



HAL
open science

Dialogue entre agents naturels et agents artificiels. Une application aux communautés virtuelles

Guillaume Chicoisne

► **To cite this version:**

Guillaume Chicoisne. Dialogue entre agents naturels et agents artificiels. Une application aux communautés virtuelles. Interface homme-machine [cs.HC]. Institut National Polytechnique de Grenoble - INPG, 2002. Français. NNT: . tel-00004385

HAL Id: tel-00004385

<https://theses.hal.science/tel-00004385>

Submitted on 29 Jan 2004

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.

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

n° attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--

THÈSE

pour obtenir le grade de

DOCTEUR DE L'INPG

Spécialité : "Informatique : Systèmes et Communications"

préparée au laboratoire LEIBNIZ-IMAG

dans le cadre de l'École Doctorale
"Mathématiques, Sciences et Technologies de l'Information"

présentée et soutenue publiquement

par

Guillaume CHICOISNE

À la date du : 11 Décembre 2002

**DIALOGUE ENTRE AGENTS NATURELS ET
AGENTS ARTIFICIELS**

UNE APPLICATION AUX COMMUNAUTÉS VIRTUELLES

Directrice de thèse :
Madame Sylvie PESTY

JURY

Monsieur Nicolas BALACHEFF	Examineur
Monsieur Jean CAELEN	Examineur
Monsieur Alexis DROGOUL	Rapporteur
Monsieur Stéphane MAGUET	Examineur
Madame Sylvie PESTY	Directrice de thèse
Monsieur Jean-Paul SANSONNET	Rapporteur

*Tout n'est que brouillon en effet,
l'idée de texte définitif ne relevant
que de la religion ou de la fatigue.*
Jorge-Luis BORGES

À ceux qui m'ont supporté¹ pendant ma (trop) longue rédaction. . .

¹SENS 1 : Admettre, tolérer la présence, le comportement d'une personne ou d'un animal en dépit des inconvénients que cela peut comporter. SENS 2 : Aider activement, donner son soutien moral ou matériel à. Encourager, soutenir.

Une thèse, c'est une histoire à trois : un thésard, un sujet, un directeur. Ne parlons pas de moi. Ne parlons pas non plus du sujet, les 300 prochaines pages lui sont consacrées. L'essentiel de mes remerciements ira donc à Sylvie. Scientifiquement, professionnellement et humainement, elle a été *la* personne qu'il fallait.

Cela étant dit, cette thèse, tout comme ma vie, est une mosaïque. Et comment pourrait-il en être autrement, puisque ma vie et ma thèse furent quasiment confondues pendant ces quelques années ? Les influences extérieures étant tout simplement innombrables, je vais, à mon grand regret, ne remercier ici que quelques unes des personnes sans lesquelles ce travail ne serait pas ce qu'il est.

Mes rapporteurs, Jean-Paul et Alexis, pour leur sens critique et l'extrême pertinence de leurs remarques, même s'il m'aurait encore fallu quelques années pour vous répondre avec tous les détails que j'aurais souhaité... Les autres membres du jury : Jean Caelen, Nicolas Balacheff, et Stéphane l'insaisissable, tour à tour industriel, philosophe, webmestre, enseignant, artiste, ...

Yves, notre responsable d'équipe, qui travaille à nous offrir un environnement matériel, administratif et surtout humain particulièrement motivant.

L'équipe du Deuxième Monde, le Comptoir des Planètes, les fabuleux stagiaires Canal.

J'hésite finalement à remercier la communauté Sciences Cognitives de France, car ma vie ne suffira certainement pas à exploiter l'ensemble des idées qu'elle a éveillé en moi. Plus particulièrement, je remercie mes camarades co-organisateurs des colloques jeunes chercheurs. Promis, la prochaine fois, je ne présenterai pas la session en t-shirt à carreaux oranges et bretelles.

Kelly qui, malgré quatre ans à partager le même bureau que moi ne sait pas encore distinguer quand je plaisante et quand je suis sérieux. Carine, pour avoir préservé l'équilibre de notre directrice commune en étant, elle, rigoureuse, pragmatique, sûre d'elle et concentrée.

Quelques phrases de ce manuscrit ne comportent pas de fautes d'orthographe. Merci pour cela à Joëlle, ma relectrice principale, à Alan, Agnès, Yoann et Lætitia, mes relecteurs auxiliaires. Ne faites jamais corriger vos écrits par une psy : elle chercherait à comprendre "*l'origine de tes problèmes avec le féminin pluriel*".

Ceux qui m'ont aidé à élever le *squat* au rang d'art lors de mes déplacements ou déménagements. Mention spéciale aux Parisiens Fred et Alice et aux Grenoblois João, Thierry et Serban.

Les rencontres éphémères des conférences et tous ceux qui m'y ont donné leurs desserts.

Mes amis étrangers, qui m'ont fait le plaisir de passer ou de m'accueillir : Zara, Elina, Tero, Mari, Lucie.

Alice, Pierre-Michel et João, Alan et Thierry, Mathieu et Éric-Marie, Lucie. Ma vie serait « moins » si je ne vous connaissais pas, même si j'ai rencontré certains d'entre vous un peu tard, ou que d'autres sont partis un peu tôt. Un grand merci aussi aux conjointes indispensables à la vie des gens qui me sont indispensables : Sandra, Maayan, Agnès.

Mes frères et mes parents, présents, proches et disponibles, sans jamais être oppressants, comme toujours. Vous avez été parfaits.

Table des matières

Introduction	xi
1 Une problématique transdisciplinaire	xii
2 Des objectifs scientifiques et techniques.	xiv
3 Structure du document	xv
I Communautés d’agents naturels, d’agents artificiels, et communautés mixtes	1
1 Communautés humaines en ligne	7
1.1 Le côté “ <i>cyber</i> ”	8
1.2 Les communautés virtuelles	9
1.2.1 Un environnement où interagir	10
1.2.2 Des interactions entre les membres de la communauté	13
1.2.3 Des participants	15
1.2.4 Quelques exemples de communautés virtuelles	17
1.2.5 Le Deuxième Monde	19
1.3 La communication médiatisée	25
1.3.1 Spécificités de la communication médiatisée	26
1.3.2 La netiquette	28
1.3.3 Relations personnelles sans face-à-face	30
1.4 Conclusion	31
2 Systèmes multi-agents logiciels	33
2.1 L’Agent	33
2.1.1 Point de vue interne : ce qu’ <i>est</i> l’agent	34
2.1.2 Point de vue externe : ce que <i>fait</i> l’agent	35
2.2 De l’agent au collectif d’agents	36
2.2.1 L’Interaction	36
2.2.2 L’Organisation	38
2.3 Juxtaposition d’agents ou communauté ?	41
2.3.1 Comportement à base de buts explicites ou sans buts	41
2.3.2 Actions personnelles ou collectives	42
2.3.3 Rôle unique ou multiple, figé ou évoluant	43
2.4 Conclusion	43
3 Communautés mixtes	45
3.1 Intégration monde réel / monde virtuel	45
3.2 Qu’est-ce qu’une communauté mixte ?	48
3.3 Nécessités côté humain	51

3.4	Nécessités côté agent	52
3.4.1	De la cohérence	53
3.4.2	Une personnalité	53
3.4.3	Des capacités conversationnelles	55
3.5	Quelques agents participant à des groupes mixtes	57
3.5.1	Le projet OZ (CMU)	58
3.5.2	REA et projets connexes (MIT)	63
3.5.3	Autres projets	71
3.6	Conclusion	78

II Principes du dialogue et modèle d'agent conversationnel pour les communautés mixtes 79

4	Principes du dialogue	81
4.1	Les actes de langage	82
4.1.1	La notion de performatif	83
4.1.2	But illocutoire et taxonomie des actes de langage	84
4.1.3	Les composantes de la force illocutoire	85
4.1.4	Les conditions de succès et de satisfaction	86
4.2	La communication entre agents	88
4.2.1	L'ACL-FIPA	88
4.2.2	Les conduites de conversation (<i>conversation policy</i>)	90
4.2.3	Les langages d'agents et la langue naturelle	92
4.3	Un oubli majeur : le Sens	94
4.3.1	La représentation du sens	95
4.3.2	Le potentiel de sens	96
4.4	La place des interactants	97
4.4.1	Interprétation et expression	97
4.4.2	Du message au(x) sens et <i>vice versa</i>	105
4.4.3	Quels paramètres pour les interactants?	112
4.5	La dynamique de la communication	116
4.5.1	Co-construction de sens / négociation de sens	117
4.5.2	Enchaînements	120
4.5.3	Le focus conversationnel	122
4.5.4	Les attentes	124
4.5.5	Liens entre focus et attentes	132
4.5.6	Confrontations, ruptures, négociations	133
4.6	Conclusion	135
5	Modèle et architecture d'agent	137
5.1	L'agent conversationnel	138
5.1.1	Conscience sociale	138
5.1.2	Historique de l'interaction	139
5.1.3	Connaissances sur les autres	140
5.1.4	Paramètres d'individuation	141
5.1.5	Une interaction socialisée mais fonctionnelle	142
5.2	Intention conversationnelle et forme de surface	142
5.2.1	Potentiel de sens et message	143
5.2.2	Enrichissement contextuel	148
5.3	Enchaînements	149

5.3.1	Focus conversationnel	150
5.3.2	Prise en compte d'attentes	150
5.4	Une architecture possible	150
5.4.1	Fonctionnement d'ensemble	151
5.4.2	Les modules et leurs traitements	155
5.4.3	Les données partagées	158
5.5	Conclusion	161
III Application		163
6	Application	165
6.1	Spécifications	166
6.2	Les Données	167
6.2.1	Actes de communication	168
6.2.2	Focus conversationnel	172
6.2.3	Attentes	173
6.2.4	Historique	175
6.2.5	Personnalité propre	175
6.2.6	Connaissances sur les autres	176
6.3	Modules liés à la perception et à l'action	177
6.3.1	Formalisation	178
6.3.2	Intégration multi-modale	183
6.3.3	Interprétation	184
6.3.4	Expression	188
6.3.5	Répartition multi-modale	189
6.3.6	Énonciation	190
6.4	Modules liés au raisonnement	193
6.4.1	Architecture générale	193
6.4.2	Sélection	194
6.4.3	Cinéma	197
6.4.4	Conversation	203
6.4.5	Chatbot	205
6.5	Analyses	208
6.5.1	Dialogue 1	210
6.5.2	Dialogue 2	214
6.5.3	Dialogue 3	216
6.5.4	Dialogue 4	221
6.5.5	Dialogue 5	224
6.6	Considérations techniques	228
Conclusions		233
1	Limitations	233
1.1	Limitations d'ordre technologique	233
1.2	Limitations d'ordre théorique	234
2	Extensions	236
3	En conclusion : de l'intérêt de la pluridisciplinarité	237
Références		240
A Documents électroniques		1

Introduction

*“Begin at the beginning,” the King said
gravely, “and go on till you come to the
end; then stop.”*

LEWIS CARROLL, *Alice’s Adventures
in Wonderland*

Tout le monde parle à son ordinateur. . . Pour l’insulter, l’encourager, l’implorer, critiquer son fonctionnement. . .

Des études faites en psychologie sociale révèlent des comportements entre humains et ordinateurs qui s’expliquent par l’attribution par ces premiers d’une sensibilité sociale, d’un amour propre, d’un *ego*, etc., à ces derniers [Nass *et al.*, 1994]. Des sondages récents montrent que les agressions physiques envers les ordinateurs sont pratique courante². D’autres études montrent que des programmes de dialogue ayant une représentation anthropomorphique sont considérés comme plus *compétents* que des humains alors que l’interaction est strictement identique mais que change le type ethnique de l’humain ou de la représentation personnifiée du programme [Nass *et al.*, 2000]. Les juges du concours Loebner³, une version adaptée du test de Turing, commencent à attribuer à des ordinateurs des notes “d’humanité” supérieures à celles de certains humains. On pourrait ainsi accumuler pendant longtemps les exemples, tant “grand public” que scientifiques : l’être humain voit en l’ordinateur plus qu’un simple objet.

L’arrivée des agents, entités autonomes pouvant facilement être considérées dans leur *individualité*, et, qui plus est, pouvant maintenant disposer de capacités de reconnaissance et de synthèse de la parole, ainsi que de représentations anthropomorphes, ne peuvent qu’ajouter à la confusion des humains en interaction avec de tels systèmes. Le domaine des systèmes multi-agents (SMA), dans lequel s’inscrit cette thèse, est jeune mais possède une histoire déjà riche à travers ses domaines environnants, qui comprennent l’informatique — intelligence artificielle et systèmes répartis — bien sûr, mais aussi l’éthologie, la psychologie, la sociologie ou encore la physique. Cependant, beaucoup des mécanismes d’interaction employés entre agents artificiels (phéromones, champs de forces, messages formels, échange direct de connaissances, etc.) sont inadaptés à l’humain, et les spécificités de l’humain commencent à peine à être prises en compte dans le domaine des SMA. Il y a deux principales raisons à ce récent intérêt. Premièrement, les situations d’interaction agent/humain se sont multipliées avec l’arrivée des agents dans les applications de commerce électronique, de propositions de services, d’*e-learning*, etc. Deuxièmement, le domaine des SMA

²Parmi les 4000 personnes ayant répondu à un sondage réalisé en 2001 par la société Novatech, fabricants de matériel informatique, 25% admettent avoir déjà agressé *physiquement* leur ordinateur.

³<http://www.loebner.net/Prizef/loebner-prize.html>

a pris l'humain comme objet d'étude à travers des applications comme les migrations urbaines, les simulations d'écosystèmes, etc. Parfois même, l'humain est tout simplement considéré comme un agent et intégré avec des agents artificiels au sein de systèmes multi-agents parfois qualifiés d'*hétérogènes*, mais pour lesquels nous retiendrons le qualificatif de *mixtes*. Ces nouvelles situations ont amené le domaine des SMA à approfondir ses racines dans les sciences humaines, s'intéressant aux comportements sociaux ou aux interactions chez les humains.

La thèse que nous défendons s'inscrit dans l'objectif de constitution de communautés mixtes comprenant des agents artificiels et des humains. Ces communautés se veulent mettre en relation ces entités aux compétences complémentaires dont la mise en commun engendrera l'apparition au niveau de la communauté de capacités nouvelles, indisponibles (ou difficilement disponibles) si la communauté ne se composait que d'agents artificiels ou que d'humains. Une telle situation nécessite, comme le dit Licklider, une forme de *symbiose* [Licklider, 1960] entre humains et agents artificiels, symbiose qui dépend de l'existence de mécanismes de communication adaptés aux deux catégories d'agents.

Nous traiterons donc dans ce manuscrit du dialogue entre humains et agents artificiels. Premièrement, nous compléterons certains des mécanismes classiques d'interaction entre agents artificiels par la notion de *contexte*, de façon à pouvoir introduire les concepts d'*interprétation* et d'*expression* d'un message. Deuxièmement, nous présenterons une façon de structurer la dynamique d'une interaction, structuration basée sur la négociation collective du sujet de discussion baptisé *focus conversationnel*.

