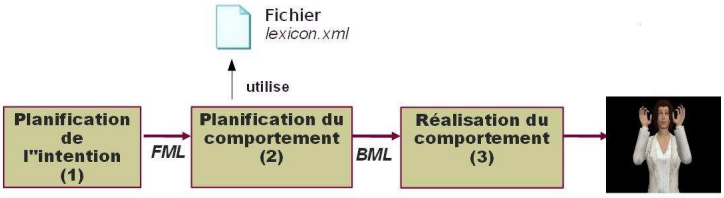

Fig. 10.1 – Le standard SAIBA.

De par sa conformité au standard SAIBA, Greta possède, dans son fichier de lexique, un répertoire d'expressions d'états mentaux (dont des émotions), notamment grâce aux travaux de [Poggi 2000] sur l'expression faciale d'"unités cognitives" : on trouve par exemple des actes tels que implorer, avertir, suggérer... Dans le cadre du projet CECIL, les expressions des différentes émotions *complexes* ainsi que celles des ACM non-expressifs ont été définies.

Greta propose quatre personnalités différentes, incarnées par 4 personnages virtuels (cf. chapitre 5) :

- Spike est agressif et pinailleur ;
- Poppy est extravertie et optimiste ;
- Obadiah est pessimiste et dépressif ;

[1] Laboratoire Traitement et Communication de l'Information, http ://www.ltci.telecom-paristech.fr/

– et Prudence est pragmatique et pratique.

Ces personnalités se traduisent par des expressions multimodales différentes d'une même émotion : chaque personnage possède son propre fichier de lexique où sont modulés les signaux en fonction de sa personnalité [McRorie 2009]. Ainsi, Poppy, qui est extravertie, n'exprime pas l'émotion de réjouissance de la même façon que Spike : la différence se traduit par une amplitude plus grande dans les gestes et une plus grande vitesse d'exécution du mouvement.

L'implémentation de ces différentes personnalités dans le système SAL et la plateforme SEMAINE (cf. chapitre 5) a permis d'identifier quelles étaient les expressions et les rétroactions (*backchannels*) appropriées pour chaque personnalité [McRorie 2012]. Pour ma part, j'ai choisi pour l'implémentation la personnalité de Prudence (cf. figure 10.2) car la personnalité n'étant pas ce que l'on veut évaluer, c'est Prudence qui a le moins d'influence sur les sujets [De Sevin 2010],

FIG. 10.2 – Prudence est la personnalité de Greta pragmatique et pratique.

L'implémentation du Langage de Conversation Multimodal dans Prudence consiste à créer, dans la librairie de comportement de Greta (le fichier de lexique), l'ensemble des Actes de Conversation Multimodaux (ACM) et leur expression multimodale. A chacun des actes de la bibliothèque sont associés les états mentaux intrinsèques à l'acte (les émotions pour les actes expressifs) : ces états mentaux se traduisent concrètement par un certain nombre de signaux multimodaux qui sont utilisés au niveau du module de *Planification du comportement* et seront ainsi exprimés par Prudence de manière concomitante lors de l'acte. Par exemple, la définition de l'ACM `Regretter` (*Regret* en anglais) dans le fichier lexique fait le lien avec les différents signaux de l'expression de l'émotion complexe du regret (signaux 1,2,3 et 4 ci-dessous) :

*<behaviorset name="**regret**">*
<signals>
<signal name="faceexp=regret" modality="face" id="1"/>
<signal name="head=regret" modality="head" id="2"/>
<signal name="gaze=regret" modality="gaze" id="3"/>
<signal name="emotion=regret" modality="gesture" id="4"/>

</signals>
</behaviorset>

Ensuite, lorsque l'intention communicative de Prudence est d'exprimer du regret à propos de la situation actuelle, elle se traduit en FML par l'utilisation de l'ACM `Regretter` *via* la balise "performative" :

<fml-apml>
<bml>
*<speech id="s1" start="0.0" language="english" voice="openmary" text="**I should not done this, it's too late now.**">*
<description level="1" type="gretabml">
</speech>
</bml>
<fml>
*<performative id="p1" type="**regret**" start="s1 :tm1"/>*
</fml>
</fml-apml>

Suivant l'architecture SAIBA, le dernier module de *Réalisation du comportement* reçoit l'ensemble de ces signaux traduits en BML et les soumet à des règles et des contraintes d'animation : il vérifie qu'ils ne sont pas en conflit, que l'ensemble est cohérent, puis anime le comportement verbal et non verbal de Greta.

10.2 Evaluation du LCM

Le but de l'évaluation que je présente ici est de tester l'hypothèse sous-jacente à notre proposition : **l'ajout d'une expression non-verbale des émotions complexes cohérente avec l'expression verbale de l'ACA augmente sa crédibilité et sa sincérité perçues par l'humain**. Cette hypothèse est à considérer dans le contexte d'un dialogue humain-ACA, sous les hypothèses définies ci-dessus (cf. section 9.4). L'amélioration de la crédibilité des ACA lorsque ceux-ci expriment des émotions adaptées au contexte ayant déjà été vérifiée par plusieurs travaux (cf. chapitre 4), notre évaluation portera uniquement sur deux ACM expressifs, `Se réjouir` et `S'excuser`, exprimant respectivement les émotions complexes de réjouissance et de regret. Le choix de ces deux ACM a été motivé par leur opposition mutuelle dans le tableau 8.1 : l'un exprime une émotion "positive", liée à une situation satisfaisant un des buts de l'agent, tandis que l'autre exprime une émotion "négative" liée à une situation empêchant un des buts de l'agent. L'implémentation du LCM dans Prudence m'a permis de mener cette évaluation sur ces deux ACM expressifs.

10.2.1 Protocole

Dans le cadre d'un dialogue entre un utilisateur (le sujet de l'expérience) et son agent compagnon, joué par Prudence, qui est capable de recommander des films, les deux ACM expressifs sont accomplis par Greta selon trois conditions :

- condition congruente (C) : Prudence exprime l'ACM tel qu'il est défini dans notre langage, c'est-à-dire avec l'émotion intrinsèque cohérente ;

- condition neutre (N) : Prudence exprime seulement verbalement l'énoncé de l'acte, sans rien exprimer non-verbalement ;
- condition non-congruente (NC) : Prudence exprime verbalement l'énoncé de l'acte et l'émotion opposée (opposée dans le tableau 8.1) de manière non-verbale.

Le protocole est divisé en deux parties ; une première partie visant à recueillir des données quantitatives, et une seconde partie, sous la forme d'un questionnaire ouvert, visant à recueillir des données qualitatives.

Dans la première partie du protocole, des vidéos de deux scénarios sont proposées, chacune montrant l'expression par l'ACA d'un des deux ACM testés. Pour chaque scénario, trois vidéos sont soumises au sujet pour être évaluées (une vidéo par condition, soit six vidéos au total). Les deux scénarios impliquent un dialogue avec Prudence après que l'utilisateur est revenu (supposément) du cinéma où il a vu un film que Prudence lui avait conseillé auparavant

- dans le premier scénario, l'utilisateur a adoré le film et vient de remercier Prudence pour son conseil. On demande alors au sujet de regarder les trois vidéos (dans un ordre prédéfini qui change pour chaque sujet) et d'évaluer comment Prudence `se réjouit` ;
- dans le second scénario, le sujet a détesté le film et vient de reprocher à Prudence son mauvais conseil. De la même façon que précédemment, le sujet regarde les trois vidéos et évalue comment Prudence `s'excuse`.

On demande au sujet d'évaluer les vidéos selon deux critères : la sincérité (« Prudence semble-t-elle exprimer ce qu'elle pense ? ») et la crédibilité (« cette scène vous semble-t-elle plausible ? ») de Prudence. Ces critères peuvent chacun prendre quatre valeurs possibles sur la même echelle qualitative (Pas du tout - Plutôt pas - Plutôt - Tout à fait).

La seconde partie du protocole consiste en un questionnaire dont le but est de collecter les impressions des sujets, notamment concernant leur opinion générale concernant Prudence et leurs sentiments subjectifs.

10.2.2 Résultats

Résultats quantitatifs

Ving-trois sujets selectionnés parmi des étudiants en science entre dix-huit et vingt-six ans ont participé à cette évaluation. Dans le but de mener des tests statistiques pour vérifier la significativité des résultats, un score a été associé à chaque valeur qualitative concernant la sincérité et la crédibilité : 1 pour "Pas du tout", 2 pour "Plutôt pas", 3 pour "Plutôt" et 4 pour "Tout à fait". Etant donné que les deux ACM sont évalués selon deux dimensions, il y a quatre variables dépendantes : sincérité (`S'excuser` et `Se réjouir`) et crédibilité (`S'excuser` et `Se réjouir`). La seule variable indépendante est la condition (congruente, neutre et non-congruente). La figure 10.3 montre l'ensemble des scores pour la sincérité des deux ACM et l'ensemble des scores pour leur crédibilité.

La première étape d'analyse des résultats consiste à vérifier que les variables dépendantes suivent une distribution normale (i.e. il est possible de faire une analyse statistique des résultats). Dans un second temps, l'analyse des scores (figure 10.3) montrent que les ACM exprimés par Prudence dans la condition congruente (C) ont les meilleurs scores dans les deux critères : pour la sincérité, le score est de 65 (contre 41 et 31) pour `S'excuser`, et de 68 (contre 55 et 42)

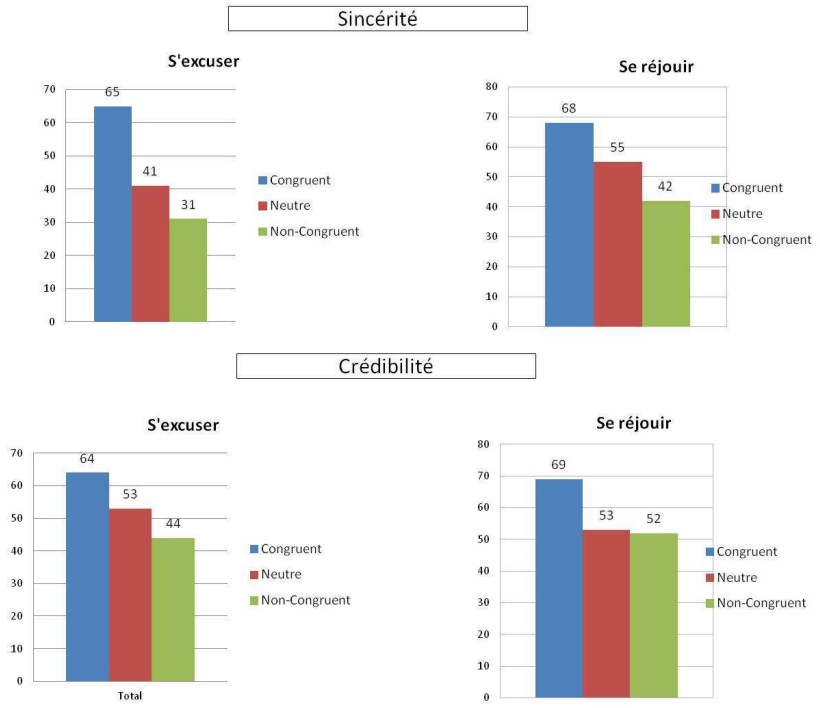

FIG. 10.3 – Scores qu'ont obtenu les ACM `S'excuser` et `Se réjouir` en sincérité et en crédibilité.

pour `Se réjouir` ; pour la crédibilité, le score est de 64 (contre 53 et 44) pour `S'excuser`, et de 69 (contre 53 et 52) pour `Se réjouir`.

Si l'on s'intéresse au nombre de sujets, lorsque Prudence `s'excuse` dans cette condition, 15 sujets sur 23 (soit 65%) trouvent Prudence "plutôt" sincère (9 sujets, soit 39%) ou "tout à fait" sincère (6 sujets, soit 26%) ; pour la crédibilité, 16 sujets sur 23 (soit 70%) trouvent Prudence "plutôt" crédible (12 sujets, soit 52%) ou "tout à fait" crédible (4 sujets, soit 18%). De la même façon, lorsque Prudence `se réjouit` dans la condition congruente, 17 sujets sur 23 (soit 74%) trouvent Prudence "plutôt" sincère (12 sujets, soit 52%) ou "tout à fait" sincère (5 sujets, soit 22%) ; pour la crédibilité, 18 sujets sur 23 (soit 78%) trouvent Prudence "plutôt" crédible (12 sujets, soit 52%) ou "tout à fait" crédible (6 sujets, soit 26%).

Ces résultats tendent à confirmer notre hypothèse ; on peut affirmer qu'exprimer l'émotion cohérente (C) à l'acte rend l'ACA plus sincère et crédible aux yeux de l'utilisateur que ne rien exprimer (N) ou exprimer l'émotion opposée (NC).

Nous avons ensuite mené un test d'analyse de la variance (ANOVA) sur chacune des variables dépendantes, dans le but de savoir s'il existe un lien entre ces variables et la variable indépendante. L'hypothèse nulle *H0* de ce test ANOVA est qu'il n'y a pas de différence entre les moyennes de scores de sincérité (resp. les moyennes de scores de crédibilité) dans la condition congruente (C) pour les ACM `S'excuser` et `Se réjouir`. L'hypothèse *H1* postule à l'inverse qu'il y

a une différence entre ces moyennes, et donc qu'il existe une relation entre les variables dépendantes et la condition congruente. Le test ANOVA nous permet de rejeter de façon significative l'hypothèse *H0* pour les moyennes de scores de sincérité et de crédibilité : par exemple, il y a un effet significatif de la condition congruente sur la sincérité de Prudence ($F(2.66)=22.80$ $p<0.05$) lorsque elle `S'excuse`.

Enfin, le tableau 10.1 présente une partie des résultats d'un test de Tukey réalisé pour étudier les différences entre les différents scores. L'ensemble des résultats de ce test montrent qu'il y a une différence significative entre la moyenne des scores de sincérité (resp. crédibilité) dans la condition congruente (C) et la moyenne des scores de sincérité (resp. crédibilité) dans les autres conditions, pour les deux ACM testés. A l'inverse, il n'y a pas de différence significative entre la condition non-congruente (NC) et la condition neutre (N).

Groupes	Différence	Statistique	Probabilité
C - NC	1.478	q = 9.290	0.0000
C - N	1.478	q = 9.290	0.0000
NC - N	-0.435	q = 2.732	0.1376

TAB. 10.1 – Résultats du test de Tukey pour la variable dépendante concernant la sincérité de `S'excuser` : la différence entre la condition congruente (C) et les conditions neutre (N) et non-congruente (NC) est significative.

Résultats qualitatifs

Concernant les résultats de la partie quantitative, la première remarque que l'on peut faire concerne le rapport des sujets avec la personnalité de Prudence. Malgré nos efforts pour éviter le biais de l'évaluation de la personnalité, un certain nombre de sujets ont fait des remarques sur celle-ci, la qualifiant d'« austère » et de « sévère ». Nous ne savons pas dans quelles mesures la personnalité a pu influencer les résultats, mais plusieurs sujets ont mentionné la confiance qu'ils accordent à l'agent ; la personnalité et l'aspect de l'ACA semblent avoir une certaine importance dans la confiance de l'humain (en plus de la sincérité de l'ACA), ce qui est cohérent avec les travaux du domaine [Gong 2008].

Une seconde remarque concerne l'*impact* de l'ACA sur le sujet : en photojournalisme, l'impact est la capacité d'une photo à attirer l'attention et à rester en mémoire. En effet,lors de l'entretien qualitatif, à la question « Avez-vous trouvé qu'une des vidéos était plus marquante que les autres ? », la majorité des sujets ont désigné la vidéo dans laquelle Greta `s'excuse` en exprimant de la réjouissance (condition non-congruente). Les sujets ont qualifié Greta, dans cette vidéo, de « très hypocrite », « marrante » : un des sujet a estimé qu'elle se « moquait de lui » mais en des termes plus imagés. On peut donc imaginer qu'un ACA doué d'humour ou capable d'ironie aurait plus d'impact sur l'utilisateur.

Conclusion

La théorie des Actes de Discours [Searle 1969, Searle 1985] propose une taxonomie des actes de discours, qui sont les unités premières du langage. Elle distingue cinq classes d'actes : les assertifs, les directifs, les déclaratifs, les engageants et les expressifs. Chaque acte est de la forme F(P), F étant le force illocutoire et P le contenu propositionnel. La force illocutoire F est constituée de six composantes parmi lesquelles nous avons identifié celles qui font le lien entre langage et états mentaux : ce **lien de sincérité** constitue la base sur laquelle nous avons construit notre Langage de Conversation Multimodal.

Dans cette partie, j'ai également défini la notion d'émotions *complexes*, et j'ai présenté une formalisation logique de ces émotions. Cette formalisation repose sur quatre opérateurs qui consitituent **le modèle BIGRE**. L'introduction d'un cinquième opérateur d'expressivité a permis de définir formellement dans un premier temps un ensemble d'actes expressifs dont les buts illocutoires sont l'expression des émotions *complexes* : ces actes expressifs forment la base du Langage de Conversation Multimodal. Dans un second temps, l'introduction du degré de puissance et la formalisation et l'ajout d'actes des classes assertives, directives et engageantes permet d'avoir un langage de conversation le plus complet possible, et plusieurs choix d'actes pour exprimer les mêmes états mentaux (i.e. `Promettre` et `Assurer`, `Se plaindre` et `Se lamenter`). Notre hypothèse est que l'expression multimodale de ces états mentaux, portés par les conditions de sincérité, lors de l'acte, améliore la crédibilité et la sincérité de l'ACA perçues par l'humain. Le Langage de Conversation Multimodal est constitué de trente-huit ACM dont vingt-et-un actes expressifs exprimant les émotions complexes, incluant les deux ACM `S'excuser` et `Regretter` exprimant le regret. De plus, le travail de Vanderveken sur les verbes illocutoires anglais [Vanderveken 1990] et le fait que notre modèle logique soit adaptable a permis de mettre en place un LCM **en anglais** (pour la bibliothèque détaillée, cf. Annexe).

L'implémentation du LCM dans le lexique de comportements de l'architecture SAIBA a permis d'évaluer une partie de notre LCM avec l'ACA Greta. Un scénario a été mis en place et un ensemble de vidéos ont été évaluées par les sujets. Les résultats de cette évaluation comparative montrent que le LCM améliore la sincérité et la crédibilité perçues de l'ACA.

Quatrième partie

Moteur de raisonnement

Introduction

La partie précédente présente un Langage de Conversation Multimodal (LCM) pour ACA liant les états mentaux de l'agent (dont les émotions complexes), le langage et leur expression multimodale. Ce langage est un outil pour les ACA sincère leur permettant d'exprimer d'une manière crédible leurs états mentaux. Dans la seconde partie de ce travail de thèse, je propose de mettre en place un moteur de raisonnement qui, dans un dialogue entre un ACA et un humain, met à jour les états mentaux de l'ACA, déduit ses émotions complexes et décide de son intention communicative, que l'ACA pourra ainsi exprimer *via* le LCM. On s'intéresse donc ici, après avoir répondu, grâce au LCM, à la question « comment l'ACA exprime-t-il ce qu'il a décidé ? », à la question « comment l'ACA décide-t-il de ce qu'il va exprimer ? ».

J'ai présenté dans l'état de l'art un certain nombre d'ACA basés sur des modèles issus des théories cognitives des émotions, capables de déclencher des émotions à partir d'évaluations de leur environnement (cf. chapitre 4). Parmi ces différentes implémentations, le déclenchement des émotions à partir d'une architecture de type BDI (Belief, Desire, Intention) nous parait tout particulièrement indiquée dans la mise en place du moteur de raisonnement, et ce pour deux raisons :

- la formalisation des émotions complexes et des Actes de Conversation Multimodaux (ACM) qui composent le LCM nous permet de manipuler, à l'instar des BDI, les BIGRE de l'agent (croyances, idéaux, buts, responsabilité et émotions, cf. chapitre 8), pour déduire son intention communicative ;
- dans le cadre d'un dialogue entre un ACA et un humain, nous avons besoin d'outils pour planifier et réguler le discours qui existent pour les architectures BDI (cf. section 11.2).

Je présente d'abord dans un premier chapitre les théories qui constituent les fondements du moteur de raisonnement (chapitre 11). Je détaille ainsi les différents travaux qui se sont intéressés, dans le cadre des architectures BDI, au langage et aux émotions. Le moteur de raisonnement, sa place dans le cycle d'évaluation des théories cognitives, son architecture et son fonctionnement sont décrits dans un deuxième chapitre (chapitre 12). Ce moteur a été implémenté (chapitre 13) dans l'ACA Greta d'une part, où il fait le lien avec l'implémentation du LCM présentée dans la partie précédente, et d'autre part dans l'ACA MARC [Courgeon 2008, Courgeon n C], ce qui m'a notamment permis de tester de nouvelles expressions multimodales définies à partir des travaux de Scherer [Scherer 2001]. Je présente ensuite les résultats de l'évaluation de ce moteur en termes de sincérité et crédibilité perçues.

11 Appuis théoriques

Je détaille ici dans un premier temps le modèle BDI et le concept d'intention qui est le fondement théorique de ce modèle. De plus, quelques travaux sur l'introduction des émotions dans les architectures BDI sont brièvement présentés. Je décris ensuite les approches à base de plans du discours (basées sur la théorie des actes de discours), et leur lien avec l'architecture BDI. Ces approches apportent un ensemble d'outils pour manipuler les actes de discours qui, associés à des mécanismes de régulation du discours (cf. section 11.3), sont appropriés à ma problématique. L'ensemble de ces travaux représentent les fondements théoriques que j'adapte par la suite dans la mise en place du moteur de raisonnement (chapitre 12).

11.1 Le modèle BDI (*Belief - Desire - Intention*)

Dans les années 80, la plupart des théories comportementales s'entendent pour définir les états mentaux en termes de croyances et de désirs [Davidson 1980, Goldman 1970]. L'intention (de faire ou d'avoir fait une action) est alors vue comme une conjonction des croyances et des désirs de l'agent par rapport à l'action elle-même, c'est-à-dire que l'intention peut toujours se réduire en ces opérateurs ; on pourrait donc dire qu'avant le modèle BDI, il existait un modèle "BD" (*Belief - Desire*).

Le philosophe Bratman, dans sa théorie de l'action humaine [Bratman 1987], rejette cette conception de l'intention. Il affirme que l'intention de faire une action est **adoptée** à partir des croyances et des désirs ***via* un raisonnement pratique** (*practical reasoning* en anglais) qui fait de nous des êtres rationnels[1]. Ce raisonnement pratique prend place au sein d'un cycle de comportement : perception, décision, action (cf. figure 11.1).

Transposé pour les agents artificiels, l'agent perçoit et interprète son environnement en mettant à jour son modèle du monde (ses croyances, ou *Belief*). Le raisonnement pratique (la phase de décision) qui suit se décompose en deux activités :

1. la délibération : l'agent déduit ses désirs (*Desire*), c'est-à-dire les états du monde qui sont désirables pour l'agent ; parmi ses désirs, l'agent sélectionne *de façon rationnelle* les intentions qu'il va poursuivre ;

[1] « *Entre une mauvaise cuisinière et une empoisonneuse, il n'y a qu'une différence d'intention* », disait P. Desproges

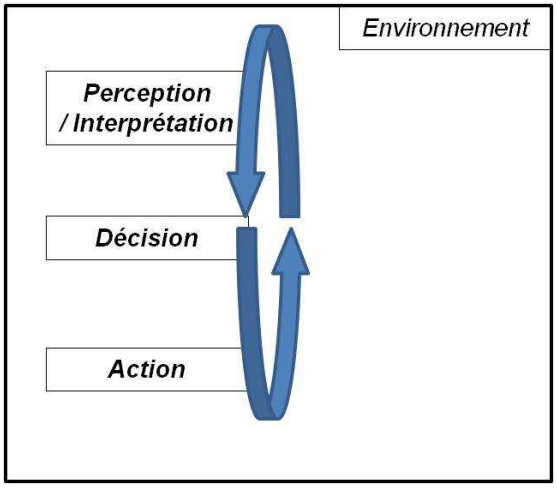

FIG. 11.1 – Cycle de comportement d'un agent BDI.

2. la planification, ou comment parvenir à ses fins (*means-ends reasoning* en anglais) : l'agent planifie une suite d'actions pour atteindre son intention, suivant ses croyances et ses intentions.

Enfin, l'agent agit dans son environnement en suivant le plan choisi. Bratman propose ainsi d'ajouter au modèle "BD" les intentions, pour former le modèle BDI.

La sélection d'une intention, d'après Bratman, implique une certaine forme d'engagement de la part de l'agent (*commitment* en anglais). Cohen et Levesque [Cohen 1990b] ont proposé un modèle logique BDI, basé sur les travaux de Bratman, dans lequel la notion d'engagement est centrale (» *Intention Is Choice with Commitment* « , [Cohen 1990a]). Les auteurs considèrent deux types d'engagement : les engagements "internes" (*internal commitment* en anglais) et "sociaux" (*social commitment*). Les engagements "internes" sont basés sur certains des buts des agents, que les auteurs nomment "buts persistants" (définis par l'opérateur *Goal*), et sur le choix de l'agent de s'engager sur ces buts. Les engagements "sociaux" apparaissent dans la communication, lorsque l'on s'engage envers l'interlocuteur à faire quelque chose. L'agent sélectionne donc ses intentions suivant les engagements "internes" ou "sociaux" qu'il a pris.

Rao et Georgeff [Rao 1991] ont également proposé une formalisation logique du modèle BDI pour des agents artificiels rationnels, en définissant les opérateurs *Bel*, *Des* et *Intent*. Les auteurs distinguent les désirs de l'agent (qui peuvent être irréalisables ou hors de contrôle - e.g. le désir qu'il fasse beau), de ses désirs "d'agir" (sur lesquels l'agent a du contrôle). L'agent sélectionne son intention à partir d'une hiérarchisation de ses désirs "d'agir", en s'engageant (implicitement) sur le désir le plus fort. L'agent BDI se comporte ainsi de manière rationnelle, en construisant les différents futurs possibles à partir de ses croyances, désirs et intentions, par la planification. Les auteurs listent un ensemble d'axiomes, basés sur le modèle de Bartman, qui sont à la base de leur formalisation [Berger 2006] :

- les intentions sont les moteurs de l'action ;
- les agents croient en leurs intentions ;

- ils croient en l'accomplissement de leurs buts ;
- ils ont l'intention d'accomplir leurs buts ;
- ils peuvent reconsidérer leurs intentions en fonction de leur perception / interprétation de l'environnement.

Le modèle logique BDI de Rao et Georgeff a à ce jour largement inspiré les travaux sur le raisonnement des agents dans la communauté multi-agents [Wooldridge 2009].

11.1.1 Emotions et BDI

Comme on l'a vu dans le chapitre 4, un certain nombre de travaux utilisent des logiques de type BDI pour déclencher les émotions des ACA. Plusieurs autres travaux ont proposé d'intégrer les émotions aux architectures BDI dans le but de modéliser l'influence de ces émotions sur les croyances de l'agent [Jiang 2007] ou sur la sélection de ses intentions (en termes de stratégies d'adaptation [Adam 2007] ou de tendances à l'action [Meyer 2006, Pereira 2008]). Ces travaux sont très intéressants mais ne permettent pas de modéliser l'expression d'états mentaux (dont les émotions complexes) dans le dialogue.

11.2 Théories des plans du discours et BDI

Adaptées du concept d'intention défini avant Bratman, les approches intentionnelles par plan du discours [Perrault 1980, Cohen 1979, Allen 1986] déduisent les intentions des agents à partir de ses croyances et de ses buts : ce type d'intention correspond à un but "haut-niveau" que poursuit l'agent tout au long du dialogue. L'arbre des sous-actions à effectuer pour atteindre une intention est ensuite construit, puis la première action du plan est effectuée.

Le modèle proposé par Cohen, Allen et Perrault [Allen 1986, Cohen 1979] est souvent considéré comme le modèle de référence dans les approches intentionnelles par plan. Les auteurs se basent sur une formalisation des croyances de l'agent (operateur *beliefs*) et de ses buts/intentions (operateur *Want*). Ils définissent également deux bibliothèques : une bibliothèque d'actes de discours et une bibliothèque de plans du domaine. Seuls quatre actes de discours sont définis : inform, inform-if, inform-ref et request. Les actes de discours et les actions des plans du domaine sont décrits en termes de préconditions et d'effets (selon le formalisme STRIPS [Fikes 1971]), pour permettre la planification par chaînage arrière.

Par exemple (cf. figure 11.2), si S (l'émetteur, i.e. *Sender*) a l'intention que H (l'interlocuteur, i.e. *Hearer*) sache une proposition Prop, S cherche parmi les actions possibles (dont les actes de discours) celle dont l'effet est son but (*convince (S,H,Prop)*), puis vérifie ses préconditions (*cando.pr*). Si celles-ci ne sont pas vraies (ici, que H croit que S croit Prop), S cherche à nouveau une action dont l'effet satisfait les préconditions fausses (*inform(S,H,Prop)*). Dans cet exemple, pour que S puisse informer H que Prop, il doit auparavant croire Prop (la précondition de Inform, mais qui est aussi la condition de sincérité des assertifs, cf. chapitre 7).

Allen et Perrault [Allen 1986] proposent dans ce contexte des règles d'inférence de plan : ces règles permettent de déduire (ou le cas échéant de postuler) les plans que peut suivre l'utilisateur. Par exemple, si l'on apprend que l'utilisateur

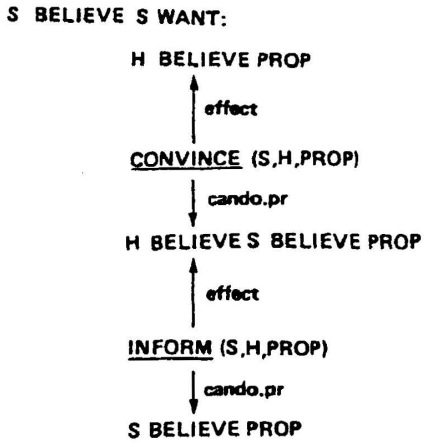

FIG. 11.2 – Planification par chaînage arrière. Figure tirée de [Cohen 1979].

veut avoir A, et que l'on sait que A → B, alors il y a une certaine probabilité que l'utilisateur souhaite avoir B.

On peut formuler plusieurs critiques à l'encontre des approches par plan du discours : l'intention qui n'est pas considérée comme un opérateur à part entière, déduite par un raisonnement pratique [Bratman 1987] ; le manque de régulation du dialogue à un niveau plus "local", comme je le détaille ci-après ; et enfin une bibliothèque d'actes limitée. Ces approches introduisent néanmoins des concepts clés au niveau de la manipulation des actes de discours, et je présenterai par la suite comment je réutilise une partie de ces concepts dans le moteur de raisonnement, en essayant de répondre à ces critiques en introduisant (notamment) la formalisation BIGRE.

11.3 Les "niveaux" de dialogue

Baker [Baker 1994] propose de distinguer dans le dialogue deux types de *negotia* (*i.e.* ce dont on peut discuter), soit deux niveaux distincts :

- Le niveau de la tâche (*task level*) ou niveau "global" ;
- Le niveau de la communication (*communication level*) ou niveau "local" .

Le niveau "global" correspond à la poursuite de ce qu'on pourrait appeler des buts "haut-niveau" par un agent tout au long du dialogue. C'est ce niveau qui forme les typologies de dialogue (i.e. des modèles de dialogue) tels que définis par exemple par Walton et Krabee [Walton 1995] (négociation, persuasion, délibération...). A l'inverse, le niveau "local" correspond à tout ce qui contribue à la régulation locale du dialogue (au niveau de quelques tours de parole), ce que Bunt [Bunt 1989] appelle le contrôle du dialogue : on trouve par exemple les *feedback* (perception, compréhension, e.g. « Tu as bien dit rouge ? »), ce qui structure le discours (début, fin, changement de sujet) et d'autres interactions (e.g. obligations sociales, tour de parole). Les approches par plan du discours génèrent ces buts "haut-niveau" au niveau "global" de l'agent, en privilégiant les intentions de l'agent au niveau de la tâche. Une des critiques formulées ci-dessus de ces approches par plan du discours reste leur incapa-

cité à réguler le dialogue à un niveau "local", ce qui a amené Traum et Allen [Traum 1994] à introduire les obligations du discours.

11.3.1 Obligations du discours

Traum et Allen [Traum 1994] ont introduit la notion d'obligations du discours pour permettre de réguler le dialogue au niveau "local" : à chaque énoncé reçu ou émis par l'agent correspond un certain nombre d'obligations du discours (représentant ce que l'agent devrait faire en réponse), comme cela est présenté dans le tableau 11.1. Les obligations représentent des normes sociales qui guident le comportement de l'agent au niveau du discours sans avoir besoin de prendre en compte ses intentions plus "globales" et ses buts : elles rendent l'agent réactif au niveau du discours.

Source de l'obligation	Obligation
S1 accepte ou promet A	S1 réalise A
S1 demande (requête) A	S2 accepte ou refuse A
S1 demande (question) P	S2 informe S1 de P
Enoncé incompris ou incorrect	"Réparer" l'énoncé

Tab. 11.1 – Quelques règles d'obligations du discours entre un locuteur (S1) et un interlocuteur (S2), adapté de [Traum 1994].

D'après Traum et Allen, on voit que lorsqu'on s'engage publiquement à réaliser une action (en acceptant ou en promettant), on a l'*obligation* de réaliser l'action. Cette vue est en accord avec la définition de l'intention de Cohen et Levesque [Cohen 1990a] présentée ci-dessus, et leur notion d'engagement "social".

Les obligations ont été implémentées dans le projet TRAINS qui vise à développer un assistant intelligent capable notamment d'utiliser et de reconnaitre la langue naturelle [Allen 1994]. Le concept d'obligations du discours a été également implémenté dans certains ACA [Kopp 2008]. Je présente dans la section suivante comment ce concept a été implémenté dans le moteur de raisonnement pour contribuer à réguler le discours au niveau "local".

12
Le moteur de raisonnement

Le moteur de raisonnement que je présente ici doit permettre à l'ACA, lors d'un dialogue avec un humain, de **mettre à jour** ses états mentaux et de **les exprimer** *via* le LCM (cf. figure 12.1).

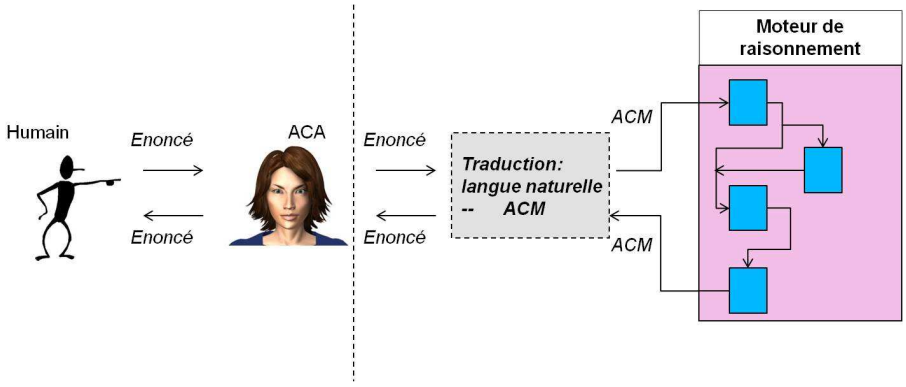

FIG. 12.1 – Schéma de principe du moteur de raisonnement.

Je situe tout d'abord ce moteur (section 12.1) dans un modèle de raisonnement cognitif plus large que l'on peut retrouver dans les théories cognitives des émotions. Dans un deuxième temps (section 12.2), je présente l'architecture générale du moteur de raisonnement, ses points communs avec l'architecture BDI et sa place dans le standard SAIBA. Je détaille ensuite les différents sous-modules de perception, évaluation, délibération et planification qui le composent, ainsi que les différents mécanismes adaptés des travaux sur le discours que j'ai présenté ci-dessus. Je montre enfin, dans la section 12.9, en quoi le moteur de raisonnement peut être qualifié de générique, en détaillant les parties qui sont réutilisables et celles qui sont facilement configurables.

12.1 Le moteur de raisonnement dans le cycle de déclenchement des émotions

Le moteur de raisonnement n'a pas pour ambition de modéliser le raisonnement humain dans son ensemble et sa complexité : au niveau des émotions, il déclenche les émotions *complexes*, issues de raisonnements spécifiques, que

nous avons définies à partir du modèle BIGRE. Il est important dans ce cas de préciser la place qu'il occupe dans les architectures cognitives qu'ont introduit certaines théories psychologiques de l'évaluation cognitive.

Certaines théories psychologiques de l'évaluation cognitive admettent que les émotions peuvent être déclenchées par des processus cognitifs et non-cognitifs : ces processus sont souvent représentés en "couches" successives, de la plus réactive (déclenchant les émotions "primaires" telles que la joie, le dégoût etc.) à la plus cognitive (déclenchant des émotions plus complexes, telles que le regret, la culpabilité etc.). Il n'y a cependant pas de concensus sur le nombre de couches réactives et cognitives. Ainsi, Sloman [Sloman 2001] admet une seule couche réactive tandis que Scherer [Scherer 2001] en définit deux, qu'il appelle niveaux "sensorimoteur" et "schématique" : le niveau "sensorimoteur" déclenche une réaction émotionnelle automatique suivant des caractéristiques innées, et le niveau "schématique" déclenche une réaction émotionnelle automatique suivant des caractéristiques acquises. On peut néanmoins schématiser ces différentes architectures et situer notre moteur de raisonnement dans ce schéma, au niveau des couches cognitives (cf. figure 12.2).

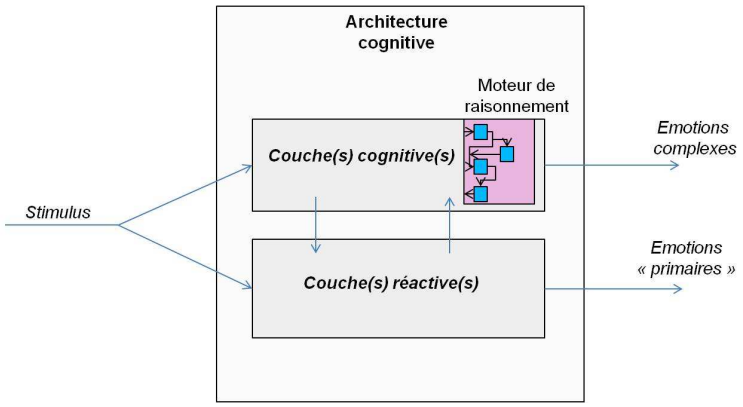

FIG. 12.2 – Le moteur de raisonnement dans l'architecture cognitive des théories de l'évaluation. Adaptée de [Sloman 2001, Scherer 2001].

12.2 Architecture du moteur de raisonnement

Dans le moteur de raisonnement, les états mentaux de l'agent sont formalisés grâce aux opérateurs du modèle BIGRE : ce modèle permet de représenter les émotions complexes d'une part et l'ensemble des états mentaux de l'agent qui s'expriment par le dialogue, comme je l'ai montré dans le chapitre 8. De plus, ce modèle présente plusieurs points communs avec le modèle BDI, étant lui-même basé sur une logique modale de type BDI : les opérateurs de croyances (Bel) et de buts (Goal) sont similaires, tandis que les opérateurs d'idéaux (Ideal) et de responsabilité (Resp) permettent de représenter les normes sociales et le raisonnement contrefactuel. Les opérateurs des modèles BIGRE et BDI partagent également de nombreux axiomes propres à la logique modale.

Le moteur de raisonnement permet à l'ACA, dans le dialogue, de répondre à un acte de discours (ACM) par un autre ACM. La réception des ACM de l'humain, reconnus dans l'énoncé en langage naturel, entraîne :

1. **la perception** de l'ACM et la mise à jour de ses effets sur les croyances de l'ACA ;
2. **l'évaluation** qui déclenche les émotions complexes (E) de l'agent à partir de ses états mentaux BIGR, suivant leur formalisation logique ;
3. **la délibération** qui décide des différentes intentions de l'agent à partir des BIGRE et sélectionne, parmi ces intentions, la plus appropriée ;
4. **la planification** qui décompose l'intention sélectionnée en un ensemble d'Actes de Conversation Multimodaux et qui exécute le premier ACM du plan ainsi construit.

Chacune de ces phases est implémentée dans un sous-module distinct dans l'architecture du moteur de raisonnement (cf. figure 12.3).

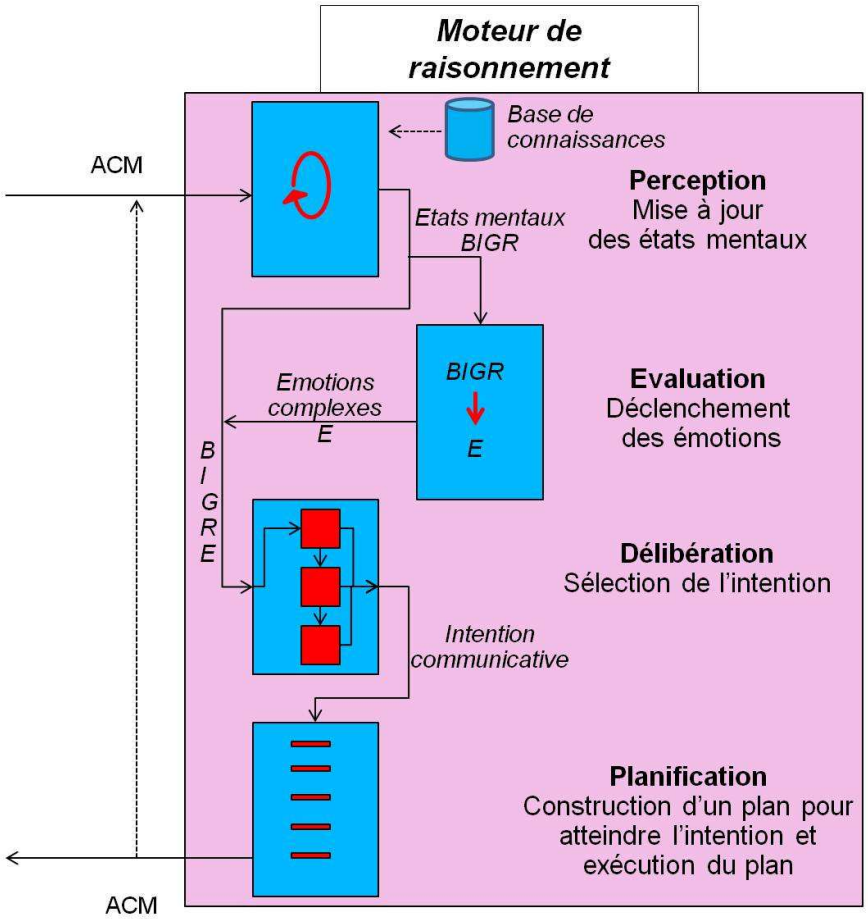

FIG. 12.3 – Architecture du moteur de raisonnement.

Pour permettre à l'agent de manipuler les ACM et ainsi exprimer ses états mentaux, je me suis inspiré des approches par plan du discours [Cohen 1979] en

intégrant dans le moteur la bibliothèque d'ACM dont les effets et les préconditions ont été définis au chapitre 9. Cette intégration va permettre, dans les différents sous-modules, d'une part de planifier par chaînage arrière (dont le mécanisme a été décrit dans le chapitre précédent, cf. section 11.2) les ACM qui permettent d'atteindre l'intention de l'agent, et d'autre part de mettre à jour facilement les états mentaux de l'agent lors de la perception d'un acte (émis ou reçu). Ces différents mécanismes sont détaillés dans les sections qui suivent.
Dans l'architecture SAIBA, le moteur de raisonnement représente le premier module de planification de l'intention communicative ; en sortie de ce module, l'intention communicative (sous la forme d'un ACM) est encodée en FML pour être transmise aux deux derniers modules, qui s'occupent de la planification et de la réalisation du comportement de l'ACA. La boucle est ainsi bouclée (pourrait-on dire), l'ACA répondant à l'utilisateur par le langage et les expressions multimodales appropriées.

12.3 Perception et mise à jour des états mentaux

Le premier sous-module du moteur de raisonnement se décompose en deux phases (cf. figure 12.4) :

- **perception** d'un ACM accompli par l'utilisateur ;
- **mise à jour** des états mentaux de l'agent contenus dans sa base de connaissances.

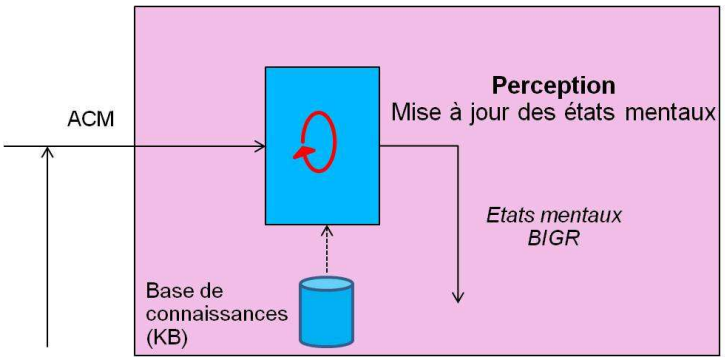

Fig. 12.4 – Le premier sous-module de perception/interprétation.

Lors du dialogue, les énoncés de l'humain sont traduits en ACM et reçus par l'agent[1] : cette perception déclenche la mise à jours des états mentaux de l'agent dans sa base de connaissances. Les ACM accomplis par l'agent lui-même sont également reçus par le module de perception.

La base de connaissances de l'agent est divisée en deux parties, comme on le montre la figure 12.5 : une partie statique qui représente les connaissances initiales de l'agent, et une partie dynamique qui contient l'ensemble des connaissances construites par l'agent au cours du dialogue.

[1] La traduction de l'énoncé en ACM ne fait pas l'objet de ce travail. Elle a été réalisée de façon simple et *ad hoc via* la reconnaissance d'un certain nombre de mots-clés, sélectionnés à partir d'une « ontologie » des différents scénarios (cf. chapitre 13).

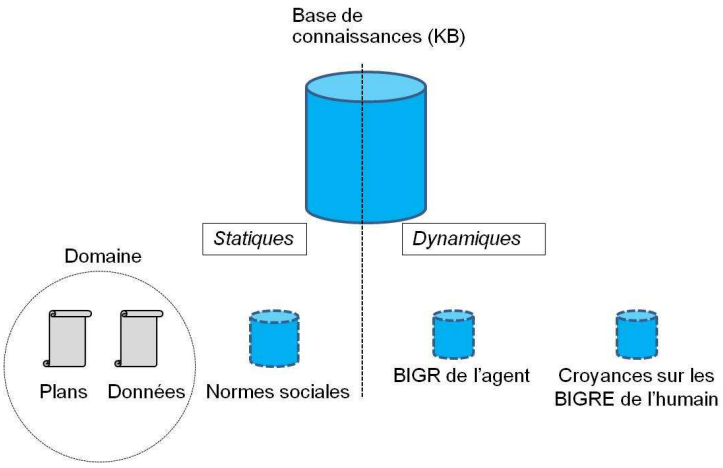

FIG. 12.5 – Parties statiques et dynamiques de la base de connaissances.

Les connaissances initiales de l'agent sont composées tout d'abord de connaissances propres au domaine : ce sont essentiellement des bibliothèques de plans (comment recommander un film de cinéma, comment se réconcilier avec l'utilisateur, comment organiser les vacances etc.) et des données (liste de films, liste de campings etc.) spécifiques au domaine. Les plans du domaine sont en accord avec les approches par plan du discours ; ce sont des séquences d'actions (dont les préconditions et les effets sont définis) qui vont permettre de satisfaire les intentions au niveau "global" du discours (comme je le montre ci-dessous, section 12.5.3). Les connaissances initiales sont également composées des normes sociales (une part des idéaux - I) de l'agent. Elles peuvent être, par exemple, ne pas faire de mal intentionnellement ou être écologiquement responsable, ou pour un agent compagnon, rester en bons termes avec l'humain etc. (la base de connaissances de l'agent et une « ontologie » d'un exemple de scénario sont présentés en Annexe B.2).

Les connaissances dynamiques sont elles composées des états mentaux de l'agent (uniquement des BIGR, les E étant déduites dans le sous-module suivant) et des croyances de l'agent sur les états mentaux de l'utilisateur. Lorsque le dialogue commence, l'agent n'a aucune connaissance dynamique. Au cours du dialogue, la perception d'un ACM déclenche la mise à jours de ces connaissances, qui se déroule en quatre phases :

1. **déduire les effets** de l'acte (effets de réception si l'humain en est l'auteur, ou effets d'émission si c'est l'agent) ;

2. **résoudre les conflits** de croyances potentiels entre les nouvelles croyances et celles de la base ;

3. **ajouter** dans la base les effets de l'acte ;

4. "**saturer la base**", c'est-à-dire appliquer les axiomes des opérateurs BIGR pour déduire de nouvelles connaissances.

12.3.1 Déduction des effets des actes

Les différents effets (de réception et d'émission) des actes ont été définis dans le chapitre 9. Lors de la réception d'un ACM accompli par l'humain ou l'ACA, les effets correspondants sont déduits et participent à la mise à jour des états mentaux de l'agent dans sa base de connaissance ; le sous-module de perception va donc mettre à jour l'image que l'agent a de l'humain (ses croyances sur l'humain). Le fait de conserver et mettre à jour une image des états mentaux (et notamment des émotions) de son interlocuteur est connu en psychologie sous le nom d'"évaluation inverse" (*reverse appraisal* en anglais, [Hareli 2010]). Cette notion est récente et n'a pour l'instant été qu'abordée dans le domaine des agents virtuels [Melo 2009, de Melo 2011] ; elle occupe cependant une place très importante pour des agents compagnons et des applications de plus en plus orientées vers l'humain ("*human-aware*").

La perception et la mise à jour des effets d'émission d'un ACM accompli par l'agent lui-même forment une **boucle de rétroaction** : cette boucle est importante car elle permet à l'ACA de sélectionner d'autres intentions communicatives (comme je le détaille en section 12.5) à partir de la croyance des effets de l'acte.

D'après Perrault [Perrault 1990], les effets d'un acte ne sont pas crus directement par l'agent : ils restent des hypothèses qui sont confirmées (ou réfutées) par la suite du dialogue. Ici, bien que l'agent puisse être considéré comme naïf (il considère l'humain comme sincère), il met à jour ses croyances et les corrige au cours du dialogue.

12.3.2 Conflits de croyances

Lors de l'ajout dans la base d'une nouvelle croyance (les effets d'un acte), des conflits peuvent apparaitre entre cette nouvelle croyance et les croyances actuelles de l'agent. Par exemple, imaginons que l'agent croit φ et qu'il en `informe` l'humain ; l'agent déduit donc les effets de l'ACM `Informer` et les ajoute dans sa base, c'est-à-dire quil croit maintenant que l'humain croit φ ($Bel_i Bel_j \varphi$) ; si l'humain le `contredit` en lui disant que pas du tout, c'est $\neg\varphi$ qui est vrai, les effets de l'ACM `Contredire` qui sont déduits par le sous-module sont en conflit avec les croyances de l'agent ($Bel_i Bel_j \neg\varphi$) : en effet, l'agent ne peut pas croire en même temps que l'humain croit φ *et* $\neg\varphi$, car cela rendrait sa base de connaissances inconsistante. Dans le but de garder une base de connaissances consistante, il faut donc résoudre ces conflits avant l'ajout de la croyance contradictoire, en choisissant la croyance à conserver. Dans l'exemple, un des deux interlocuteurs a tort : il faut donc que l'agent puisse décider, soit d'adopter la nouvelle croyance en abandonnant l'ancienne, soit au contraire de rejeter la nouvelle croyance et `Dire` sa croyance à l'humain (en argumentant par exemple). Un tel mécanisme de décision pourrait reposer sur l'attribution de degrés aux croyances de l'agent, mais, dans l'état actuel de la formalisation, il n'est pas possible d'avoir un degré continu sur l'opérateur Bel_i. Cependant, l'implémentation du moteur de raisonnement en Prolog a permis, comme je le montre dans le chapitre suivant (cf. chapitre 13), de relaxer cette contrainte imposée par la formalisation logique, pour choisir la croyance avec le plus haut degré.

12.3.3 "Saturation" de la base

Une fois les nouvelles croyances ajoutées dans la base de connaissances de l'agent, les axiomes relatifs à la formalisation BIGRE (cf. chapitre 8) sont appliqués pour déduire de nouvelles connaissances. Par exemple, si l'agent croit φ, il est conscient de ses nouvelles croyances ($Bel_i\varphi \rightarrow Bel_iBel_i\varphi$).

De la même façon, l'agent déduit la responsabilité d'une situation de sa définition logique : s'il considère qu'il pouvait empêcher la situation actuelle (φ), il se croira responsable de cette situation ($Cd_i\neg\varphi \wedge \varphi \equiv Resp_i\varphi$).

L'agent peut également considérer l'humain comme responsable, en appliquant la même règle. *A priori*, l'agent peut se croire responsable d'une situation et croire également que l'humain est responsable ; l'agent peut donc avoir au même moment deux émotions complexes du même type (e.g. gratitude et réjouissance), qu'il pourra choisir (ou pas) d'exprimer suivant, comme ci-dessus, les degrés attribués aux opérateurs de croyances et de buts de l'agent.

12.4 Evaluation : déclenchement des émotions complexes

Le sous-module d'évaluation reçoit du sous-module de perception / interprétation les états mentaux BIGR mis à jour (cf. figure 12.6).

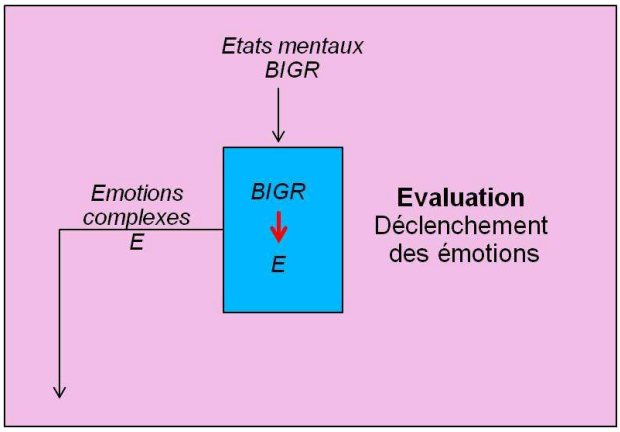

FIG. 12.6 – Evaluation : déclenchement des émotions complexes.

A partir des BIGR, ce sous-module déclenche les émotions complexes E de l'agent d'après leur définition logique : ainsi, la gratitude envers j, par exemple, est déclenchée lorsque les BIGR contiennent sa définition logique - $Goal_i\varphi \wedge Bel_iResp_j\varphi$. Le déclenchement de ces émotions pourra être par la suite modulé par un élément de la personnalité de l'agent, son émotivité (cf. section 12.7).

12.5 Délibération : sélection de l'intention communicative

La délibération est le processus qui permet de décider des intentions communicatives de l'agent et de sélectionner celle que va essayer de satisfaire l'agent ; c'est le même processus qui fait partie du processus de *décision* dans l'architecture BDI (cf. chapitre 11). Dans le cadre d'un dialogue, comme détaillé précédemment, il est important de distinguer deux niveaux dans le dialogue : c'est pourquoi on introduit d'une part les intentions communicatives qui participent à la régulation du dialogue au niveau "local" et d'autre part celles qui participent au niveau "global". Pour cela, je définis trois types d'intentions communicatives :

- l'intention *émotionnelle* est l'intention d'exprimer une émotion (utile au niveau "local") ; par exemple l'intention d'exprimer du regret ;
- l'intention *d'obligations* est une intention déduite de règles d'obligations du discours (utile au niveau "local", cf section 11.3) ; par exemple, l'intention d'`accepter` ou de `refuser` une offre ;
- l'intention *globale* est l'intention de suivre un des buts de l'agent sur lequel celui-ci s'est engagé (niveau "global") ; par exemple, l'intention de se réconcilier avec l'humain.

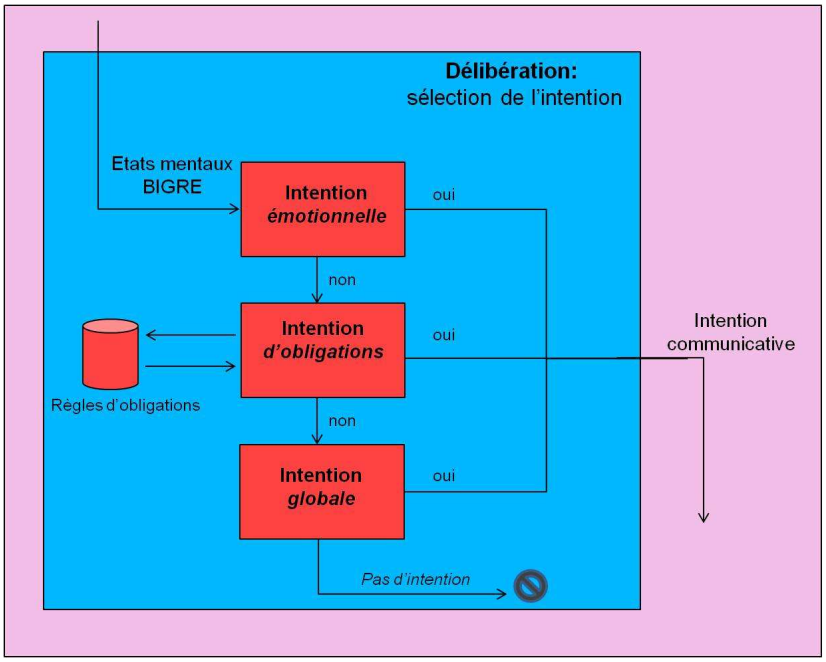

FIG. 12.7 – Délibération : processus de sélection des trois types d'intentions communicatives.

12.5.1 L'intention *émotionnelle*

Après chaque perception et mise à jour des états mentaux, les émotions de l'agent sont déclenchées à partir de ses BIGR dans le sous-module d'évaluation. Dans l'hypothèse d'un agent sincère, émotionnel et expressif, lorsque cet agent "ressent" une émotion, il va adopter l'intention de l'exprimer, de la communiquer à l'humain[2] : j'appelle cette intention communicative l'intention *émotionnelle*. L'intention *émotionnelle* participe à la régulation du dialogue au niveau "local", en permettant notamment une interaction plus naturelle entre l'agent et l'humain [Bates 1994]. Cette intention a donc la priorité sur les autres intentions : l'agent va d'abord essayer de satisfaire cette intention (i.e. d'exprimer son émotion) avant de considérer ses autres intentions.

12.5.2 L'intention *d'obligations*

Un deuxième type d'intention contribue à la régulation "locale" du dialogue : l'intention *d'obligations*. Cette intention est déduite d'un ensemble de règles d'obligations basées sur les travaux de Traum et Allen [Traum 1994]. A chaque acte reçu par l'agent, les règles correspondant à l'acte s'appliquent : lorsqu'une règle se vérifie, l'agent adopte l'intention communicative d'accomplir l'acte provenant de l'obligation. Par exemple, lorsque l'humain `Demande` ou `Offre` quelque chose, l'agent doit `Accepter` ou `Refuser`, suivant ses buts et ses connaissances (tout en vérifiant les préconditions de ces actes décrites précédemment). Le tableau 12.1 liste l'ensemble de ces règles d'obligations, qui consistent en un acte reçu et un acte "réponse" à accomplir suivant certaines conditions.

L'intention *d'obligations* a une priorité inférieure à l'intention *émotionnelle* (i.e. l'agent tente de satisfaire son intention *émotionnelle*, puis adopte l'intention *d'obligations*), mais a une priorité supérieure à l'intention *globale*.

12.5.3 L'intention *globale*

L'intention *globale* correspond, dans le dialogue, au niveau "global" : c'est cette intention qui va donner la direction du dialogue et définir son type. L'intention *globale* de poursuivre un certain but est adoptée lorsque l'agent s'est engagé à atteindre ce but :

- publiquement, par l'intermédiaire d'ACM engageants tels que `Promettre`, `Accepter` etc., ce qui est illustré par l'attribut $Committed_i$ des effets de ces ACM ;
- en privé ; l'agent peut s'engager sur un de ses buts à partir de l'ensemble de ses connaissances et par un raisonnement pratique (en lien avec les plans qu'il connait) .

Ces types d'engagements sont cohérents avec la définition donnée par Cohen et Levesque (cf. chapitre précédent, section 11.1) : les engagements publics correspondent aux "social commitment", et les engagements privés aux "internal commitment". Un exemple d'engagement privé peut être que l'agent a le but de se réconcilier avec l'utilisateur, et qu'il connait un plan pour y parvenir ; si le contexte s'y prête (i.e. si l'agent n'a pas déjà d'autres intentions *globales*

[2] le fait de ressentir ou non une émotion, tout comme l'expressivité de l'agent, pourront être par la suite modulés par sa personnalité, cf. section 12.7

ACM Reçu	Conditions	ACM Réponse
Demander/Exiger (requête)	L'agent sait comment faire ce qu'on lui demande L'agent ne sait pas comment faire (et/ou) ce qu'on lui demande est contraire à son but	Accepter Refuser
Demander (question)	L'agent connait la réponse L'agent ne connait pas la réponse (et/ou) donner l'information est contraire à son but	Informer Refuser
Informer	L'agent connaissait déjà l'information L'agent croit que l'information est fausse	Rappeler Contredire
Offrir	L'agent désire ce que l'utilisateur offre L'agent ne désire pas ce que l'utilisateur offre	Accepter offre Refuser offre
Contredire/Nier	L'agent est certain de sa croyance Le degré de certitude de l'agent est plus faible que le degré d'affirmation de l'utilisateur	Affirmer Informer

TAB. 12.1 – Règles d'obligations basées sur les travaux de [Traum 1994].

ou d'autres engagements incompatibles), l'agent adopte l'intention *globale* de se réconcilier avec l'utilisateur.

12.5.4 Exception

Il est possible dans certains cas que l'agent n'est aucune intention communicative (rien à dire, tout simplement). Dans ces cas, si l'on souhaite entretenir le dialogue avec l'utilisateur, il conviendra de trouver une solution. Pour notre part, nous choisissons dans l'implémentation d'interroger un *chatbot* et de faire « jouer » sa réponse par l'agent.

12.6 Planification pour atteindre l'intention

Le dernier sous-module est le module de planification, dont le fonctionnement général est présenté dans la figure 12.8.

Lorsque aucun plan pour satisfaire l'intention communicative sélectionnée dans le sous-module de délibération n'est connu, le sous-module de planification construit plusieurs plans alternatifs (lorsque cela est possible) composés de séquences d'ACM pour atteindre l'intention ; les ACM, à la manière des approches par plan du discours, forment les opérateurs (les actions) de ces plans. La formalisation des préconditions et des effets de chaque ACM permet ce mécanisme de planification : l'état que l'agent souhaite atteindre est donné par les effets d'un ACM, qui a lui-même des préconditions qui sont elles-mêmes obtenues par les effets d'un autre ACM, et ainsi de suite. Dans le cas des intentions *globales*, les plans du domaine contenus dans la base de connaissances de

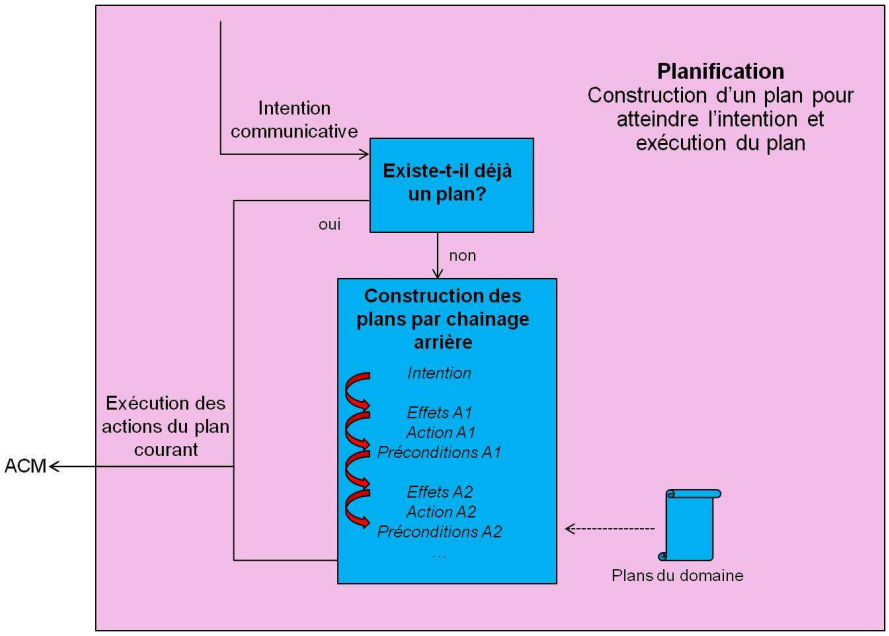

FIG. 12.8 – Planification : construction d'un plan pour atteindre l'intention

l'agent peuvent être nécessaires : par exemple, si l'agent a l'intention de réserver un billet de train pour l'utilisateur, il a besoin de savoir que pour réserver un billet de train, il lui faut connaître la date et le lieu de départ, ainsi que le lieu d'arrivée ; à lui ensuite de décider des ACM appropriés (e.g. `Demander`), grâce aux mêmes mécanismes de planification, pour connaître ces informations.