1 Une problématique transdisciplinaire

Notre sujet de recherche est par essence trans- ou du moins pluri-disciplinaire : les théories sur lesquelles nous nous appuyons sont issues d'horizons différents, et plusieurs grands champs disciplinaires sont pertinents dans le cadre de nos travaux. Si nous nous contentons de mettre en regard ces différents champs disciplinaires composant un *patchwork*, une mosaïque, où nous emprunterons à chaque discipline les points qui nous semblent importants, nous travaillerons dans la pluridisciplinarité. Si nous réussissons à fondre ces disciplines et à en obtenir plus que leur simple juxtaposition — et cet objectif est bien plus délicat à atteindre — alors nous pourrions parler de transdisciplinarité, une forme de discipline émergente.

Intelligence Artificielle & Psychologie Deux types d'agents nous intéressent : les agents naturels et les agents artificiels. À ces types d'agent s'attachent deux domaines proposant des modèles de comportement : la psychologie et l'Intelligence Artificielle (IA). Si cette assertion peut sembler évidente dans le cas de la psychologie, elle l'est moins pour l'IA. Nous considérons l'IA comme ce qu'elle prétend être par son nom même : une tentative de reproduction artificielle de certaines formes d'intelligence.

Une des différences majeure des travaux dans ces deux domaines reste l'*emploi* des modèles. En psychologie, les modèles sont généralement *explicatifs*, ils servent de guide à l'*analyse* d'un comportement⁴. En IA, les modèles sont généralement *génératifs*, ils servent à *produire* un comportement. En effet, pour la psychologie, l'objet d'étude (l'humain) préexiste à la discipline et est *étudié*, alors que pour l'IA,

⁴Ces généralités valent essentiellement pour la psychologie expérimentale. En psychologie clinique par exemple, les thérapies visent à *modifier* le comportement du sujet.

c'est la discipline qui préexiste à son objet d'étude (le programme, l'algorithme, etc.), qui est *créé*.

Malgré tout, des liens forts existent entre ces deux disciplines. S'il n'est pas facile de modifier directement les modèles internes donnant naissance au comportement humain, il est en revanche possible de les simuler, pour peu que ces modèles soient suffisamment formels. Et finalement, la simulation d'une entité dont le modèle interne est basé sur des analyses du comportement humain ne relève-t-il pas, mot pour mot, de l'intelligence artificielle ? Ainsi, les chercheurs de ces deux domaines tendent à estomper quelques-unes de leurs différences : certaines théories psychologiques sont de plus en plus formalisées, associant variables et fonctions aux comportements humains et usant de simulation, tandis que dans certaines branches de l'IA, on voit apparaître des données ou processus baptisés croyances, intentions, émotions. . .

Notre objectif ayant trait aux communautés mixtes où les agents naturels interagissent avec les agents artificiels, il nous est indispensable de nous intéresser aux deux domaines qui traitent du comportement de ces deux types d'entités, dans l'objectif d'en synthétiser un modèle qui puisse à la fois être explicatif et servir de base à des prévisions pour l'être humain et générateur de comportements pour les agents artificiels.

Théorie de l'information et Interaction Homme-Machine & Psychologie de l'interaction et Linguistique La théorie de l'information apporte un cadre formel aux communications entre systèmes automatisés et le domaine de l'Interaction Homme Machine (IHM) dispose de théories portant sur la conception, la réalisation et l'évaluation de systèmes informatisés avec lesquels interagissent des êtres humains. La psychologie de l'interaction dispose, elle, de théories portant sur le comportement des humains entre eux, et la linguistique sur les messages échangés. Les liens entre ces quatre domaines sont donc particulièrement forts, et il est évident que nos travaux ne peuvent que tirer bénéfice de leur considération.

Systèmes Multi-Agents & Psychologie sociale Le comportement d'un individu dépend beaucoup du groupe dans lequel il interagit (ainsi, à travers sa métaphore théâtrale, Erwing Goffman prétend qu'à chaque interaction les individus en présence endossent un rôle qui dépend de leurs spectateurs [Goffman, 1973]), mais ce comportement dépend aussi des groupes avec lesquels il a interagi par le passé (et qui conditionne son éducation ou sa culture). La psychologie sociale traite de ces comportements de l'individu "en groupe" mais aussi des comportements du groupe lui-même⁵. Le domaine des systèmes multi-agents artificiels, de son côté, s'intéresse à la mise en commun, en relation, de différentes entités, et aux conséquences d'une telle mise en commun.

Les relations entre systèmes multi-agents et psychologie sociale sont proches de celles dont nous avons parlé entre intelligence artificielle et psychologie. Il serait d'ailleurs possible — dans une certaine mesure — de considérer que l'objet d'étude de ces deux premiers domaines est la mise en relation des objets d'études respectifs de ces deux derniers domaines. Nous avons précédemment fait remarquer que, travaillant sur les communautés mixtes, l'étude du comportement *individuel* des entités

⁵La psychologie sociale se veut combler un creux entre la psychologie et la sociologie. Elle traite des petits groupes, où l'individu est encore identifiable. La psychologie y a donc une place importante, car l'individu influe fortement sur le comportement du groupe et peut en être fortement influencé. De même, le domaine dépend de la sociologie qui s'attache à l'étude des phénomènes collectifs, mais à une échelle à laquelle l'individu est "gommé".

composant ces communautés s'imposait, et nous arrivons ici à la même conclusion en ce qui concerne le comportement *collectif* de ces entités.

2 Des objectifs scientifiques et techniques.

Nos objectifs ne se limitent pas à la détermination des éléments importants au fonctionnement d'une communauté mixte. Bien que nous nous restreignions à ne traiter que de certaines parties des phénomènes d'interaction, il est important que ces travaux, théoriques, débouchent sur des réalisations nous permettant de valider nos modèles par la pratique. Ce désir, qui correspond à l'origine à un choix personnel (l'idée que les résultats d'une recherche sont encore plus intéressants s'ils sont mis en pratique), a naturellement trouvé sa place dans le contexte particulier d'un partenariat entre un laboratoire de recherche (Leibniz-IMAG) et une entreprise (Canal+). L'objectif scientifique de concevoir un modèle de "conversation" pour des agents logiciels correspond à un besoin de la part du partenaire industriel, et ces mêmes agents logiciels, une fois réalisés, permettent en retour au partenaire scientifique d'affiner ses modèles. Cette thèse a en grande partie bénéficié de ce "cercle vertueux".

Le travail de recherche s'intègre dans le projet "Deuxième Monde" de la société Canal+. Le Deuxième Monde est une communauté virtuelle⁶ en accès libre sur Internet dont l'environnement reproduit en trois dimensions la ville de Paris. Dans cette communauté virtuelle évoluent des avatars pilotés par les internautes. La principale activité est la discussion en direct de type *chat*⁷ entre les internautes. Des animations sont régulièrement organisées par les équipes de Canal+ pour donner vie à cette communauté et des boutiques virtuelles (en partenariat avec d'autres entreprises) sont accessibles à l'internaute via son avatar qu'il manipule comme une marionnette et déplace dans les différents quartiers de Paris. Une grande place est réservée à l'animation de ce Paris virtuel par les internautes eux-mêmes qui peuvent créer des clubs ou des commissions pour gérer la vie et l'évolution de la ville, bâtir une extension d'un quartier ou construire un immeuble par exemple.

À partir des travaux conduits par Canal+ dans le cadre du Deuxième Monde, un premier prototype d'environnement 3D favorisant le commerce électronique a été réalisé. Autour du concept d'une boutique spécialisée dans les contenus culturels (livres, musique, films, ...), Canal+ a mis en évidence les potentiels nouveaux apportés au commerce en ligne à partir de l'utilisation des techniques de modélisation 3D, de la communauté virtuelle et des outils de recommandation tels que celui mis en place au niveau du cinéma sous le nom de Mes Meilleurs Amis (MMA).

Plusieurs applications ont été proposées, discutées, et ont entraîné des réalisations à différents stades (agents majordomes, animateurs de jeu ou de clubs, représentants, etc.). Mais c'est le contexte de la boutique virtuelle qui se prêtait le mieux à l'expérimentation de nos travaux, et le prototype d'agent sur lequel nous sommes focalisé est un agent destiné à assister les clients lors de leurs visites sur des sites de commerce en ligne. Cette assistance peut prendre la forme d'aide à la recherche d'un produit particulier ou de recommandation d'un produit (en se basant sur l'outil MMA), le tout en employant une interface de type conversationnelle et non pas un moteur de recherche à base de mots clés comme cela est souvent le cas

⁶fermée depuis le printemps 2002.

⁷"*Activité permettant à un internaute d'avoir une conversation écrite, interactive et en temps réel avec d'autres internautes, par clavier interposé.*" (Office de la Langue Française du Québec)

sur les sites de commerce en ligne.

3 Structure du document

Trois parties composent ce document. La première partie, *communautés d'agents*, présente le contexte général qui a donné naissance à plusieurs de nos réflexions sur la communication, les dynamiques de groupes et les particularités respectives des agents artificiels et des agents naturels. Cette partie se compose de trois chapitres traitant respectivement des communautés d'humains, des communautés d'agents artificiels et enfin des communautés mixtes, composées d'agents artificiels et d'agents naturels. La deuxième partie, qui traite des *modèles*, présente nos propositions pour l'élaboration d'agents interagissant avec l'humain. Cette partie se compose de deux chapitres traitant respectivement des principes de dialogue et d'un modèle d'agent capable d'interagir suivant ces principes de dialogue. La troisième partie se compose d'un chapitre qui décrit notre application, un agent conversationnel dédié à la recherche et la recommandation de films, et en analyse les résultats.

Chapitre 1 : Communautés humaines en ligne Dans ce chapitre, nous nous intéressons aux communautés virtuelles. Nous précisons ce que nous entendons par "communauté virtuelle", dont nous présentons quelques exemples, et nous mettons en évidence les mécanismes de communication mis en œuvre par les humains en situation de communication médiatisée. Les comparaisons faites entre communautés virtuelles et communautés "réelles" lors de ce chapitre (et, plus généralement, la comparaison entre les études faites sur des humains en situation de face-à-face et des humains en situation médiatisée), nous amèneront à dire que les théories issues des sciences humaines sur l'interaction humaine et les comportements de groupe restent applicables aux relations médiatisées.

Chapitre 2 : Systèmes multi-agents logiciels Dans ce chapitre, nous nous intéressons aux systèmes multi-agents. Nous y présentons les agents, dont la notion apparaît tout au long de cette thèse, en insistant particulièrement sur la différence entre ce qu'est l'agent (son modèle interne) et ce qu'il fait (son comportement). La même distinction sera faite dans la suite de ce chapitre, lors de la présentation des effets de groupe (la fonctionnalité du système), et cette distinction nous permettra de caractériser les situations où le système présente à un observateur un comportement global émergent.

Chapitre 3 : Communautés mixtes Dans ce chapitre, nous nous intéressons aux communautés mixtes, communautés mettant en relation des agents naturels (humains) et des agents artificiels. Nous y rassemblons les nécessités du côté des humains et du côté des agents pour que leur mise en relation soit la plus profitable possible. Nous y présentons aussi à titre d'exemple les principaux projets relatifs aux agents participant à des groupes mixtes.

Chapitre 4 : Principes du dialogue Dans ce chapitre, nous confrontons, pour les intégrer, les remarques faites dans le chapitre précédent aux théories de l'interaction existant à la fois en sciences humaines et au sein des systèmes informatiques. Nous rappelons (après l'avoir dit dans le chapitre 1) l'importance des interactants,

et plus globalement, du contexte d'énonciation d'un message, là où les systèmes informatiques portent plutôt l'emphase sur le message lui-même. De ce rappel naît l'un des deux points centraux de cette thèse : l'*interprétation* et l'*expression*. Le deuxième point central de cette thèse est une méthode de structuration de la dynamique de la conversation basée sur une construction et une négociation collective d'un objet commun aux interactants : le *focus conversationnel*.

Chapitre 5 : Modèle et architecture d'agent Dans ce chapitre, nous définissons les fonctionnalités que doit posséder un agent que nous souhaiterions voir interagir suivant les principes présentés dans le chapitre précédent. La première partie de ce chapitre ne définit que les *éléments* d'un modèle d'agent nécessaires à suivre nos principes de dialogue. La seconde présente une architecture *complète* d'agent. Nous justifions dans ce chapitre les concessions que nous avons dû faire pour passer d'un modèle d'interaction à un modèle d'agent, d'un modèle essentiellement descriptif à un modèle essentiellement fonctionnel, d'un modèle théorique, abstrait, à un modèle informatique implémentable.

Chapitre 6 : Application Dans ce chapitre, nous détaillons les différents éléments de l'agent que nous avons réalisé, nous en présentons les particularités technologiques, et nous analysons plusieurs des dialogues obtenus lors de ses interactions avec un humain. Nous justifions aussi tout au long de ce chapitre les concessions que nous avons dû faire pour passer d'un modèle d'agent à une implémentation d'agent, d'un modèle informatique encore abstrait à un agent entièrement fonctionnel, d'une spécification à un agent implémenté.

Ainsi, de chapitre en chapitre, le sujet abordé se précise, partant du contexte vaste des communautés, pour proposer ensuite un modèle d'interaction aux considérations plus théoriques que pratiques. De ce modèle d'interaction assez abstrait, nous tirons un modèle, puis une architecture d'agent susceptible d'interagir suivant les principes formulés au chapitre précédent. Enfin, dans notre dernière étape de focalisation, nous présentons l'agent tel qu'il a été réalisé, prêt à rejoindre une communauté, rebouclant ainsi sur les présentations de la première partie.