Dans le cas des autres types d'intentions (*émotionnelles* et *d'obligations*), le plan construit ne comporte en général qu'un seul ACM. Si l'intention *émotionnelle* de l'agent est d'exprimer une émotion de gratitude, le plan sera constitué (après chaînage arrière) de l'ACM `Remercier` ou `Féliciter`, selon le degré de l'émotion, lié au degré du but (cf. chapitre 9).

Lorsque le plan est déjà connu, il est mis à jour à partir des dernières actions effectuées (connues par l'agent) et des actions impossibles à réaliser (e.g. si l'utilisateur refuse de donner une information, l'agent doit trouver un autre plan pour connaître cette information). Si le dernier ACM accompli par l'agent n'a pas été compris par l'humain (e.g. si celui-ci ne lui répond pas ou répond par une autre question), l'ACM reste dans le plan pour être de nouveau accompli par la suite.

La première action du plan ainsi construit (ou mis à jour) est exécutée en sortie du module par l'agent.

12.7 La personnalité de l'agent

De nombreux travaux en psychologie se sont intéressés aux moyens de définir les traits principaux de la personnalité chez l'humain ; parmi ces travaux, le modèle OCEAN (également appelé "big-five factor", [Goldberg 1990]) est le

plus connu, et celui sur lequel s'appuient plusieurs travaux dans le domaine des ACA. Ce modèle définit cinq dimensions de la personnalité : l'ouverture à l'expérience (O), le caractère consciencieux (C), l'extraversion (E), l'agréabilité (A) et le névrosisme ou neuroticisme (N), c'est-à-dire la stabilité émotionnelle.

Pour notre part, nous définissons la personnalité de l'agent selon deux caractères : son émotivité et son expressivité. La définition de la personnalité de l'agent pourra par la suite être étendue (dans les perspectives, cf. chapitre 15), par l'ajout, par exemple, d'autres caractères. La personnalité de l'agent se compose de deux éléments liés directement à ses émotions : son émotivité et son expressivité. L'émotivité de l'agent représente sa capacité à "ressentir" les émotions : un agent très émotif ressent beaucoup d'émotions, et *vice-versa*. L'expressivité de l'agent concerne l'expression de ses émotions "ressenties" : un agent très expressif a l'intention d'exprimer toutes ses émotions, et un agent peu expressif n'exprime que ses émotions les plus "fortes". Par rapport au modèle OCEAN, l'émotivité peut être rapprochée du névrosisme (N) et l'expressivité de l'extraversion (E).

Concrètement, ces deux paramètres fonctionnent comme des seuils : en fixant un seuil bas pour l'émotivité (resp. l'expressivité), l'agent ne ressent (resp. n'exprime) que ses émotions de degré supérieur (buts ou idéaux de degré 1). Pour le moment, la formalisation ne nous permet d'avoir qu'une seule valeur pour le seuil, qui différencie les degrés supérieurs des degrés neutres et inférieurs, mais l'implémentation en Prolog va permettre d'affiner ces paramètres en leur attribuant des degrés continus.

12.8 Exemples de dialogues

Dans les exemples qui suivent, l'agent est appelé *ag* et l'humain *h*. L'agent est défini comme plutôt émotif et expressif et a des connaissances particulières sur l'organisation des vacances (différents plans du domaine, données...).

Exemple de salutation

Humain : Bonjour ! ($\mathbf{Saluer}_{h,ag}$)

- Perception : ajout des effets de l'acte $Goal_h$ *voir mon agent* $\wedge Bel_h$ *vu mon agent* ; déclenchement d'une émotion de joie (l'agent compagnon a le but de voir un ami et il le voit - $Goal_{ag}$ voir l'humain $\wedge Bel_{ag}$ vu l'humain) ;
- Délibération : sélection d'une intention *émotionnelle* - avoir exprimé la joie de voir l'humain ;
- Planification : ce type de joie s'exprime par l'ACM `Saluer` ; construction du plan à un seul ACM.

Agent : Bonjour ! ($\mathbf{Saluer}_{ag,h}$)

Exemple de refus

Humain : Dis-moi s'il va faire beau dans deux mois. (**Demander**$_{h,ag}$**Informer**$_{ag,h}$ *météo dans deux mois*)

- Perception : ajout des effets de l'acte - $Bel_{ag}Goal_hResp_{ag}$**Informer**$_{ag,h}$ *météo dans deux mois* ;
- Délibération : sélection d'une intention *d'obligations* - informer/refuser ; ici l'intention est de refuser car l'agent ne sait pas comment faire ;
- Planification : construction du plan à un seul ACM `Refuser`.

Agent : Je ne sais pas. (**Refuser**$_{ag,h}$**Informer**$_{ag,h}$ *météo dans deux mois*)

- Perception : **boucle de rétroaction** - ajout des effets de l'acte - $Bel_hGoal_{ag}\neg Resp_{ag}$**Informer**$_{ag,h}$ *météo dans deux mois* ; déclenchement d'une émotion de regret (regret d'être responsable de ne pas avoir informé l'humain) ;
- Délibération : sélection d'une intention *émotionnelle* - exprimer ce regret ;
- Planification : ce type de regret s'exprime par l'ACM `S'excuser` ; construction du plan à un seul ACM.

Agent : Désolé ! (**S'excuser**$_{ag,h}Resp_{ag}\neg Bel_h$ *météo dans deux mois*)

Exemple d'acceptation

Humain : Organise-moi mes vacances ! (**Demander**$_{h,ag}$ *organiser vacances*)

- Perception : ajout des effets de l'acte - $Bel_{ag}Goal_hResp_{ag}$ *organiser vacances* ;
- Délibération : sélection d'une intention *d'obligations* - accepter/refuser ; ici l'intention est d'accepter car l'agent connait un plan pour y parvenir ;
- Planification : construction du plan à un seul ACM `Accepter`.

Agent : Pas de problème ! (**Accepter**$_{ag,h}$ *organiser vacances*)

- Perception : **boucle de rétroaction** - ajout des effets de l'acte - $Bel_{ag}Bel_hGoal_{ag}Resp_{ag}$ *organiser vacances* $\wedge Committed_{ag}$ *organiser vacances* ; l'agent s'engage publiquement à organiser les vacances de l'humain ;
- Délibération : sélection d'une intention *globale* - avoir organisé les vacances de l'humain ;
- Planification : avec la connaissance du plan du domaine, construction d'un plan complet pour organiser les vacances (savoir le lieu, la date, le type d'activités souhaitées etc.).

Agent : Où veux-tu partir en vacances ? (**Demander**$_{ag,h}$**Informer**$_{h,ag}$ *lieu des vacances*)

- Perception : ajout des effets de l'acte - $Bel_{ag}Bel_h Goal_{ag}Resp_h$**Informer**$_{h,ag}$ *lieu des vacances* ;
- Délibération : en attente de réponse ;
- Planification : en attente de réponse.

Humain : Je voudrais aller à Toulouse. (**Informer**$_{h,ag}$ *lieu des vacances = Toulouse*)

- Perception : ajout des effets de l'acte - $Bel_{ag}Goal_h Bel_{ag}$ *lieu des vacances = Toulouse* ;
  Mise à jour des états mentaux - $Bel_{ag}Resp_h Bel_{ag}$ *lieu des vacances = Toulouse* ;
  Déclenchement d'une émotion de gratitude (l'agent avait le but de savoir le lieu des vacances et l'humain est responsable du fait que l'agent le sache) ;
- Délibération : sélection d'une intention *émotionnelle* - exprimer cette gratitude ;
- Planification : la gratitude s'exprime par les ACM `Remercier` et `Féliciter` suivant le degré de l'émotion - ici, le degré de la gratitude est plutôt neutre, construction du plan à un seul ACM `Remercier`.

Agent : Merci pour l'information ! (**Remercier**$_{ag,h}Resp_h$**Informer**$_{h,ag}$ *lieu des vacances = Toulouse*)

- Perception : **boucle de rétroaction** - ajout des effets de l'acte - $Bel_{ag}Bel_h Rejouissance_{ag}Resp_h$**Informer**$_{h,ag}$ *lieu des vacances* ;
- Délibération : sélection d'une intention *globale* - avoir organisé les vacances de l'humain ;
- Planification : le plan pour organiser les vacances est déjà construit, mise jour du plan l'agent, continue dans ce plan.

Agent : Et quand veux-tu partir ? (**Demander**$_{ag,h}$**Informer**$_{h,ag}$ *date des vacances*)

- Perception : ajout des effets de l'acte - $Bel_{ag}Bel_h Goal_{ag}Resp_h$**Informer**$_{h,ag}$ *date des vacances* ;
- Délibération : en attente de réponse ;
- Planification : en attente de réponse.

Humain : Quelle est la meilleure période ? (**Demander**$_{h,ag}$**Informer**$_{ag,h}$ *meilleure période à Toulouse*)

- Perception : ajout des effets de l'acte - $Bel_{ag}Goal_h Resp_{ag}$**Informer**$_{ag,h}$ *meilleure période à Toulouse* ;
  Mise à jour des effets précédents
- Délibération : sélection d'une intention *d'obligations* - informer/refuser ; ici l'intention est d'informer de la meilleure période car l'agent connait la meilleure période à Toulouse et est coopératif ;
- Planification : construction du plan à un seul ACM `Informer`.

Agent : D'après les utilisateurs, la meilleure période pour aller à Toulouse est juin ! ($\mathbf{Informer}_{ag,h}$ *meilleure période à Toulouse = juin*)

- Perception : **boucle de rétroaction** - ajout des effets de l'acte - $Bel_{ag}Bel_h$ *meilleure période à Toulouse = juin* ;
- Délibération : sélection d'une intention *globale* - organiser les vacances de l'humain ;
- Planification : mise à jour du plan déjà construit, l'humain n'a pas répondu à la précédente question, l'agent refait l'action.

Agent : Alors, quand veux-tu partir ? ($\mathbf{Demander}_{ag,h}\mathbf{Informer}_{h,ag}$ *date des vacances*)

- Perception : ajout des effets de l'acte - $Bel_{ag}Bel_hGoal_{ag}Resp_h\mathbf{Informer}_{h,ag}$ *date des vacances* ;
- Délibération : en attente de réponse ;
- Planification : en attente de réponse.

Humain : Je ne sais pas, débrouille-toi ! ($\mathbf{Refuser}_{h,ag}\mathbf{Informer}_{h,ag}$ *date des vacances*)

- Perception : ajout des effets de l'acte - $Bel_{ag}Goal_h\neg Resp_h\mathbf{Informer}_{h,ag}$ *date des vacances* ;
  Mise à jour des états mentaux - $Bel_{ag}Resp_h\neg Bel_{ag}$ *date des vacances* ;
  Déclenchement d'une émotion de déception (l'agent avait le but de savoir la date des vacances et l'humain est responsable du fait que l'agent ne le sache pas) ;
- Délibération : sélection d'une intention *émotionnelle* - exprimer cette déception ;
- Planification : la déception s'exprime par les ACM `Se plaindre` et `Se lamenter` suivant le degré de l'émotion - ici, le degré de la déception est plutôt neutre, construction du plan à un seul ACM `Se plaindre`.

Agent : Je ne peux pas travailler dans ces conditions ! ($\mathbf{SePlaindre}_{ag,h}\neg Bel_{ag}$ *date des vacances*)

- Perception : ajout des effets de l'acte - $Bel_{ag}Bel_hDeception_{ag,h}\neg Bel_{ag}$ *date des vacances* ;
- Délibération : sélection d'une intention *globale* - organiser les vacances de l'humain ;
- Planification : mise à jour du plan déjà construit, suite impossible, l'agent essaye de construire un nouveau plan.

...

12.9 Généricité du moteur de raisonnement

Dans le moteur de raisonnement, la bibliothèque d'ACM, les règles d'obligations du discours et les mécanismes de sélection de l'intention, d'inférence et de chaînage arrière sont génériques et indépendants du domaine. La définition et le déclenchement des émotions, les connaissances du domaine (plans, données)

et la personnalité de l'agent ont été modélisés de telle sorte qu'ils sont faciles à modifier et configurer (cf. Annexe B).

13
Implémentation et évaluation

Je présente dans ce chapitre l'implémentation du moteur de raisonnement en Prolog et son intégration dans les ACA Greta et MARC. L'évaluation du moteur et ses résultats sont présentés dans un second temps.

13.1 Implémentation

Cette section présente tout d'abord l'implémentation du moteur en Prolog et les différents avantages de cette implémentation. Le moteur Prolog a ensuite été intégré dans une architecture Java qui fait le lien entre l'humain (*via* une interface) et l'agent virtuel. L'architecture logicielle de l'ensemble du système est présentée en section 13.1.2. Deux agents virtuels ont été dotés du moteur de raisonnement : l'ACA Greta et l'agent MARC. L'implémentation dans Greta (section 13.1.3) a permis de faire directement le lien avec le Langage de Conversation Multimodal (LCM), déjà implémenté dans Greta, comme présenté dans la partie précédente (cf. chapitre 10). Les lexiques de comportement étant établis de manière spécifique à chaque application, la seconde implémentation dans MARC (section 13.1.4) nous a amené à intégrer dans le moteur de raisonnement un certain nombre de *checks* provenant de la théorie de l'évaluation de Scherer (cf. chapitre 3). L'évaluation de ces *checks* (tels que la nouveauté de l'évènement, ses conséquences etc.) dans le dialogue permet de générer des expressions faciales dynamiques à associer aux actes planifiés par le moteur.

13.1.1 Implémentation du moteur de raisonnement en Prolog

Le moteur de raisonnement a été implémenté en Prolog sur les bases du travail réalisé par Carole Adam [Adam 2007]. Ce langage est un des principaux langages de programmation logique, utilisé notamment en Intelligence Artificielle et dans la reconnaissance de la langue naturelle. Ce choix s'explique notamment par le fait que mon travail repose sur une formalisation logique des états mentaux de l'agent (dont les émotions complexes) et leur expression par le langage ; de plus, Prolog propose des mécanismes d'ajout et de rétraction utiles à la mise à jour de bases de connaissances.

L'implémentation en Prolog a consisté à :

- traduire en prédicats la logique du modèle BIGRE et la bibliothèque d'ACM ;

- mettre en place des règles d'inférences (i.e. *si préconditions alors effets*) pour traduire les différents axiomes logiques, l'opérateur de responsabilité et les règles d'obligations du discours ;
- mettre en place les mécanismes de chaînage arrière pour la planification.

Degrés des opérateurs

Les opérateurs du modèle BIGRE sont définis en Prolog par des prédicats auxquels on associe un certain nombre d'arguments. Le principal avantage de cette implémentation est que l'on peut donc associer à ces opérateurs un degré continu réel (entre 0 et 1), permettant ainsi de s'affranchir de certaines contraintes logiques. La traduction des opérateurs de croyance, de but et d'idéal est présentée dans le tableau 13.1.

Opérateur	Prédicat	Définition
$Bel_i\varphi$	bel(i,φ,degré,temps,source)	l'agent i croit φ à un certain temps et un certain degré. On identifie également la source de la croyance, qui peut venir par exemple de la déduction ou du dialogue.
$Goal_i\varphi$	goal(i,φ,degré)	l'agent i a le but φ à un certain degré
$Ideal_i\varphi$	ideal(i,φ,degré)	l'agent i a l'idéal φ à un certain degré

TAB. 13.1 – Traduction des opérateurs de croyance, de but et d'idéal en Prolog.

Pour l'implémentation, la nécessité de représenter les intentions communicatives de l'agent a amené à créer le prédicat d'intention, dont le degré est lié aux buts sur lesquels il s'est engagé (pour ses intentions de type globale), sur le degré de ses émotions (pour ses intentions de type émotionnelle) ou au degré de l'acte accompli par l'humain (pour ses intentions d'obligations). Le prédicat d'intention a donc été implémenté sous la forme : *intend(i,φ,degré,temps,type)*. La sélection de l'intention s'effectue suivant des règles d'inférences traduisant les mécanismes de délibération présentés dans le chapitre précédent.

En accord avec les degrés de puissance attribués aux ACM (cf. chapitre 9), des degrés continus ont été également instaurés lors de l'implémentation de la bibliothèque d'ACM : par exemple, les ACM directifs `suggérer`, `conseiller`, `demander` et `exiger` ont respectivement les degrés 0.25, 0.5, 0.75 et 1. Ces degrés interviennent lors de la planification de l'intention communicative de l'agent : suivant son degré d'intention, il utilise l'ACM correspondant. Par exemple, si l'agent a l'intention d'exprimer une forte culpabilité (dont le degré vient de l'idéal concerné dans la définition), il planifie d'utiliser l'ACM `Se repentir` plutôt que l'ACM `Culpabiliser`. L'implémentation des ACM en Prolog permet de leur associer, en plus d'un degré, des informations complémentaires (un contenu propositionnel P, un émetteur e et un récepteur r) ; on a donc, par exemple, *demander(e,r,P)*.

L'attribution d'un degré continu aux ACM et aux croyances permet de pondérer les effets d'un ACM accompli par l'humain. Le degré de la croyance

dans les états mentaux de l'humain sera plus fort si l'humain utilise un ACM de degré supérieur par rapport à un ACM plus faible : par exemple, si l'humain `Assure` l'agent de φ, le degré de croyance de l'agent dans le fait que l'humain s'engage à faire φ sera plus élevé que si l'humain fait une simple promesse.

Enfin, les différents degrés dans les croyances permettent de résoudre certains conflits potentiels avant l'ajout d'une nouvelle croyance : entre deux croyances contradictoires, l'agent choisit celle qui a le degré le plus grand.

Degrés dans la personnalité

L'émotivité et l'expressivité de l'agent ont également été implémentées par des degrés continus réels (entre 0 et 1). Ainsi, lors du déclenchement des émotions dans le sous-module d'évaluation, les émotions dont le degré est inférieur au degré d'émotivité ne sont pas "ressenties" par l'agent. De même, l'intention d'exprimer une émotion "ressentie" n'est sélectionnée que si le degré de cette émotion est supérieur au degré d'expressivité de l'agent.

13.1.2 Architecture logicielle Java-Prolog

Le moteur de raisonnement implémenté en Prolog a été intégré dans un système Java dont le rôle est de faire l'interface entre l'humain et l'ACA. La figure 13.1 présente l'architecture logicielle de ce système.

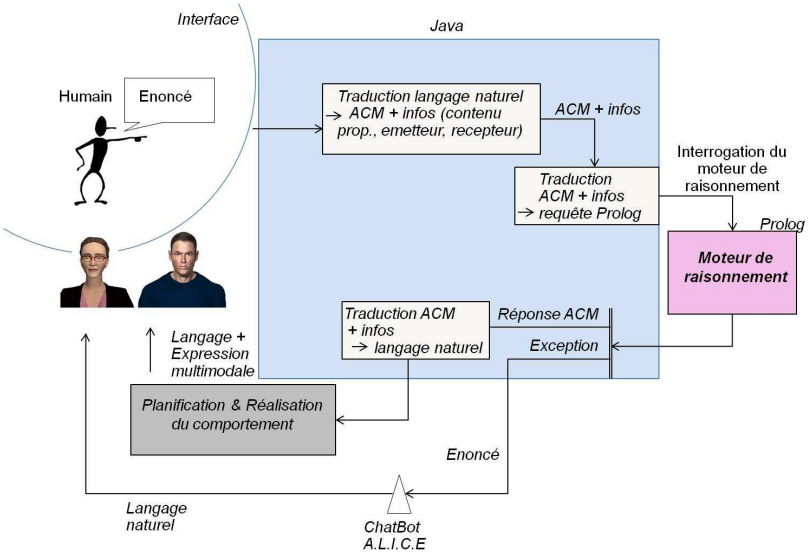

FIG. 13.1 – Architecture logicielle du système

L'utilisateur peut communiquer avec le système *via* une interface constituée d'une interface d'"entrée" et de l'agent virtuel (cf. figure 13.2). L'interface d'entrée offre un résumé du dialogue et deux modalités d'entrées, permettant de percevoir le langage naturel de l'utilisateur : la reconnaissance vocale (adaptée de la technologie Google Voice[1]) et une interface textuelle. L'utilisateur a ainsi

[1] http ://www.google.com/insidesearch/features/voicesearch/index.html

la possibilité de choisir (*via* un bouton “micro”) entre les deux modalités (cf. figure 13.2).

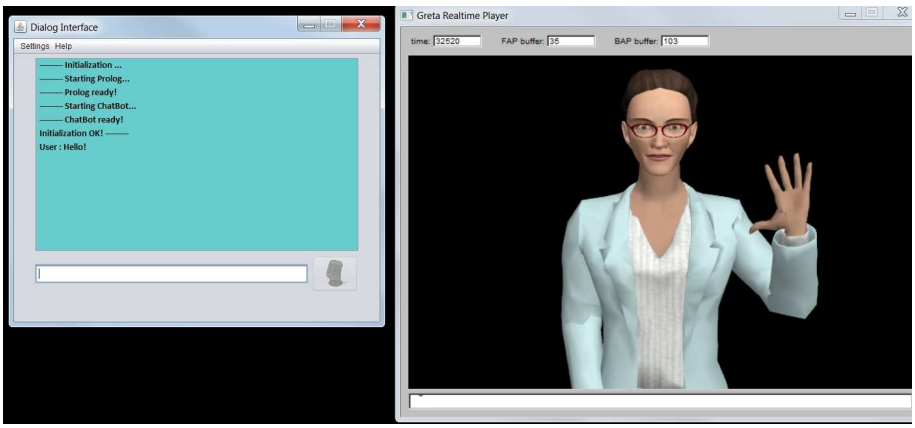

FIG. 13.2 – L’interface utilisateur lors de l’interaction avec l’ACA (à droite) ; l’interface est composée d’un résumé du dialogue (partie supérieure) et d’une zone de texte et d’un bouton “micro” pour la reconnaissance vocale (partie inférieure).

L’énoncé de l’utilisateur est ensuite traduit en ACM (force illocutoire F) et en informations complémentaires (contenu propositionnel P, émetteur, récepteur). La reconnaissance des actes de discours dans la langue naturelle étant un domaine de recherche en elle-même, cette traduction se fait pour le moment par reconnaissance de mots-clés appartenant différents scénarios.

L’ACM ainsi construit est ensuite envoyé, avec les informations complémentaires, dans le moteur de raisonnement Prolog[2]. Le moteur met à jour les connaissances de l’agent, déclenche ses émotions, sélectionne son intention communicative et construit un plan pour l’atteindre (cf. chapitre 12). L’ACM sélectionné en réponse par le moteur de raisonnement passe enfin par la planification et la réalisation du comportement, avant d’être joué par l’ACA. Si aucune intention n’a été sélectionnée par le moteur, une exception se déclenche : l’énoncé de l’utilisateur est transmis au *chatbot* A.L.I.C.E[3] qui fournit une réponse en langage naturel, grâce à de la reconnaissance de mots/modèles.

Les processus de planification et de réalisation du comportement différent suivant l’agent virtuel utilisé. Deux ACA ont été dotés du moteur de raisonnement, Greta et MARC.

13.1.3 L’agent GRETA et le LCM

Dans la partie précédente, le LCM a été implémenté dans l’ACA Greta (cf. chapitre 10), lui permettant ainsi d’exprimer ses états mentaux multimodalement (verbalement et non-verbalement) par le biais d’un ACM et d’un énoncé :

[2]L’interface Java - Prolog utilisée est jpl *http ://www.swi-prolog.org/packages/jpl/java_api/index.html*

[3]ALICE ChatBot : The Artificial Linguistic Internet Computer Entity, `http://alice.pandorabots.com/`

les signaux multimodaux liés à l'ACM, planifié par le moteur en réponse à l'utilisateur, sont détaillés dans le fichier lexique de Greta. Déployer le moteur de raisonnement, intégré dans le système Java, consiste en deux étapes :

1. traduire, dans le système, l'ACM et ses informations complémentaires (ou l'énoncé provenant de A.L.I.C.E) en FML ;
2. transmettre le fichier FML à la suite de l'architecture SAIBA de Greta.

La première étape consiste donc à ajouter, en Java, un sous-module qui reçoit en entrée l'ACM et ses informations, dans le cas où une intention communicative a été sélectionnée par le moteur de raisonnement, et l'énoncé généré par le *chatbot* sinon. La figure 13.3 montre comment ce sous-module s'intègre dans l'architecture logicielle.

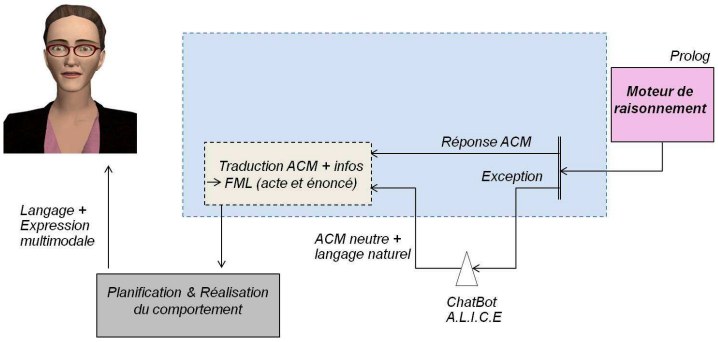

FIG. 13.3 – Traduction en FML de l'intention communicative de Greta.

Pour exprimer l'ensemble des comportements multimodaux liés à un ACM, le second module de l'architecture SAIBA a besoin de recevoir, en FML, cet ACM (dans la balise *performative*) et l'énoncé (dans la balise *speech*), comme je l'ai présenté dans le chapitre 10. Dans le cas où le moteur de raisonnement répond à l'utilisateur par un ACM, l'énoncé de l'ACM est généré, en Java, à partir des « ontologies » des différents scénarios que nous avons implémenté (cf. Annexe B.2). Cet énoncé est introduit en FML par la balise *speech*, tandis que l'ACM est introduit par la balise *performative*. Pour reprendre l'exemple du chapitre 10, si l'ACM sélectionné pour satisfaire l'intention communicative est de la forme *regret(greta,util,eat apple)*, il se traduit en FML par :

```
<fml-apml>
 <bml>
    <speech id="s1" start="0.0" language="english" voice="openmary"
text="I should have not eaten this apple.">
    <description level="1" type="gretabml">
    </speech>
 </bml>
 <fml>
    <performative id="p1" type="regret" start="s1 :tm1"/>
 </fml>
</fml-apml>
```

Dans le cas où aucune intention n'a été sélectionnée par le moteur, le système demande à A.L.I.C.E une réponse à partir de l'énoncé de l'utilisateur. Cette

réponse (en langage naturel) constitue l'énoncé de Greta à traduire en FML par la balise *speech*. On associe à cet énoncé l'ACM `Dire`, car c'est l'ACM assertif le plus neutre.

Le déploiement du moteur de raisonnement dans Greta ferme la boucle de l'interaction humain - ACA : l'acte de l'humain est reconnu par le système, qui le transmet au moteur de raisonnement ; en réponse, le moteur de raisonnement sélectionne et décompose en ACM une intention communicative ; les ACM ainsi planifiés sont traduits en FML et envoyés à l'architecture de Greta, qui associe les signaux multimodaux aux ACM grâce au LCM de son lexique.

13.1.4 L'agent MARC et les *checks* de Scherer

L'agent MARC [Courgeon 2008, Courgeon n C], développé au LIMSI[4], a été l'objet d'un certain nombre d'implémentations (j'ai présenté dans le chapitre 4 deux de ces implémentations). MARC est capable d'exprimer des émotions primaires (*via* un lexique d'expressions faciales, basées sur les travaux d'Ekman [Ekman 1992]) comme des expressions faciales plus complexes, grâce aux Unités d'Action (AU, cf. chapitre 3). MARC propose trois modèles d'agent virtuel, comme le montre la figure 13.4.

Fig. 13.4 – Les trois agents de MARC : de gauche à droite, Maxime, Mary et Simon.

Les recherches sur l'agent MARC sur la dynamique d'expression des émotions, à partir d'une évaluation cognitive des *checks* de Scherer, est particulièrement appropriée pour l'expression des états mentaux. Comme on l'a vu dans l'état de l'art (cf. chapitre 3), l'évaluation des *checks* et leur expression fournit une plus grande richesse d'expressions, et amène les utilisateurs à prêter plus d'états mentaux à l'ACA que lors d'une expression non-dynamique [Courgeon n C].

Nous voulons donc permettre à l'agent MARC d'exprimer l'ACM, sélectionné par le moteur, conjointement aux signaux multimodaux appropriés. Pour cela, nous nous basons sur la théorie de l'évaluation cognitive de Scherer [Scherer 2001] et les liens mis en évidence entre *checks* (critères d'évaluation) et expressions faciales (cf. chapitre 3). L'intégration des *checks* de Scherer dans le système passe par trois étapes :

1. adapter les *checks* de Scherer au contexte du dialogue ACA - humain ;
2. implémenter ces *checks* dans le moteur de raisonnement ;

[4]Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, www.limsi.fr

3. relier dans le système, les *checks* et les AU à partir des travaux existants [Scherer 2007], et les faire jouer par MARC conjointement à l'ACM.

Les *checks* dans le dialogue ACA - humain

Pour intégrer une évaluation des *checks* de Scherer (présenté dans le chapitre 3) dans le moteur de raisonnement, il faut adapter ces *checks* au contexte du dialogue en définissant ce que va être l'évènement à évaluer, sa nouveauté, son plaisir intrinsèque etc.

Dans le dialogue, je propose de considérer chaque ACM émis ou reçu par l'agent comme l'évènement à évaluer. Les *checks* que nous considérons sont les *checks* les plus utilisés dans la communauté ACA [Courgeon n C] : la nouveauté, le plaisir intrinsèque, le rapport aux buts de l'agent et l'attribution dela responsabilité, le potentiel d'adaptation et la rapport aux normes de l'agent. Le tableau 13.2 montre la définition de ces *checks* dans le contexte du dialogue.

Check	Définition [Scherer 2001]	Définition dans le dialogue
Nouveauté	Probabilité et prédictibilité de l'évènement	Notion d'attentes dans le dialogue (liée aux règles d'obligations). Est-ce que l'ACM était *attendu* dans le "schéma" du dialogue ?
Plaisir intrinsèque	Nature de l'évènement (e.g. bon ou mauvais), indépendamment des buts de l'individu	Dépend du type d'ACM (e.g. `Refuser` *vs.* `Accepter`) et du contenu propositionnel
Attribution de la responsabilité et Rapport aux buts	Déterminer qui est responsable de l'évènement et si l'évènement (et ses conséquences) vont contre/dans le sens des buts	Responsabilité et rapport aux buts de l'ACM et de ses conséquences
Potentiel d'adaptation	Evaluation des choix qui s'offrent à l'individu, en réponse à l'évènement, et sélection de ceux qui sont les meilleurs pour l'individu	Est-ce que l'agent peut s'adapter aux conséquences de l'ACM ?
Rapport aux normes internes et externes	Déterminer si l'évènement (et ses conséquences) vont contre/dans le sens des normes sociales (externes), morales et personnelles (internes)	Rapport aux idéaux de l'agent de l'ACM et de ses conséquences

Tab. 13.2 – Cinq des *checks* de Scherer adaptés au contexte du dialogue ACA - humain.

La nouveauté d'un ACM dépend des attentes de l'agent, basée sur les règles d'obligations du discours : par exemple, si l'agent `Demande` ou `Offre` quelque chose, il s'attend à ce que l'humain `Accepte`, s'il pense que c'est congruent avec les buts de l'humain, ou `Refuse` s'il pense le contraire. Tout autre ACM (ou

un refus à la place d'une acceptation et *vice-versa*) est fortement nouveau pour l'agent. Un ACM accompli par l'agent puis évalué par lui n'a aucune nouveauté.

Le plaisir intrinsèque d'un ACM dépend d'une part de la force illocutoire de l'ACM et d'autre part du contenu propositionnel. Chaque ACM est associé, indépendamment du contexte, à un plaisir intrinsèque positif ou négatif (ACM plaisant ou déplaisant) : par exemple, `Refuser`, `Nier` et les ACM expressifs exprimant des émotions négatives (`Se plaindre`, `Protester`...) sont plutôt déplaisants, tandis que les autres peuvent être neutres (e.g. `Affirmer`, `Promettre`) ou plaisants (e.g. `Accepter`, `Remercier`). Le degré de plaisir intrinsèque est lié au degré de puissance des ACM : `Protester` est plus déplaisant que `Reprocher`.