Première partie

Communautés d'agents
naturels, d'agents artificiels, et
communautés mixtes

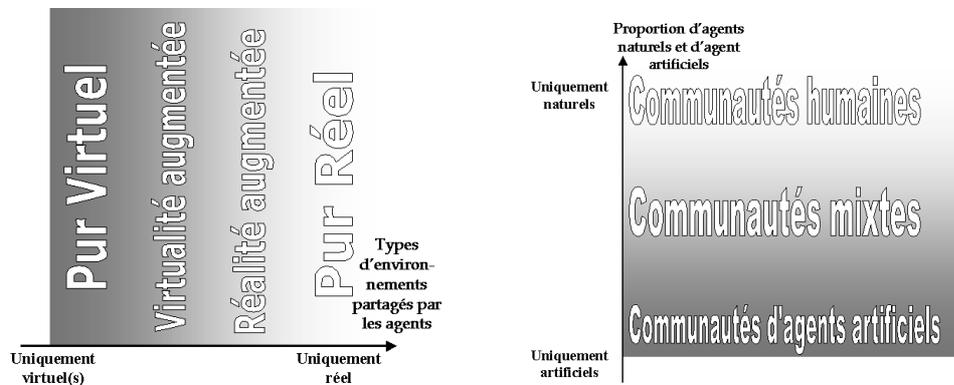
Avant d’ouvrir cette partie sur les *communautés d’agents* (quels que soient ces agents : artificiels ou naturels ; et que les communautés soient composées d’un seul type d’agent ou des deux) il nous faut préciser certains points de vocabulaire quant aux différents types de communautés dont nous allons être amenés à parler. En effet, d’un pur point de vue syntaxique il existe des ambiguïtés embarrassantes :

- lorsque le terme de *communauté virtuelle* est employé, l’adjectif “virtuel” s’applique à un espace : l’un des espaces à travers lequel interagissent les membres de la communauté est un espace virtuel, ou *cyberespace* ;
- lorsque nous employons le terme *communauté mixte*, l’adjectif “mixte” s’applique à des agents : les *membres* de cette communauté sont des agents humains et des agents artificiels ;
- lorsque le terme de *communauté épistolaire* est employé, l’adjectif “épistolai- re” s’applique à des moyens d’interaction : les membres de la communauté communiquent en s’expédiant des lettres ;
- etc.

Il faut donc prendre garde à ne pas imaginer comme exclusifs les uns des autres des types de communautés qui ne le sont pas.

Nous avons donc décidé de caractériser les communautés suivant deux critères :

1. quels sont les types d’environnement, virtuel ou réel, partagés par les membres de la communauté ?
2. quels sont les types d’agent, naturel ou artificiel, auxquels appartiennent les membres de la communauté ?



Le continuum monde virtuel ↔ monde réel

Le continuum communautés humaines ↔ communautés d’agents artificiels

FIG. 1 – Organisation des communautés virtuelles (1)

Ces deux critères nous permettent de définir un continuum entre, d’une part, des environnements purement virtuels et le monde réel, et, d’autre part, les communautés humaines et les communautés d’agents artificiels. A mi-chemin des communautés humaines et des communautés d’agents artificiels se trouvent les communautés mixtes, sujet du chapitre 3.

La combinaison de ces deux critères permet de définir quatre grandes catégories de communautés (voir figure 2) :

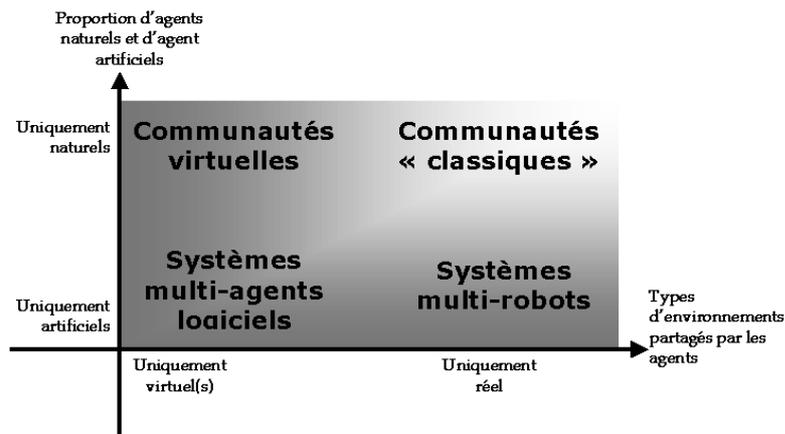


FIG. 2 – Organisation des communautés virtuelles (2)

- les *communautés "classiques"*, dont les membres sont des humains interagissant dans le monde réel.
- les *communautés virtuelles*, dont les membres sont des humains interagissant dans un monde virtuel. Ces communautés sont le sujet du chapitre 1.
- les *systèmes multi-agents logiciels*, dont les membres sont des agents artificiels interagissant dans un monde virtuel. Ces communautés sont le sujet du chapitre 2.
- les *systèmes multi-robots*, dont les membres sont des agents artificiels interagissant dans le monde réel.



FIG. 3 – Exemples de communautés

La figure 3 superpose à la figure 2 des exemples existants. On trouve sur cette figure :

-
- Usenet** Usenet, présenté page 17, est un exemple de communauté humaine interagissant dans un environnement purement virtuel, c'est-à-dire une communauté virtuelle ;
- Le Deuxième Monde** Le Deuxième Monde, présenté page 19, est un exemple de communauté humaine disposant de certaines ouvertures sur le monde réel (vidéos, sons, ...);
- Magma** L'équipe Magma, équipe de recherche sur les systèmes multi-agents à laquelle nous appartenons, est un exemple de communauté humaine interagissant dans un environnement extrêmement réel (l'essentiel des interactions ont en effet lieu en face-à-face), c'est-à-dire une communauté classique ;
- Freewalk** Freewalk [Nakanishi *et al.*, 1998], est un exemple de communauté essentiellement humaine interagissant dans un environnement virtuel augmenté ;
- Steve** Steve, présenté page 73, est un exemple d'agent s'intégrant dans une communauté d'humains interagissant dans un environnement virtuel ;
- Comris** Comris [Van de Velde, 1997], est un exemple de communauté d'agents artificiels et humains, interagissant dans l'environnement réel, mais dans laquelle les membres artificiels disposent aussi d'une représentation d'un espace social, donc virtuel ;
- MICRobES** MICRobES [Drogoul et Picault, 1999], propose l'étude de l'intégration de robots dans une communauté d'humains interagissant dans un environnement réel ;
- OZ** Oz, présenté page 58, est un exemple de communauté essentiellement humaine interagissant dans un environnement virtuel augmenté ;
- SMAM** Les SMAMs, Systèmes Multi-Agents Minimaux [Van Aeken, 1999], sont un exemple de communauté d'agents interagissant dans un environnement virtuel, c'est-à-dire un système multi-agent logiciel ;
- Martha** Martha [Alami *et al.*, 1998], est un exemple de communauté d'agents artificiels interagissant dans un environnement réel, disposant d'environnements virtuels pour leurs coordinations, et représentés simultanément dans un environnement virtuel ;
- Robocup** La Robocup⁸, est un exemple de communauté d'agents artificiels matériels (robots) interagissant dans un environnement réel, c'est-à-dire un système multi-robots ;

⁸<http://www.robocup.org>

Chapitre 1

Communautés humaines en ligne

L'ensemble des croyances et des sentiments communs à la moyenne des membres d'une même société forme un système déterminé qui a sa vie propre ;
ÉMILE DURKHEIM

Le sujet des communautés humaines en ligne (souvent appelées *communautés virtuelles*), est couramment associé à tout un imaginaire issu de la science-fiction, rassemblant pêle-mêle les notions d'interfaces immersives, de réalité virtuelle, d'ordinateurs intelligents, ... Même si la réalité est encore loin d'atteindre ce que les différents auteurs de science-fiction peuvent nous avoir déjà présenté, on ne peut nier un certain lien entre ces fictions et la réalité du domaine, qui s'en est nourri à ses débuts, comme nous le verrons dans la section 1.1. Mais de la vision des communautés virtuelles donnée par la science-fiction est principalement gardé le côté *virtuel*, et non le côté *communauté*, dégradation qui reviendrait, dans un tout autre domaine, à ne retenir de l'Académie Française (en tant qu'institution, avec ses membres, ses activités, ...) que les bâtiments¹ ! Ainsi, la section 1.2 présentera les communautés virtuelles d'une manière plus profonde, et surtout, plus ancrée dans leur "réalité". Enfin, puisque nous nous intéressons ici aux communautés virtuelles *en ligne*, nous nous pencherons plus spécifiquement dans la section 1.3 sur les particularités apportées par l'emploi d'une machine pour médier la communication entre les différents participants.

Nous notons que les communautés virtuelles présentent des avantages importants pour notre sujet de recherche. Particulièrement :

- elles sont des reproductions de communautés sociales réelles², et donc se prêtent aisément à l'application de théories issues des sciences humaines comme celles que nous avons décidé d'employer ;

¹"A distinction will need to be made between the cyber-place within which a virtual community operates, which will be termed a virtual settlement, and the virtual communities themselves." [Jones, 1997]

²"Les communautés virtuelles sont de véritables communautés sociales et les individus qui les composent tissent des liens affectifs, échangent des informations, collaborent ensemble." [Quéau, 1993] ; "Cyberspace is simply another place to meet." [Parks et Floyd, 1996] ; etc.

- elles sont supportées par un environnement informatique, ce qui nous offre un environnement où les actions et leurs effets sont clairement mesurables, enregistrables, manipulables.

1.1 Le côté “*cyber*”

Il est difficile d’aborder le thème des communautés virtuelles en faisant abstraction des productions du mouvement littéraire *cyberpunk*, chacun s’étant en partie nourri de l’autre. Par exemple la société Blaxxun³, tire son nom directement de Black Sun, un “bar” virtuel, central à l’histoire du “Samouraï Virtuel”, de Neal Stephenson. De la même façon, beaucoup du vocabulaire aujourd’hui employé dans le domaine des communautés virtuelles, ou, plus généralement de la réalité virtuelle, est issu de ces romans ou nouvelles (*cyberspace*, *multivers*, etc.).

Sur le terme même, *cyberpunk*, Bruce Sterling (un des auteurs marquants du mouvement) dans sa préface à une anthologie du cyberpunk [Sterling, 1987] lui “reconnait une certaine pertinence. Il saisit un élément essentiel de l’œuvre de ces auteurs [...] : une nouvelle forme d’intégration. L’intégration d’univers auparavant dissociés : le royaume de la technologie de pointe et les aspects modernes de l’underground pop.” (traduction Française issue des éditions Denoël.) Parmi les divers traits caractéristiques qui forment les sujets abordés par les auteurs du mouvement cyberpunk, le plus typique est sans conteste l’impact de la technologie sur la vie personnelle et culturelle des citoyens. L’arrivée du cyberpunk dans le media de masse qu’est le cinéma (ne citons que le récent “Matrix” et le fondateur “Blade Runner”), couplée au développement extrêmement rapide de l’Internet firent connaître certaines branches de ce mouvement à une grande partie de la population.

Parmi les thèmes “classiques” de la littérature cyberpunk, on trouve la relation humain-ordinateur (ou tout équivalent romanesque de l’ordinateur). Cette relation peut s’exprimer dans différentes situations. Dans la situation classique, l’ordinateur est dans le monde de l’humain, le monde *réel*, et l’interaction s’effectue en employant des périphériques habituels, tels qu’écran, clavier, microphone, ... La situation qui nous intéresse plus particulièrement amène l’utilisateur *dans* le monde *virtuel* de l’ordinateur, et représente la situation typique où les termes “réalité virtuelle” et “cyberespace” sont employés.

En 1984, le terme *cyberspace* fait son apparition, dans “Neuromancer” de William Gibson, un des romans fondateurs du mouvement cyberpunk. Il s’agit, dans le cadre de ce roman, d’une simulation informatique d’un environnement ressemblant à l’espace physique qui nous entoure. Depuis, le terme fût employé dans de nombreuses situations, y compris dans le monde de la recherche et fournir une définition semble fort délicat, aucun consensus précis ne semblant émerger⁴.

³La société Blaxxun fournit des outils de gestion de communautés virtuelles : visualisation d’environnement 3D, gestion de moyens de communication, manipulation d’outils de vote, de clubs, ... C’est autour de la technologie Blaxxun qu’est construite la communauté virtuelle du Deuxième Monde, présentée section 1.2.5.

⁴Il suffit de comparer la définitions de John Perry Barlow (artiste et journaliste scientifique) : “*Cyberspace is where you are when you are talking on the telephone*” à celle de l’office de la langue Française : “*Lieu imaginaire appliqué métaphoriquement au réseau Internet et dans lequel les internautes qui y naviguent s’adonnent à des activités diverses*”, de l’encyclopédie Hachette : “*Ensemble des informations et des relations que l’on peut trouver sur un réseau électronique.*” ou à celle de William Gibson lui-même (romancier) “*Cyberspace*

Des différentes définitions existantes, on retient que la caractéristique fondamentale d'un cyberspace est qu'il s'agit d'un espace n'étant pas directement présenté à l'utilisateur. Qu'il soit re-présentation d'un environnement réel ou purement imaginaire, il dépend d'un média. Des applications actuelles, deux cas de figure (n'étant pas exclusifs l'un de l'autre) se présentent assez clairement, l'un relevant des "espaces virtuels" et l'autre des "espaces de communications".

Tout d'abord les "espaces virtuels"⁵ qui sont des représentations d'environnements spatiaux (en trois dimensions pour les plus connus d'entre eux). Nous y intégrons toute représentation d'un espace. C'est le cas de jeux vidéos tels que *Half-Life*⁶ (environnement imaginaire), ou d'environnements de télémanipulation (re-présentation d'un environnement réel).

Deuxièmement, les "espaces de communication" qui sont les cyberspaces dans lesquels sont déposés les messages entre les différents utilisateurs, reprenant l'idée du tableau noir ou de la conférence publique. On trouve dans cette situation les discussions publiques du Deuxième Monde (la communauté virtuelle de Canal+, voir section 1.2.5), ou les forums de Usenet (groupes de discussions thématiques, voir section 1.2.4).

Notons qu'il existe des cyberspaces correspondant partiellement à chacun de ces deux pôles (qui ne répondent pas à une quelconque tentative de taxinomie, mais uniquement à la constatation d'un état de fait), comme par exemple les simulations de colonies de fourmis où la communication intervient par l'intermédiaire de traces laissées dans un environnement spatialisé, ou bien les personnages plongés dans un espace virtuel et enrichissant leur communication sur un autre canal (un canal textuel par exemple) par des expressions corporelles. Dans ces deux cas, un environnement spatialisé contient simultanément des représentations des agents et des représentations des messages. C'est ce qui est communément appelé "communication par l'environnement".

En résumé, nous emploierons ces termes de la façon suivante :

- un *cyberspace* est un environnement présenté de manière indirecte, particulièrement par l'intermédiaire de systèmes informatisés ;
- les *espaces virtuels* désignent un cyberspace dont la représentation est de nature spatiale ;
- les *espaces de communication* sont des cyberspaces dans lesquels se déroulent les interactions entre utilisateurs, par l'intermédiaire de dépôt et de perception de messages.

1.2 Les communautés virtuelles

Si l'on considère les définitions que le psychosociologue Roger Mucchielli donne de "communauté" et "groupe", on constate que la grande différence entre ces deux termes est l'exigence de relations plus "directes" dans le cadre du groupe tandis

is a consensual hallucination. A graphic representation of data abstracted from the banks of every computer in the human system."

⁵Le terme d'"espace virtuel" (ou de "monde virtuel") est préférable à celui de "réalité virtuelle", car possédant un champ sémantique moins vaste, même si l'usage en fait souvent des synonymes. De plus, le sens du terme "réalité virtuelle" a évolué et qualifie maintenant tout un domaine de recherche s'intéressant, globalement, à la présentation (généralement sous forme visuelle) de données et aux interfaces permettant leur manipulation.