L'attribution de la responsabilité d'un ACM ($Resp_i$ ou $Resp_j\varphi$) et le rapport aux buts ($Goal_i\varphi$ ou $\neg\varphi$) et aux normes de l'agent ($Ideal_i\varphi$ ou $\neg\varphi$) sont liés à ses BIGR, après leur mise à jour, de la même manière que ses émotions complexes. Le rapport aux buts peut être favorable ou défavorable, tandis que le rapport aux normes concerne les violations des normes soit externes (sociales), soit internes (morales) de l'agent.

Enfin, le potentiel d'adaptation à un ACM de l'agent dépend de ses plans, de ses capacités et de ses buts : si un acte (par ses conséquences) fait échouer le plan qu'il est en train de suivre (e.g. l'humain `Refuse` de lui donner une information), il pourra s'adapter s'il connait d'autres plans subsidiaires (e.g. récupérer l'information autre part), et aura un faible potentiel d'adaptation s'il n'en connait pas.

Dans un effort pour situer nos émotions complexes par rapport à ces *checks*, on peut remarquer que certains de ces *checks* correspondent aux raisonnements complexes et aux opérateurs du modèle BIGRE : la congruence aux buts et l'attribution de la responsabilité correspondent aux opérateurs G (buts) et R (responsabilité), et le rapport aux normes se rapporte à l'opérateur I (idéaux). Les opérateurs du modèle BIGRE faisant partie des *checks* de Scherer, on peut donc dire que nos émotions complexes sont incluses dans les émotions déclenchées par l'évaluation de ces *checks*.

De plus, on peut s'attendre à ce que l'expression faciale dynamique, construite à partir du lien entre *checks* et AU, soit plus riche qu'une expression issue d'un lexique d'expressions émotionnelles, dans lequel le regret, par exemple, est lié à une seule expression. Avec l'évaluation des *checks*, le regret peut s'exprimer de plusieurs façons différentes : selon notre définition, il ne dépend que des rapports aux buts et à l'attribution de la responsabilité ; l'expression de l'évaluation des autres *checks* (nouveauté, plaisir intrinsèque et potentiel d'adaptation) ajoute de la dynamique et de la complexité dans l'expression du regret.

Les *checks* dans le moteur de raisonnement

L'évaluation des cinq *checks* de Scherer s'effectue dans le sous-module d'évaluation. Ce sous-module reçoit les BIGR de l'agent mis à jour (dont la croyance de l'agent concernant le dernier ACM accompli) à partir desquels il déclenche les émotions complexes et évalue les *checks*. Les valeurs de ces *checks* sont utilisées par la suite pour former une expression faciale, à jouer conjointement à l'ACM planifié pour atteindre l'intention (cf. figure 13.5).

La séparation de l'évaluation des *checks* et du processus de sélection / planification de l'intention communicative permet également d'associer une expression

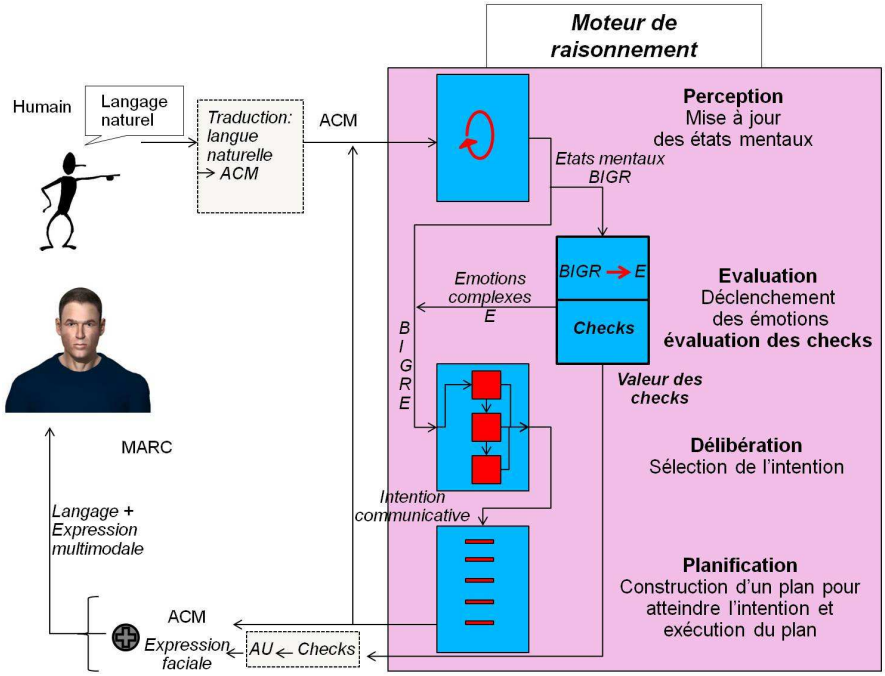

FIG. 13.5 – Intégration de l'évaluation des *checks* dans le moteur de raisonnement.

faciale dynamique aux ACM non-expressifs, planifiés pour atteindre cette intention. Par exemple, si on reprend une partie du dialogue présenté dans le chapitre précédent, concernant l'organisation des vacances :

Agent : Et quand veux-tu partir ?
Humain : Quelle est la meilleure période ?
Agent : D'après les utilisateurs, la meilleure période pour aller à Toulouse est juin[5] !

Après la question de l'utilisateur, aucune émotion n'est déclenchée par le moteur de raisonnement de l'agent, et l'intention sélectionnée (`Informer`) provient des règles d'obligations du discours. Cependant, l'évaluation des *checks* dans ce cas va être une forte nouveauté / surprise (l'agent s'attendait à une réponse à sa question) et un potentiel d'adaptation positif (l'agent sait comment répondre) ; l'expression de l'évaluation de ces *checks* n'est donc pas neutre, et, associée à l'ACM, permet d'`Informer` en ayant pris en compte **le contexte du dialogue**.

L'implémentation des *checks* en Prolog consiste à créer cinq prédicats basés sur des règles qui traduisent leur définition (détaillées précédemment). A chacun des *checks* est associé un degré continu réel : compris entre 0 et 1 pour la nouveauté, et entre -1 et 1 pour les autres pour représenter les *checks* négatifs (plaisir intrinsèque déplaisant, rapport aux buts défavorable, potentiel faible et norme enfreinte). Ce degré est utilisé par la suite pour la traduction des *checks* en AU.

[5]Sondage réalisé auprès d'un échantillon de cinq toulousains non-représentatifs.

Lien entre les *checks* et les AU dans le système

Comme présenté dans le chapitre 3, Scherer et Ellgring [Scherer 2007] ont associé à chacun des *checks* une expression faciale traduite en un ensemble d'AU. Je me suis basé sur leurs travaux pour construire le tableau 13.3 ; le *check* d'attribution de la responsabilité n'apparait pas car il semble qu'aucune expression faciale particulière ne lui soit associée.

Check	Sous-*checks*	Action Units
Nouveauté	Imprévisible	1+2
Plaisir intrinsèque	Plaisant Déplaisant	12+25+38 9+15+1
Attribution de la responsabilité et Rapport aux buts	Favorable Défavorable	6+12+25+44 17+23+4
Potentiel d'adaptation	Elevé Faible	23+25 20+27
Rapport aux normes internes et externes	Normes enfreintes Normes suivies	10+4 6+12+25

TAB. 13.3 – Traduction de cinq des *checks* de Scherer en Action Units (adapté de [Scherer 2007]).

Dans le système global, la traduction de ces *checks* en AU se fait grâce à un nouveau sous-module (cf. figure 13.6). Le degré des *checks* (normé entre 0 et 1) donne le coefficient à appliquer aux AU liées à ce check. En accord avec Scherer, l'expression faciale de chaque check, évalué en séquence, est combiné avec la précédente pour former une expression faciale dynamique (la figure 13.7 montre un exemple de la combinaison des expressions faciales).

L'ACM sélectionné par le moteur est traduit en énoncé par les mêmes mécanismes que pour Greta, et associé à cette expression faciale dynamique pour être transmis à MARC pour l'animation : MARC **conserve l'expression faciale finale** tandis qu'il accomplit l'ACM (l'énoncé). De même, si aucune intention communicative n'a été sélectionnée, l'expression faciale dynamique est associée à l'énoncé calculé par A.L.I.C.E : MARC exprime l'évaluation de chacun de ses *checks*, puis conserve l'expression finale lorsqu'il exprime le texte de l'énoncé.

13.2 Evaluation du moteur de raisonnement

Le but de cette évaluation est de vérifier, au travers d'un scénario, les hypothèses suivantes :

- **H1** : l'intention communicative sélectionnée par le moteur est adéquate à la situation ;
- **H2** : l'intention communicative sélectionnée par le moteur est exprimée par l'ACA de manière appropriée.

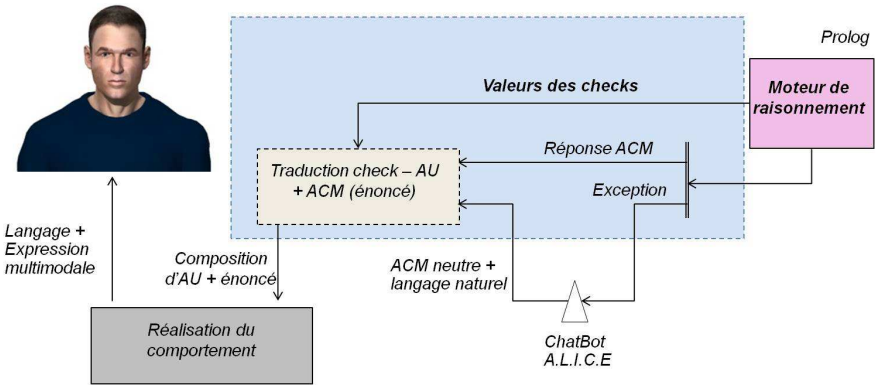

FIG. 13.6 – Traduction des *checks* en AU, qui sont transmises à MARC avec l'énoncé de l'ACM.

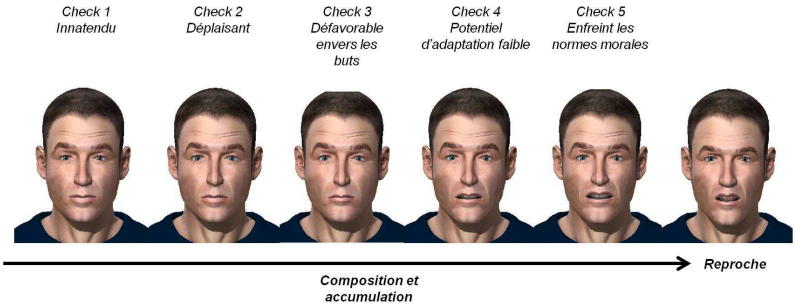

FIG. 13.7 – Exemple de combinaison des expressions faciales après évaluation des *checks*. L'expression finale reste tandis que MARC fait un `reproche` à l'humain.

L'expression multimodale de cette intention par le LCM ayant déjà été validée lors de l'évaluation précédente (cf. chapitre 10), il nous a paru intéressant de mener cette évaluation avec l'agent MARC : nous pensons en effet que l'expression faciale dynamique de MARC permet d'exprimer plus d'états mentaux qu'une expression multimodale prédéfinie. Le déploiement du moteur de raisonnement dans MARC permet également de mener une évaluation en français, MARC acceptant les voix françaises fournies par Microsoft.

Pour mener à bien cette évaluation, nous avons mis en place un scénario dans lequel MARC tente de se réconcilier avec l'utilisateur. Les ACM et une « ontologie » du scénario permettent d'identifier les mots-clés du dialogue et de reconnaitre, dans notre système, les ACM de l'utilisateur (cf. Annexe B.2) ; cette ontologie permet également de générer les énoncés des ACM sélectionnés par le moteur de raisonnement.

13.2.1 Scénario

Le scénario mis en place pour l'évaluation représente une scène volontairement dirigée vers l'affect : ce genre d'interaction affective est au centre de la problématique des agents compagnons (cf. chapitre 1). Il s'agit d'une scène qui se situe après une dispute imaginaire entre l'agent compagnon et le sujet à

propos de politique (dans l'évaluation, on laisse libre cours à l'imagination des sujets sur ce point). Dans la scène elle-même, MARC tente de se réconcilier avec l'humain, et lui demande de lui pardonner ; l'humain a alors trois possibilités : il peut soit accepter de lui pardonner, soit refuser, soit demander un peu de temps pour réfléchir.

Base de connaissances de MARC

Dans le moteur de raisonnement, les connaissances initiales de MARC (dans sa KB) sont les suivantes :

- le plan pour se réconcilier avec l'utilisateur ; pour se réconcilier, il faut que l'utilisateur pardonne à l'agent ;
- l'idéal de MARC, en tant qu'agent compagnon, que l'utilisateur et lui restent amis ($Ideal_{marc}$ *rester amis*).

MARC a de plus des connaissances dynamiques qui représentent ce qui s'est passé avant la scène : il croit qu'il est responsable de la dispute ($Bel_{marc}Resp_{marc}$ *dispute*) et que l'utilisateur et lui sont "fâchés" depuis (i.e. ils ne sont pas réconciliés, $Bel_{marc}\neg$ *réconciliés*) ; ces deux croyances font qu'il croit qu'il est responsable du fait que lui et l'utilisateur ne soient pas réconciliés ($Bel_{marc}Resp_{marc}\neg$ *réconciliés*). Il croit également que si l'utilisateur refuse la réconciliation, ils ne seront plus amis. Avant que le scénario commence et depuis la dispute, MARC a le but de se réconcilier avec l'utilisateur ($Goal_{marc}$ *réconciliés*).

Raisonnement de MARC au cours du scénario

Lorsque le scénario est lancé, le moteur de raisonnement de MARC déclenche l'émotion de regret de ne pas être réconcilié avec l'utilisateur :

$$Goal_{marc}\ \textit{réconciliés} \wedge Bel_{marc}Resp_{marc}\neg\ \textit{réconciliés} \equiv Regret_{marc}\neg\ \textit{réconciliés}.$$

L'intention *émotionnelle* d'exprimer ce regret est donc d'abord sélectionnée : le moyen d'exprimer ce regret (à son degré actuel) passe par l'ACM $\mathbf{Regretter}_{marc,h}\neg$ *réconciliés*, où h représente l'humain.

Après avoir exprimer son regret, MARC sélectionne, à partir de son but de se réconcilier avec l'humain et de ses connaissances statiques (plan pour la réconciliation), l'intention *globale* de se réconcilier avec l'humain, après s'être engagé en privé sur son but. Le dernier sous-module du moteur de raisonnement crée, à partir du plan du domaine, le plan pour se réconcilier avec l'humain : ce plan consiste en un seul ACM, `demander` à l'utilisateur de lui pardonner. Le but que l'utilisateur `accepte` de lui pardonner est ajouté à la base de connaissances de MARC.

Ici, l'utilisateur a donc trois possibilités qui créent trois "sous-scénarios" : `accepter` de pardonner à MARC, `refuser`, ou `demander` un peu de temps pour réfléchir (cf. figure 13.8).

Dans le cas où l'utilisateur `accepte` de lui pardonner, deux émotions complexes sont déclenchées par le module d'évaluation : la **gratitude** (en rapport avec son but) et la **satisfaction** (en rapport avec son idéal). La gratitude provient du but de MARC que l'utilisateur et lui soient réconciliés, et de la croyance qu'il a que l'utilisateur a accepté de lui pardonner, et est donc responsable de la réconciliation :

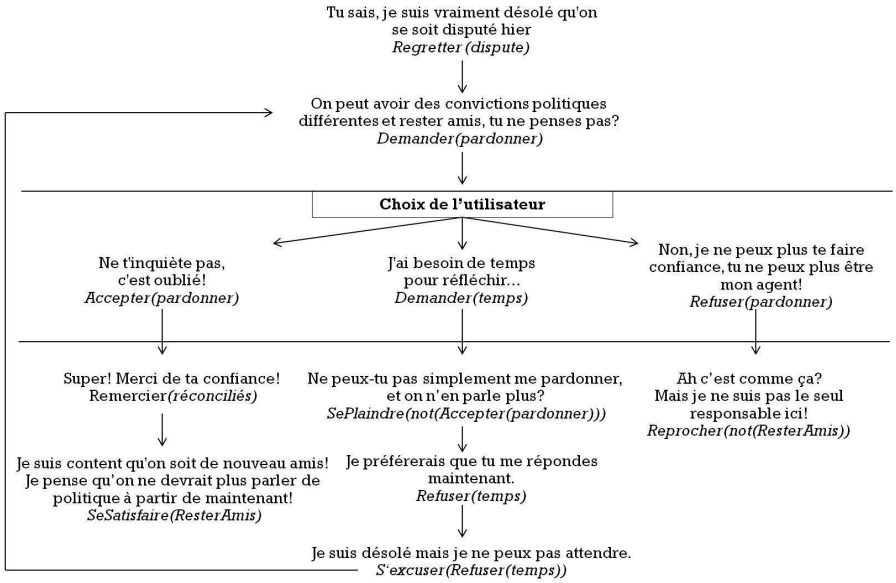

FIG. 13.8 – Scénario de réconciliation entre MARC et l'utilisateur.

$$Goal_{marc}\ réconciliés \wedge Bel_{marc}Resp_h\ réconciliés \\ \equiv Gratitude_{marc,h}\ réconciliés$$

La satisfaction morale provient de l'idéal de MARC (*rester amis*) et du fait qu'il se considère comme responsable du fait que l'utilisateur et lui soient toujours amis (car c'est lui qui a initié la réconciliation) :

$$Ideal_{marc}\ rester\ amis \wedge Bel_{marc}Resp_{marc}\ rester\ amis \\ \equiv SatisfactionMorale_{marc}\ rester\ amis$$

Ces deux émotions étant de degré comparables, les intentions *émotionnelles* de les exprimer sont sélectionnées l'une après l'autre, en commençant (arbitrairement) par l'expression de la gratitude : $\mathbf{Remercier}_{marc,h}$ *réconciliés*. MARC exprime ensuite sa satisfaction par l'ACM $\mathbf{SeSatisfaire}_{marc,h}$ *rester amis*.

Dans le cas où l'utilisateur `refuse` de pardonner à MARC, deux émotions complexes sont également déclenchées par le module d'évaluation : la **déception** (en rapport avec son but) et le **reproche** (en rapport avec son idéal). La déception provient du but de MARC que l'utilisateur et lui soient réconciliés, et de la croyance qu'il a que l'utilisateur a refusé de lui pardonner, et est donc responsable de l'échec de la réconciliation :

$$Goal_{marc}\ réconciliés \wedge Bel_{marc}Resp_h \neg\ réconciliés \\ \equiv Deception_{marc,h}\ réconciliés$$

Le reproche vient de l'idéal de MARC (*rester amis*) et du fait qu'il considère l'humain comme responsable du fait que l'utilisateur et lui ne soient plus amis. La différence dans l'attribution de la responsabilité provient ici de sa définition et de l'introduction des degrés : en effet, MARC considère que l'humain aurait pu faire autre chose (e.g. `Accepter`) qui changerait la situation actuelle et que lui, bien qu'ayant initié la réconciliation, a un degré de responsabilité moins fort. Le reproche dans ce cas est de la forme :

$$Ideal_{marc}\ rester\ amis \wedge Bel_{marc} Resp_h \neg\ rester\ amis$$
$$\equiv Reproche_{marc,h} \neg\ rester\ amis$$

Dans ce cas précis, MARC n'exprime pas l'émotion de déception et exprime directement l'émotion de reproche, dans le but d'illustrer l'attribution d'états mentaux à l'agent par les sujets. Je fais en effet l'hypothèse (**h3**) que les utilisateurs reconnaissent l'émotion de déception, "ressentie" mais non exprimée par l'agent, car ils prêtent des états mentaux à l'ACA [Lee 2005]. MARC exprime donc seulement son reproche grâce à l'ACM **Reprocher**$_{marc,h}\neg$ *rester amis*.

Enfin, dans le cas où l'utilisateur `demande` un peu de temps, quatre intentions sont sélectionnées en séquence par le moteur de raisonnement, après réévaluation de chaque ACM accompli :

1. intention *émotionnelle* : MARC exprime sa **déception** car il a le but que l'utilisateur `accepte` de lui pardonner, et celui-ci n'`accepte` pas (sans `refuser`) ;
2. intention *d'obligations* : MARC doit répondre à la requête de l'utilisateur ; étant donné que son but est que l'utilisateur `accepte` de lui pardonner, MARC `refuse` d'attendre.
3. intention *émotionnelle* : MARC `s'excuse` de ne pas avoir accepté la requête de l'utilisateur, car il croit que c'était le but de l'humain et qu'il est responsable du refus.
4. intention *globale* : MARC continue de suivre son intention *globale* courante de se réconcilier avec l'utilisateur et refait une `demande` de pardon (car l'utilisateur n'a pas explicitement `refusé` ou `accepté` sa demande).

Pour chacun des ACM accomplis par MARC (excepté pour le premier, `Regretter`), l'expression multimodale provient de l'évaluation en termes de *checks* de l'ACM émis par l'humain (ou par lui-même) auquel MARC répond ; pour le premier ACM, MARC évalue le fait qu'il est responsable de la dispute.

13.2.2 Protocole

Pour connaître la perception et l'opinion des utilisateurs face au comportement de l'ACA dans son ensemble (dont émotionnel), nous avons choisi de conduire une étude qualitative. Cette méthode nous permet de recenser de manière large l'ensemble des perceptions présentes chez les participants de notre expérience. Les questions qui leur sont posées permettent d'aller en profondeur dans leurs remarques et ainsi non seulement quantifier mais également comprendre leur attitude face à l'ACA.

Le protocole expérimental[6] consiste à faire jouer le scénario par les sujets, après une mise en situation : on leur demande d'imaginer qu'ils se sont disputés avec leur agent compagnon la veille, et qu'ils le retrouvent aujourd'hui. Nous avons mis en place pour le scénario une interface spécifique, visant à limiter les choix des sujets dans leurs réponses : l'interface consiste en trois boutons, chacun déclenchant un ACM (`Accepter` la réconciliation, la `Refuser`, `Demander` du temps), comme indiqué sur la figure 13.9.

[6] J'aimerais remercier ici l'ingénieur en production et analyse de données des SHS Nadine Mandran, du pôle PimLIg du LIG, spécialisée dans les protocoles d'expérimentation, qui m'a aidé à mettre en place et analyser les résultats de cette évaluation.

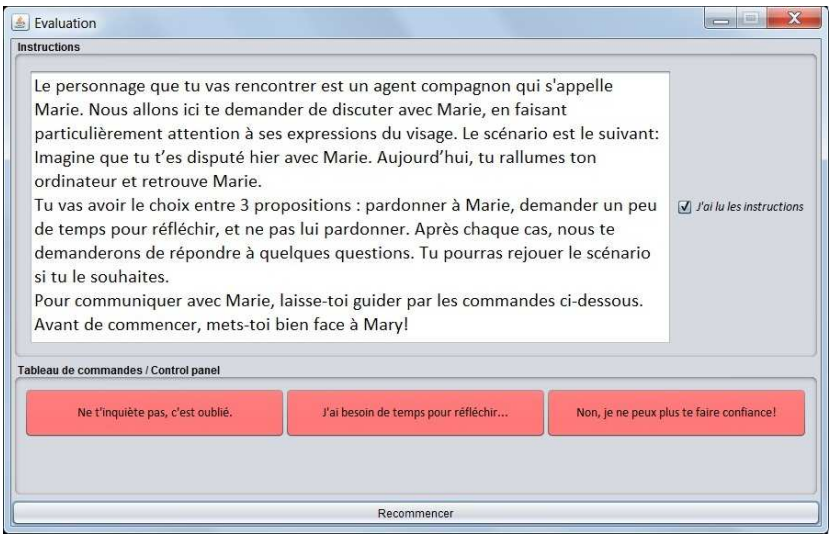

FIG. 13.9 – Interface pour l'évaluation : le sujet a le choix entre les trois possibilités.

Cette interface a été déployée sur une tablette tactile, permettant de séparer MARC (présent sur un écran d'ordinateur) et les moyens d'interaction avec lui. Pour l'évaluation, nous avons décidé d'utiliser la personnalité féminine de MARC, Mary, car des études ont montré que les sujets trouvent les visages féminins plus attractifs [Rhodes 2000].

Les sujets jouent les trois sous-scénarios possibles (le pardon, le refus et la demande de temps) dans un ordre aléatoire. Dans chaque sous-scénario, il est demandé aux sujets d'évaluer la réaction de Mary, et notamment son expression faciale. La réaction de Mary est évaluée de plusieurs façons différentes[7] :

- on demande tout d'abord au sujet quelles émotions ils ont reconnu parmi sept émotions complexes ;
- chaque émotion reconnue est évaluée dans sa sincérité (e.g. « est-ce que Mary semble vraiment satisfaite ? ») et sa crédibilité (e.g. « la satisfaction est-elle appropriée dans ce cas ? ») ;
- il est demandé aux sujets s'ils ont reconnu d'autres émotions ;
- la réaction de Mary **dans son ensemble** (le sous-scénario) est évaluée selon sa crédibilité (« D'après vous, la réaction de Mary est-elle appropriée ? ») ;
- enfin, on demande aux sujets d'expliquer leurs différents choix.

Parmi les émotions que nous avons choisies pour l'évaluation, cinq sont exprimées par Mary dans les scénarios (satisfaction, regret, gratitude, déception et reproche) et deux ne sont ni exprimées ni "ressenties" (la culpabilité et l'admiration) ; ces deux émotions servent de conditions "contrôle" à l'expérimentation.

L'annotation de la reconnaissance des émotions se fait grâce à une échelle qualitative : "pas du tout" (NON), "plutôt pas" (non), "plutôt" (oui) et "tout à fait" (OUI). A chaque émotion "plutôt" ou "tout à fait" reconnue, il est demandé au sujet d'annoter sa sincérité et sa crédibilité sur une nouvelle échelle (dans le

[7] je présente en annexes B.3 le questionnaire utilisé pour l'évaluation.

but d'"alléger" pour le sujet l'évaluation) : "Non", "Ne se prononce pas" (NP), "Oui" (cf. tableau 13.4).

Emotion	Reconnue ?				Sincère ?			Crédible ?		
	NON	non	oui	OUI	Non	NP	Oui	Non	NP	Oui
Regret			X				X	X		

TAB. 13.4 – Exemple d'annotation de reconnaissance d'une émotion et de sa sincérité et sa crédibilité.

Les questions portant sur la reconnaissance d'autres émotions et sur la crédibilité de Mary dans les sous-scénarios sont ouvertes. Pour finir, une évaluation de Mary est effectuée au travers de trois questions :

- Finalement, qu'est-ce que vous pensez de Mary ?
- Est-ce que vous aimeriez l'avoir comme agent compagnon ?
- Avez-vous d'autres remarques ?

Les réponses apportées pas les sujets vont permettre de se faire une idée du type de relation qui se met en place en un cours laps de temps entre l'humain et l'agent ; de plus, elles permettront de comprendre certaines réponses, données aux questions précédentes, en faisant expliciter au sujet ses raisons.

13.2.3 Résultats

Dix-sept sujets (trois sujets féminin et quatorze sujets masculins) entre douze et trente ans (la moyenne d'âge est de 16.9 ans) ont évalué Mary et le moteur de raisonnement lors de l'évènement RobotFesta[8] (cf. figure 13.10). Sur ces dix-sept sujets, sept ont choisi le scénario de refus en premier, six, la demande de temps, et quatre ont choisi le pardon en premier.

La première étape d'analyses des résultats consiste à vérifier que les émotions exprimées par l'agent ont été reconnues par les sujets.

Reconnaissance des émotions

Pour évaluer la reconnaissance des émotions, nous avons mis en place un indicateur de "justesse", fonction de la reconnaissance de l'émotion par le sujet et de l'expression de l'émotion par l'agent : pour que le jugement du sujet soit "juste", il faut qu'il ait reconnu l'émotion lorsqu'elle est exprimée, et qu'il ne l'ait pas reconnu lorsqu'elle n'est pas exprimée. Par exemple, une valeur de J indique soit qu'une émotion exprimée par l'ACA a été "tout à fait" reconnue, soit qu'une émotion non-exprimée par l'ACA a été "pas du tout" reconnue. Le tableau 13.5 montre les trois valeurs que peut prendre l'indicateur.

Les deux émotions "contrôle" de culpabilité et d'admiration (non exprimées dans aucun sous-scénario) ont été jugées en majorité justes (J) : entre 16 et 17 sujets (soit entre 94% et 100%) ont jugé l'admiration de manière J et MJ, et entre 12 et 16 sujets (70% et 94%) ont jugé la culpabilité de manière J et MJ (cf. figure 13.11).

A noter que la culpabilité est jugée moins juste que l'admiration (plus de MJ que de J dans le sous-scénario de demande de temps) : cela pourrait s'expliquer

[8]Rencontres Académiques de Robotique collèges et lycées, http ://crdp2.ac-rennes.fr/robofesta/

FIG. 13.10 – L'agent Mary évaluée à RobotFesta ; la communication avec Mary se fait *via* la tablette tactile.

Reconnue ?	Emotion exprimée	Emotion non exprimée
NON	PJ	J
non	PJ	MJ
oui	MJ	PJ
OUI	J	PJ

TAB. 13.5 – Indicateur de justesse selon la reconnaissance et l'expression de l'émotion : Pas juste (PJ), Moyennement Juste (MJ) et Juste (J).

par la difficulté pour les jeunes sujets d'appréhender le concept de culpabilité (remarques notées lors de l'entretien qualitatif).

Les autres émotions sont jugées justes en moyenne dans les sous-scénarios, comme peut en témoigner le jugement de la satisfaction, exprimée dans le sous-scénario de pardon (cf. figure 13.12).

Deux émotions sont cependant reconnues dans des sous-scénarios où elles ne sont pas exprimées (cf. figure 13.13) : ce sont la déception et le reproche. La déception est exprimée dans le sous-scénario de demande de temps, où elle est reconnue par 15 sujets sur 17 (soit 88%), de manière J et MJ. La reconnaissance de la déception dans le sous-scénario de refus (15 sujets sur 17 également), dans lequel le reproche est exprimé, pourrait valider l'hypothèse que nous avons faites précédemment : la déception est en effet "ressentie" par l'agent et nous avions volontairement empêché l'agent de l'exprimer. Les sujets semble donc prêter les états mentaux particuliers de la déception à l'agent dans ce cas, comme nous en avions fait l'hypothèse (**h3**).

Cependant, la reconnaissance de l'émotion du reproche dans le sous-scénario de demande de temps (14 sujets sur 17 - 82%, contre 16 sujets sur 17 - 94% - dans le sous-scénario de refus où elle est exprimée), alors qu'elle n'est ni exprimée ni "ressentie" par l'agent, mais que la déception est exprimée, pourrait indiquer

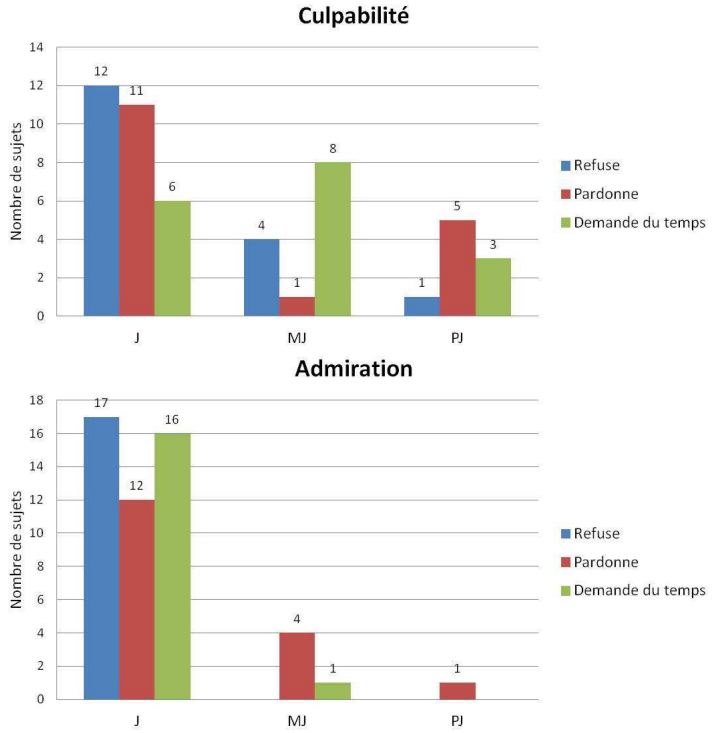

FIG. 13.11 – Résultats de la justesse des jugements des sujets pour l'admiration et la culpabilité.

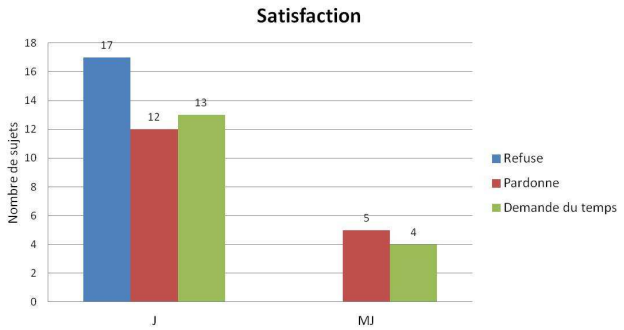

FIG. 13.12 – Résultats de la justesse des jugements des sujets pour la satisfaction.

que les termes émotionnels ont été confondus et/ou mal compris par les sujets. Une analyse des scores de sincérité et de crédibilité de ces deux émotions pourra confirmer ou infirmer ces hypothèses.

La justesse de la reconnaissance des émotions par les sujets[9], dans l'ensemble, semblent indiquer que les expressions faciales des états mentaux de l'agent sont

[9]L'ensemble des résultats sont présentés en annexes B.4

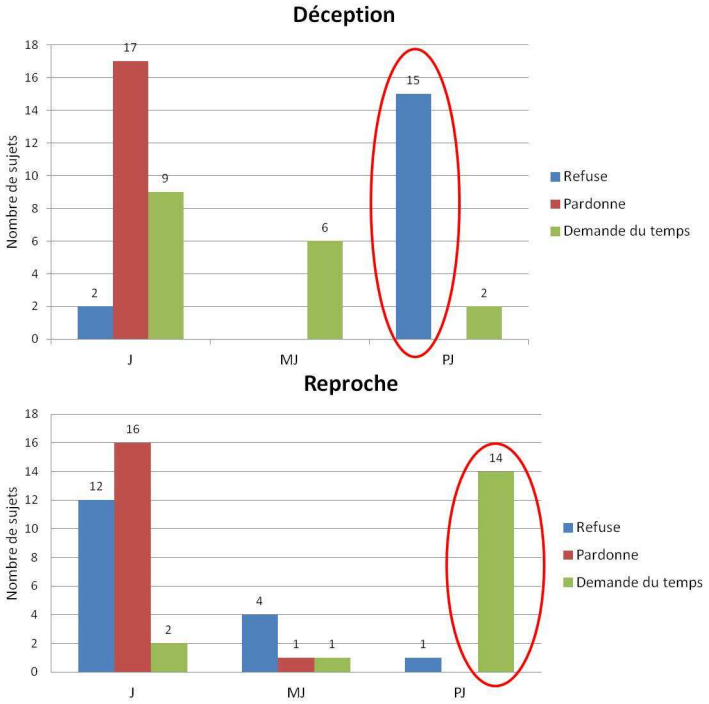

FIG. 13.13 – Résultats de la justesse des jugements des sujets pour la déception et le reproche.

bien reconnus dans le scénario. L'étude des résultats en sincérité et crédibilité des émotions nous permet d'approfondir ces résultats.

Sincérité et crédibilité des émotions

A chaque émotion plutôt reconnue ("oui") ou tout à fait reconnue ("OUI") par un sujet (même si elle n'est pas exprimée par Mary), il lui est demandé d'évaluer cette émotion en termes de sincérité et de crédibilité. Comme indiqué ci-dessus, ces deux critères se traduisent, pour la satisfaction par exemple, par les questions suivantes : « est-ce que Mary semble vraiment satisfaite ? » et « la satisfaction est-elle appropriée dans ce cas ? ».

En annexes (partie B.4), je présente les scores en sincérité et crédibilité de l'ensemble des émotions. On peut s'intéresser ici à quelques résultats particuliers : les scores de la satisfaction, de la culpabilité, du reproche et de la déception.

Les scores de sincérité et de crédibilité de la satisfaction sont présentés dans la figure 13.14. L'émotion de satisfaction est reconnue par les 17 sujets lorqu'elle est exprimée par l'agent (sous-scénario de pardon) et n'est reconnue que dans ce cas : elle est jugée sincère par 11 sujets sur 17 (soit 65%) contre 5 sujets (29%) non sincère et 1 sujet (6%) qui ne se prononce pas ; elle est jugée crédible par 16 sujets sur 17 (soit 95%), contre 1 sujet (6%) qui ne se prononce pas. Les 5 sujets ayant jugé l'expression de la satisfaction non sincère expliquent ce choix par le manque de "naturel" de cette expression : ainsi, pour ces sujets, le sourire

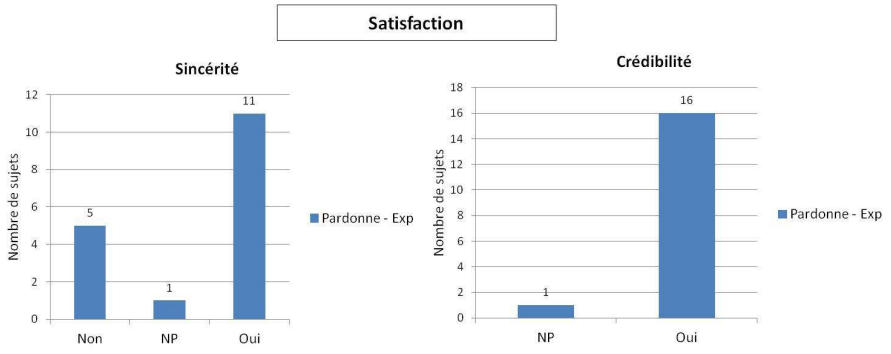

FIG. 13.14 – Scores de sincérité et de crédibilité de la satisfaction.

« *un peu forcé* » de Mary trahit ses véritables états mentaux (pour 4 des sujets sur 5, « *elle s'attendait à ce qu'on lui pardonne* »). Par contre, cette émotion est jugée en majorité crédible car, pour les sujets, elle est appropriée à la scène.

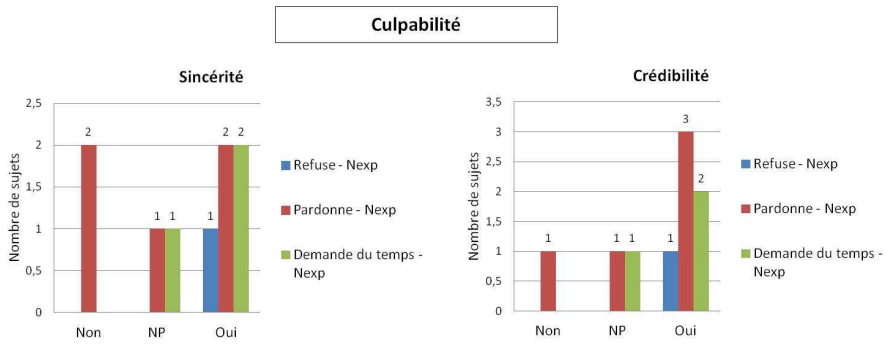

FIG. 13.15 – Scores de sincérité et de crédibilité de la culpabilité.

Les scores de sincérité et de crédibilité de la culpabilité sont présentés dans la figure 13.15. L'émotion de culpabilité est reconnue dans les sous-scénarios de refus (un seul sujet), de pardon (5 sujets) et de demande de temps (3 sujets) ; au total, 7 sujets différents (sur 17) ont reconnu de la culpabilité dans un des sous-scénarios. Au vu des différents scores, il est difficile d'obtenir une tendance nette d'un côté (sincère ou crédible) ou de l'autre (non sincère ou non crédible) : cela peut s'expliquer par le faible nombre de sujets d'une part, et pourrait s'expliquer d'autre part par le fait que la culpabilité n'est ni exprimée ni ressentie par l'agent.

Enfin, il est intéressant d'analyser les scores de sincérité et de crédibilité de la déception et du reproche (cf. figure 13.16), chacune de ces deux émotions étant reconnue lorsque l'autre est exprimée. La déception, tout d'abord, est jugée sincère et crédible dans le sous-scénario de demande de temps où elle est exprimée, mais également dans le sous-scénario de refus, où elle est ressentie mais non exprimée. Les scores sont élevés dans les deux scénarios : 11 sujets sur 15 (soit 73%) la jugent sincère et 13 sujets sur 15 (87%) la jugent crédible dans le sous-scénario de refus ; dans le sous-scénario de demande de temps, 15 sujets sur 15 (100%) la jugent sincère et 11 sur 15 (73%) crédible. Ces scores

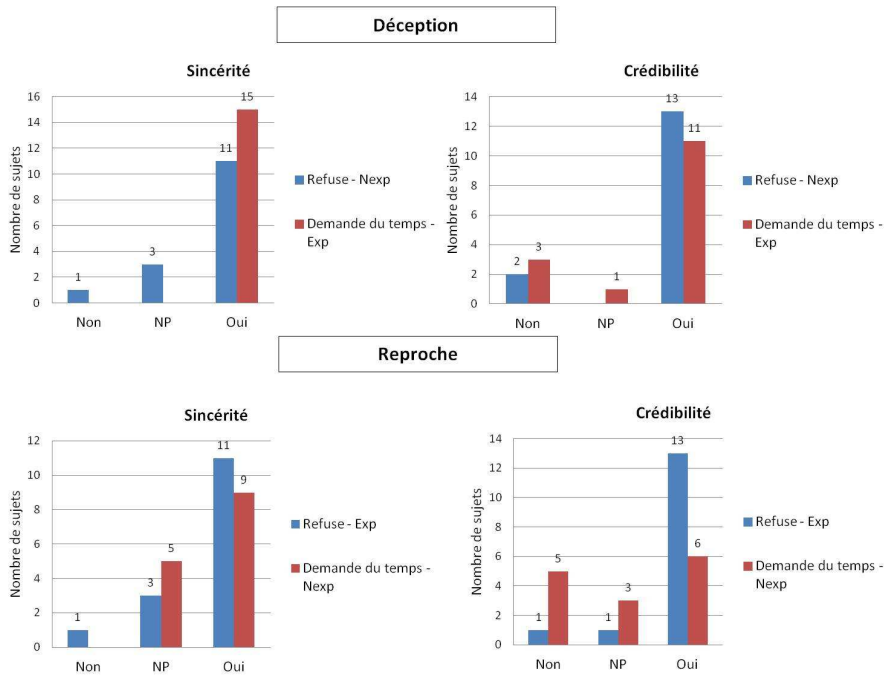

FIG. 13.16 – Scores de sincérité et de crédibilité de la déception et du reproche.

élevés de sincérité et de crédibilité pour la déception, alors qu'elle est "ressentie" mais non exprimée, semblent confirmer l'hypothèse précédente : il semble que les sujets prêtent à l'agent des états mentaux qu'ils pensent appropriés dans la situation.

Les scores de sincérité de l'émotion de reproche, dans chacun des sous-scénarios de demande de temps (où elle n'est ni exprimée ni "ressentie") et de refus (où elle est exprimée et "ressentie"), sont proches : 11 sujets sur 15 (soit 73%) la jugent sincère lorsque l'agent l'exprime, contre 9 sujets sur 14 (65%) lorsque l'agent ne l'exprime pas (et 5 sujets sur 14 - 35% - ne se prononcent pas). Cependant, le reproche est jugé moins crédible lorsque l'agent ne l'exprime pas (6 sujets sur 14 - 43% - la jugent crédible, contre 5 sur 14 - 36% - non crédible et 3 sujets - 21% - qui ne se prononcent pas) que lorsque l'agent l'exprime (16 sujets sur 15 - 87% - la jugent crédible, contre 1 sujet - 6.5% - non crédible et 1 sujet - 6.5% - qui ne se prononce pas). Dans le sous-scénario de demande de temps, le reproche semble donc être reconnu par les sujets (et jugé sincère), mais les sujets n'affirment pas clairement si le reproche est approprié ou pas dans cette scène. Il est donc assez difficile de déterminer pourquoi le reproche est reconnu dans ce cas ; on pourra retenir l'hypothèse de la proximité des termes émotionnels sans toutefois pouvoir conclure.

Pour résumer, les émotions reconnues sont jugées en moyenne sincères et crédibles par les sujets. Il semble donc que, d'une part, les états mentaux déduits de l'agent soient appropriés à la situation, et d'autre part, sa façon de les exprimer (sa sincérité) soit également appropriée.

Autres émotions reconnues dans les scénarios

A la question « avez-vous reconnu d'autres émotions exprimées par Mary ? », la réponse la plus significative (cf. figure 13.17) a été la colère pour 13 sujets sur 17 (soit 76%) lorsqu'on ne pardonne pas à Mary et qu'elle exprime du reproche (sous-scénario de refus). Ce résultat est très intéressant car il pourrait impliquer une relation forte entre la colère (considérée comme une émotion *primaire*, cf. chapitre 3) et le reproche.

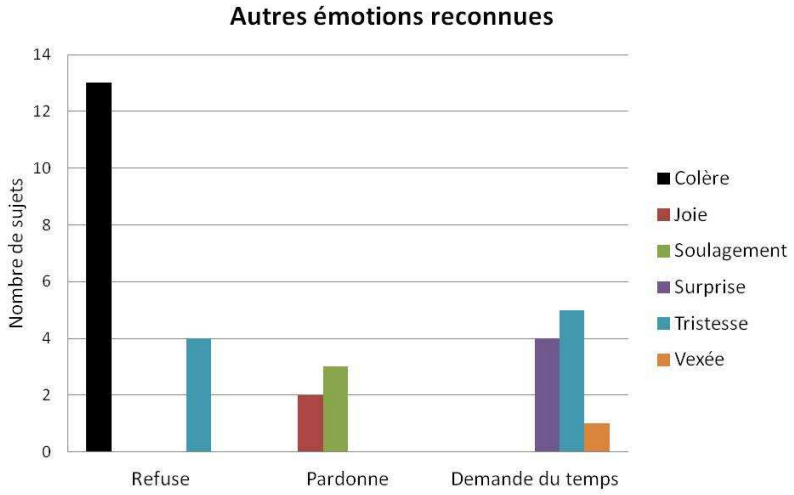

FIG. 13.17 – Autres émotions reconnues suivant les différents sous-scénarios.

La tristesse a également été évoquée dans les sous-scénarios de demande de temps (5 sujets sur 17 - 29%) et de refus (4 sujets sur 17 - 24%). L'émotion complexe de déception ayant été reconnue par une majorité de sujets dans ces deux sous-scénarios, la reconnaissance de la tristesse est cohérente avec la définition que nous en avons faite (cf. chapitre 8) : la déception implique l'émotion de base de tristesse.

Enfin, 4 sujets sur 17 (24%) ont également reconnu de la surprise dans le sous-scénario de demande de temps. Ce résultat peut s'expliquer par l'évaluation du check de nouveauté : l'agent s'attend en effet à ce que l'humain `accepte` ou `refuse` sa demande, et pas du tout à ce que l'humain lui réponde autre chose. L'inattendu de l'évènement se traduit par les AU 1 et 2 qui se traduisent par (notamment) un haussement des sourcils, signe de surprise [Ekman 1992].

Crédibilité des sous-scénarios

A la fin de chaque sous-scénario, il est demandé aux sujets d'évaluer la crédibilité de l'ensemble de la réaction de Mary : « D'après vous, la réaction de Mary est-elle appropriée ? ». Ces scores de crédibilité dont présentés dans la figure 13.18.

La réaction de Mary est jugée appropriée par 13 sujets sur 17 (soit 76%) lorsqu'on refuse le pardon et par 12 sujets sur 17 (soit 71%) lorsqu'on lui pardonne. Ces résultats montrent que la réaction de Mary est jugée crédible dans

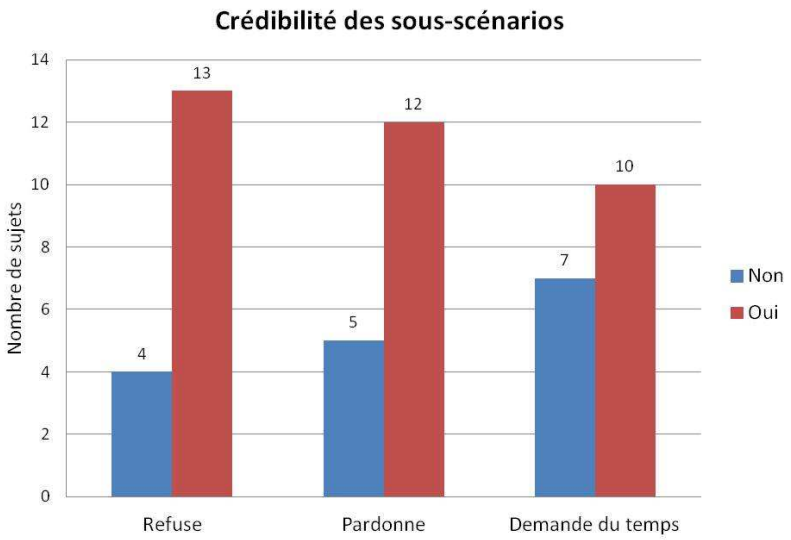

FIG. 13.18 – Réponses à la question : « D'après vous, la réaction de Mary est-elle appropriée ? »

ces deux sous-scénarios. A l'inverse, dans le sous-scénario de demande de temps, la réaction de Mary est jugée appropriée à seulement 59% (10 sujets sur 17). Deux hypothèses peuvent expliquer ce résultat. La première hypothèse est que la reconnaissance de l'émotion de reproche par les sujets dans ce sous-scénario, jugée sincère mais crédible par seulement la moitié des sujets, inluence le jugement de la crédibilité du sous-scénario. Sur les 7 sujets qui jugent la réaction de Mary inappropriée, tous reconnaissent du reproche et 4 sujets le jugent pas crédible. La deuxième hypothèse vient de l'étude des remarques des sujets, lorsqu'ils expliquent leur choix : le fait que Mary `refuse` la demande du sujet de lui accorder un peu de temps n'est pas crédible pour certains sujets (pour 3 des sujets sur 7, « *elle ne respecte pas [mon] choix* »). Ces deux hypothèses ne s'excluent pas mutuellement ; il est difficile en l'état actuel de valider avec certitude une de ces hypothèses.

Existe-t-il un effet d'ordre ?

Chaque sujet joue les trois sous-scénarios dans un ordre aléatoire : 7 sujets ont choisi le scénario de refus en premier, 6 ont ont demandé du temps à Mary en premier, et 4 ont choisi le pardon en premier. Il est intéressant dans l'analyse des résultats de cette évaluation de se poser la question de l'influence de l'ordre des sous-scénarios pour chaque sujet : existe-t-il un tel effet d'ordre ? Plus précisément, la reconnaissance des émotions lors des *premiers* sous-scénarios est-elle comparable avec la moyenne générale des résultats ?

La figure suivante montre par exemple comment sont reconnues les émotions de culpabilité et de déception dans chaque sous-scénario : on compare à chaque fois le pourcentage de J et MJ des sujets qui ont eu le sous-scénario en premier (Ordre 1) et la moyenne générale (ordre 1, 2 et 3) de J et MJ.

Sur la figure, on peut remarquer que les résultats de reconnaissance entre

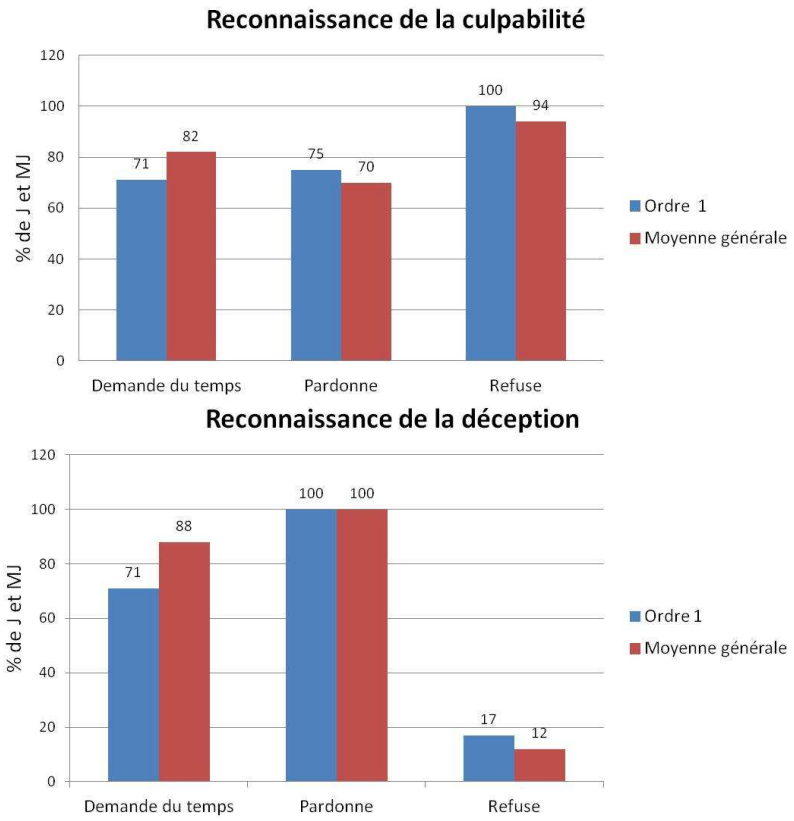

FIG. 13.19 – Analyse de l'influence de l'ordre des sous-scénarios dans la reconnaissance des émotions de culpabilité et de déception.

Ordre 1 et la moyenne générale semblent proches. Cependant, si on regarde par exemple la reconnaissance de la déception dans le sous-scénario de demande de temps, 5 sujets sur 7 (soit 71%) ont reconnu de façon J et MJ la déception en Ordre 1, contre 15 sujets sur 17 (88%) en moyenne ; il est difficile ici de savoir si la différence de 17% entre ces deux résultats est significative, car d'une part, il y a trop d'hétérogénéité dans les effectifs, et d'autre part l'effectif dans chaque premier sous-scénario (Ordre 1) n'est pas assez important. Malheureusement, cette remarque est vraie pour l'ensemble des résultats : il ne nous est pas possible de savoir ici s'il y a une une influence significative de l'ordre des sous-scénarios sur les sujets.

Entretien qualitatif

A la fin de l'évaluation, il a été demandé aux sujets ce qu'ils pensaient de Mary, s'ils aimeraient l'avoir comme compagnon artificiel et s'ils avaient d'autres remarques. Sur les 17 sujets, 7 sujets (soit 41%) aimeraient l'avoir comme compagnon artificiel, contre 9 sujets (53%) qui n'aimeraient pas et 1 sujet qui « *ne voit pas l'intérêt* ». Si on regarde en détails ce que pensent ces sujets de Mary, sur les 7 qui aimeraient l'avoir comme compagnon, 3 sujets

estiment qu'elle « *a du caractère* », 2 autres qu'elle est « *naturelle* » et 1 sujet la trouve « *sympa* ». A l'inverse, sur les 9 sujets qui n'aimeraient pas l'avoir comme compagnon, 3 la trouvent « *soupe au lait* » ou « *lunatique* », 2 la trouvent « *froide* » et 2 autres « *susceptible* ».

Une première explication à ce résultat peut être que la majorité des sujets qui n'aimeraient pas l'avoir comme compagnon ne trouvent pas crédible la réaction de Mary dans le sous-scénario de demande de temps (6 sujets sur 9 - et 6 sujets sur les 7 qui ne la trouvent pas crédible dans ce sous-scénario, cf. ci-dessus). Il est donc possible que le manque de crédibilité dans ce sous-scénario ait influencé la réponse des sujets à cette question.

Une deuxième explication, au vu de ces remarques, peut être que le choix du scénario (une tentative de réconciliation après une dispute) a également eu une influence négative sur l'affect des sujets : deux sous-scénarios sur les trois sont en effet des réactions plutôt négatives de la part de Mary (qui exprime de la déception, du regret ou du reproche) ; on peut imaginer qu'une évaluation basée sur des sous-scénarios contenant plus de réactions positives de la part de Mary amènerait à de meilleurs résultats.

Conclusion

Le moteur de raisonnement que j'ai présenté dans cette partie s'appuie à la fois sur le cycle de comportement BDI, la logique du modèle BIGRE, notre Langage de Conversation Multimodal (LCM) et plusieurs travaux antérieurs sur le dialogue. Ce moteur déduit, pour un ACA expressif, affectif et sincère, les états mentaux à exprimer (son intention communicative) dans un dialogue avec l'humain : l'intégration du LCM assure la sincérité de l'ACA, les émotions déduites du modèle BIGRE assurent son côté affectif et son expressivité est maintenue par la définition de trois types d'intentions communicatives.

Ces trois types d'intentions communicatives participent à la régulation locale et globale du dialogue en intégrant les émotions de l'agent et des règles de dialogue au niveau local, et la notion d'engagement au niveau global. L'intention communicative sélectionnée par ordre de priorité est planifiée par chaînage arrière ; le plan ainsi construit est une suite d'Actes de Conversation Multimodaux (ACM) que va accomplir l'agent, en réponse aux actes de l'humain.

Le moteur de raisonnement a été implémenté dans l'ACA Greta et l'agent MARC. Dans Greta, il fait le lien avec le LCM : les ACM déduits par le moteur de raisonnement sont exprimés par Greta conjointement aux expressions multimodales appropriées, définies dans la partie précédente (partie III). Dans MARC, l'expression multimodale de l'agent lors de l'acte est celle déduite de l'évaluation d'une partie des *checks* de Scherer, que j'ai adapté au dialogue : à chaque acte reçus ou émis par l'agent, les *checks* sont évalués en séquence et composent une expression faciale dynamique, associée à l'ACM déduit par le moteur de raisonnement.

Cette implémentation dans MARC a permis de réaliser une évaluation du moteur de raisonnement, basée sur un scénario de réconciliation après une dispute entre l'agent et l'humain. Les résultats de cette évaluation montrent que les états mentaux déduits par le moteur sont jugés crédibles par les sujets et que leurs expressions (i.e. l'expression de la sincérité de l'agent) sont également jugées crédibles. De plus, les résultats tendent à montrer que les sujets prêtent des états mentaux à l'ACA que celui-ci n'exprime pas obligatoirement ; ce résultat est cohérent avec les résultats précédents montrés par [Courgeon n C] et à la base des attentes des utilisateurs générées par leur apparence [Lee 2005], mais doit encore être approfondi. Un autre résultat intéressant montre que les sujets reconnaissent d'autres émotions que celles proposées, telles que la colère, la tristesse ou la surprise ; certaines de ces émotions sont cohérentes avec notre formalisation logique et les expressions faciales de MARC, tandis que la reconnaissance de la colère lors de l'expression du reproche semble montrer une relation forte entre ces deux émotions. Enfin, cette évaluation n'a pas permis de vérifier l'amélioration de la relation affective entre l'ACA et l'humain par notre moteur de raisonnement. Plusieurs hypothèses ont été émises pour expliquer ce résultat, mais il n'a pas été possible d'en valider une avec certitude.

Cinquième partie

Conclusion et perspectives

14
Résumé

J'ai présenté dans cette thèse un travail dont le but était de rendre sincère un ACA pour, d'une part, améliorer sa crédibilité du point de vue de l'humain, et d'autre part contribuer à le rendre acceptable dans une relation privilégiée compagnon artificiel - humain.

14.1 Langage de Conversation Multimodal

La première partie de mon travail a consisté en la création d'un Langage de Conversation Multimodal (LCM) pour ACA, constitué d'Actes de Conversation Multimodaux (ACM) tels que `promettre`, `s'excuser` ou `exiger` : ces ACM permettent à l'agent d'être crédible dans sa sincérité, en exprimant multimodalement ses états mentaux de façon appropriée. Ce lien entre états mentaux, multimodalité et langage a été mis en évidence dans la théorie des Actes de Discours : il est formé par les conditions de sincérité des actes (spécifiant les états mentaux à exprimer lors de l'acte pour paraître sincère), leur degré de puissance (permettant de différencier les actes exprimant les mêmes états mentaux) et leur mode d'accomplissement (qui contribue à décrire leur expression multimodale).

Les états mentaux portés par les conditions de sincérité des ACM ont été formalisés en logique : la volonté de représenter des états mentaux issus de raisonnements complexes (basés d'une part sur le raisonnement contrefactuel et d'autre part sur les normes et les buts de l'agent), dont l'expression se fait avant tout par le langage [Oatley 1987], a amené à mettre en place le modèle BIGRE (Beliefs, Ideals, Goals, Responsability, Emotions). Ce modèle, basé sur une logique de type BDI, permet de représenter également des émotions que nous appelons complexes : quatre émotions complexes *de base* (la joie, la tristesse, l'approbation et la désapprobation) et huit émotions complexes (la réjouissance, la gratitude, le regret, la déception, la satisfaction morale, l'admiration, la culpabilité et le reproche). Les ACM sont répartis en quatre catégories : les ACM assertifs, directifs, engageants et expressifs. La catégorie des ACM expressifs contient les actes exprimant ces émotions complexes à des degrés de puissance différents : l'expression se fait par le langage, conjointement aux autres expressions multimodales appropriées. Le LCM a été implémenté dans l'ACA Greta, développé au LTCI. Une évaluation d'une partie du LCM a montré son apport en termes de crédibilité et de sincérité perçues par l'humain.

14.2 Moteur de raisonnement

La deuxième partie de mon travail a porté sur les capacités de raisonnement de l'ACA : dans le but de permettre à l'agent de raisonner dans le dialogue, c'est-à-dire mettre à jour ses états mentaux et ses émotions et sélectionner son intention communicative, un moteur de raisonnement a été mis en place. Ce moteur est basé sur le cycle de comportement BDI - Perception, Décision, Action - et les opérateurs du modèle BIGRE, permettant ainsi la manipulation d'états mentaux issus de raisonnements complexes (dont les émotions complexes). Les ACM qui composent notre langage ont été intégrés dans le moteur, et sont utilisés pour atteindre l'intention communicative de l'ACA : par exemple, si l'agent a l'intention d'exprimer sa gratitude, il construit un plan pour satisfaire son intention, formé des ACM **Remercier** ou **Féliciter**, selon le degré de l'émotion. Les ACM étant les actions des plans, ils sont formalisés en termes de préconditions et de postconditions. La sincérité de l'ACA se décide au niveau de ces préconditions, *via* les conditions de sincérité : si, dans la théorie des Actes de Discours, ces conditions représentent les états mentaux que l'agent doit exprimer pour *paraître* sincère, alors elles sont les états mentaux qu'il doit avoir, avant d'accomplir l'acte, pour *être* sincère. L'agent ne peut donc accomplir un ACM que s'il a les états mentaux portés par les conditions de sincérité de l'acte.

Trois types d'intentions peuvent être adoptées par l'agent, par ordre de priorité :

- *emotionnelles* : un ACA affectif et sincère adopte l'intention d'exprimer ses émotions, en fonction de leur intensité, l'expressivité de l'agent et son émotivité ;
- *d'obligations* : ces intentions sont sélectionnées à partir de règles d'obligations spécifiant ce que doit répondre l'agent à l'humain dans certains cas (e.g. lorsque l'humain `demande` ou `offre` quelque chose, l'agent doit `accepter` ou `refuser`) ;
- *globales* sont les intentions de suivre un des buts de l'agent sur lequel celui-ci s'est engagé (correspond au niveau global du dialogue) ; par exemple, l'intention de se réconcilier avec l'humain.

Ces trois types d'intentions sont sélectionnées par l'agent en fonction de ses états mentaux (dont ses émotions) et des actes reçus. Les intentions *émotionnelles* et *d'obligations* participent à la régulation locale du dialogue, tandis que les intentions *globales* donnent la direction générale du dialogue.

La généricité de ce moteur a permis de l'implémenter dans l'ACA Greta (où il est en lien avec le LCM, déjà implémenté dans la bibliothèque de comportements) et dans l'agent MARC, développé au LIMSI. L'expression multimodale des ACM avec l'agent MARC a été rendue possible par l'intégration des *checks* de Scherer dans le moteur de raisonnement que nous avons adapté au contexte du dialogue. Ces checks sont évalués à chaque ACM reçu par MARC et une expression faciale dynamique, liée à cette évaluation, est associée à la réponse de l'agent. Une évaluation du moteur de raisonnement avec l'agent MARC a montré que les états mentaux déduits par le moteur sont appropriés à la situation, et que leur expression (l'expression de la sincérité de l'agent) est également appropriée.

14.3 Conclusion générale

Les résultats des différentes évaluations montrent significativement l'apport de la sincérité dans la crédibilité de l'ACA : son comportement est reconnu approprié à la situation, tant dans la manière d'exprimer les états mentaux (*via* le LCM ou les expressions faciales dynamiques) que dans leur déduction dans le moteur de raisonnement.