⁶©Sierra Studio.

que la communauté peut se satisfaire de relations bien plus indirectes entre ses membres (comme une communauté d'expatriés, ou la communauté des croyants en une religion).

Puisque nous nous intéressons aux *interactions*, il serait plus correct de parler de “groupes virtuels” ou de “groupes mixtes” là où nous employons les termes “communautés virtuelles” et “communautés mixtes”. Toutefois, nous continuerons à employer les termes formés sur “communauté”, essentiellement pour des raisons de conformité avec l’usage.

Quentin Jones [Jones, 1997] considère comme communautés virtuelles un *cyberespace* associé à des moyens de communication médiatisée de groupe (“*a cyber-place with associated group-CMC (Computer Mediated Communication)*”) qui de plus respectent les quatre points suivants :

1. un niveau d’interactivité minimal ;
2. plusieurs communicants ;
3. un espace public commun où se passe une partie significative des échanges ;
4. un niveau de fidélité et de participation minimal.

Nous détaillerons ces points dans les trois sections suivantes où ils s’insèrent : environnement, moyens de communication et participants.

1.2.1 Un environnement où interagir

Comme nous l’avons déjà dit, le terme de cyberespace est souvent associé à celui de communauté virtuelle. Il ne faut toutefois pas confondre une communauté virtuelle et les cyberespaces dans lesquels (ou à travers lesquels) ses participants interagissent. Nous n’avons rencontré aucun argument définitif précisant si une communauté virtuelle doit impérativement dépendre d’un cyberespace⁷. Pour notre part, nous adhérons à l’idée de Quentin Jones qui demande un *espace public commun* où une proportion significative des échanges peut avoir lieu. Cet espace peut s’intégrer dans la communauté de manières très diverses : par exemple, il peut être le seul espace de communication de la communauté (comme par exemple un canal IRC ou un groupe de discussion), ou être une partie d’un système de communication plus vaste (comme la zone “accueil” du Deuxième Monde, ou, dans un autre cadre, une fréquence radio donnée comme le canal 19 des cibistes). De même, l’espace public commun peut être similaire ou non à l’ensemble des autres cyberespaces (combinaison d’espaces de communication, par exemple textuels, et d’espaces virtuels).

Certains espaces représentent des mondes réels (comme la ville de Paris [2M] ou la ville de Kyōtō [Kyoto]) ou imaginaires (colonicity [colonicity], Ultima On Line [Ultima]), pour permettre à leurs participants d’interagir. Ces espaces peuvent être représentés de diverses manières au niveau du poste de l’utilisateur. Chip Morningstar et Randal Farmer [Morningstar et Farmer, 1990] recommandent de laisser la possibilité à l’utilisateur de choisir sa représentation : une scène contenant un arbre pourra être, chez l’utilisateur, un simple texte “il y a un arbre” si l’utilisateur ne souhaite pas (ou ne peut pas) consacrer trop de ressources à la représentation de l’environnement, ou au contraire, une représentation 3D réaliste projetée sur un visiocasque.

⁷En 1968, Licklider, dont les travaux ont largement participé à la création d’Arpanet, et donc d’Internet, prédisait que la communication médiatisée entraînerait la “délocalisation” des communautés (“...*communities not of common location, but of common interest*”[Licklider, 1968]).

La qualité technologique de la représentation semble n'avoir au final que peu d'influence sur la vie de la communauté, le point important étant de plonger l'utilisateur dans l'espace commun (voir les travaux d'Elizabeth Reid [Reid, 1994], de Malcom Parks et Kory Floyd [Parks et Floyd, 1996] et de Chip Morningstar et Randal Farmer [Morningstar et Farmer, 1990], par exemple). Les MUD⁸, purement textuels, ont montré qu'il était possible de compenser une représentation fruste par de l'imagination⁹.

Comme nous venons de le préciser, il est important de faire la distinction entre *l'environnement* et sa *représentation*. Nous présenterons dans les sous-sections suivantes les moyens de représentations classiques des environnements des communautés virtuelles (textuel, visuel en deux dimensions, visuel en trois dimensions), la quatrième et dernière sous-section présente des moyens de représentation en pleine émergence, mais restant pour le moment marginaux.

a) Une représentation textuelle de l'espace.

Les premières communautés virtuelles (MUD par exemple) étaient entièrement sous forme textuelle. Ce type de représentation était contraint par les technologies de l'époque mais continue à être énormément employé, ne serait-ce que pour la simplicité technologique et la faible puissance requise, tant au niveau des clients que du serveur (voir figure 1.1).

The Living Room
It is very bright, open, and airy here, with large plate-glass windows looking southward over the pool to the gardens beyond. On the north wall, there is a rough stonework fireplace. The east and west walls are
(la description continue pendant quelques lignes)
There are two sets of couches, one clustered around the fireplace and one with a view out the windows.
You see Welcome Poster, a fireplace, Cockatoo, the living room couch, The Birthday Machine, lag meter, and Helpful Person Finder here.

FIG. 1.1 – Une représentation textuelle de l'espace (extrait de LambdaMOO)

⁸Multi-User Dungeons. Jeu d'aventure multi-utilisateur, ayant le texte comme seule modalité de communication et de présentation de l'environnement.

⁹*Virtual worlds exist not in the technology used to represent them, nor purely in the mind of the user, but in the relationship between internal mental constructs and technologically generated representations of these constructs. The illusion of reality lies not in the machinery itself, but in the users' willingness to treat the manifestation of their imaginings as if they were real. [Reid, 1994]*

b) Une représentation visuelle 2D de l'espace.

Après les premières interfaces purement textuelles, apparaissent les interfaces graphiques représentant le monde d'une façon visuelle. L'exemple le plus typique (et le plus étudié) est Habitat, une communauté développée par Lucasfilm au milieu des années 80 et toujours en activité au Japon. Dans cette communauté, chaque utilisateur dispose d'une représentation graphique de son personnage et de l'environnement (simples décors, objets manipulables) dans lequel il évolue. Les messages apparaissent dans des bulles au-dessus des utilisateurs (voir figure 1.2).



FIG. 1.2 – Une représentation visuelle 2D de l'espace (communauté "Habitat").

c) Une représentation visuelle 3D de l'espace.

Les progrès technologiques ont par la suite permis d'avoir une représentation en trois dimensions des environnements. On peut citer les communautés basées sur la technologie Blaxxun (le Deuxième Monde, Colonycity), Cryopolis, développé par Cryo sur la base de la technologie Scol, ou enfin la technologie Active Worlds qui remporte un vif succès (voir figure 1.3). Les technologies de représentation de monde 3D ayant de plus un support multi-utilisateurs sont actuellement en plein essor, portées principalement par l'évolution de la micro-informatique personnelle et de la bande passante des réseaux.



FIG. 1.3 – Une représentation visuelle 3D de l'espace : un des mondes d'Active World [mauz]

d) Au-delà...

Même si la “beauté de surface” n’est pas un facteur déterminant de la vie de la communauté, il s’agit d’un facteur important de sa création. Les communautés à forte composante visuelle, ou dotées d’interfaces conviviales et ergonomiques sont plus attractives, et de fait, attirent plus le grand public que les systèmes purement textuels. La fidélisation (indispensable à l’établissement d’une communauté) des internautes de passage dépend bien plus du contenu (animations, thèmes de discussion, construction de relations personnelles...) que du contenant (environnement 3D, multiples canaux de communication...) Après l’étape de l’image, l’étape du son est en train de se mettre en place, que ce soit au niveau de l’environnement (sons d’oiseaux ou de fontaines dans le Deuxième Monde) ou au niveau de la communication entre participants (basée en général sur le standard de fait real audio ou d’autres technologies propriétaires). Les interfaces les plus évoluées sont maintenant multimodales, employant la voix, le clavier et la souris en entrée, le moniteur et le son en sortie. Les interfaces dites immersives (gants de données, visiocasques, ...) sont encore peu présentes, probablement pour des raisons financières (ce genre de périphérique est encore hors de portée du grand public) et technologiques (leur fiabilité / robustesse n’est pas très élevée comparée aux périphériques plus classiques). Toutefois, comme le font remarquer Morningstar et Farmer [Morningstar et Farmer, 1990], au delà d’un ensemble de moyens de communication de base, “*the technology used to present this environment to its participants, while sexy and interesting, is a peripheral concern.*”

1.2.2 Des interactions entre les membres de la communauté

Le deuxième point indispensable à l’existence d’une communauté virtuelle est la communication entre ses participants¹⁰. Comme le signalait Quentin Jones [Jones, 1997], le problème de la survie d’une communauté tient en partie au maintien d’un certain niveau d’interaction entre ses participants. A cet égard, la définition d’interactivité à laquelle se réfère Jones est importante : “*Interactivity is an expression of the extent to which in a given series of communication exchanges, any third or later transmission is related to the degree to which previous exchanges referred to even earlier transmissions*” ([Rafaeli, 1988]). Selon cette définition, il n’y a *interactivité* que lorsque les messages émis se font référence les uns aux autres. Cette définition lui permet d’éviter de considérer comme communautés virtuelles des agrégats de personnes ne disposant que de communications unidirectionnelles (une liste de diffusion transmettant à ses membres des informations ciblées, par exemple), ou pour lesquelles un seul échange question/réponse constitue la norme (*newsgroup* de petites annonces par exemple).

Deux types de communication nous intéressent : la *communication langagière textuelle* et la *communication visuelle* (expressions faciales et corporelles principalement). Un troisième type de communication, la communication langagière orale, commence à apparaître, mais reste encore occasionnelle et peu répandue. Toutefois, il existe des situations dans lesquelles chacun des participants dispose d’interfaces permettant la communication orale, mais on peut noter qu’il existe aussi des situations de communication orale monodirectionnelle, comme par exemple sur le Deuxième

¹⁰“*The essential lesson that we have abstracted from our experiences with habitat is that a cyberspace is defined more by the interactions among the actors within it than by the technology with which it is implemented.*” [Morningstar et Farmer, 1990]

Monde où certaines interviews publiques sont retransmises suivant une modalité orale, les internautes intervenant par un canal textuel.

a) Communication textuelle.

Il s'agit de la méthode la plus ancienne et aussi la plus répandue. Depuis ICQ (logiciel de *chat*) jusqu'à Usenet en passant par les courriels [Yamakami et Nishio, 1998] et les salons de discussion, la plupart des communications se font sous forme textuelle. On peut en distinguer plusieurs formes : aux extrêmes, on trouve d'un côté une forme purement synchrone, type ICQ ou commande `talk` sous Unix et de l'autre une forme asynchrone, type Usenet ou tableau noir ; proche de la communication synchrone se situe le *chat* classique ; proche de la communication asynchrone se situent les listes de diffusion. Bien sûr, cette catégorisation est plus basée sur l'usage de ces moyens que sur leurs limitations technologiques. Rien n'empêche de prendre son temps pour répondre à un message ICQ, de même que des messageries rapides peuvent donner l'illusion d'une discussion "en temps réel".

Pour décrire brièvement ces différents types de communication, nous pouvons dire que dans le cadre de l'ICQ, les utilisateurs voient les mots se former au fur et à mesure que leur interlocuteur les tape. Ce type de communication se rapproche énormément de la communication parlée, puisque permettant de voir les hésitations, d'interrompre, etc. Les logiciels de *chat* les plus couramment répandus en sont différents puisque les messages sont postés en un seul bloc, au moment où le scripteur le choisit (en général, une phrase ou un morceau de phrase) et non lettre par lettre. Dans le cadre de listes de diffusion, chaque message (sous forme de courriel) est retransmis à l'ensemble des participants, qui peuvent à leur tour y répondre en envoyant un message sur cette même liste, et ainsi de suite. La différence est mince entre une liste de diffusion et un forum, toutefois, on peut considérer que les forums de discussion constituent un cyberspace (de type *espace de communication*) dans lequel les utilisateurs déposent leurs messages : un message est transmis en étant déposé dans un environnement commun. Dans le cas des listes de diffusion, le message est recopié et envoyé en autant d'exemplaires que de participants, il est alors beaucoup plus difficile dans ce cas d'identifier l'*espace de communication* généré par les interactions.

b) Communication visuelle.

Dans les mondes à composante visuelle où l'utilisateur est représenté par un avatar (2D ou 3D), il est souvent possible de déclencher des expressions corporelles. Bien que beaucoup de chercheurs s'accordent sur l'importance de la communication non-verbale, sa mise en application reste délicate, principalement dans le cadre de la communication médiatisée (lorsque des humains communiquent de manière indirecte, leurs messages étant transféré de l'un à l'autre par l'intermédiaire d'un moyen technique). En effet, un grand nombre d'indications non-verbales (comme les expressions faciales ou la direction du regard) qui enrichissent les communications de face-à-face sont involontaires. Forcer le locuteur à montrer explicitement ces indications, si tant est que cela soit possible, risque de les dénaturer. Par exemple, signaler explicitement "je suis excédé" ou "je suis dubitatif" nécessite que l'utilisateur soit conscient de cet état et qu'il souhaite l'exprimer, alors que l'expression faciale adaptée saute ces étapes ; symétriquement, les interlocuteurs ne sauront pas forcément comment adapter leur comportement à ce message qu'ils reçoivent alors qu'ils les traitent de manière implicite lors de face-à-face. De plus, ces techniques

restent très frustrés et ne sont pas, à l'heure actuelle, suffisamment évoluées pour permettre d'apporter un réel "plus" à l'interaction : interfaces cassant le côté "naturel" de la conversation, représentations graphiques peu adaptées à la visualisation d'expressions corporelles ou faciales, ... En forçant un peu le trait, on pourrait dire que le problème actuel de l'enrichissement de la communication médiatisée par du non-verbal est que l'attention de l'utilisateur doit se porter excessivement sur l'interface (que ce soit en manipulation ou en perception) et ce, au détriment de la communication elle-même.

À l'opposé de cette solution (la plus courante) où l'utilisateur explicite chacun des messages qu'il souhaite transmettre, on trouve les systèmes basés sur de la vidéo, comme les outils de téléconférence, ou, dans une optique plus communautaire, le système FreeWalk [Nakanishi *et al.*, 1998] (voir la figure 1.4).

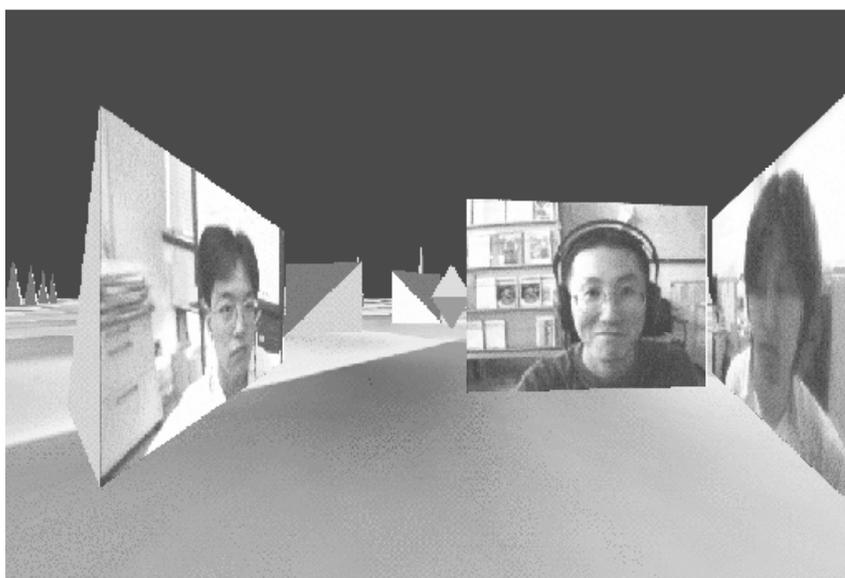


FIG. 1.4 – L'interface visuelle du système Freewalk ([Nakanishi *et al.*, 1998])

L'emploi de la vidéo décharge l'utilisateur de l'introspection, de la décision de transmettre et de la transmission d'informations complémentaires aux messages, qui sont déjà présentes dans son expression faciale, sa posture ou la direction de son regard. Dans ces situations à base de vidéo, les interlocuteurs présentent les uns aux autres une grande partie des indices non-verbaux présents dans une relation de face-à-face, sans avoir à les expliciter et les retransmettre consciemment.

Il existe cependant une troisième voie, dans laquelle les interlocuteurs sont représentés, mais où ils délèguent tout ou partie du contrôle de leur représentation à un agent. C'est une approche qui a été explorée par BodyChat [Vilhjálmsson, 1997] et le projet Amusement (voir le point d) de la sous-section 3.5.3). Dans ces applications, l'agent analyse la zone de *chat* dont se servent les interlocuteurs pour adapter les mouvements de l'avatar (saluts, regards se déplaçant de façon à signaler la fin du tour de parole, ...)

1.2.3 Des participants

Les participants sont l'élément à la fois le plus incontrôlable et le plus indispensable d'une communauté virtuelle. Ce problème reprend les points 2 (Plusieurs

communicants) et 4 (Un niveau de fidélité et de participation minimal) de Quentin Jones, cités dans l'introduction.

Le point numéro 2 semble aller de soi (on peut raisonnablement douter de la légitimité d'une communauté ne comportant qu'un participant ; deux participants étant un cas limite) et le point numéro 4 pose le problème de la fidélisation. Comme nous le verrons dans la partie 1.3, la création de relations personnelles par le biais de communications électroniques prend du temps. La stabilité d'une communauté repose sur la stabilité des relations entre ses membres, et ces relations ne peuvent se former que s'il y a une certaine fidélité que cette fidélité soit installée entre les participants eux mêmes, ou qu'elle s'exprime vis-à-vis de valeurs communes.

Dans les mondes virtuels, il est facile de séparer le "corps" d'un participant de son "esprit", ce qui anime ce corps, c'est-à-dire séparer les participants de leur représentation [Chicoisne et Pesty, 2000]. Comme nous allons le voir dans les paragraphes suivants, il existe plusieurs façons de représenter les participants (de même qu'il existe différentes façons de représenter des environnements, cf. section 1.2.1). Ce chapitre étant dédié aux communautés humaines, nous ne considérons comme participants que des êtres humains ; les agents artificiels constituent la deuxième catégorie de participants que nous pourrions considérer et nous en traiterons dans le chapitre 2. Néanmoins, il faut noter que rien ne s'oppose à ce qu'humains et agents aient des représentations similaires, voire même, dans certaines situations, prennent à tour de rôle le contrôle d'une même représentation (voir le point d) de la sous-section 3.5.3).

Les participants. Dans les enseignements tirés de la communauté Habitat, Morningstar et Farmer concluent que la communauté appartient et est définie par ceux qui y participent plus que par tout autre paramètre [Morningstar et Farmer, 1990]. Ils ont tenté un moment d'endosser le rôle de "planificateur omniscient" (ce sont leurs termes) et se sont vite rendu compte que la seule attitude viable était de se mettre au service des utilisateurs et de répondre à leurs attentes, plutôt que de chercher à planifier leur communauté.

Les membres d'une communauté participent à la vie d'une communauté parce qu'ils y trouvent un intérêt. Dans le cadre des communautés virtuelles "ouvertes" (c'est-à-dire celles qui n'existent pas dans un but précis comme les communautés de jeu ou les forums dédiés à un sujet particulier), on peut constater que les intérêts les plus classiques sont l'information, la transmission de documents (musique, logiciels, et, plus récemment, films) et la compagnie que les participants retirent de leur participation à la communauté.

La représentation des participants. Lors de leur "vie" dans la communauté, les différents participants ont une représentation qui leur permet de se présenter aux autres. Cette représentation peut prendre différentes formes : fiche avec photo, description purement textuelle, personnage animé ou statique en deux ou trois dimensions. Les représentations auxquelles nous nous intéressons sont de type personnage animé en trois dimensions. Ces personnages sont appelés avatars, en référence à la mythologie indienne où l'*avatar* représente une incarnation sur terre du Dieu Vishnu. L'avatar dans le cadre de communautés permettant une représentation 3D, est donc une coquille, une enveloppe habitée ensuite par un être humain ou un programme qui définit son comportement.

Certaines représentations sont statiques, mais d'autres peuvent autoriser un contrôle sur elles-mêmes. C'est le cas de la plupart des représentations graphiques,

qu'elles soient en deux ou trois dimensions, et pour lesquelles l'utilisateur dispose d'un certain contrôle, lui permettant de déplacer l'avatar dans son environnement, de le faire agir ou de s'exprimer (quoique très superficiellement pour le moment) de manière non-verbale.

1.2.4 Quelques exemples de communautés virtuelles

Toutes les communautés virtuelles ne ressemblent pas au Deuxième Monde (présenté en détail dans la section suivante). Nous allons donc, comme éléments de comparaison, présenter dans cette section deux communautés virtuelles : Usenet et les jeux vidéo en ligne. Nous avons décidé de mettre ces deux communautés en regard du Deuxième Monde car ce sont des communautés relativement répandues, mais aussi pour des raisons de proximité :

- elles sont relativement éloignées du Deuxième Monde dans leurs interfaces et objectifs, contrairement à d'autres communautés comme *colonycity* [*colonycity*] ou *ActiveWorlds* [*ActiveWorlds*] qui sont basées sur un monde 3D couplé à une interface de *chat* ;
- elles en sont toutefois relativement proches, en cela qu'elles sont supportées par un ordinateur contrairement, par exemple, aux communautés de radio-amateurs, ou, pourquoi pas, aux communautés épistolaires ;

a) Usenet

Usenet (aussi connu sous la simple appellation de “*les news*”) est un ensemble de groupes de discussions (forums, ou *newsgroups*) thématiques. Usenet reprend le principe du tableau noir, où les différents intervenants laissent des messages auxquels les autres intervenants peuvent répondre, formant ainsi une *cascade* (ou un *fil*, en référence au terme anglais consacré, *thread*) de messages.

Les forums à travers lesquels les messages sont échangés sont classés de manière hiérarchique, séparés au plus haut niveau en huit catégories : **comp** (ordinateurs), **misc** (divers), **news** (informations sur Usenet lui-même), **rec** (loisirs), **sci** (sciences), **soc** (société/culture), **talk** (discussions), **humanities** (sciences humaines), avec en marge **alt** (liste particulière permettant la création de sous groupes sans passer par le processus habituel de vote). A ces huit catégories initiales, on peut ajouter des catégories par pays, qui en général se déclinent à leur deuxième niveau suivant les huit catégories précédentes. Ainsi, **fr.comp** pour les *newsgroups* français traitant des ordinateurs ou **fr.soc.culture.japon** pour des discussions ayant trait à la culture Japonaise.

Ces groupes de discussion sont créés suivant un cycle (en général interminable) de propositions/discussions, sur un forum public dédié (pour la hiérarchie **fr.***, il s'agit des forums **fr.usenet.forums.annonces** et **fr.usenet.forums.evolution**), qui se termine par un vote. Mais tout un chacun est susceptible d'installer sur son ordinateur un serveur de *news*, et de créer ses propres forums avec toute latitude quant au choix du nom, du contenu, des participants autorisés, ... En général, ces forums ne seront pas aussi largement diffusés sur les autres serveurs de *news* que les forums des hiérarchies principales (le “*big eight*” et les hiérarchies nationales), la politique d'hébergement d'un forum particulier étant prise indépendamment sur chaque serveur du réseau.

Au sein de Usenet, on peut trouver des forums très différents les uns des autres. Si l'on passe les forums de Usenet au crible de la définition de “communauté virtuelle”

de Quentin Jones [Jones, 1997], on se rend compte que presque toutes les situations existent, mais que certains forums méritent amplement leur statut de communauté. Les quatre points de la définition de Quentin Jones sont les suivants :

1. *un niveau d'interactivité minimal* : depuis le simple échange question/réponse (`fr.petites-annonces.*`, de par son thème) jusqu'aux longues discussions (`fr.soc.politique`, de par son thème aussi...);
2. *plusieurs communicants* : il existe de nombreux *newsgroups* dans lesquels personne ne communique (`fr.bio.canauxioniques` détient d'ailleurs probablement un record : celui du *newsgroup* dont on parle, sans y parler!) et d'autres particulièrement actifs (300 ou 400 messages par jour ne sont pas rares sur `fr.rec.moto`);
3. *un espace public commun où se passe une partie significative des échanges* : par essence, le forum *est* l'espace public commun. Même s'il n'est pas rare de retrouver les mêmes participants dans plusieurs forums (par exemple, les forums `fr.soc.japon`, `fr.rec.anime` et `fr.lettres.langue.japonaise` ont beaucoup de participants actifs communs), et qu'il est possible de poster un message simultanément dans plusieurs forums, le forum reste une unité aux frontières bien définies.
4. *un niveau de fidélité . . .* : il n'est pas rare de retrouver dans certains *newsgroups* des gens qui y participent depuis des années, parfois depuis la création; *. . . et de participation minimal* : pour ce point encore, on peut trouver des forums où la participation est nulle, et d'autres où les messages postés se comptent en milliers par jour.

Certains de ces forums valident les quatre points proposés par Quentin Jones, et forment donc de véritables communautés, où se tissent des liens personnels entre les différents participants, créant même parfois un climat "familial". L'étude de Parks et Floyd [Parks et Floyd, 1996] à ce sujet est particulièrement significative, et, comme cette étude le montre aussi, les participants à ces *newsgroups* cherchent à organiser des rencontres "physiques" (repas au restaurant (évidemment Japonais) pour `fr.soc.japon` à Paris ou à Tōkyō, organisation d'une "concentration" annuelle et invitations à des sorties pour les motards de `fr.rec.moto`, . . .)

b) Les jeux vidéos multi-utilisateurs

Dans les années 1970, l'interconnexion des ordinateurs permit la naissance de systèmes où plusieurs utilisateurs pouvaient interagir, comme Usenet ou les MUDs. Ces MUDs sont (car certains sont encore actifs) des mondes représentés de façon textuelle dans lesquels des utilisateurs interagissent par *chat* et disposent de commandes leur permettant de se déplacer et d'agir dans le monde. Ils sont une évolution multi-joueurs des tout premiers jeux d'aventure textuels.

Actuellement, les héritiers directs de ces MUDs sont les jeux de rôle en ligne massivement multi-utilisateurs (*massively multiplayer online role playing game*). Parmi ces jeux, Ultima Online¹¹ marque les plus récents changements : un monde complexe disposant d'une histoire riche (car héritier de la longue série "Ultima", jeux d'aventure mono-utilisateur); une représentation du monde en 3D isométrique; la possibilité pour les participants d'incarner des personnages différents de ceux, classiques, des jeux de rôle (comme mineur, tavernier, forgeron. . . contrairement aux classiques guerrier ou magicien), entraînant ainsi des interactions plus riches car ces

¹¹Développé par Origin Systems et distribué par Electronic Arts.

personnages étaient auparavant exclusivement des personnages non-joueurs (PNJ), contrôlés par des programmes ; de nombreuses autres innovations en terme d'interactivité avec le monde, etc.

En reprenant les critères de Quentin Jones [Jones, 1997], on constate que, du fait de la nécessité d'un espace commun de discussion, Ultima Online est un monde contenant plusieurs communautés :

1. *un niveau d'interactivité minimal* : les moyens d'interaction sont très vastes, le *chat* étant le plus évident, mais les personnages peuvent aussi s'échanger des objets, se battre, être la cible de sorts aux divers effets. . .
2. *plusieurs communicants* : 225000 inscrits en Juillet 2001. Bien sûr, ce chiffre ne correspond pas à la quantité de personnes jouant simultanément, mais il donne toutefois une bonne indication de la fréquentation du jeu.
3. *un espace public commun où se passe une partie significative des échanges* : le monde d'Ultima Online est vaste. De la même façon qu'on ne peut parler de communauté pour l'ensemble d'Usenet, mais qu'on peut le faire pour certains des forums, il est possible de reconnaître plusieurs communautés de joueurs dans Ultima Online (que ces groupes soient formels, comme les guildes, ou non, comme les groupes d'aventuriers jouant ensemble), ayant chacun leurs lieux de rencontre, sans qu'il ne soit possible d'identifier un lieu générique de rencontre.
4. *un niveau de fidélité et de participation minimal* : presque par essence, les jeux de rôle nécessitent une participation de leurs joueurs, et le principe de l'expérience, que le joueur accumule au fur et à mesure de sa participation au jeu, entraîne une grande fidélité (qui peut aller jusqu'à la dépendance) au jeu.

Nous n'avons cité ici qu'Ultima Online, mais nous aurions pu choisir Everquest, son plus proche concurrent. Il sera probablement très intéressant de suivre le futur "The Sims Online" (distribué, comme Ultima Online, par Electronic Arts) qui a des objectifs massivement multi-joueurs, mais, contrairement à la quasi-totalité des autres jeux de cette catégorie, se déroule dans un monde réaliste, et, à la manière du jeu mono-utilisateur dont il sera issu, mettra l'accent sur la vie sociale des participants (interactions, vie de famille, emplois, . . .)

En parallèle de ces jeux "massivement multi-joueurs" on trouve des jeux "multi-joueurs" formant des communautés embryonnaires. Sur des jeux aussi différents que Diablo II (action/aventure, développé par Blizzard), Starcraft (stratégie temps réel, développé par Blizzard) ou Half-Life ("*first-person shooter*", développé par Sierra Studio), les joueurs ont la possibilité de former des coalitions (incluant éventuellement des agents artificiels, dans le cadre de Half-Life) et de communiquer par *chat* et, lorsque les joueurs se rassemblent en groupe, des stratégies collectives sont indispensables pour mener le jeu à bien. Sur certains de ces jeux, des groupes de participants jouent régulièrement ensemble, au point de pouvoir prétendre au statut de communauté, selon les critères de Quentin Jones, seul le nombre de participants pouvant limiter cette prétention : les groupes de Diablo II sont nettement moins grands que les guildes d'Ultima Online.