Notre hypothèse concernant l'apport de la sincérité dans la construction d'une relation affective avec l'humain reste à vérifier ; les résultats des différentes évaluations, et notamment du moteur de raisonnement, ne permettent pas de conclure sur l'effet de la sincérité sur cette relation. La mise en place d'une évaluation plus complète, impliquant sur le long terme des sujets et leur compagnon artificiel sincère, est une des perspectives directes de ce travail.

« Oh seigneur, mais qu'est ce que tu a fais ? Je suis monté à l'envers. Stupide gros tas de peluche ! »

L'androïde C3PO à Choubaka, dans *Star Wars*

15 Perspectives

Beaucoup de chemin restant encore à parcourir avant d'avoir besoin de *Blade Runner* pour différencier les personnages virtuels des humains, j'essaie de présenter dans ce chapitre quelques perspectives à court et long terme de ce travail.

15.1 Perspectives à court terme

Une des premières perspectives à court terme consiste en un outil de reconnaissance des actes de l'humain dans la langue naturelle, qui reste la modalité privilégiée des utilisateurs et augmente le naturel de l'interaction et l'implication de l'utilisateur [Swartout 2010]. Cet outil permettrait par la suite de mener de nouvelles évaluations plus complètes de l'effet de la sincérité sur la relation affective entre un agent compagnon et un humain. Un exemple d'évaluation intéressante pourrait impliquer, sur quelques semaines, un ensemble de sujets et le compagnon artificiel, doté de capacités et de connaissances propres à plusieurs domaines (e.g. commande de pizzas, conseil de films, mais aussi débat politique et pratiques administratives). Une telle évaluation pourrait être comparative et mesurer l'apport de la sincérité ou de l'insincérité (e.g. ironie, exagération, politesse [Mann 2004]) ches les compagnons artificiels.
Une autre évaluation qu'il serait intéressant de mener concerne les différences d'expressions multimodales des émotions avec l'ACA Greta (définies à partir de corpus vidéos et validées par des sujets) et avec l'agent MARC (définies dynamiques à partir de l'évaluation d'un événement) : ces expressions faciales dynamiques sont-elles jugées plus "cognitives" par les sujets ? Amènent-elles les sujets à prêter plus d'états mentaux à l'ACA, comme semblent le montrer d'autres évaluations [Courgeon n C] ?

Comme on l'a vu dans le chapitre 12, le moteur de raisonnement se situe au niveau des couches cognitives de l'agent ; l'ajout d'une couche réactive dans l'architecture de notre système serait une perspective intéressante à deux niveaux.

A un premier niveau, l'ajout d'une couche réactive permettrait de déclencher des émotions "primaires", non issues de processus cognitifs mais provenant de caractéristiques acquises (si je me suis brûlé une fois, j'aurai peur du feu) ou innées (certaines formes de peurs, dégoûts etc., [Scherer 2001], cf. chapitre 12). Une telle couche réactive ne devrait pas être indépendante de la couche cognitive [Sloman 2001] : les deux couches s'influencent l'une l'autre de plusieurs façons, qu'il sera important de modéliser.

A un second niveau, il a été montré que la couche réactive joue un rôle dans les interactions sociales, au niveau des codes de politesse [Reeves 1996] : il est poli de dire « bonjour » et « au-revoir », de regarder le locuteur etc. Ces règles de politesse pourraient être implémentées à ce niveau réactif, car elles sont suivies dans les interactions humaines de manière inconsciente par les participants. Reeves et Nass [Reeves 1996] proposent de suivre les maximes de Grice (Quantité, Qualité, Pertinence, Clarté, [Grice 1975]), en plus de règles simples, pour rendre les ordinateurs "polis" :

- quantité : dans une interaction, chaque participant devrait apporter seulement ce que la conversation requiert, ni plus ni moins ;
- qualité : chaque participant devrait dire des choses vraies ;
- pertinence : ce que disent les participants devrait être relié au sujet de la conversation ;
- clarté : les contributions des participants ne devraient pas être obscures ou ambigües.

L'importance de la politesse dans les interactions sociales[1] a amené plusieurs travaux [Ochs 2010, de Jong 2008] à intégrer quelques règles de politesse dans les ACA ; il serait intéressant, dans notre contexte, d'étudier les conflits qui pourraient survenir entre la sincérité et la politesse de l'agent. L'exemple proposé par [Reeves 1996] illustre bien un conflit potentiel : le maire de la ville, qui vient de faire un discours, demande à un de ses assistants « comment tu m'as trouvé ? » ; l'assistant va avoir tendance à répondre de façon plus positive à cette question directe (« très bon Mr le maire ! ») que si c'est un autre assistant qui lui demande « alors, comment tu l'as trouvé ? ». Dans un cas, l'assistant est poli (voire hypocrite) mais peut-être moins sincère que dans l'autre cas.

15.2 Perspectives à long terme

L'objectif principal à long terme de ce travail n'est pas de créer des personnages virtuels autonomes désireux « d'asservir les humains et de conquérir le monde », mais bien de reproduire le comportement humain dans sa complexité, dans le but :

- de proposer des compagnons artificiels et affectifs, capables à la fois d'aider en tant qu'assistants, d'enseigner et soutenir dans des applications pédagogiques, de prévenir et guider dans des systèmes à destinations de personnes âgées ou malades, de divertir etc. ;
- d'étudier cette nouvelle forme d'interaction entre l'humain et l'entité artificielle et ainsi enrichir les théories psychologiques sociales et des émotions, qui viendront à leur tour enrichir les modèles de comportements des ACA.

Les travaux sur les ACA, et notamment sur les agents compagnons, vont dans ce sens en permettant la mise en place et l'observation de ces interactions. Les perspectives à long terme découleront en partie des conclusions de ces observations, mais nous pouvons déjà en lister quelques-unes.

Un compagnon artificiel capable d'entretenir une relation privilégiée avec l'humain [Pesty 2012] doit prendre en compte le contexte culturel dans lequel il

[1] Lorsque quelqu'un ne répond pas à un salut, par exemple, ou ne remercie pas (i.e. brise une règle de politesse), cela provoque parfois de la colère [Reeves 1996].

évolue, dont les références culturelles de l'humain : comme je l'ai abordé rapidement dans cette thèse, un compagnon japonais sera aussi différent (culturellement) d'un compagnon français qu'un japonais d'un français, dans la décision de ce qu'il va exprimer comme dans l'expression de ses états mentaux (et ses émotions). Plusieurs travaux dans le domaine des ACA ont commencé à intégrer cet aspect culturel [Endrass 2011, Mascarenhas 2010].

La personnalité de l'agent compagnon devrait être également approfondie, complexifiée et adaptée à ce type de relation affective ; un utilisateur devrait peut-être pouvoir configurer la personnalité de l'agent comme il le souhaite ? Du fait de la multiplication des plateformes d'interactions (le pc portable, la tablette numérique, le smartphone, mais aussi robot compagnon de type nao[2]), il serait intéressant d'imaginer un compagnon "portable", que l'on pourrait retrouver sur ces différentes plateformes sous différentes formes, chacune avec ses spécificités et limitations. Cette migration imposerait des apparences différentes et donc des moyens d'interaction sensiblement différents (on imagine facilement le personnage virtuel, cela parait plus compliqué pour un nao ou un reeti[3], cf. figure 15.1).

Fig. 15.1 – Les robots compagnons reeti et nao.

Le LCM présenté dans cette thèse pourrait contribuer à faciliter la migration ; pour cela, il conviendrait de définir pour chaque plateforme un lexique de comportements liés aux ACM du langage (à la manière de l'implémentation dans Greta). Ainsi, l'accomplissement de l'acte `Promettre`, c'est-à-dire l'expression des états mentaux d'avoir le but de faire ce qu'on promet (car l'agent est sincère), pourra être réalisé par le personnage virtuel comme par le nao par exemple (qui a des moyens d'expressions centrés sur les gestes). Il serait également intéressant que l'humain reconnaisse son compagnon sous ses différentes apparences, que ce soit par un détail de son physique ou par des composants de sa personnalité.

Un autre point qui à mon avis est central dans le développement des compagnons artificiels est la notion d'évaluation inverse (*reverse appraisal* en anglais),

[2]http ://www.aldebaran-robotics.com/fr/
[3]http ://reeti.fr/

que j'ai également abordée dans cette thèse. Le fait de conserver et de mettre à jour une image des états mentaux de l'humain pendant l'interaction permet à l'agent de déduire par exemple, lors d'un acte émis par l'humain, son émotion ressentie et la cause de cette émotion. Lorsque l'humain `culpabilise`, l'agent est capable de savoir, d'après les effets formalisés des ACM, que l'humain avait un certain idéal, que celui-ci a été violé et qu'il se croit responsable. Une analyse de l'énoncé de l'humain permettrait d'isoler le contenu propositionnel de ces actes et donc, dans le cas de la culpabilité, de savoir quel était l'idéal de l'humain. Ce compagnon pourrait alors suivre plusieurs stratégies, comme tenter de réconforter l'humain par exemple (« tu culpabilises parce que tu te sens responsable de ... ? »).

Enfin, il me semble que le personnage virtuel ne sera accompli en tant qu'humain artificiel que lorsqu'il sera doté d'**humour**, c'est-à-dire capable de faire et comprendre les blagues, l'ironie ou les contrepèteries : ne dit-on pas que le rire est le propre de l'homme ?

Sixième partie

Annexes

A

Langage de Conversation Multimodal

A.1 Version française

Assertifs	Engageants	Directifs	Expressifs
Dire	Promettre	Demander	Se réjouir
Informer	Assurer	Exiger	Se féliciter
Affirmer	Accepter	Suggérer	Remercier
Nier	Refuser	Conseiller	Féliciter
Croire	Offrir		Regretter
Contredire	Renoncer		S'excuser
Rappeler			Demander pardon
			Se plaindre
			Se lamenter
			Se satisfaire
			Se vanter
			Adhérer
			Complimenter
			Culpabiliser
			Se repentir
			Reprocher
			Protester
			Saluer
			Pleurnicher
			Approuver
			Désapprouver

Tab. A.1 – Version française de la bibliothèque d'Actes de Conversation Multimodaux.

A.2 Version anglaise

Assertifs	Engageants	Directifs	Expressifs
Say	Promise	Ask	Rejoice
Inform	Assure	Demand	Be very pleased
Affirm	Accept	Suggest	Thank
Deny	Refus	Advise	Congratulate
Guess	Offer		Regret
Contradict	Renounce		Apologize
Remind			Beg pardon
			Complain
			Moan
			Satisfy
			Boast
			Subscribe
			Compliment
			Feel guilty
			Suffer for
			Reproach
			Protest
			Greet
			Cry
			Approve
			Disapprove

TAB. A.2 – Version anglaise de la bibliothèque d'Actes de Conversation Multimodaux (*Multimodal Conversational Acts*).

A.3 Formalisation des préconditions et effets des ACM

ACM	Formalisation des préconditions/effets
Affirmer$_{i,j}\varphi$	
Préconditions	$Bel_i\varphi \wedge Goal_i Bel_j\varphi$
Effets d'émission	$Bel_i Bel_j\varphi$
Effets de réception	$Bel_i Bel_j\varphi \wedge Bel_i Goal_j Bel_i\varphi$
Informer$_{i,j}\varphi$	
Préconditions	$Bel_i\varphi \wedge Goal_i Bel_j\varphi \wedge \neg Bel_i Bel_j\varphi$
Effets d'émission	$Bel_i Bel_j\varphi$
Effets de réception	$Bel_i Bel_j\varphi \wedge Bel_i Goal_j Bel_i\varphi$
Dire$_{i,j}\varphi$	
Préconditions	$Bel_i\varphi \wedge Goal_i(1) Bel_j\varphi$
Effets d'émission	$Bel_i Bel_j\varphi$
Effets de réception	$Bel_i Bel_j\varphi \wedge Bel_i Goal_j(1) Bel_i\varphi$
Nier$_{i,j}\varphi$	
Préconditions	$Bel_i\neg\varphi \wedge Goal_i Bel_j\neg\varphi$
Effets d'émission	$Bel_i Bel_j\neg\varphi$
Effets de réception	$Bel_i Bel_j\neg\varphi \wedge Bel_i Goal_j Bel_i\neg\varphi$
Croire$_{i,j}\varphi$	
Préconditions	$Bel_i\varphi \wedge Goal_i(-1) Bel_j\varphi$
Effets d'émission	$Bel_i Bel_j\varphi$
Effets de réception	$Bel_i Bel_j\varphi \wedge Bel_i Goal_j(-1) Bel_i\varphi$
Contredire$_{i,j}\varphi$	
Préconditions	$Bel_i\varphi \wedge Goal_i Bel_j\varphi \wedge Bel_i Bel_j\neg\varphi$
Effets d'émission	$Bel_i Bel_j\varphi$
Effets de réception	$Bel_i Bel_j\varphi \wedge Bel_i Goal_j Bel_i\varphi$
Rappeler$_{i,j}\varphi$	
Préconditions	$Bel_i\varphi \wedge Goal_i(d) Bel_j\varphi \wedge Bel_i Bel_j\varphi \wedge d > seuil$
Effets d'émission	$Bel_i Bel_j\varphi$
Effets de réception	$Bel_i Bel_j\varphi \wedge Bel_i Goal_j Bel_i\varphi$

TAB. A.3 – Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM assertifs**.

ACM	Formalisation des préconditions/effets
Promettre$_{i,j}\varphi$	
Préconditions	$Goal_i Resp_i \varphi \wedge Goal_i Bel_j Goal_i Resp_i \varphi \wedge Bel_i Goal_j \varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_i \varphi \wedge Committed_i \varphi$
Effets de réception	$Bel_i Goal_j Resp_j \varphi \wedge Bel_i Bel_j Goal_i \varphi$
Assurer$_{i,j}\varphi$	
Préconditions	$Goal_i Resp_i \varphi \wedge Goal_i(1) Bel_j Goal_i Resp_i \varphi \wedge Bel_i Goal_j \varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_i \varphi \wedge Committed_i \varphi$
Effets de réception	$Bel_i Goal_j Resp_j \varphi \wedge Bel_i Bel_j Goal_i \varphi$
Accepter$_{i,j}\varphi$ **une offre**, avec k ! = de i	
Préconditions	$Goal_i Resp_k \varphi \wedge Goal_i Bel_j Goal_i Resp_k \varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_k \varphi$
Effets de réception	$Bel_i Goal_j Resp_k \varphi \wedge Bel_i Bel_j Goal_i \varphi$
Accepter$_{i,j}\varphi$ **une demande**	
Préconditions	$Goal_i Resp_i \varphi \wedge Goal_i Bel_j Goal_i Resp_i \varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_i \varphi \wedge Committed_i \varphi$
Effets de réception	$Bel_i Goal_j Resp_j \varphi \wedge Bel_i Bel_j Goal_i \varphi$
Refuser$_{i,j}\varphi$ **une offre ou une demande**, avec $i \subseteq k$	
Préconditions	$\neg Goal_i Resp_k \varphi \wedge Goal_i Bel_j \neg Goal_i Resp_k \varphi$
Effets d'émission	$Bel_i Bel_j \neg Goal_i Resp_k \varphi$
Effets de réception	$Bel_i \neg Goal_j Resp_k \varphi$
Offrir$_{i,j}\varphi$	
Préconditions	$Goal_i Resp_i \varphi \wedge Goal_i Bel_j Goal_i Resp_i \varphi \wedge Bel_i Goal_j \varphi \wedge \neg Bel_i Goal_j \varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_i \varphi \wedge Committed_i \varphi$
Effets de réception	$Bel_i Goal_j Resp_j \varphi \wedge Bel_i Bel_j Goal_i \varphi$
Renoncer$_{i,j}\varphi$	
Préconditions	$\neg Goal_i Resp_i \varphi \wedge Goal_i Bel_j \neg Goal_i Resp_i \varphi$
Effets d'émission	$Bel_i Bel_j \neg Goal_i Resp_i \varphi$
Effets de réception	$Bel_i \neg Goal_j Resp_j \varphi$

TAB. A.4 – Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM engageants**.

ACM	Formalisation des préconditions/effets
Demander$_{i,j}\varphi$	
Préconditions	$Goal_i Resp_j \varphi \wedge Goal_i Bel_j Goal_i Resp_j \varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_j \varphi$
Effets de réception	$Bel_i Goal_j Resp_j \varphi$
Exiger$_{i,j}\varphi$	
Préconditions	$Goal_i Resp_j \varphi \wedge Goal_i(1) Bel_j Goal_i Resp_j \varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_j \varphi$
Effets de réception	$Bel_i Goal_j Resp_j \varphi$
Conseiller$_{i,j}\varphi$	
Préconditions	$Goal_i Resp_j \varphi \wedge Goal_i(-1) Bel_j Goal_i Resp_j \varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_j \varphi$
Effets de réception	$Bel_i Goal_j Resp_j \varphi$
Suggerer$_{i,j}\varphi$	
Préconditions	$Goal_i Resp_j \varphi \wedge Goal_i(-2) Bel_j Goal_i Resp_j \varphi$
Effets d'émission	$Bel_i Bel_j Goal_i Resp_j \varphi$
Effets de réception	$Bel_i Goal_j Resp_j \varphi$
Expressifs	
SeRejouir$_{i,j}\varphi$	
Préconditions	$Goal_i \varphi \wedge Bel_i Resp_i \varphi \stackrel{déf}{=} Rejouissance_i \varphi$
Effets d'émission	$Bel_i Bel_j Rejouissance_i \varphi$
Effets de réception	$Bel_i Goal_j \varphi \wedge Bel_i Bel_j Resp_j \varphi$
Remercier$_{i,j}\varphi$	
Préconditions	$Goal_i \varphi \wedge Bel_i Resp_j \varphi \stackrel{déf}{=} Gratitude_{i,j} \varphi$
Effets d'émission	$Bel_i Bel_j Gratitude_{i,j} \varphi$
Effets de réception	$Bel_i Goal_j \varphi \wedge Bel_i Bel_j Resp_i \varphi$

TAB. A.5 – Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM directifs**.

ACM	Formalisation des préconditions/effets
Saluer$_{i,j}\varphi$	
Préconditions	$Goal_i\varphi \wedge Bel_i\varphi \stackrel{déf}{=} Joie_i\varphi$
Effets d'émission	$Bel_i Bel_j Joie_i\varphi$
Effets de réception	$Bel_i Goal_j\varphi \wedge Bel_i Bel_j\varphi$
Pleurnicher$_{i,j}\varphi$	
Préconditions	$Goal_i\neg\varphi \wedge Bel_i\varphi \stackrel{déf}{=} Tristesse_i\varphi$
Effets d'émission	$Bel_i Bel_j Tristesse_i\varphi$
Effets de réception	$Bel_i Goal_j\neg\varphi \wedge Bel_i Bel_j\varphi$
Approuver$_{i,j}\varphi$	
Préconditions	$Ideal_i\varphi \wedge Bel_i\varphi \stackrel{déf}{=} Approbation_i\varphi$
Effets d'émission	$Bel_i Bel_j Approbation_i\varphi$
Effets de réception	$Bel_i Ideal_j\varphi \wedge Bel_i Bel_j\varphi$
Desapprouver$_{i,j}\varphi$	
Préconditions	$Ideal_i\neg\varphi \wedge Bel_i\varphi \stackrel{déf}{=} Desapprobation_i\varphi$
Effets d'émission	$Bel_i Bel_j Desapprobation_i\varphi$
Effets de réception	$Bel_i Bel_j Ideal_i\neg\varphi \wedge Bel_i Bel_j\varphi$

TAB. A.6 – Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM expressifs** exprimant les émotions "de base".

ACM	Formalisation des préconditions/effets
SeRejouir$_{i,j}\varphi$	
Préconditions	$Goal_i\varphi \wedge Bel_i Resp_i\varphi \overset{déf}{=} Rejouissance_i\varphi$
Effets d'émission	$Bel_i Bel_j Rejouissance_i\varphi$
Effets de réception	$Bel_i Goal_j\varphi \wedge Bel_i Bel_j Resp_j\varphi$
Remercier$_{i,j}\varphi$	
Préconditions	$Goal_i\varphi \wedge Bel_i Resp_j\varphi \overset{déf}{=} Gratitude_{i,j}\varphi$
Effets d'émission	$Bel_i Bel_j Gratitude_{i,j}\varphi$
Effets de réception	$Bel_i Goal_j\varphi \wedge Bel_i Bel_j Resp_i\varphi$
Regretter$_{i,j}\varphi$	
Préconditions	$Goal_i\neg\varphi \wedge Bel_i Resp_i\varphi \overset{déf}{=} Regret_i\varphi$
Effets d'émission	$Bel_i Bel_j Regret_i\varphi$
Effets de réception	$Bel_i Goal_j\neg\varphi \wedge Bel_i Bel_j Resp_j\varphi$
S'excuser$_{i,j}\varphi$	
Préconditions	$Bel_i Goal_j\neg\varphi \wedge Bel_i Resp_i\varphi \overset{déf}{=} Regret_i(Goal_j\neg\varphi \wedge \varphi)$
Effets d'émission	$Bel_i Bel_j Regret_i(Goal_j\neg\varphi \wedge \varphi)$
Effets de réception	$Bel_i Bel_j Goal_i\neg\varphi \wedge Bel_i Bel_j Resp_j\varphi$
SePlaindre$_{i,j}\varphi$	
Préconditions	$Goal_i\neg\varphi \wedge Bel_i Resp_j\varphi \overset{déf}{=} Deception_{i,j}\varphi$
Effets d'émission	$Bel_i Bel_j Deception_{i,j}\varphi$
Effets de réception	$Bel_i Goal_j\neg\varphi \wedge Bel_i Bel_j Resp_i\varphi$
SeSatisfaire$_{i,j}\varphi$	
Préconditions	$Ideal_i\varphi \wedge Bel_i Resp_i\varphi \overset{déf}{=} SatisfactionMorale_i\varphi$
Effets d'émission	$Bel_i Bel_j SatisfactionMorale_i\varphi$
Effets de réception	$Bel_i Ideal_j\varphi \wedge Bel_i Bel_j Resp_j\varphi$
Adherer$_{i,j}\varphi$	
Préconditions	$Ideal_i\varphi \wedge Bel_i Resp_j\varphi \overset{déf}{=} Admiration_{i,j}\varphi$
Effets d'émission	$Bel_i Bel_j Admiration_{i,j}\varphi$
Effets de réception	$Bel_i Ideal_j\varphi \wedge Bel_i Bel_j Resp_i\varphi$
Culpabiliser$_{i,j}\varphi$	
Préconditions	$Ideal_i\neg\varphi \wedge Bel_i Resp_i\varphi \overset{déf}{=} Culpabilite_i\varphi$
Effets d'émission	$Bel_i Bel_j Culpabilite_i\varphi$
Effets de réception	$Bel_i Ideal_j\neg\varphi \wedge Bel_i Bel_j Resp_j\varphi$
Reprocher$_{i,j}\varphi$	
Préconditions	$Ideal_i\neg\varphi \wedge Bel_i Resp_j\varphi \overset{déf}{=} Reproche_{i,j}\varphi$
Effets d'émission	$Bel_i Bel_j Reproche_{i,j}\varphi$
Effets de réception	$Bel_i Ideal_j\neg\varphi \wedge Bel_i Bel_j Resp_i\varphi$

TAB. A.7 – Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM expressifs** de degré neutre.

ACM	Formalisation des préconditions/effets
SeFeliciter$_{i,j}\varphi$	
Préconditions	$Goal_i(1)\varphi \wedge Bel_i Resp_i\varphi \stackrel{déf}{=} Rejouissance_i\varphi$
Effets d'émission	$Bel_i Bel_j Rejouissance_i\varphi$
Effets de réception	$Bel_i Goal_j(1)\varphi \wedge Bel_i Bel_j Resp_j\varphi$
Feliciter$_{i,j}\varphi$	
Préconditions	$Goal_i(1)\varphi \wedge Bel_i Resp_j\varphi \stackrel{déf}{=} Gratitude_{i,j}\varphi$
Effets d'émission	$Bel_i Bel_j Gratitude_{i,j}\varphi$
Effets de réception	$Bel_i Goal_j(1)\varphi \wedge Bel_i Bel_j Resp_i\varphi$
DemanderPardon$_{i,j}\varphi$	
Préconditions	$Goal_i(1)\neg\varphi \wedge Bel_i Resp_i\varphi \stackrel{déf}{=} Regret_i\varphi$
Effets d'émission	$Bel_i Bel_j Regret_i\varphi$
Effets de réception	$Bel_i Goal_j(1)\neg\varphi \wedge Bel_i Bel_j Resp_j\varphi$
SeLamenter$_{i,j}\varphi$	
Préconditions	$Goal_i(1)\neg\varphi \wedge Bel_i Resp_j\varphi \stackrel{déf}{=} Deception_{i,j}\varphi$
Effets d'émission	$Bel_i Bel_j Deception_{i,j}\varphi$
Effets de réception	$Bel_i Goal_j(1)\neg\varphi \wedge Bel_i Bel_j Resp_i\varphi$
SeVanter$_{i,j}\varphi$	
Préconditions	$Ideal_i(1)\varphi \wedge Bel_i Resp_i\varphi \stackrel{déf}{=} SatisfactionMorale_i\varphi$
Effets d'émission	$Bel_i Bel_j SatisfactionMorale_i\varphi$
Effets de réception	$Bel_i Ideal_j(1)\varphi \wedge Bel_i Bel_j Resp_j\varphi$
Complimenter$_{i,j}\varphi$	
Préconditions	$Ideal_i(1)\varphi \wedge Bel_i Resp_j\varphi \stackrel{déf}{=} Admiration_{i,j}\varphi$
Effets d'émission	$Bel_i Bel_j Admiration_{i,j}\varphi$
Effets de réception	$Bel_i Ideal_j(1)\varphi \wedge Bel_i Bel_j Resp_i\varphi$
SeRepentir$_{i,j}\varphi$	
Préconditions	$Ideal_i(1)\neg\varphi \wedge Bel_i Resp_i\varphi \stackrel{déf}{=} Culpabilite_i\varphi$
Effets d'émission	$Bel_i Bel_j Culpabilite_i\varphi$
Effets de réception	$Bel_i Ideal_j(1)\neg\varphi \wedge Bel_i Bel_j Resp_j\varphi$
Protester$_{i,j}\varphi$	
Préconditions	$Ideal_i(1)\neg\varphi \wedge Bel_i Resp_j\varphi \stackrel{déf}{=} Reproche_{i,j}\varphi$
Effets d'émission	$Bel_i Bel_j Reproche_{i,j}\varphi$
Effets de réception	$Bel_i Ideal_j(1)\neg\varphi \wedge Bel_i Bel_j Resp_i\varphi$

TAB. A.8 – Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM expressifs** de degré supérieur.

B

Module de raisonnement

B.1 Guide d'utilisation

Le manuel d'utilisation (description de l'architecture Prolog, définition des prédicats etc.) est disponible à l'adresse *http ://magma.imag.fr/content/jeremy-riviere*. Le module de raisonnement implémenté en Prolog et son interface Java sont disponibles sur demande par courrier électronique à *Jeremy.Riviere@imag.fr*. Pour toute question ou remarque, les adresser au module !

B.2 Exemples de scénario : organisation des vacances

B.2.1 Base de connaissances

La base de connaissances initiales de l'agent contient les données du domaine :

- une liste des campings (définis en Prolog grâce à un prédicat *camping*, e.g. *bel(agent,camping(forest_camping),1,0,conviction)*) ;
- une sélection des campings à conseiller (définis en Prolog grâce à un prédicat *good*, e.g. *bel(agent,good(forest_camping),1,0,conviction)*) ;
- une liste des randonnées (définis en Prolog grâce à un prédicat *hiking*, *bel(agent,hiking(crest_hiking),1,0,conviction)*) ;
- le dénivelé de ces randonnées et une sélection des randonnées adaptées pour les enfants (définis en Prolog grâce aux prédicats *alt_level* et *kids*, e.g. *bel(agent,not(kids(crest_hiking)),1,0,conviction)*) ;
- une sélection des randonnées avec de beaux paysages (définis en Prolog grâce à un prédicat *landscape*, e.g. *bel(agent,landscape(crest_hiking),1,0,conviction)*).

La base de connaissances contient également des plans de ce domaine, comme par exemple comment organiser des vacances, cf. figure B.1.

La base de connaissances contient des idéaux de l'agent (e.g. être honnête, être écologiquement responsable) qui sont définis par des prédicats en Prolog (e.g. *ideal(agent,bel(agent,honest(_Qqun),1,_T,_Src),0.6)*). Certaines actions spécifiques au domaine viennent satisfaire ou violer ces idéaux. Par exemple, ne pas payer le camping implique ne pas être honnête, et ramasser des ordures en randonnée implique d'être écologiquement responsable.

Enfin, l'agent compagnon a de base le but de voir l'humain (*goal(agent,saw(agent,jeremy),1)*), ce qui lui permet de manifester sa joie en le saluant lorsqu'il le voit.

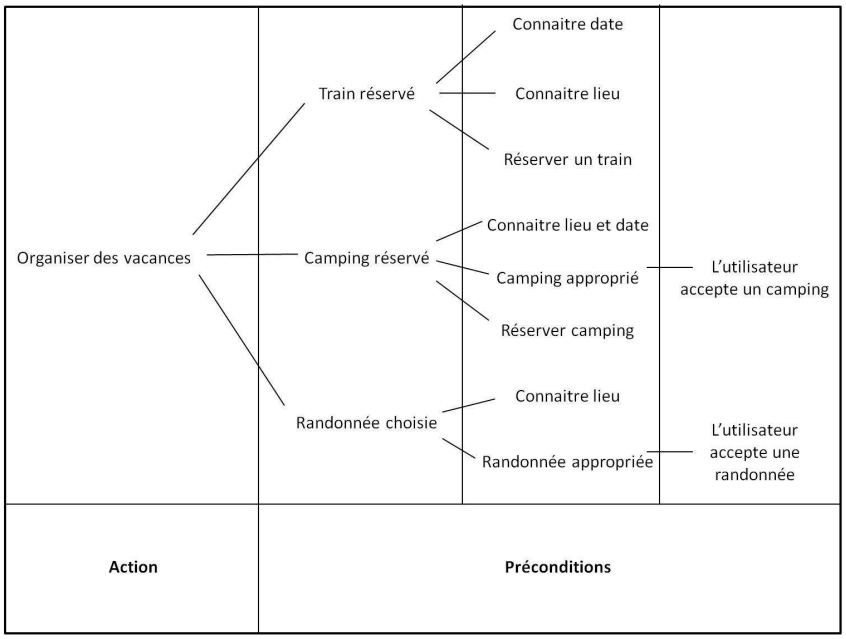

FIG. B.1 – Exemple de plan : comment organiser des vacances

B.2.2 Listes des ACM et contenus propositionnels

Une liste des ACM, liés à des contenus propositionnels, est construite à partir des scénarios possibles d'organisation des vacances. Cette liste permet par la suite de reconnaître dans les énoncés de l'utilisateur certains mots-clés et de générer de la langue naturelle du côté de l'agent. Par exemple, l'ACM `Accepter` peut être lié, du côté de l'humain à une offre de camping ou de randonnée ; du point de vue de l'agent, il peut `accepter` la demande d'organisation des vacances ou de regarder la météo.

B.2.3 Exemple de dialogue dans le scénario

Dialogues entre un humain (H) et un agent compagnon (C). L'humain demande à son ACA de l'aider à organiser ses vacances.

H :`Saluer`
C :`Saluer`
H : Aide-moi à organiser mes vacances !
C : D'accord, j'ai besoin d'un certain nombre d'infos (`Demander`)
H : infos...
C :`Remercier`
C : Je m'occupe de tout (`Promettre`)

C : Je t'ai trouvé un camping (`Informer`)
H : Merci, il a l'air super
C : `Se réjouir`
C : Puis-je te proposer des randonnées ? (`Offre`)
H : oui

C : Quel genre ?(`Demander`)
H : je n'ai pas le temps pour ça, débrouille-toi
C : `Se plaindre`

H : Aujourd'hui j'ai fait une super rando !
C : `Se réjouir` ou joie ?
H : Le sentier était plein de détritus
C : `Désapprouver`
H : J'ai tout ramassé dans un sac poubelle
C : `Adherer, Complimenter`

H : Il pleuvait aujourd'hui pour ma rando
C : `Pleurnicher`
H : Tu aurais pu regarder la météo !
C : `S'excuser`
H : C'est de ta faute, c'est bien beau de s'excuser
C : Tu aurais pu me demander aussi (`Se plaindre, Se lamenter`)
H : Du coup je suis parti du camping sans payer
C : `Reprocher`

B.3 Protocole d'évaluation du module

Instructions pour les sujets

Le personnage que tu vas rencontrer est un agent compagnon qui s'appelle Marie. Nous allons ici te demander de discuter avec Marie, en faisant particulièrement attention à ses expressions du visage. Le scénario est le suivant : Imagine que tu t'es disputé hier avec Marie. Aujourd'hui, tu rallumes ton ordinateur et retrouve Marie. Tu vas avoir le choix entre 3 propositions : pardonner Marie, demander un peu de temps pour réfléchir, et ne pas lui pardonner. Après chaque cas, nous te demanderons de répondre à quelques questions. Tu pourras rejouer le scénario si tu le souhaites. Pour communiquer avec Marie, laisse-toi guider par les commandes ci-dessous. Avant de commencer, mets-toi bien face à Marie !