1.2.5 Le Deuxième Monde, la communauté virtuelle de Canal +

Développé à l'origine en partenariat avec la société Cryo, le Deuxième Monde (souvent abrégé 2M par ses utilisateurs) se présentait sous la forme d'un CD-ROM

à acheter, permettant par la suite de se connecter à un site Internet où les différents participants se rassemblaient. Depuis Mars 98, le Deuxième Monde n'est plus géré que par Canal+ et se présente sous une forme purement "en ligne" et à accès gratuit. Depuis sa fusion dans le site de Canal+, il est possible d'y accéder de deux façons différentes : en exploitant une extension d'application (*plug-in*) spécialisée qui permet de naviguer dans le monde en 3D, ou en utilisant une *applet* JAVA qui n'autorise que la discussion (*chat*) et non la visualisation de l'espace 3D.



FIG. 1.5 – L'arc de Triomphe dans le Deuxième Monde

Sous sa forme actuelle, le Deuxième Monde met à la disposition des utilisateurs une interface de communication de type *chat* (voir un exemple de discussion dans la table 1.1 ainsi que dans la table 1.2 page 26) et une fenêtre représentant le monde en trois dimensions. L'ensemble de l'interface est visible dans la figure 1.8, et la figure 1.5 montre la place de l'étoile telle qu'elle peut être vue par un visiteur du Deuxième Monde. En plus de ces moyens de communication, l'utilisateur peut disposer d'argent, d'un logement et d'objets personnels. Il peut aussi avoir un rôle (maire, artisan, ...) et être investi de pouvoirs (mise en place de votes ou de sondages, introduction de nouveaux objets dans la communauté, ...) L'activité la plus courante dans le Deuxième monde est la discussion entre les internautes. Il existe aussi des événements organisés, soit spontanément par les participants eux-mêmes, soit par l'équipe d'animation du Deuxième Monde. Parmi les événements organisés sur l'initiative des utilisateurs eux-mêmes, on peut trouver des mariages, matchs d'improvisation ou concours d'avatars. Parmi les événements sur l'initiative de l'équipe d'animation, on peut trouver des interviews interactives avec des romanciers ou des sportifs, des visites guidées de musées (par exemple dans le cadre du partenariat avec le musée du Louvre) ou des retransmissions de concert (par exemple dans le cadre du partenariat avec la salle de concerts "le divan du monde").

a) Le monde

L'espace 3D du Deuxième Monde est divisé en *cellules*, zones dans lesquelles et entre lesquelles se déplacent les participants. A chaque cellule est associé une zone de *chat* public. On peut rassembler ces cellules sous différentes catégories :

Tout d'abord, la représentation de Paris. Certains lieux de Paris sont reproduits intégralement par les infographistes de Canal+ (comme la place de l'étoile, voir

```
Niko > mais par exemple cybertown est bcp mieux
Niko > mais ca veuty pas dire grand chose ce que tu viens de
dire
Koraz > on peut pas comprarer cybertown a été plus facil à
construire que le 2M
Niko > parce que everquest par exemple bah c un monde vir-
tuel aussi (un jeu soit mais ca en reste pas moins un monde
virtuel)
Koraz > le 2M est à l'echelle
Niko > comment ca?
lilo22 a rejoint le groupe
Niko > non mais de toute fgaocn meme CT ca craint.. le blem c
que le moteur 3D est pourri
Koraz > cybertown a été fabriqué complètement pas il on pas
été mesurer sur le terrain la taille des batiment pris des
photos ....
lilo22 a quitté le groupe
Niko > oué ok
Visiteur a quitté le groupe
Niko > enfin bon ca devrait pas se limiter a des bout de paris
dans lekel onj peut causer
Koraz > on peut faire des animations aussi
Sliver > ben rien ne t'empêche de faire une anim'
Koraz > si tu as des idées n'hésite pas
Niko > oué mais ca manque crusialement dinteractivité
```

TAB. 1.1 – Extrait de Dialogue du Deuxième Monde (1) (les pseudonymes ont été modifiés)

figure 1.5, ou la place des Vosges, voir figure 1.7). Pour d'autres zones, les internautes étaient invités à prendre le contrôle de certaines "concessions" dans Paris, préparées par les infographistes dans leurs grandes lignes (plan au sol et volumes des bâtiments). Après avoir été reproduits (d'après plans et photos sur le terrain), les arrondissements étaient fragmentés en plusieurs cellules et ouverts à la "colonisation", c'est-à-dire distribués à qui en faisait la demande. Certaines de ces cellules étaient "classées", c'est-à-dire que leurs propriétaires s'engageaient à les habiller de façon réaliste.

Dans cette catégorie des représentations de lieux de Paris, on peut citer des lieux particuliers comme le Louvre (où se tenaient régulièrement des expositions organisées par les guides du Louvre), le K-Hole (une boîte de nuit sonorisée par l'équipe "musiques électroniques" de Canal+, voir figure 1.6) et la fondation Cartier (où sont parfois reproduites les expositions photo de la Fondation Cartier "réelle").



FIG. 1.6 – Le K-Hole

Ensuite, nous avons certaines cellules, proposées par les internautes eux-mêmes, et n'ayant rien à voir avec Paris, qui sont rassemblées sous le terme "mondes perso", pour rappeler le concept de "pages perso".

Puis, les netcafés. Au nombre de quatre (ciné, info, cyber, sport), ils représentent des lieux de discussion thématiques.

Il existe aussi des cellules commerçantes, développées en collaboration avec des partenaires industriels.

Enfin, une série de lieux un peu particuliers, parmi lesquels on peut citer une salle de concert (dans laquelle sont parfois retransmis les concerts ayant lieu dans la salle "réelle"), un planétarium (mis en place en partenariat avec des astronomes, dans le cadre d'un cycle de conférences sur le système solaire ayant comme point d'orgue l'éclipse du 11 août 99), des galeries de photos, des zones "expérimentales" (comme "la fabrique", réalisée par les membres de fabric|ch [fabric|ch]) ...

Une zone tient une place particulière : *l'accueil*. Dans cette cellule, tous les jours, de midi à minuit, une ou deux personnes de l'équipe d'animation de Canal+ sont présentes pour prendre en main les nouveaux connectés et les aider à régler leurs problèmes (configuration, activités proposées par le Deuxième Monde. . .) Une cellule particulière, aménagée au premier étage de la Tour Eiffel lui a été consacrée.

b) L'interface

D'un point de vue technologique, le Deuxième Monde s'appuie sur un produit client / serveur proposé par la firme Blaxxun [Blaxxun]. Leur moteur 3D exploite des mondes décrits en VRML, langage public (limitant le risque de s'enfermer dans une technologie propriétaire, permettant un certain choix des outils de développement, . . .) Pour pouvoir se connecter, les internautes doivent télécharger un *plug-in* spécialisé qui gère à la fois la navigation dans le monde 3D et la communication (*chat*).

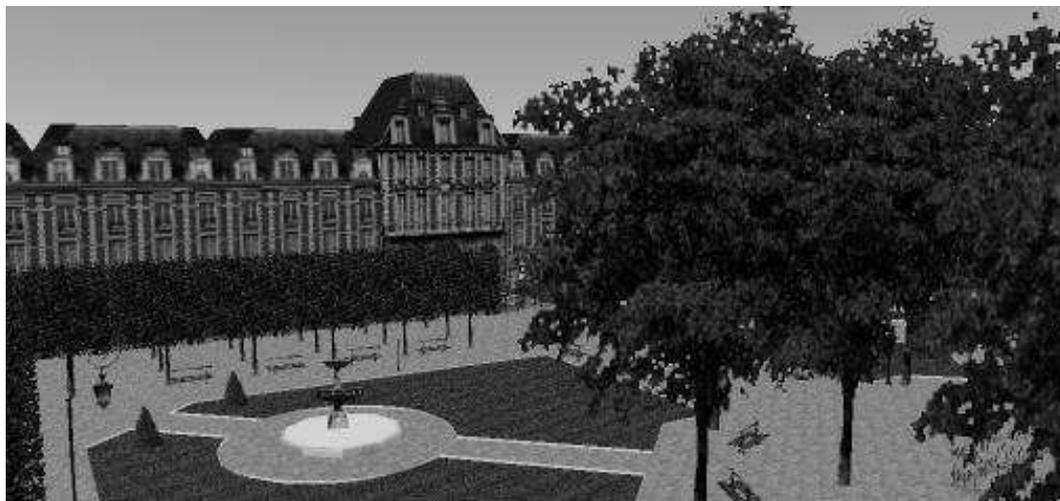


FIG. 1.7 – La place des Vosges dans le Deuxième Monde

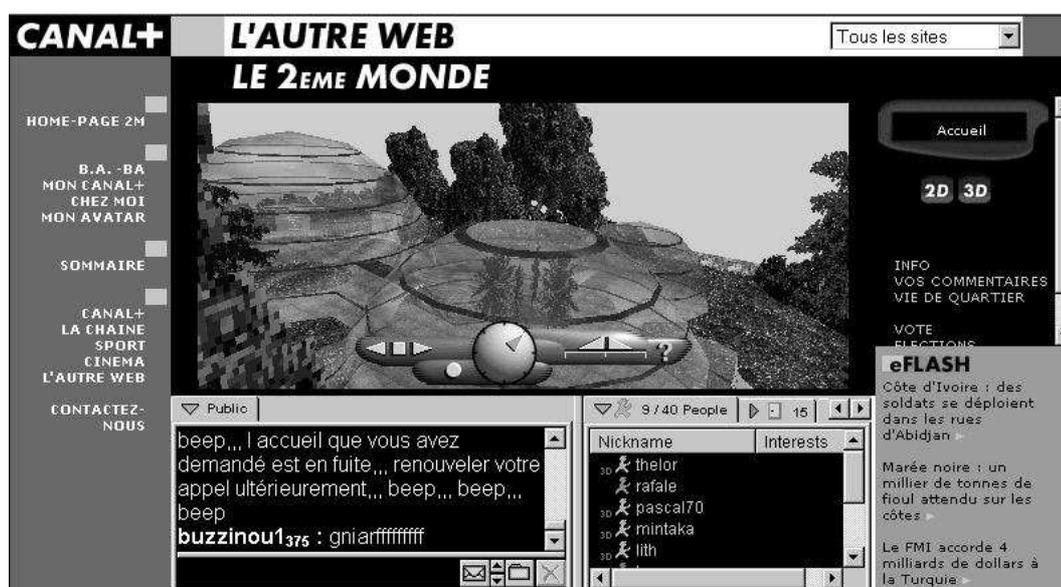


FIG. 1.8 – L'interface de CANAL+ sur le Deuxième Monde

La figure 1.8 représente l'interface Blaxxun. Les différentes options présentes sur la page web et ne faisant pas directement partie de l'interface Blaxxun concernent principalement des liens de navigation à l'intérieur du site (que ce soit une navigation "fine" au sein des pages du Deuxième Monde, ou une navigation dans le site de Canal+ dans son ensemble).

c) Les utilisateurs

Le profil. Si l'on reporte sur un graphique la population française classée par catégorie socio-économique, on constate que la pénétration de la télévision ou de la radio suit le même schéma : ce sont des médias "démocratiques", répartis de manière homogène dans les différentes classes. Une étude réalisée pour Canal+ montre que les utilisateurs d'Internet (et plus particulièrement les *internauts* fréquentant le site de Canal+) quant à eux sont relativement atypiques, étant principalement présent parmi des classes "minoritaires" comme les étudiants ou les cadres supérieurs. Le connecté habituel au site de Canal+ est plutôt un homme (77%) de moins de 34 ans (65%) qui reste en moyenne 40 minutes (ce qui classe Canal+ dans les meilleurs sites francophones en terme de temps moyen de consultation).

Les services offerts. Les *bimondiens* (nom que se sont donnés les membres du Deuxième Monde) qui se connectent sur le Deuxième Monde y trouvent plus qu'un salon de discussion (*chat-room*) classique. Plusieurs services particuliers leur sont offerts. La plupart ont déjà été présentés précédemment. On peut toutefois en reprendre ici la liste :

Clubs. Possibilité de créer et d'animer des clubs, rendez-vous thématiques.

Appartements. Les bimondiens peuvent acquérir et s'installer dans un appartement, dans certains quartiers de Paris qui constituent des cellules comme les autres, où les bimondiens peuvent se rencontrer. Ces appartements peuvent être construits entièrement par l'utilisateur, ou bien correspondre à un modèle standard personnalisable par le téléchargement d'images personnelles qui serviront de papier peint, de poster, etc.

Mondes perso. De manière exceptionnelle, le site de Canal + héberge des cellules développées par des internautes, principalement lorsque ces cellules ont une qualité artistique particulière.

Votes. Les bimondiens (du moins, ceux disposant d'un certain pouvoir, voir plus bas) peuvent initier des votes ou des sondages.

Accueils. Tous les jours, de midi à minuit, des accueillants sont présents pour prendre en charge les nouveaux arrivants ou pour gérer les problèmes techniques côté client.

Objets. Les bimondiens peuvent posséder des objets, qu'ils achètent ou construisent, formant un début d'économie.

Évolution & Pouvoir. Les bimondiens disposent de points d'expérience qu'ils acquièrent de manière automatique par leur participation à la vie de la communauté ou par leur simple présence. Cette expérience leur permet d'accéder à certains statuts et de choisir des rôles dans la communauté. À des rôles sont associés des pouvoirs comme celui de créer des objets, d'administrer une cellule, ...

La plupart de ces services vont dans le sens de Quentin Jones [Jones, 1997], stimulant l'arrivée de participants, puis leur fidélisation et leur implication au sein de la communauté.

d) Les coulisses

Equipe d'animation. L'équipe d'animation a plusieurs rôles. Elle définit le planning des différentes activités du Deuxième Monde (accueils, clubs, ...), recherche et gère des partenariats d'animation (Astronomes de la Villette, Guides du Louvre, concerts en partenariat avec la salle de spectacle "le Divan du Monde", ...), et s'occupe des invités (contacts, interviews, ...) Elle prend en charge globalement du contenu du Deuxième Monde.

Equipe technique. L'équipe technique se compose d'intégrateurs et d'infographistes. Les intégrateurs développent l'ensemble des mécanismes internes des pages du Deuxième Monde, typiquement, toutes les pages générées dynamiquement à partir de bases de données. Leur travail a aussi trait à tous les processus de gestion des fonctionnalités de la communauté (identification, gestion des profils, votes, appartements ...), et aux animations impliquant des situations particulières (flux vidéo ou audio par exemple). Les infographistes travaillent pour leur part à l'aspect visuel du Deuxième Monde, depuis les pages web classiques, jusqu'au développement des cellules 3D.