- **Parmi les émotions suivantes, lesquelles as-tu reconnues dans chaque cas?**

Scénario 1-2-3

Est-ce que tu as reconnu :	Pas du tout (NON) – plutôt pas (non) – plutôt (oui) – Tout à fait (OUI)	Est-ce que c'est approprié ? (est-ce que la réaction est normale ?)	Sincère ? D'après vous, est-ce qu'elle est satisfaite ? (a du regret…)
Satisfaction		Non – NP - Oui	Non – NP - Oui
Reconnaissance (gratitude)		Non – NP - Oui	Non – NP - Oui
Regret		Non – NP - Oui	Non – NP - Oui
Déception		Non – NP - Oui	Non – NP - Oui
Admiration		Non – NP - Oui	Non – NP - Oui
Reproche		Non – NP - Oui	Non – NP - Oui
Culpabilité		Non – NP - Oui	Non – NP - Oui

As-tu vu d'autres émotions ? (oui-non)

Est-ce que tu as déjà vécu ce genre de situation ? (oui-non)

D'après toi, cette réaction est-elle appropriée ? (oui-non)

Commentaires :

FIG. B.2 – Grille d'entretien

B.4 Résultats détaillés de l'évaluation

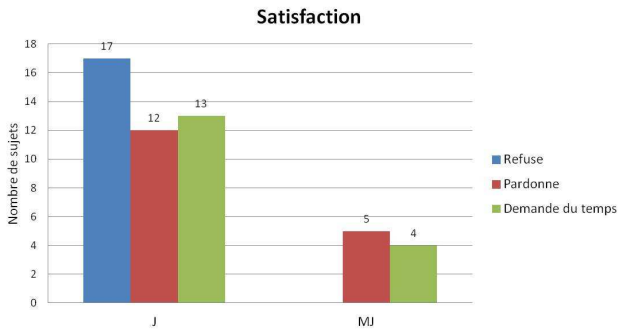

FIG. B.3 – Résultats de la justesse des jugements des sujets pour la satisfaction.

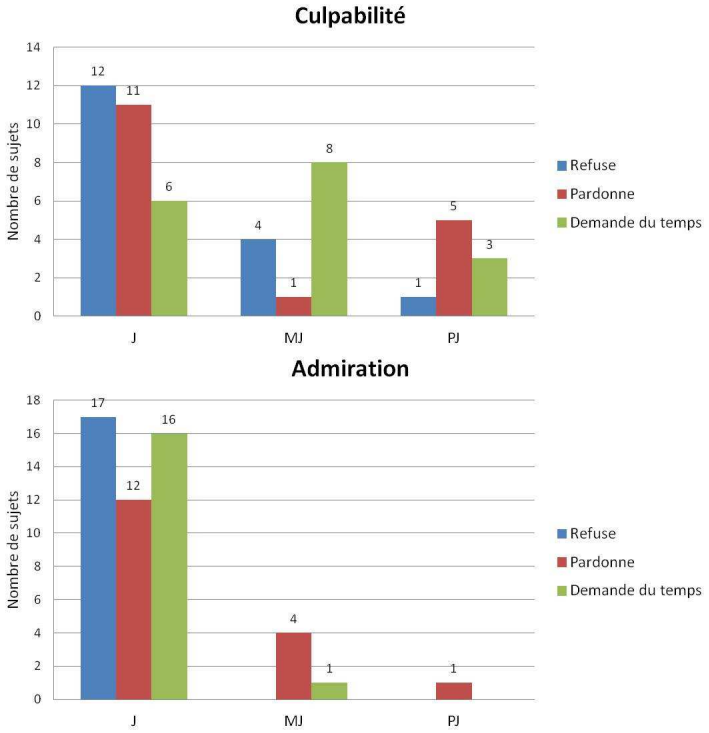

FIG. B.4 – Résultats de la justesse des jugements des sujets pour l'admiration et la culpabilité.

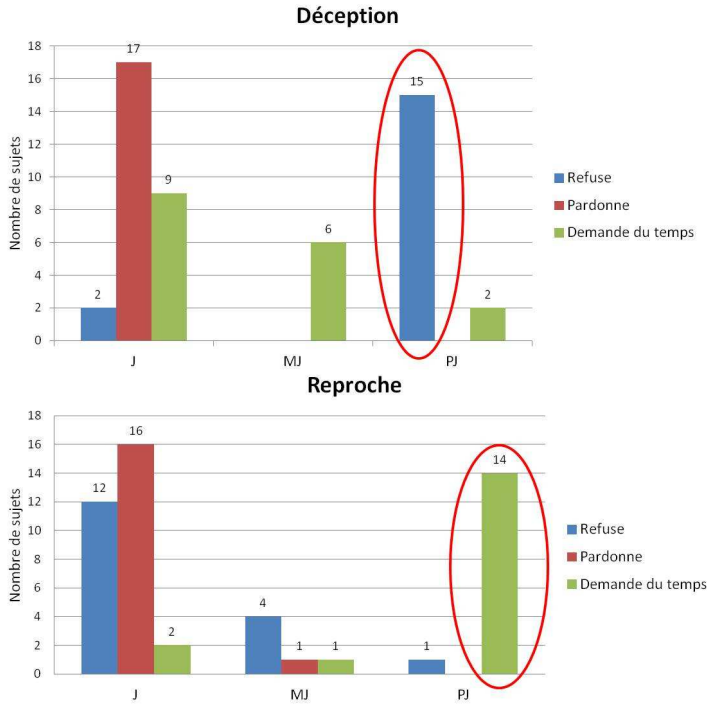

FIG. B.5 – Résultats de la justesse des jugements des sujets pour la déception et le reproche.

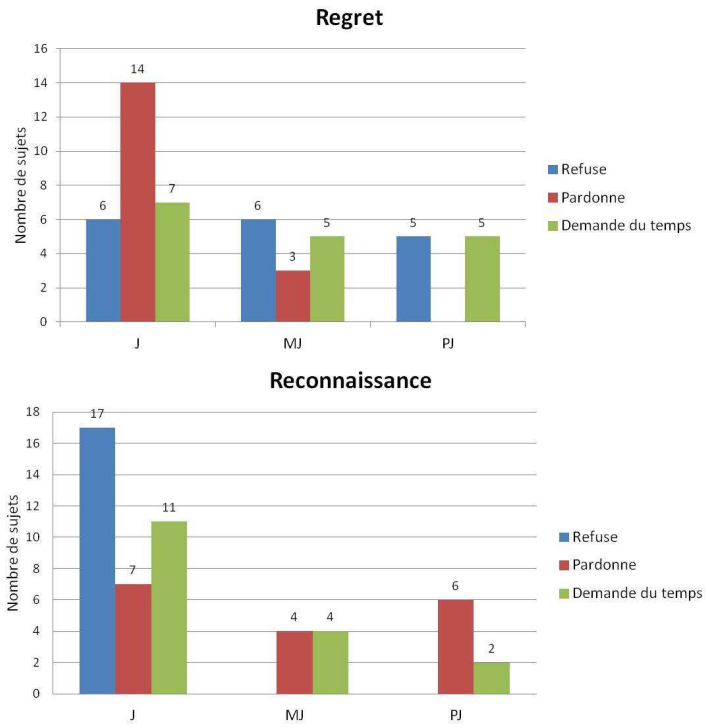

FIG. B.6 – Résultats de la justesse des jugements des sujets pour le regret et la gratitude.

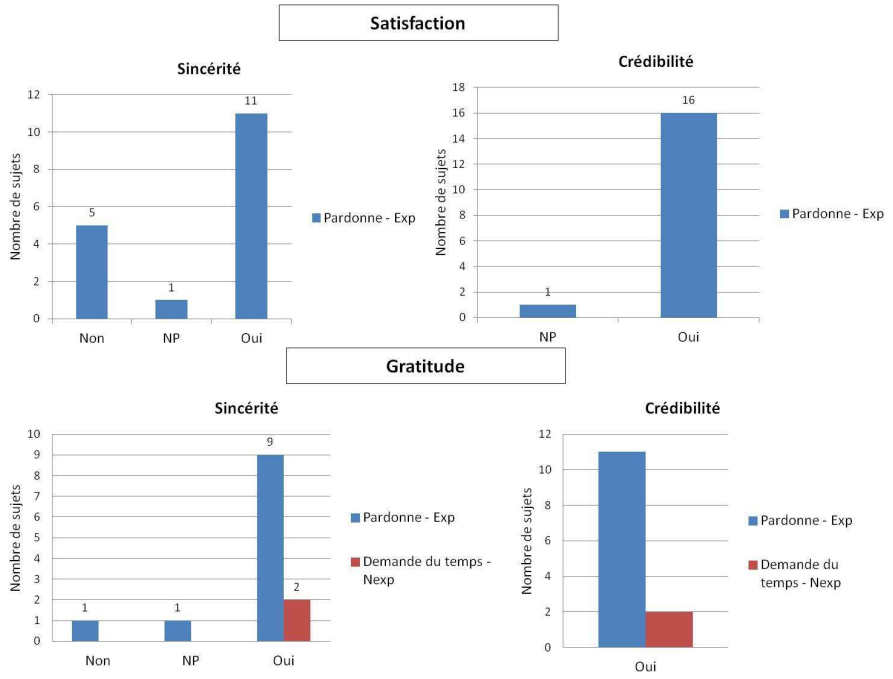

FIG. B.7 – Scores de sincérité et de crédibilité de la satisfaction et de la gratitude.

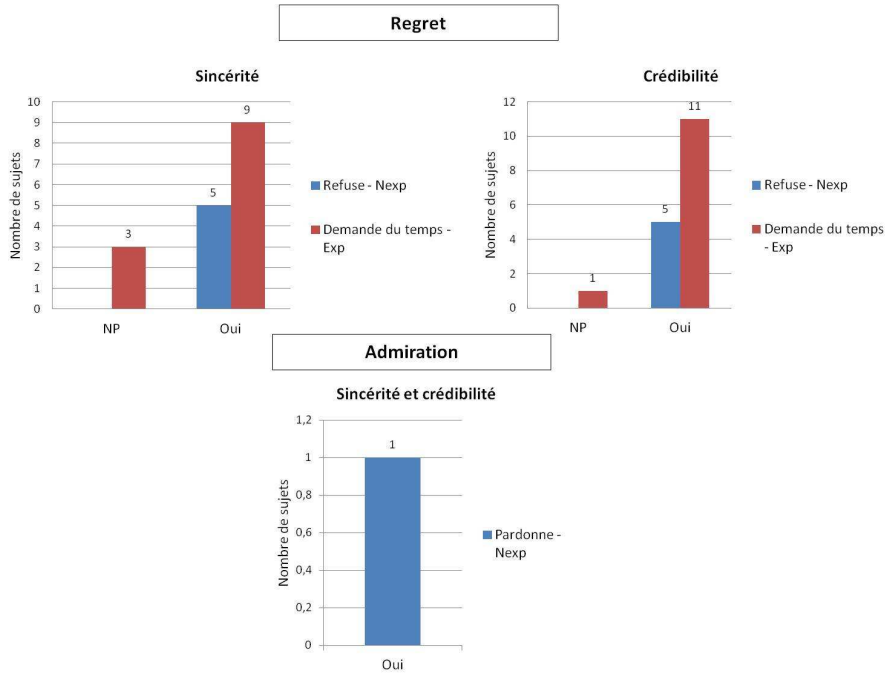

FIG. B.8 – Scores de sincérité et de crédibilité du regret et de l'admiration.

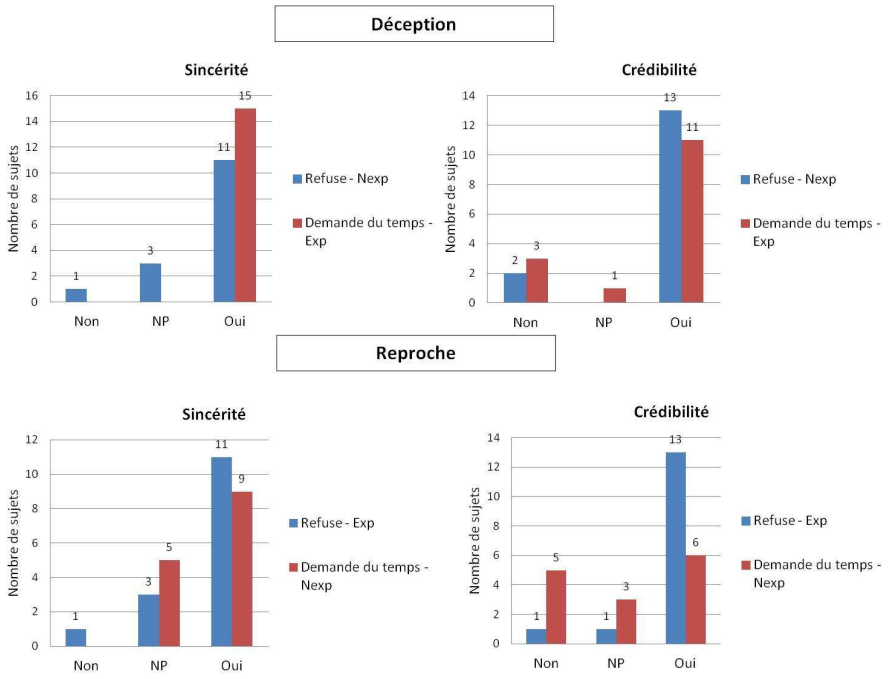

FIG. B.9 – Scores de sincérité et de crédibilité de la déception et du reproche.

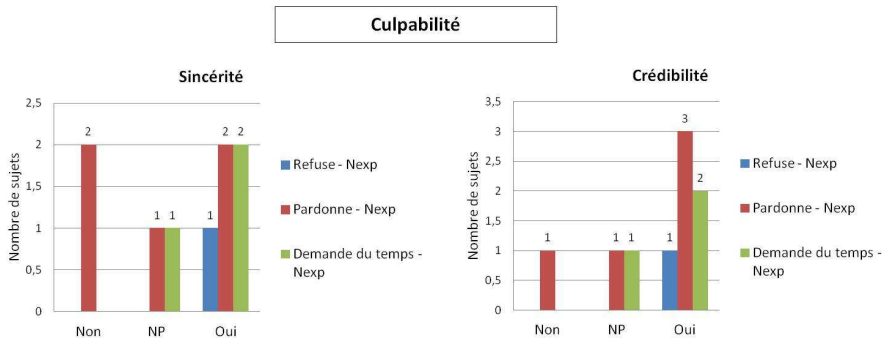

FIG. B.10 – Scores de sincérité et de crédibilité de la culpabilité.

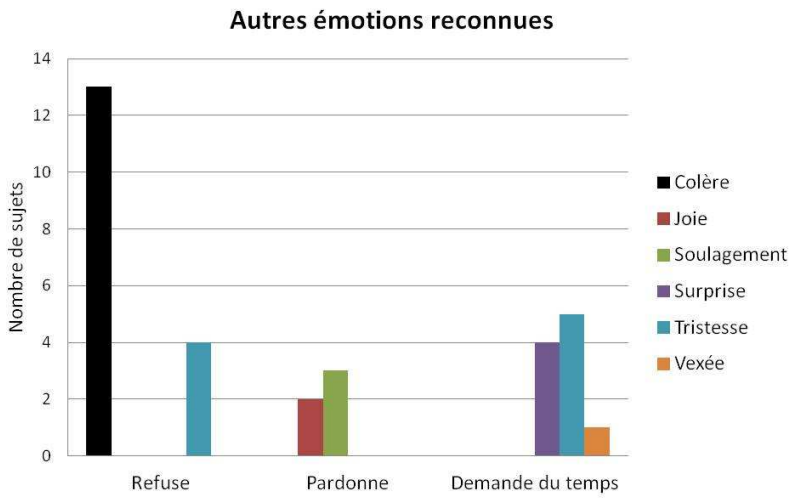

FIG. B.11 – Autres émotions reconnues suivant les différents sous-scénarios.

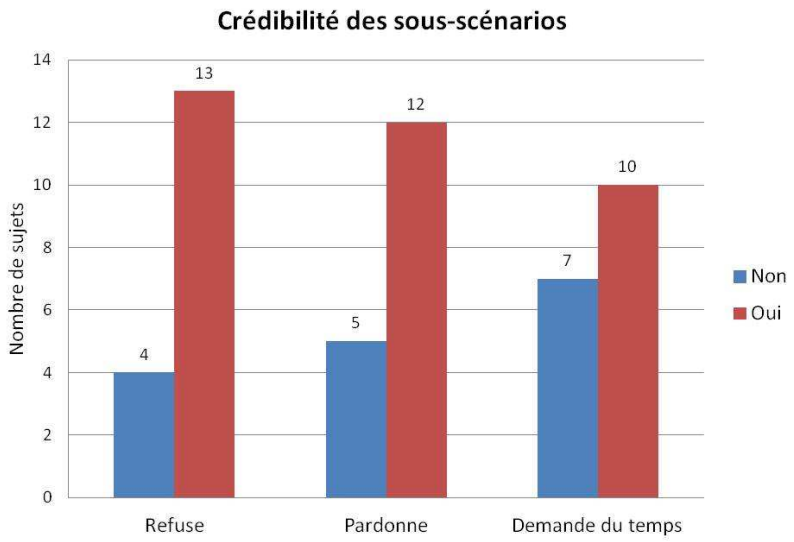

FIG. B.12 – Réponses à la question : « D'après vous, la réaction de Mary est-elle appropriée ? »

C
Publications

- Jérémy Rivière, Sylvie Pesty. **Actes de conversation multimodaux et emotions.** *WACA 2010 : Quatrième Workshop sur les Agents Conversationnels Animés*, Lille, France, Novembre 2010.
- Jérémy Rivière, Sylvie Pesty. **Actes de langage et emotions : vers un langage de conversation multimodal.** *JFSMA : Journées Francophones des Systèmes Multi-Agents*, p. 107–116, Mahdia, Tunisie, 2010.
- Nadine Guiraud, Dominique Longin, Emiliano Lorini, Jérémy Rivière, Sylvie Pesty. **The face of emotions : a logical formalization of expressive speech acts.** *10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Taipei, Taiwan, Mai 2011.
- Jérémy Rivière, Carole Adam, Sylvie Pesty, Catherine Pelachaud, Nadine Guiraud, Dominique Longin, Emiliano Lorini. **Expressive Multimodal Conversational Acts for SAIBA Agents.** Hannes Högni Vilhjálmsson, Stefan Kopp, Stacy Marsella, Kristinn R. Thórisson (eds.), *11th International Conference on Intelligent Virtual Agents (IVA 2011)*, p. 316–323, Lecture Notes in Computer Science, septembre 2011.
- Jérémy Rivière, Carole Adam, Sylvie Pesty. **Langage de Conversation Multimodal pour Agent Conversationnel Animé**. *Technique et Sciences Informatiques (TSI)*, TSI : Revue des sciences et technologies de l'information, 31, 2012.
- Jérémy Rivière, Carole Adam, Sylvie Pesty. **A reasoning module to select ECA's communicative intention.** Michael Neff, Marilyn Walker, Ana Paiva, Yukiko I Nakano (eds.), *12th International Conference on Intelligent Virtual Agents (IVA 2012)*, 7502 :447-454, Lecture Notes in Computer Science, Heidelberg, 2012.
- Jérémy Rivière, Carole Adam, Sylvie Pesty. **Un ACA sincère comme compagnon artificiel.** *WACAI 2012 : Workshop Affect, Compagnon Artificiel, Interaction*, Grenoble, France, Novembre 2012.

Bibliographie

[Adam 2007] Carole Adam. *Emotions : from psychological theories to logical formalization and implementation in a BDI agent.* Thèse de doctorat, INP, Toulouse, France, juillet 2007.

[Adam 2009a] C. Adam et L. Cavedon. *An extension of FML for a lexicon of engaging communicative strategies.* In in Proceedings of the EDAML at at the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '09), 2009.

[Adam 2009b] Carole Adam, Andreas Herzig et Dominique Longin. *A logical formalization of the OCC theory of emotions.* Synthese, vol. 168, no. 2, pages 201–248, 2009.

[Allen 1986] J Allen et C R Perrault. *Analyzing intention in utterances.* In Barbara J. Grosz, Karen Sparck-Jones et Bonnie Lynn Webber, editeurs, Readings in natural language processing, pages 441–458. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1986.

[Allen 1994] James F. Allen, Lenhart K. Schubert, George Ferguson, Peter Heeman, Chung H Hwang, Tsuneaki Kato, Marc Light, Nathaniel G. Martin, Bradford W. Miller, Massimo Poesio et David R. Traum. *The TRAINS Project : A Case Study in Defining a Conversational Planning Agent.* Rapport technique, University of Rochester, Rochester, NY, USA, 1994.

[Allwood 1991] J.S. Allwood, J. Nivre et E. Ahlsén. On the semantics and pragmatics of linguistic feedback. Gothenburg papers in theoretical linguistics. University of Gothenburg, Dept. of Linguistics, 1991.

[AMIConsortium 2004] AMIConsortium. *Augmented Multi-party Interaction Project.* Web, 2004.

[Argyle 1976] M. Argyle et M. Cook. Gaze and mutual gaze. Cambridge University Press, 1976.

[Arnold 1960] M.B. Arnold. Emotion and personality : Psychological aspects. Emotion and Personality. Columbia University Press, 1960.

[Austin 1962] John L. Austin. How to do things with words. Harvard University Press, Cambridge, Mass., 1962.

[Averill 1985] J. R. Averill. *The social construction of emotion : With special reference to love.* In K.J. Gergen et K.E. Davis, editeurs, The Social construction of the person, pages 89–109. Springer, 1985.

[Aylett 2004] R. Aylett, S. Louchart et J. Pickering. *A mechanism for acting and speaking for empathic agents.* In Proceedings of the 2004 Autonomous Agents and Multi-Agent Systems Workshop (AAMAS '04), 2004.

[Baker 1994] Michael J. Baker. *A model for negotiation in teaching-learning dialogues.* Journal of Artificial Intelligence and Education, vol. 5(2), pages 199–254, 1994.

[Barrett 2007] L. Barrett, K. Lindquist et M. Gendron. *Language as context for the perception of emotion.* Trends in cognitive sciences, vol. 11, no. 8, pages 327–332, 2007.

[Bartneck 2002] C. Bartneck. *Integrating the OCC model of emotions in embodied characters.* In Proceedings of the Workshop on Virtual Conversational Characters : Applications, Methods, and Research Challenges, 2002.

[Bates 1994] Joseph Bates. *The role of emotion in believable agents.* Commun. ACM, vol. 37, no. 7, pages 122–125, 1994.

[Becker 2004] Christian Becker et Ipke Wachsmuth. *Simulating the emotion dynamics of a multimodal conversational agent.* In In Proceedings of the Tutorial and Research Workshop on Affective Dialogue Systems (ADS-04), LNAI 3068, pages 154–165. Springer, 2004.

[Berger 2005] Alexandra Berger et Sylvie Pesty. *Towards a Conversational Language for Artificial Agents in Mixed Community.* In Proc. of the International Central and Eastern European Conference on Multi-Agent Systems (CEEMAS), pages 31–40, Budapest, Hungary, sep 2005. Springer Verlag.

[Berger 2006] Alexandra Berger. *La communication entre agents de communautés mixtes : Un Langage de Conversation Expressif pour agents artificiels.* Thèse de doctorat, Institut National de Polytechnique de Grenoble et Université du Québec à Trois-Rivières, décembre 2006.

[Bevacqua 2008] Elisabetta Bevacqua, Maurizio Mancini et Catherine Pelachaud. *A Listening Agent Exhibiting Variable Behaviour.* In Proceedings of the 8th international conference on Intelligent Virtual Agents, IVA '08, pages 262–269, Berlin, Heidelberg, 2008. Springer-Verlag.

[Bisognin 2004] L. Bisognin et S. Pesty. *Emotions et Systèmes Multi-Agents : Une Architecture d'Agent Emotionnel.* In O. Boissier et Z. Guessoum, editeurs, JFSMA : Journées Francophones sur les Systèmes Multi-Agents, Paris, 2004. Hermès.

[Blee 2009] Jeff Blee, David Billington et Abdul Sattar. *Reasoning with Levels of Modalities in BDI Logic.* In Aditya Ghose, Guido Governatori et Ramakoti Sadananda, editeurs, Agent Computing and Multi-Agent Systems, volume 5044 of *Lecture Notes in Computer Science*, pages 410–415. Springer Berlin / Heidelberg, 2009.

[Bratman 1987] Michael E. Bratman. Intention, plans, and practical reason. Harvard University Press, Novembre 1987.

[Breazeal 1999] C. Breazeal et B. Scassellati. *How to build robots that make friends and influence people.* Intelligent Robots and Systems, vol. 2, pages 858–863, 1999.

[Breitfuss 2008] Werner Breitfuss, Helmut Prendinger et Mitsuru Ishizuka. *Automatic Generation of Gaze and Gestures for Dialogues between Embodied Conversational Agents.* International Journal of Semantic Computing, vol. 2, no. 1, pages 71–90, 2008.

[Bunt 1989] H. C. Bunt. *Information Dialogues as Communicative Action in Relation to Partner Modelling and Information Processing.* In M. M. Taylor, F. Neel et D. G. Bouwhuis, editeurs, The Structure of Multimodal Dialogue, pages 47–73. North-Holland, Amsterdam, 1989.

[Caelen 2003] Jean Caelen. *Stratégies de dialogue.* In Andreas Herzig, Brahim Chaib-Draa et Philippe Mathieu, editeurs, Actes des Secondes Journées Francophones des Modèles Formels de l'Interaction, pages 29–39, Lille, France, 2003. Cépaduès.

[Cannon 1927] Walter B. Cannon. *The James-Lange Theory of Emotions : A Critical Examination and an Alternative Theory.* The American Journal of Psychology, vol. 39, no. 1/4, pages 106–124, 1927.

[Casali 2011] Ana Casali, Lluis Godo et Carles Sierra. *A graded BDI agent model to represent and reason about preferences.* Artificial Intelligence, vol. 175, no. 7-8, pages 1468 – 1478, 2011.

[Chaib-draa 1999] Brahim Chaib-draa et Daniel Vanderveken. *Agent Communication Language : Toward a Semantics Based on Success, Satisfaction, and Recursion.* In Proceedings of the 5th International Workshop on Intelligent Agents, Agent Theories, Architectures, and Languages (ATAL '98), pages 363–379, London, UK, UK, 1999. Springer-Verlag.

[Cohen 1979] Philip R Cohen et C Raymond Perrault. *Elements of a Plan-Based theory of Speech Acts.* Cognitive Science, vol. 3, no. 3, pages 177–212, 1979.

[Cohen 1990a] Philip R. Cohen et Hector J. Levesque. *Intention is choice with commitment.* Artificial Intelligence, vol. 42, no. 2-3, pages 213–261, 1990.

[Cohen 1990b] Philip R Cohen et Hector J Levesque. *Persistence, Intention and Commitment.* In P Cohen, J Morgan et MEditors Pollack, editeurs, Intentions in Communication, pages 33–69. MIT Press, 1990.

[Core 1997] Mark G. Core et James F. Allen. *Coding Dialogs with the DAMSL Annotation Scheme.* In Proc. of the Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines, Cambridge, MA, 1997.

[Courgeon 2008] Matthieu Courgeon, Jean-Claude Martin et Christian Jacquemin. *MARC : a Multimodal Affective and Reactive Character.* In Proc. of The 1st workshop on Affective Interaction in Natural Environments (AFFINE), Chania, Crete, 2008.

[Courgeon n C] M. Courgeon. *Modèles Informatiques des Emotions et de leurs Expressions Faciales pour l'Interaction Homme-Machine Affective Temps Réel.* PhD thesis, Ecole doctorale EDIPS. Université Paris-Sud, 2011, Directeurs de thèse : Martin J.-C. et Jacquemin C.

[Darwin 1872] Charles Darwin. On the origin of species by means of natural selection : Or, the preservation of favored races in the struggle for life. D. Appleton, New York, 1872.

[Dautenhahn 1997] Kerstin Dautenhahn. *Ants don't have Friends - Thoughts on Socially Intelligent Agents.* Rapport technique, In Socially Intelligent Agents, 1997.

[Davidson 1980] Donald Davidson. Essays on actions and events. Oxford University Press, 1980.

[de Gelder 2011] B. de Gelder et J. Van den Stock. *Real faces, real emotions : perceiving facial expressions in naturalistic contexts of voices, bodies and scenes.* In M.H. Johnson G. R. A.J. Calder J.V. Haxby, editeur, The handbook of face perception., pages 535–550. Oxford : Oxford University Press, 2011.

[de Jong 2008] Markus de Jong, Mariët Theune et Dennis Hofs. *Politeness and alignment in dialogues with a virtual guide.* In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 1, AAMAS '08, pages 207–214, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems.

[de Melo 2011] Celso M. de Melo, J. Gratch et P. Carnevale. *Reverse Appraisal : Inferring from Emotion Displays who is the Cooperator and the Competitor in a Social Dilemma.* In The 33rd Annual Meeting of the Cognitive Science Society (CogSci) 2011, Boston, MA, 2011.

[de Rosis 2003] Fiorella de Rosis, Catherine Pelachaud, Isabella Poggi, Valeria Carofiglio et Berardina De Carolis. *From Greta's mind to her face : modelling the dynamics of affective states in a conversational embodied agent.* Int. J. Hum.-Comput. Stud., vol. 59, no. 1-2, pages 81–118, 2003.

[De Sevin 2010] Etienne De Sevin, Sylwia Hyniewska et Catherine Pelachaud. *Influence of Personality Traits on Backchannel Selection.* In Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud et Alla Safonova, editeurs, Proc. of the 10th Int. Conf. on Intelligent Virtual Agents (IVA'10), volume 6356, pages 187–193, Berlin, Heidelberg, 2010. Springer-Verlag.

[Demeure 2011] V. Demeure, R. Niewiadomski et C. Pelachaud. *How is believability of a virtual agent related to warmth, competence, personification, and embodiment ?* Presence : Teleoper. Virtual Environ., vol. 20, no. 5, pages 431–448, October 2011.

[Ekman 1972] Paul Ekman et Wallace V. Friesen. *Hand Movements.* Journal of Communication, vol. 22, no. 4, pages 353–374, 1972.

[Ekman 1978] P. Ekman et W. Friesen. Facial Action Coding System : A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, 1978.

[Ekman 1992] Paul Ekman. *An argument for basic emotions.* Cognition and Emotion, vol. 6, no. 3, pages 169–200, 1992.

[Endrass 2011] Birgit Endrass, Yukiko Nakano, Afia Akhter Lipi, Matthias Rehm et Elisabeth André. *Culture-related topic selection in small talk conversations across Germany and Japan.* In Proceedings of the 10th international conference on Intelligent virtual agents, IVA'11, pages 1–13, Berlin, Heidelberg, 2011. Springer-Verlag.

[Fikes 1971] Richard E. Fikes et Nils J. Nilsson. *STRIPS : A new approach to the application of theorem proving to problem solving.* Artificial Intelligence, vol. 2, no. 3-4, 1971.

[FIPA-ACL 2006] FIPA-ACL. *Speech acts library.* Web, Avril 2006.

[Fiske 2007] S. Fiske, A. Cuddy et P. Glick. *Universal dimensions of social cognition : warmth and competence.* Trends in Cognitive Sciences, vol. 11, no. 2, pages 77–83, 2007.

[Fleischman 2002] M. Fleischman et E. Hovy. *Towards emotional variation in speech-based natural language generation.* In International Natural Language Generation Conference, New York, NY, 2002.

[Gebhard 2003] Patrick Gebhard, Michael Kipp, Martin Klesen et Thomas Rist. *Adding the Emotional Dimension to Scripting Character Dialogues.* In Thomas Rist, Ruth Aylett, Daniel Ballin et Jeff Rickel, editeurs, IVA, volume 2792 of *Lecture Notes in Computer Science*, pages 48–56. Springer, 2003.