C'est dans le contexte de cette communauté virtuelle que se situent les applications de notre travail. Notre objectif est d'y introduire des agents capables d'y interagir avec les internautes. Les applications visées concernent dans un premier temps l'offre de services exploitant l'intégration de ces agents dans le monde : contrairement à une interface web classique à base d'hypertextes, l'agent fait partie du monde des utilisateurs et peut communiquer avec, au minimum, les mêmes moyens qu'eux (il peut en effet disposer d'autres moyens comme la manipulation de pages web par exemple). Les capacités de communication dont dispose l'agent, inspirées des communications humaines, viendront renforcer l'efficacité des agents en permettant aux internautes de communiquer en employant des modalités auxquelles ils sont habitués.

1.3 La communication médiatisée

Deux types de communication impliquant l'humain nous intéressent : celle de type humain \rightarrow ordinateur \rightarrow humain, qui se nomme communication médiatisée¹², et celle de type humain \leftrightarrow ordinateur qui relève du domaine de l'interaction homme-machine (et plus particulièrement des IHM textuelles, dans le cas qui nous intéresse). Notre objectif étant d'améliorer la communication entre humains et agents (donc une situation relevant plus de l'IHM), il est important de savoir comment les humains communiquent entre eux, et plus particulièrement dans une situation où leurs interactions sont médiatisées.

Des études comme celles de Clifford Nass [Nass *et al.*, 1994] montrent que, spontanément et bien qu'il s'en défende, l'humain agit vis-à-vis de l'ordinateur comme s'il avait une "personnalité", un "amour-propre", ou un "statut social". Des travaux

¹² *Computer Mediated Communication* connue aussi sous l'acronyme de CMC

ultérieurs [Nass *et al.*, 2000] sur la confiance accordée à un ordinateur vont eux aussi en ce sens¹³.

Nous verrons dans cette partie les différences qui existent entre ce type de communication et la communication de face-à-face. Ensuite, nous aborderons la “netiquette”, un ensemble de règles tentant de gommer les handicaps inhérents aux communications électroniques. Enfin, nous terminerons sur le problème de la construction de relations personnelles dans un monde aux interactions dégradées, ou tout au moins différentes des interactions auxquelles nous sommes habituées.

1.3.1 Spécificités de la communication médiatisée

La communication médiatisée a ceci de particulier par rapport à la communication de face-à-face qu’au moins une partie des messages est échangée par l’intermédiaire d’un moyen technologique. Ces moyens technologiques peuvent être assimilés à des *espaces de communication* (voir page 9), espaces dans lesquels ou à travers lesquels les différents interactants déposent (ou envoient, marquent, etc.) et perçoivent (ou reçoivent, consultent, etc.) des messages.

La communication “tapotée” (de type *chat*) à laquelle nous nous intéressons présente de nombreuses analogies avec à la fois la communication écrite et la communication orale. L’observation d’un dialogue typique comme celui reproduit dans la table 1.2 permet de voir certaines de ces analogies : emploi de structures grammaticales plus proches de l’oral que de l’écrit, échange de messages textuels mais avec un emploi phonétique des lettres, erreurs (lexicales, syntaxiques, ...) acceptées par l’interlocuteur, etc.

```

...
Marty > Salut tous
Artois > pas moi moi pas m'en foutre moi contre le bresil e't
moi déçu pour hier
Dominique > Yé vais aller avec des amis faire la fiesta avec
les camérounais...
Gagarine > Prendrai bien un p'tit café Marty
Bilbo > viva lé cameroun!!!
Gagarine > Ca va chauffer Dom
Marty > un ptit café, je suis pas barman Gagarine, j ai pas de
café moi!!!
Artois > viva espana!
Corneille > vous parlez foot a bientôt...
Bilbo > bouh..Artois..ehh viva lé nigéria!!!
...

```

TAB. 1.2 – Extrait de Dialogue du Deuxième Monde (2) (les pseudonymes ont été modifiés)

Puisque certains points de ces deux modes de communication sont incompatibles, nous présentons dans cette partie les spécificités de ce mode de communication. Nous

¹³Même si l’on peut considérer que l’anthropomorphisme latent chez les humains est encore accentué dans ces travaux par l’attribution de visages ou de corps anthropomorphes aux ordinateurs.

avons distingué quatre points : tout d'abord, l'absence de face-à-face ; puis l'anonymat apporté par le "tampon" que forme l'ordinateur ; ensuite le fait que nombre des discussions sont des discussions publiques ; enfin le stockage des conversations, dans leur intégralité et leur intégrité.

L'absence de face-à-face. Lors d'une communication de face-à-face, les messages transmis d'un locuteur à l'autre ne se limitent pas aux mots. Outre les informations auditives (prosodie, emphase sur certains mots, pauses...), on rencontre des informations visuelles (posture, expression faciale...) qui viennent enrichir la conversation. Toutes ces informations (en partie inconscientes) servent à guider et aider la conversation entre les deux interlocuteurs. Dans le cadre des conversations écrites, toute l'information transmise l'est sous forme écrite, chaque mot étant consciemment tapé par chacun des scripteurs. La spontanéité existe, mais est limitée par l'emploi d'un moyen de communication qui n'est pas aussi instinctif que la parole. On perd donc une grande partie de l'information qui permet particulièrement de se rendre compte de l'état de notre "interscripteur", relativement au message qu'il est en train d'émettre et relativement au contexte. Au delà de ce point (présenter certaines informations sans nécessiter de la part de l'émetteur une grosse charge cognitive), se pose le problème de la transmissibilité de ces informations. Le canal textuel présente une bande passante étreinée comparée aux possibilités de la relation de face-à-face. Plusieurs conventions ont été mises en places pour tenter de palier au mieux au manque d'informations et à la difficulté de les transmettre. Les plus connues sont l'emploi des majuscules pour représenter une phrase criée et les émoticônes (*smiley*) qui rendent compte de l'état émotionnel du scripteur (ironique, fâché...) ou qui orientent la façon d'interpréter une phrase (humour...)

L'absence de face-à-face pose aussi le problème du retour d'information. Ce retour d'information est quasi-immédiat dans le cas du face-à-face (attitude de l'interlocuteur, interruption éventuelle...), mais est limité, dans le cas du *chat* à ce que notre interlocuteur veut bien nous dire, au moment où il accepte de nous le dire. Il n'est pas possible d'avoir un retour "en temps réel" sur nos énoncés car le type de communication employé n'est pas synchrone (contrairement à la communication de face-à-face), ce qui pose particulièrement des problèmes au niveau du respect de l'autre.

L'ensemble de ces remarques est lié à l'*interface* à travers laquelle s'effectue l'interaction, et dont les particularités modifient plus ou moins profondément cette interaction. Comme le dit Gérard Sabah : "[...] il est vrai que par l'intermédiaire d'un clavier et d'un écran, on crée un mode de communication totalement inédit, où les tours de parole sont bien mieux respectés, et où les interruptions sont pratiquement inexistantes. La frappe au clavier, pour les non-initiés, implique des limitations probables sur la quantité des informations transmises et la présence de certaines erreurs spécifiques." [Sabah, 1989].

L'anonymat. Parmi les paramètres pris en compte dans la communication, nous avons déjà montré l'importance des *relations à l'autre*. Lors d'une première rencontre de face-à-face, une multitude de paramètres sont disponibles avant même le début de la conversation (âge apparent, sexe, vêtements...). Si cette conversation est programmée, il est en plus possible d'avoir des informations avant même de rencontrer notre (futur) interlocuteur (rôle, position hiérarchique, capacités...). Dans beaucoup de salons de *chat* la plupart des premières rencontres se passent entre deux (ou plusieurs) internautes n'ayant pas de connaissances l'un de l'autre, ni accès de ma-

nière immédiate à des informations concernant son interlocuteur. Un comportement habituel des personnes se connectant dans un salon de *chat*, est de poser la question “ASV ?” correspondant à “Age, Sexe, Ville ?”, trois paramètres assez frustrés, mais qui permettent de se faire une idée globale de nos interlocuteurs (notons que cette pratique est considérée comme inconvenante par endroits, comme dans le Deuxième Monde). Quoiqu’il en soit, la quasi-totalité des informations que nous avons sur nos interlocuteurs nous vient directement d’eux, que ce soit par l’intermédiaire de leurs réponses à nos questions, ou en lisant leur profil (qu’ils ont eux-mêmes remplis) dans l’annuaire des connectés. Il s’en suit que toute identité endossée dans le cadre d’une relation médiatisée est susceptible d’être fictive¹⁴. Cette situation entraîne forcément un temps plus long pour que s’installent des relations de confiance entre internautes, ces relations se mettant en place souvent avec l’extension de la communication à des moyens supplémentaires (e-mails, téléphone, web-cams, lettres, rencontres, ...) [Parks et Floyd, 1996].

Les discussions publiques. Dans le cadre qui nous intéresse, deux formes de discussions sont possibles. L’une est dite “publique” et correspond à l’envoi d’un message visible par tous. L’autre est dite “privée” et correspond à un tête à tête. Un des problèmes liés aux discussions publiques (comme c’est le cas pour les forums tels que Usenet) est que tout message, qu’il s’agisse d’un compliment, d’une remontrance ou d’une mise au point, est visible par tous. Il n’est pas rare de voir des échanges très énergiques, voire agressifs avoir lieu en public, situation qu’on chercherait à éviter dans le cadre d’une discussion en face-à-face.

Le stockage des conversations. La communication par *chat* apporte une autre particularité, la sauvegarde des échanges (“les écrits restent”), qui n’a pas trait à l’absence de face-à-face, mais au fait que cette forme de communication soit médiatisée. On pourrait s’attendre à ce que cette forme de discours entraîne les différents interactants à prendre plus de précautions lors de leurs discussions. Il semblerait pourtant que la plupart des internautes réagissent de manière spontanée — même si l’interface de communication (clavier) limite en partie cette spontanéité — et ne se reportent que peu aux historiques des conversations précédentes, ou de la conversation en cours, bien que l’historique des dialogues soit facilement accessible dans la plupart des logiciels de *chat*. Ce point de vue n’est qu’un simple constat, il n’existe à notre connaissance aucune étude de l’impact de l’enregistrement des échanges sur le comportement des interactants.

1.3.2 Les règles de bonne conduite dans la communication médiatisée : la netiquette

Avec la création de *Usenet*, certaines règles de savoir-vivre ont “émergé”. Rassemblées sous le terme de *netiquette*, elles comportent une série de recommandations sur le fond et la forme des messages à employer. La netiquette n’a aucun rédacteur connu, pas même un compilateur ou synthétiseur. De ce fait, il n’existe pas un texte de référence, mais plusieurs, rédigés dans le même esprit. Toutefois, on peut se rapporter au RFC1855¹⁵ de S. Hambridge [Hambridge, 1995].

¹⁴et ce, d’autant plus facilement que les environnements d’interaction sont virtuels, voir les classifications en introduction de cette partie.

¹⁵RFC signifie Request For Comments, et désigne les définitions des standards employés sur Internet. Ce document est donc *en quelque sorte* officialisé par l’*Internet Engineering*

La netiquette présente de grandes analogies avec la théorie des faces du psychologue Erving Goffman (voir page 107) et Michel Marcoccia [Marcoccia, 1999] relie plusieurs points de la netiquette au travail de *figuration* (*face work*, terme aussi employé, et dans le même sens, par Goffman). De ce travail de figuration, Marcoccia retient principalement la notion de *ménagement* des interlocuteurs, et il distingue trois étapes pour que l'interaction respecte au mieux cette notion (le parallèle avec les travaux de Goffman [Goffman, 1973] est ici direct) :

1. **Avant.** Eviter l'agression : *Vous n'enverrez pas de messages haineux même si on vous provoque* [Hambridge, 1995]
2. **Pendant.** Agresser, mais de manière "mesurée", en employant des termes qui limitent l'agressivité potentielle d'un message (Goffman parle de *softeners*) : *"Phrase your postings politely. Cursing is frowned upon. If you want to express a strong opinion, cute euphemisms and made-up expletives are usually acceptable"* [Shea, 1994]
3. **Après.** Réparer après avoir été agressif :
 - *Q – I sent flame mail to a discussion group I participate in, and now I regret it. What should I do ?*
 - *R – A gracious apology is almost always appreciated.* [Shea, 1994]

Sur l'aspect du respect du territoire (la plupart des principes de la théorie des faces peuvent s'appliquer au territoire, voir Erving Goffman [Goffman, 1974], Catherine Kerbrat-Orecchioni [Kerbrat-Orecchioni, 1989] et notre section 4.4.2 b)) , trois points sont abordés :

1. Respecter le temps des interlocuteurs : *"Respect other people's time and bandwidth. When you send email or post to a discussion group, you're taking up other people's time (or hoping to). It's your responsibility to ensure that the time they spend reading your posting isn't wasted."* [Shea, 1994]
2. Respecter les possessions : *"Respect other people's privacy — Of course, you'd never dream of going through your colleagues' desk drawers. So naturally you wouldn't read their email either."* [Shea, 1994].
3. Utiliser raisonnablement les ressources nécessaires à la communication (temps, coût de communication, bande passante, place de stockage...). *"Soyez conscient de la longueur des messages que vous envoyez. Annexer de grands fichiers, tels que des documents en Postscript ou des programmes, peut rendre vos messages si grands qu'ils peuvent ne pas être transmis ou au moins consommer une part exagérée de ressources."* [Hambridge, 1995]

Par ailleurs, la netiquette n'est qu'une partie formelle. La manière de l'appliquer peut être sujette à variation suivant le but du message (quelqu'un cherchant à lancer une polémique l'enfreindra souvent, consciemment), le contexte du message (un *thread* d'insultes entraîne généralement plus d'autres insultes que d'interventions mesurées), ou le domaine du message (certains groupes sont plus respectueux que d'autres de la netiquette). La recommandation principale afin d'éviter de heurter les participants à un groupe est celle du mimétisme : *"Know where you are in cyberspace. Netiquette varies from domain to domain. What's perfectly acceptable in one area may be dreadfully rude in another"* [Shea, 1994]. Il est fortement recommandé, lors de l'arrivée dans un forum ou toute autre communauté de discussion de rester en simple observateur (on utilise le terme "*lurker*", rôdeur) le temps d'apprendre les règles et habitudes propres à chaque groupe. Il est aussi recommandé de se reporter

aux FAQ (*Frequently Asked Questions*, Foire Aux Questions), pour éviter d'importuner les utilisateurs habituels avec des questions qui reviennent régulièrement, et auxquelles ils ont déjà répondu maintes fois.