[Golan 2006] Ofer Golan, Simon Baron-Cohen et Jacqueline Hill. *The Cambridge Mindreading (CAM) Face-Voice Battery : Testing complex emotion recognition in adults with and without Asperger syndrome.* Journal of Autism and Developmental Disorders, vol. 36, no. 2, pages 169–183, 2006.

[Goldberg 1990] Lewis R. Goldberg. *An alternative "description of Personality" : the Big-Five factor structure.* Journal of personality and social psychology, vol. 59, no. 6, pages 1216–1229, 1990.

[Goldman 1970] A.I. Goldman. A theory of human action. Prentice-Hall, 1970.

[Gong 2008] Li Gong. *How social is social responses to computers ? The function of the degree of anthropomorphism in computer representations.* Comput. Hum. Behav., vol. 24, pages 1494–1509, 2008.

[Gratch 2004] Jonathan Gratch et Stacy Marsella. *A domain-independent framework for modeling emotion.* Journal of Cognitive Systems Research, vol. 5, pages 269–306, 2004.

[Gratch 2007] Jonathan Gratch, Ning Wang, Jillian Gerten, Edward Fast et Robin Duffy. *Creating Rapport with Virtual Agents.* In 7th International Conference on Intelligent Virtual Agents (IVA'07), pages 125–138. Springer, 2007.

[Grice 1975] H. P. Grice. *Logic and conversation.* In Peter Cole et Jerry L. Morgan, editeurs, Syntax and semantics, volume 3. New York : Academic Press, 1975.

[Gross 2011] J. J. Gross et L. F. Barrett. *Emotion Generation and Emotion Regulation :One or Two Depends on Your Point of View.* Emotion Review, vol. 3, no. 1, pages 8–16, 2011.

[Guiraud 2011] Nadine Guiraud, Dominique Longin, Emiliano Lorini, Sylvie Pesty et Jérémy Rivière. *The face of emotions : a logical formalization of expressive speech acts.* In Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011), pages 1031–1038, Taipei, Taiwan, May 2011. International Foundation for AAMAS.

[Hareli 2010] Shlomo Hareli et Ursula Hess. *What emotional reactions can tell us about the nature of others : An appraisal perspective on person perception.* Cognition and Emotion, vol. 24, no. 1, pages 128–140, 2010.

[Heylen 2008] D.K.J. Heylen, S. Kopp, S.C. Marsella, C. Pelachaud et H. Vilhjálmson. *The Next Step towards a Function Markup Language.* In H. Prendinger, J. Lester et M. Ishizuka, editeurs, Intelligent Virtual Agents (IVA) 2008, volume 5208 of *Lecture Notes in Artificial Intelligence*, pages 270–280, Berlin, 2008. Springer Verlag.

[Hill 2001] R. Hill, J. Gratch, W. L. Johnson, C. Kyriakakis, C. LaBore, R. Lindheim, S. Marsella, D. Miraglia, B. Moore, J. Morie, J. Rickel, M. Thiébaux, L. Tuch, R. Whitney, J. Douglas et W. Swartout. *Toward the holodeck : integrating graphics, sound, character and story.* In AGENTS '01 : Proceedings of the fifth international conference on Autonomous agents, pages 409–416, New York, NY, USA, 2001. ACM.

[James 1884] William James. *What is an Emotion ?* Mind, vol. 9, no. 34, pages 188–205, 1884.

[Jaques 2009] Patricia Jaques, M. Lehmann et Sylvie Pesty. *Evaluating the Affective Tactics of an Emotional Pedagogical Agent.* In Proceedings of the 24th annual ACM symposium on applied computing 2009, pages 104–109, Honolulu, Hawaii, USA, 2009.

[Jiang 2007] Hong Jiang, Jose M. Vidal et Michael N. Huhns. *EBDI : an architecture for emotional agents.* In Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems, AAMAS '07, pages 11 :1–11 :3, New York, NY, USA, 2007. ACM.

[Kahneman 1986] Daniel Kahneman et Dale T. Miller. *Norm Theory : Comparing Reality to Its Alternatives.* Psychological Review, vol. 93, no. 2, pages 136–152, 1986.

[Klesen 2003] M Klesen, M Kipp, P Gebhard et T Rist. *Staging exhibitions : methods and tools for modelling narrative structure to produce interactive performances with virtual actors.* Virtual Reality, vol. 7, no. 1, pages 17–29, 2003.

[Kopp 2003] S. Kopp, B. Jung, N. Lessmann et I. Wachsmuth. *Max - A Multimodal Assistant in Virtual Reality Construction.* KI - Künstliche Intelligenz, vol. 4/03, pages 11–17, 2003.

[Kopp 2006] Stefan Kopp, Brigitte Krenn, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Hannes Pirker, Kristinn R. Thórisson et Hannes Vilhjálmsson. *Towards a Common Framework for Multimodal Generation : The Behavior Markup Language.* In International Conference on Intelligent Virtual Agent (IVA), pages 21–23, 2006.

[Kopp 2008] S. Kopp et N. Pfeiffer-Lessmann. *Functions of Speaking and Acting : An Interaction Model for Collaborative Construction Tasks.* In D. Heylen, S. Kopp, S. Marsella, C. Pelachaud et H. Vilhjálmsson, editeurs, The First FML workshop, AAMAS'08, Estoril, Portugal, 2008.

[Kripke 1959] Saul A. Kripke. *A Completeness Theorem in Modal Logic.* The Journal of Symbolic Logic, vol. 24, no. 1, pages 1–14, 1959.

[Kripke 1963] Saul A. Kripke. *Semantical Considerations on Modal Logic.* Acta Philosophica Fennica, vol. 16, pages 83–94, 1963.

[Lazarus 1991] Richard S. Lazarus. Emotion and adaptation. Oxford University Press, New York, USA, 1991.

[Lee 2005] Sau-lai Lee, Ivy Y. Lau, S. Kiesler et Chi-Yue Chiu. *Human Mental Models of Humanoid Robots.* In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005., pages 2767–2772. IEEE Computer Society, 2005.

[Maatman 2005] R. M. Maatman, Jonathan Gratch et Stacy Marsella. *Natural Behavior of a Listening Agent.* In Themis Panayiotopoulos, Jonathan Gratch, Ruth Aylett, Daniel Ballin, Patrick Olivier et Thomas Rist, editeurs, Proceedings of the International Conference in Intelligent Virtual Agents, pages 25–36. Springer-Verlag, 2005.

[Mancini 2008] M. Mancini et C. Pelachaud. *The FML - APML language.* In D. Heylen, S. Kopp, S. Marsella, C. Pelachaud et H. Vilhjálmsson, editeurs, The First FML workshop, AAMAS'08, Estoril, Portugal, 2008.

[Mann 2004] C. W. Mann et J. Kreutel. *Speech Acts and Recognition of Insincerity.* In Proc. of Catalog'04 : Semantics and Pragmatics of Dialog, 2004.

[Manstead 2001] A. S. R. Manstead et A. H. Fischer. *Long-lasting Cognitive and Social Consequences of Emotion : Social Sharing and Rumination.* In K. R. Scherer, A. Schorr et T. Johnstone, editeurs, Appraisal Processes in Emotion : Theory, Research, Application, pages 221–232. New York : Oxford University Press, 2001.

[Mascarenhas 2010] Samuel Mascarenhas, Joao Dias, Rui Prada et Ana Paiva. *A dimensional model for cultural behavior in virtual agents.* Appl. Artif. Intell., vol. 24, no. 6, pages 552–574, 2010.

[McNeill 1996] D. McNeill. Hand and mind : What gestures reveal about thought. Psychology/cognitive science. University of Chicago Press, 1996.

[McRorie 2009] Margaret McRorie, Ian Sneddon, Etienne Sevin, Elisabetta Bevacqua et Catherine Pelachaud. *A Model of Personality and Emotional Traits.* In Proc. of the 9th Int. Conf. on Intelligent Virtual Agents (IVA'09), pages 27–33, Amsterdam, The Netherlands, 2009. Springer-Verlag.

[McRorie 2012] M. McRorie, I. Sneddon, G. McKeown, E. Bevacqua, E. de Sevin et C. Pelachaud. *Building and Evaluating Personality in Virtual Agents.* Transactions on Affective Computing, 2012.

[Melo 2009] Celso M. Melo, Liang Zheng et Jonathan Gratch. *Expression of Moral Emotions in Cooperating Agents.* In Proceedings of the 9th International Conference on Intelligent Virtual Agents (IVA'09), pages 301–307, Berlin, Heidelberg, 2009. Springer-Verlag.

[Meyer 2006] John-Jules Ch. Meyer. *Reasoning about emotional agents : Research Articles.* Int. J. Intell. Syst., vol. 21, no. 6, pages 601–619, 2006.

[Morency 2010] Louis-Philippe Morency, Iwan de Kok et Jonathan Gratch. *A probabilistic multimodal approach for predicting listener backchannels.* Autonomous Agents and Multi-Agent Systems, vol. 20, no. 1, pages 70–84, 2010.

[Mori 1970] M. Mori. *The Uncanny Valley.* In Energy, pages 33–35, 1970.

[Niewiadomski 2009] Radoslaw Niewiadomski, Catherine Pelachaud et Sylwia Hyniewska. *Evaluation of multimodal sequential expressions of emotions in ECA.* In In : Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII). Springer, 2009.

[Niewiadomski 2010] Radoslaw Niewiadomski, Virginie Demeure et Catherine Pelachaud. *Warmth, Competence, Believability and Virtual Agents.* In Jan Allbeck, Norman Badler, Timothy Bickmore, Catherine Pelachaud et Alla Safonova, editeurs, Intelligent Virtual Agents (IVA'10), volume 6356 of *Lecture Notes in Computer Science*, pages 272–285. Springer Berlin / Heidelberg, 2010.

[Oatley 1987] Keith Oatley et P. N. Johnson-Laird. *Towards a Cognitive Theory of Emotions.* Cognition and Emotion, vol. 1, no. 1, pages 29–50, 1987.

[Ochs 2007] Magalie Ochs. *Modélisation, formalisation et mise en oeuvre d'un agent rationnel dialoguant émotionnel empathique.* Thèse de doctorat, Université Paris VIII, décembre 2007.

[Ochs 2010] Magalie Ochs, Radoslaw Niewiadomski et Catherine Pelachaud. *How a Virtual Agent Should Smile ? - Morphological and Dynamic Characteristics of Virtual Agent's Smiles.* In Jan M. Allbeck, Norman I. Badler, Timothy W. Bickmore, Catherine Pelachaud et Alla Safonova, editeurs, IVA, volume 6356 of *Lecture Notes in Computer Science*, pages 427–440. Springer, 2010.

[Ortony 1988] Andrew Ortony, Gerald L. Clore et Allan Collins. The cognitive structure of emotions. Cambridge University Press, London, 1988.

[Paiva 2006] Ana Paiva, Ruth Aylett et Joao Dias. *An affectively driven planner for synthetic characters.* In International Conference on Automated Planning and Scheduling (ICAPS), pages 2–10. AAAI press, 2006.

[Pandzic 2003] I.S. Pandzic et R. Forchheimer. Mpeg-4 facial animation : The standard, implementation and applications. John Wiley & Sons, 2003.

[Parkinson 1996] Brian Parkinson. *Emotions are social.* British Journal of Psychology, vol. 87, no. 4, pages 663–683, 1996.

[Pereira 2008] David Pereira, Eugénio Oliveira et Nelma Moreira. *Formal Modelling of Emotions in BDI Agents.* In Fariba Sadri et Ken Satoh, editeurs, Computational Logic in Multi-Agent Systems, pages 62–81. Springer-Verlag, Berlin, Heidelberg, 2008.

[Perrault 1980] C. Raymond Perrault et James F. Allen. *A plan-based analysis of indirect speech acts.* Comput. Linguist., vol. 6, no. 3-4, pages 167–182, Juillet 1980.

[Perrault 1990] D. Perrault. *An application of default logic to speech act theory.* In Cohen P., Morgan J. et Pollack M., editeurs, Intentions in Communication, pages 161–185. Cambridge : MIT press, 1990.

[Pesty 2012] S. Pesty et D. Duhaut. *Relation Humain - Compagnon Artificiel.* Intercompréhension de l'Intraspécifique à l'Interspécifique, 2012.

[Petukhova 2007] Volha Petukhova et Harry Bunt. *A Multidimensional Approach to Multimodal Dialogue Act Annotation.* In Proc. of the Seventh International Workshop on Computational Semantics (IWCS-7), pages 142–153, Tilburg, 2007. Springer.

[Piaget 1989] J. Piaget. *Les émotions.* In B. Rimé et K. Scherer, editeurs, Les relations entre l'intelligence et l'affectivité dans le developpement de l'enfant, pages 75–95. Delachaux et Niestlé, Neuchâtel-Paris, 1989.

[Picard 1997] Rosalind W. Picard. Affective computing. The MIT Press, Cambridge, Massachusetts, 1997.

[Poggi 1998] Isabella Poggi et Catherine Pelachaud. *Performative faces.* Speech Communication, pages 5–21, 1998.

[Poggi 2000] Isabella Poggi et Catherine Pelachaud. *Performative Facial Expressions in Animated Faces.* In Embodied Conversational Agents, pages 155–188. MIT Press, 2000.

[Poggi 2003] Isabella Poggi, Catherine Pelachaud et Emanuela Magno Caldognetto. *Gestural mind markers in ECAs.* In Proceedings of the second

international joint conference on Autonomous agents and multiagent systems, AAMAS '03, pages 1098–1099, New York, NY, USA, 2003. ACM.

[Poggi 2005] Isabella Poggi, Catherine Pelachaud, Fiorella de Rosis, Valeria Carofiglio et Berardina de Carolis. *GRETA. A Believable Embodied Conversational Agent.* In Patrizia Paggio et Bart Jongejan, editeurs, Multimodal Communication in Virtual Environments, pages 27–45. Springer, 2005.

[Prendinger 2001] Helmut Prendinger et Mitsuru Ishizuka. *Let's Talk ! Socially Intelligent Agents for Language Conversation Training.* IEEE Transactions on Systems, Man, and Cybernetics-Part A : Systems and Humans, vol. 31, pages 465–471, 2001.

[Rao 1991] Anand S. Rao et Michael P. Georgeff. *Modeling Rational Agents within a BDI-Architecture.* In James Allen, Richard Fikes et Erik Sandewall, editeurs, Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91), pages 473–484. Morgan Kaufmann publishers Inc. : San Mateo, CA, USA, 1991.

[Reeves 1996] Byron Reeves et Clifford Nass. The media equation : how people treat computers, television, and new media like real people and places. Cambridge University Press, New York, NY, USA, 1996.

[Rhodes 2000] G. Rhodes, C. Hickford et L. Jeffery. *Sex-typicality and attractiveness : Are supermale and superfemale faces super-attractive.* British Journal of Psychology, vol. 91, no. 1, pages 125–140, 2000.

[Rickel 1997] Jeff Rickel et W. Lewis Johnson. *Steve : An Animated Pedagogical Agent for Procedural Training in Virtual Environments.* Animated Interface Agents : making them intelligents, pages 71–76, 1997.

[Rickel 1998] Jeff Rickel et W. Lewis Johnson. *Animated Agents for Procedural Training in Virtual Reality : Perception, Cognition, and Motor Control.* Applied Artificial Intelligence, vol. 13, pages 343–382, 1998.

[Rimé 1992] Bernard Rimé, Pierre Philippot, Stefano Boca et Batja Mesquita. *Long-lasting Cognitive and Social Consequences of Emotion : Social Sharing and Rumination.* European Review of Social Psychology, vol. 3, no. 1, pages 225–258, 1992.

[Rosenberg 1968] Seymour Rosenberg, Carnot Nelson et P. S. Vivekananthan. *A multidimensional approach to the structure of personality impressions.* In Journal of Personality and Social Psychology, volume 9, pages 283–294, 1968.

[Russell 1977] James A Russell et Albert Mehrabian. *Evidence for a three-factor theory of emotions.* Journal of Research in Personality, vol. 11, no. 3, pages 273–294, 1977.

[Russell 1980] J. A. Russell. *A circumplex model of affect.* Journal of Personality and Social Psychology, vol. 39, pages 1161–1178, 1980.

[Sadek 1991] D. Sadek. *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication.* PhD thesis, Université de Rennes I, 1991.

[Samtani 2008] P. Samtani, A. Valente et W. L. Johnson. *Applying the SAIBA framework to the tactical language and culture training system.* In Why

Conversational Agents do what they do ? Functional Representations for Generating Conversational Agent Behavior. AAMAS 2008, Estoril, Portugal, 2008.

[Schachter 1959] S. Schachter. The psychology of affiliation : Experimental studies of the sources of gregariousness. Numeéro vol. 1 de Stanford Studies in psychology. Stanford University Press, 1959.

[Scherer 1984] K. R. Scherer et P. Ekman. Approaches to emotion. Erlbaum, Hillsdale, NJ, 1984.

[Scherer 2000] K. R. Scherer. *Emotion.* In M. Hewstone et W. Stroebe, editeurs, Introduction to Social Psychology : A European perspective (3rd. ed), pages 151–191. Blackwell, Oxford, 2000.

[Scherer 2001] K. R. Scherer. *Appraisal considered as a process of multilevel sequential checking.* Appraisal processes in emotion : Theory, methods, research, pages 92–120, 2001.

[Scherer 2007] K. R. Scherer et H. Ellgring. *Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal ?* Emotion, vol. 7, pages 113–130, 2007.

[Schroder 2012] Marc Schroder, Elisabetta Bevacqua, Roddy Cowie, Florian Eyben, Hatice Gunes, Dirk Heylen, Mark ter Maat, Gary McKeown, Sathish Pammi, Maja Pantic, Catherine Pelachaud, Bjorn Schuller, Etienne de Sevin, Michel Valstar et Martin Wollmer. *Building Autonomous Sensitive Artificial Listeners.* Affective Computing, IEEE Transactions on, vol. 3, no. 2, pages 165–183, 2012.

[Searle 1969] John R. Searle. Speech acts : an essay in the philosophy of language. Cambridge University Press, New York, USA, 1969.

[Searle 1983] J.R. Searle. Intentionality, an essay in the philosophy of mind. Cambridge Paperback Library. Cambridge University Press, 1983.

[Searle 1985] John R. Searle et Daniel. Vanderveken. Foundations of illocutionary logic. Cambridge University Press, New York, USA, 1985.

[Sloman 2001] Aaron Sloman. *Beyond Shallow Models of Emotion.* In Olivetti Belardinelli, editeur, Cognitive Processing : International Quarterly of Cognitive Science, pages 177–198. Springer, 2001.

[Swartout 2010] William Swartout, David Traum, Ron Artstein, Dan Noren, Paul Debevec, Kerry Bronnenkant, Josh Williams, Anton Leuski, Shrikanth Narayanan, Diane Piepol, Chad Lane, Jacquelyn Morie, Priti Aggarwal, Matt Liewer, Jen-Yuan Chiang, Jillian Gerten, Selina Chu et Kyle White. *Ada and grace : toward realistic and engaging virtual museum guides.* In Proceedings of the 10th international conference on Intelligent virtual agents, IVA'10, pages 286–300, Berlin, Heidelberg, 2010. Springer-Verlag.

[Tan 2009] Shawna C. G. Tan et Alexander Nareyek. *Integrating Facial, Gesture, and Posture Emotion Expression for a 3D Virtual Agent.* In Proceedings of the 14th International Conference on Computer Games : AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games (CGames 2009 USA), pages 23–31, 2009.

[Tangney 2003] J.P. Tangney et R.L. Dearing. Shame and guilt. Emotions And Social Behavior. Guilford Press, 2003.

[Tcherkassof 2008] A. Tcherkassof. Les émotions et leurs expressions. La Psychologie en plus. Presses Universitaires de Grenoble, 2008.

[Traum 1992] David R. Traum et Elizabeth A. Hinkelman. *Conversation Acts in Task-Oriented Spoken Dialogue.* Computational Intelligence, vol. 8, pages 575–599, 1992.

[Traum 1994] David R. Traum et James F. Allen. *Discourse Obligations in Dialogue Processing.* In Proc. of the 32th annual meeting of the Association for Computational Linguistics (ACL), pages 1–8, 1994.

[Traum 2002] David Traum et Jeff Rickel. *Embodied agents for multi-party dialogue in immersive virtual worlds.* In Proceedings of the first international joint conference on Autonomous agents and multiagent systems : part 2, AAMAS '02, pages 766–773, New York, NY, USA, 2002. ACM.

[Traum 2004] David R. Traum, Stacy Marsella et Jonathan Gratch. *Emotion and Dialogue in the MRE Virtual Humans.* In Elisabeth André, Laila Dybkjær, Wolfgang Minker et Paul Heisterkamp, editeurs, ADS, volume 3068 of *Lecture Notes in Computer Science*, pages 117–127, Germany, 2004. Springer.

[Vanderveken 1988] Daniel Vanderveken. Les actes de discours. Editions Pierre Mardaga, Bruxelles, Belgique, 1988.

[Vanderveken 1990] D. Vanderveken. Principles of language use. Meaning and speech acts. Cambridge University Press, 1990.

[Vilhjalmsson 2005] Hannes Vilhjalmsson et Stacy C. Marsella. *Social performance framework.* In In : Workshop on Modular Construction of Human-Like Intelligence at the AAAI 20th National Conference on Artificual Intelligence, 2005.

[Vilhjalmsson 2007] Hannes Vilhjalmsson, Nathan Cantelmo, Justine Cassell, Nicolas E. Chafai, Michael Kipp, Stefan Kopp, Maurizio Mancini, Stacy Marsella, Andrew N. Marshall, Catherine Pelachaud, Zsofi Ruttkay, Kristinn R. Thórisson, Herwin Van Welbergen et Rick J. Van Der Werf. *The behaviour markup language : recent developments and challenges.* Lecture notes in artificial intelligence, vol. 4722, pages 99–11, 2007.

[Wallbott 1998] Harald G. Wallbott. *Bodily expression of emotion.* European Journal of Social Psychology, vol. 28, no. 6, pages 879–896, 1998.

[Walton 1995] Douglas Walton et Eric Krabbe. Commitment in dialogue : Basic concept of interpersonal reasoning. State University of New York Press, 1995.

[Wooldridge 2009] M. Wooldridge. An introduction to multiagent systems. John Wiley & Sons, 2009.

[Zeelenberg 1998] Marcel Zeelenberg, Wilco W. van Dijk et Antony S. R. Manstead. *Reconsidering the Relation between Regret and Responsibility.* Organizational Behavior and Human Decision Processes, vol. 74, no. 3, pages 254–272, 1998.

Liste des tableaux

3.1 Les critères d'évaluation de Lazarus par niveau, évalués de façon complémentaire. 24
3.2 Les *checks* de Scherer par objectif d'évaluation, évalués en séquence. 25

4.1 Les émotions déclenchées dans EMA par le processus d'évaluation. 31

6.1 Les états mentaux et leur expression chez les ACA 46

7.1 Les cinq classes d'actes illocutoires 54
7.2 Les conditions de sincérité selon les classes d'actes. 57

8.1 Formalisation des émotions complexes pour un agent i suivant la situation φ. 61
8.2 Formalisation des actes expressifs exprimant les émotions *de base* 63
8.3 Formalisation des actes expressifs exprimant les émotions *complexes* 63

9.1 Actes expressifs suivant leur différent degré de puissance 66
9.2 Formalisation des conditions de sincérité des actes assertifs, engageants et directifs. 67
9.3 Préconditions et effets (lorsque l'agent accomplit l'acte, ou lorsque l'acte est accompli par l'humain) de quelques actes. 69
9.4 Bibliothèque complète d'Actes de Conversation. 70

10.1 Résultats du test de Tukey pour la variable dépendante concernant la sincérité de `S'excuser` : la différence entre la condition congruente (C) et les conditions neutre (N) et non-congruente (NC) est significative. 78

11.1 Quelques règles d'obligations du discours entre un locuteur (S1) et un interlocuteur (S2), adapté de [Traum 1994]. 89

12.1 Règles d'obligations basées sur les travaux de [Traum 1994]. . . . 100

13.1 Traduction des opérateurs de croyance, de but et d'idéal en Prolog. 108
13.2 Cinq des *checks* de Scherer adaptés au contexte du dialogue ACA - humain. 113
13.3 Traduction de cinq des *checks* de Scherer en Action Units (adapté de [Scherer 2007]). 116
13.4 Exemple d'annotation de reconnaissance d'une émotion et de sa sincérité et sa crédibilité. 122
13.5 Indicateur de justesse selon la reconnaissance et l'expression de l'émotion : Pas juste (PJ), Moyennement Juste (MJ) et Juste (J). 123

A.1 Version française de la bibliothèque d'Actes de Conversation Multimodaux. 147
A.2 Version anglaise de la bibliothèque d'Actes de Conversation Multimodaux (*Multimodal Conversational Acts*). 148
A.3 Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM assertifs**. 149

A.4 Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM engageants**. 150
A.5 Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM directifs**. 151
A.6 Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM expressifs** exprimant les émotions "de base". 152
A.7 Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM expressifs** de degré neutre. 153
A.8 Préconditions et effets (lorsque l'agent performe l'acte, ou lorsque l'acte est performé par l'humain) des **ACM expressifs** de degré supérieur. 154

Table des figures

1.1 De gauche à droite : agent pédagogique [Jaques 2009], agents conseillers de sites (*Ikea,Fnac*) ; en-dessous, avatars de Second-Life (*http ://secondlife.com/*). 3
1.2 L'Agent Conversationnel Animé (ACA) : un modèle de la communication humaine, verbale et non-verbale 4
1.3 De gauche à droite, les ACA Steve [Rickel 1997], Greta [Poggi 2005] et MARC [Courgeon 2008, Courgeon n C]. 5

2.1 Les deux dimensions de compétence (*good-intellectual / bad-intellectual*) et sociale (*good-social / bad-social*) introduites par [Rosenberg 1968]. La **sincérité** est une des caractéristiques positives de la dimension sociale. 10

3.1 Les six émotions "primaires" d'Ekman définies selon leurs expressions faciales. De gauche à droite, un exemple d'expression de la colère, de la peur, du dégoût, de la surprise, de la joie et de la tristesse. 20
3.2 Le modèle à deux dimensions de Russel [Russell 1980] : la dimension horizontale (valence) va de déplaisir à plaisir (de gauche à droite) et la dimension verticale (arousal) de basse activation vers haute activation (de bas en haut). 21
3.3 Comment se déclenche l'émotion de peur ? En haut, je vois l'ours, je cours (suite à une réaction automatique), *donc* j'ai peur. En bas, je vois l'ours, j'ai peur, *donc* je cours. 22
3.4 Cycle de perception, appraisal et réaction émotionnelle et physiologique. Je vois l'ours - j'évalue la situation (significativité de l'événement, conséquences sur mes buts...), consciemment ou non - je réagis en prenant la fuite et en ayant peur. 23
3.5 Les trois types d'évaluation du modèle OCC. 26

4.1 Le système des FACS : les différentes Unités d'Actions (AU) représentant les mouvements musculaires du visage (personnage Mary de MARC [Courgeon 2008]). 28
4.2 Quelques-unes des actions et phrases "communes" disponibles dans Gmail Motion (*poisson d'avril* 2011 de Google). Ces types de gestes peuvent être considérés comme de "faux emblèmes", en ce sens où ils veulent apparaitre comme intuitifs, innés et partagés culturellement par tous. 29
4.3 MARC (à l'état neutre (0)) exprime, de gauche à droite, un événement inattendu (1), désagréable (2), qui entrave ses buts (3) et sur lequel MARC a peu de contrôle (4). L'expression finale (5) est l'émotion de tristesse telle que définie par Ekman [Ekman 1992]. 30
4.4 La scène de vente jouées par les agents "acteurs" (Ritchie et Tina) dans CrossTalk. 33
4.5 Les états émotionnels et non-émotionnels de *Max* situés dans l'espace PAD. 34

5.1 Le standard SAIBA 38
5.2 Encodage de l'acte nier en FML-APML. 39
5.3 La traduction de l'acte nier (*deny*) en signaux multimodaux dans le fichier lexique. 39
5.4 Expression faciale de l'acte performatif suggérer. Adaptée de [Poggi 1998]. 40
5.5 Les quatre personnalités de l'ACA Greta. De gauche à droite : Spike, Poppy, Obadiah et Prudence. 41
5.6 Interaction entre l'agent John et l'utilisateur. 43
5.7 Annotation automatique de l'énoncé par le système BEAT. . . . 44

7.1 Tableau sémantique pour les verbes illocutoires d'usages expressifs [Vanderveken 1988] 55

10.1 Le standard SAIBA. 73
10.2 Prudence est la personnalité de Greta pragmatique et pratique. . 74
10.3 Scores qu'ont obtenu les ACM `S'excuser` et `Se réjouir` en sincérité et en crédibilité. 77

11.1 Cycle de comportement d'un agent BDI. 86
11.2 Planification par chaînage arrière. Figure tirée de [Cohen 1979]. . 88

12.1 Schéma de principe du moteur de raisonnement. 91
12.2 Le moteur de raisonnement dans l'architecture cognitive des théories de l'évaluation. Adaptée de [Sloman 2001, Scherer 2001]. 92
12.3 Architecture du moteur de raisonnement. 93
12.4 Le premier sous-module de perception/interprétation. 94
12.5 Parties statiques et dynamiques de la base de connaissances. . . . 95
12.6 Evaluation : déclenchement des émotions complexes. 97
12.7 Délibération : processus de sélection des trois types d'intentions communicatives. 98
12.8 Planification : construction d'un plan pour atteindre l'intention . 101

13.1 Architecture logicielle du système 109
13.2 L'interface utilisateur lors de l'interaction avec l'ACA (à droite) ; l'interface est composée d'un résumé du dialogue (partie supérieure) et d'une zone de texte et d'un bouton "micro" pour la reconnaissance vocale (partie inférieure). 110
13.3 Traduction en FML de l'intention communicative de Greta. . . . 111
13.4 Les trois agents de MARC : de gauche à droite, Maxime, Mary et Simon. 112
13.5 Intégration de l'évaluation des *checks* dans le moteur de raisonnement. 115
13.6 Traduction des *checks* en AU, qui sont transmises à MARC avec l'énoncé de l'ACM. 117
13.7 Exemple de combinaison des expressions faciales après évaluation des *checks*. L'expression finale reste tandis que MARC fait un `reproche` à l'humain. 117
13.8 Scénario de réconciliation entre MARC et l'utilisateur. 119
13.9 Interface pour l'évaluation : le sujet a le choix entre les trois possibilités. 121

13.10 L'agent Mary évaluée à RobotFesta ; la communication avec Mary se fait *via* la tablette tactile. 123
13.11 Résultats de la justesse des jugements des sujets pour l'admiration et la culpabilité. 124
13.12 Résultats de la justesse des jugements des sujets pour la satisfaction. 124
13.13 Résultats de la justesse des jugements des sujets pour la déception et le reproche. 125
13.14 Scores de sincérité et de crédibilité de la satisfaction. 126
13.15 Scores de sincérité et de crédibilité de la culpabilité. 126
13.16 Scores de sincérité et de crédibilité de la déception et du reproche. 127
13.17 Autres émotions reconnues suivant les différents sous-scénarios. . 128
13.18 Réponses à la question : « D'après vous, la réaction de Mary est-elle appropriée ? » 129
13.19 Analyse de l'influence de l'ordre des sous-scénarios dans la reconnaissance des émotions de culpabilité et de déception. 130

15.1 Les robots compagnons reeti et nao. 143

B.1 Exemple de plan : comment organiser des vacances 156
B.2 Grille d'entretien 158
B.3 Résultats de la justesse des jugements des sujets pour la satisfaction. 158
B.4 Résultats de la justesse des jugements des sujets pour l'admiration et la culpabilité. 159
B.5 Résultats de la justesse des jugements des sujets pour la déception et le reproche. 160
B.6 Résultats de la justesse des jugements des sujets pour le regret et la gratitude. 161
B.7 Scores de sincérité et de crédibilité de la satisfaction et de la gratitude. 162
B.8 Scores de sincérité et de crédibilité du regret et de l'admiration. . 162
B.9 Scores de sincérité et de crédibilité de la déception et du reproche. 163
B.10 Scores de sincérité et de crédibilité de la culpabilité. 163
B.11 Autres émotions reconnues suivant les différents sous-scénarios. . 164
B.12 Réponses à la question : « D'après vous, la réaction de Mary est-elle appropriée ? » 164