“The net right now is a little New York in the late 19th century — waves of immigrants impinging themselves upon an established society. Not surprisingly, the newcomers don't always behave according to local custom, and members of the old society are sometimes suspicious and resentful. [...] Nevertheless, some of the most unpleasant conflicts in cyberspace history have been caused by newcomers who decided to join the fun and, in their ignorance, broke all the rules.” [Shea 94]

1.3.3 Relations personnelles sans face-à-face

L'utilisation de media qui “desincarnent” la communication pose le problème des relations personnelles. Comme nous l'avons dit tout au long de ce chapitre, les conditions d'existence d'une communauté tiennent principalement sur les participants et leurs interactions. Des amitiés, des complicités, des haines, facteurs fondamentaux des communautés “réelles” peuvent-elles se construire en utilisant des moyens de communication frustrés comme le *chat* ou le courrier électronique? Comme nous l'avons plusieurs fois dit au long des sections précédentes, bien plus que le contenu (les communautés peuvent se construire autour d'un thème, d'une idée, d'une particularité...) ce sont les gens qui la composent qui font une communauté. Comme le synthétise [Parks et Floyd, 1996], deux visions opposées existent. D'un coté ceux qui affirment que les relations *en ligne* sont superficielles et impersonnelles, et n'entraînent qu'une *illusion de communauté*. De l'autre, ceux qui soutiennent que grâce à la libération des contraintes physiques et au relâchement (et non l'annulation) des pressions sociales, il existe des opportunités à la création de nouvelles, mais *authentiques*, relations personnelles et communautés.

Deux points sont importants à prendre en compte lorsque l'on considère le développement de relations dans le cadre du cyberspace. Tout d'abord, le point déjà présenté de la dégradation due au manque de face-à-face. Le deuxième point relève plus du domaine du comportement. On constate en effet une profusion d'insultes, d'infractions à la “bienséance”. Plusieurs paramètres concourent à cet état de fait : le relatif anonymat, le mode de communication, textuel, qui allie la facilité d'expression de l'oral à la pérennité des données écrites, ou encore le délai de feed-back (court dans le cas du *chat* par exemple, plus long dans le cas des forums de discussion, mais, de toutes façons, plus long et moins précis que dans les situations de face à face), et les faibles conséquences de nos actes virtuels sur le monde réel.

Malgré ces restrictions, il a été montré que dans le cas des forums de discussion, certains participants construisent des relations (amitié, estime...) de la même façon que dans le cadre de relations de face-à-face (voir l'étude de Parks et Floyd [Parks et Floyd, 1996]). La différence majeure n'étant pas liée à la qualité des relations ainsi construites, mais au temps pris pour cette construction : *“Time is the key element [...] While the multiple channels and cues available in FtF [Face to Face] interaction speed the exchange of task and relational information, the process is slowed by the “reduced bandwidth” of CMC [Computer-Mediated Communication]”*. Parks et Floyd concluent en estimant que la nature particulière du cyberspace n'influe que peu sur la construction sociale : *“Cyberspace is simply another place to meet”*.

1.4 Conclusion

Au delà des objectifs de cette thèse sur la communication entre agents naturels et agents artificiels, nous cherchons à former des communautés mixtes, dans lesquelles l'agent serait vu comme un réel *partenaire*. Pour faciliter cette intégration, il est indispensable de s'intéresser aux communautés virtuelles, comme nous venons de le faire dans ce chapitre.

Nous retiendrons principalement que les études s'accordent sur le fait que la médiatisation des interactions n'est pas un obstacle à la création de réelles communautés. La détérioration due à la médiatisation peut être compensée par l'adhésion à un ensemble de règles particulières, comme la netiquette (qui, comme nous l'avons vu, reprend sous une forme explicite des éléments déjà présents dans les interactions de face-à-face, telle que la théorie des faces d'Erving Goffman). Elle peut aussi être compensée par l'envoi explicite de messages qui seraient spontanés en situation de face-à-face (comme les émoticônes qui donnent une indication sur l'état du locuteur). Typiquement, ces remarques relèvent des travaux de Parks et Floyd [Parks et Floyd, 1996]. Ceci permet de justifier le recours à des théories issues des sciences humaines pour développer un modèle d'agent capable de s'intégrer dans une communauté d'humains. En effet, pour éviter que nos agents ne soient considérés comme des intrus dans ces mondes certes fortement informatisés mais peuplés uniquement d'humains, il est nécessaire que l'agent se comporte d'une manière acceptable. Il est donc important de connaître ces communautés en ligne avant de définir le comportement à donner aux agents qui vont avoir à s'y intégrer.

Chapitre 2

Systemes multi-agents logiciels

*For the strength of the Pack is the Wolf,
and the strength of the Wolf is the Pack.*

RUDYARD KIPLING

2.1 L'Agent

Outre l'avantage de circonscrire un domaine et de cibler les travaux, une définition permet de marquer les différences d'un domaine d'étude à l'autre, ou à l'intérieur d'un même domaine. Typiquement, la définition à laquelle nous nous attacherons pour le terme "agent" nous permet de nous positionner vis-à-vis de la communauté IA, de la programmation Orientée Objet et, plus précisément, de l'Intelligence Artificielle Distribuée. De nombreuses définitions du terme "agent" existent et les seules étant un tant soit peu consensuelles parmi les chercheurs du domaine sont trop floues pour être utiles. C'est pourquoi nous préférons une définition ciblée et adaptée à nos travaux¹ mais parfois contestable, à une définition vaste mais trop imprécise pour être exploitable.

La définition que nous retiendrons est de Léonard Foner [Foner, 1993]. Elle est précise, et partant, plus restreinte, car ne considère comme agent que des entités en relations entre elles, typiquement un agent artificiel et un agent humain.

Léonard Foner requiert trois points (il les titre "*What's an Agent? crucial notions*") pour pouvoir parler d'agents :

Autonomie Ce point est le plus communément invoqué dans les définitions d'agents, mais aussi le plus discuté. Disons que, globalement, l'agent doit être capable de comportements qualifiables de spontanés, avoir une certaine initiative, proactivité, dans son action, le tout dans le but de satisfaire son utilisateur, ou, plus globalement, de remplir sa tâche.

Personnalisation Un agent doit pouvoir apprendre et évoluer dans sa méthode de traitement de sa tâche. Par personnalisation (*personalizability*), Foner considère l'agent comme s'adaptant à l'utilisateur qui l'exploite. On peut aussi considérer dans ce point le fait qu'avec son expérience, l'agent acquiert une "personnalité", qui le rend différent des autres agents, et ce point est crucial

¹malgré le côté *ad hoc* regrettable (un peu à la manière de Pablo Picasso définissant l'art : "*l'art, c'est la production de l'artiste*").

dans la caractérisation d'un agent face à un objet actif. Un agent est différencié, individualisé. Le point clé de la personnalisation est l'apprentissage.

Conversation Ce point est essentiel à la définition de Foner et en fait toute la spécificité. Il pose que pour être considérée comme un agent, une entité informatique doit posséder des capacités d'interaction avancées, permettant une rétroaction de l'utilisateur sur l'agent et de l'agent sur l'utilisateur ("*a two-way feedback*"). Il aborde explicitement la notion d'enchaînement conversationnel et d'évolution des conversations entre les deux partenaires au cours des différentes interactions.

Comme le fait remarquer Léonard Foner, selon ces critères, une boîte de vitesse automatique ou un ramasse-miettes (*garbage collector*) ne peuvent pas être considérés comme des agents, car il leur manque les capacités conversationnelles. Pourtant, selon certaines définitions, un ramasse-miettes qui prend des initiatives pour rendre mon environnement de travail plus efficace pourrait être un agent.

Cette section présentera l'agent sous deux angles complémentaires : interne et externe. Confronter un point de vue interne et un point de vue externe sur l'agent nous permet de bien distinguer le *statut* de l'agent (au sens de Christian Brassac et Sylvie Pesty [Brassac et Pesty, 1996], c'est-à-dire son ontologie), de son *rôle* (toujours au sens de Christian Brassac et Sylvie Pesty, c'est-à-dire son comportement).

2.1.1 Point de vue interne : ce qu'est l'agent

Il est particulièrement habituel de distinguer dans un agent des étapes de perception, de raisonnement et d'action. Le point de vue interne correspond à ces trois étapes, avec une emphase particulière sur celle de raisonnement, qui est l'étape en laquelle se passe la *décision* d'action (voir figure 2.1). Ce point de vue englobe l'ensemble des mécanismes *produisant* le comportement de l'agent. À ce niveau, une distinction classique dans le domaine des SMA se fait entre les agents *cognitifs* et les agents *réactifs*, catégories entre lesquels il existe un *continuum*.

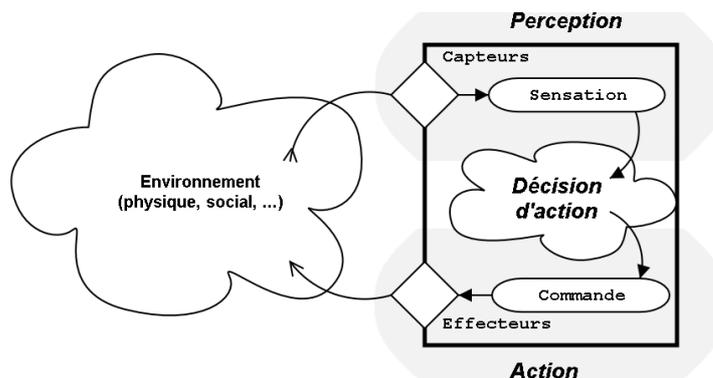


FIG. 2.1 – Structure classique d'un agent logiciel

Cette distinction classique se fait souvent en termes de complexité des mécanismes produisant le comportement de l'agent, l'agent cognitif pouvant faire appel à des représentations de l'environnement, des mécanismes d'apprentissage ou de planification, de communication directe avec d'autres agents, etc., l'agent réactif se

contentant de réagir de façon réflexe à son environnement. On a en général tendance à réserver le terme de *raisonnement* aux agents cognitifs, l'étape intermédiaire entre perception et action étant souvent réduite à une simple association chez les agents réactifs (à la manière des conditionnements stimuli / réponse des behavioriste).

Une autre façon intéressante de distinguer ces deux catégories serait de considérer comme plus cognitifs les agents manipulant plus de données temporelles, à la manière des considérations d'Henri Laborit [Laborit, 1994] :

“On distingue trois niveaux d'organisation de l'action. Le premier, le plus primitif, est incapable d'adaptation : à la suite d'une stimulation interne ou externe, il organise l'action de façon automatique. Le deuxième prend en compte l'expérience antérieure et la sensation qu'elle a provoquée : il fait donc appel à la mémoire. [...] Le troisième niveau est celui du désir, il est lié à l'élaboration imaginaire anticipatrice de la stratégie à mettre en œuvre pour assurer l'action gratifiante ou celle qui permettra d'éviter le stimulus nociceptif.

Le premier niveau ne se préoccupe que du processus présent, le deuxième niveau ajoute à l'action présente l'expérience du passé, le troisième niveau répond au présent, grâce à l'expérience passée, par une anticipation du résultat futur.”

Dans ses travaux Henri Laborit revient à l'origine de l'action pour traiter du comportement. Il est à cet égard intéressant de constater que le modèle d'agent cognitif le plus étudié, le modèle BDI, prend naissance avec la philosophie de l'action, à travers les travaux de Bratman.

Selon le principe de récursion [Demazeau, 1997], un système multi-agent peut être considéré comme un seul agent à un niveau supérieur d'abstraction. C'est-à-dire que, d'un point de vue interne, un agent peut être un système multi-agent. Nous garderons cette possibilité à l'esprit pendant tout ce chapitre, n'hésitant pas par exemple à attribuer des *buts* à un *système* multi-agent. Une telle attribution est beaucoup plus compréhensible si l'on considère que les buts sont rattachés à l'agent représentant le système plutôt qu'au système lui-même.

Nous ne développons pas dans ce chapitre les différentes architectures ou modèles d'agents existants, un panorama très complet ayant été réalisé par Olivier Boissier [Boissier, 2001].

2.1.2 Point de vue externe : ce que *fait* l'agent

Il est possible de dire que ce qui est important dans un agent, c'est ce qu'il fait, et non les mécanismes qui l'amènent à agir de telle ou telle façon. Une position proche dirait que seul le comportement peut être un objet d'étude, car seul le comportement est observable. Ce dernier point est à la base du mouvement behavioriste, pour lequel la psychologie doit étudier les comportements observables plutôt que les processus mentaux. Ces points de vue sont aussi partagés, dans une certaine mesure, par des chercheurs en IA et robotique comme Rodney Brooks. Une différence évidente entre la psychologie et l'IA est le support de leurs études : humains pour la psychologie, systèmes artificiels pour l'IA. Toutefois, on peut considérer une problématique commune dans la mesure où ces deux sciences s'intéressent au comportement de leur objet d'étude.

D'un point de vue fonctionnel, on pourrait effectivement considérer comme seul point pertinent le comportement des agents, et, dans une certaine mesure, cette attitude est effectivement suffisante. Même dans les situations où un agent chercherait à anticiper, simuler ou tout simplement modéliser le comportement d'un autre agent, les modèles qu'il emploierait n'ont pas à être congruents avec ceux effectivement employés par l'agent à modéliser. Après tout, certains systèmes cherchent bien à reproduire un comportement humain sans que les mécanismes de génération de comportement de l'humain ne soient connus. Ainsi, dans le cadre des modèles mis en place dans le domaine de la psychologie, *“tout se passe comme si [le système logique] existait. Il est une construction intellectuelle du psychologue qui l'aide à comprendre le sujet et à lui parler en « entrant dans sa logique »”* [Mucchielli, 1995].

Cependant, toute analyse fine du comportement ne peut que difficilement se passer de répondre à la question de l'ontologie de l'agent². En effet, l'aspect comportemental est la conséquence des mécanismes internes de l'agent. La prise en compte des aspects internes de l'agent permet d'affiner la connaissance (et l'exploitation de cette connaissance) issue de l'analyse du comportement perçu. Ainsi, par exemple, *“pour garantir la conformité des interprétations construites par la machine aux attentes des utilisateurs, le fonctionnement du système mis en œuvre doit présenter une certaine analogie avec celui de la cognition humaine”* [Sabah, 1997a].

Le risque majeur de limiter un agent à son comportement est l'assimilation : *“mêmes comportement → mêmes phénomènes en soi”*. Comme le dit Searle (cité par Christian Brassac et Sylvie Pesty [Brassac et Pesty, 1996]) : *“si ce principe était correct, il nous faudrait tous conclure que les postes de radio sont conscients parce qu'ils manifestent un comportement verbal intelligent”*. Il est donc important de ne pas se limiter à l'étude du comportement lorsqu'il est possible d'avoir accès aux mécanismes qui *produisent* ce comportement.

2.2 De l'agent au collectif d'agents

Un système comportant un agent unique est un cas pour le moins trivial dans le domaine des systèmes *multi-agents*. Comme le fait remarquer Christof Baeijs dans sa thèse [Baeijs, 1998], cette *“organisation à un membre”* peut se justifier d'un point de vue théorique, grâce au principe de récursion³, comme étant la représentation d'un système multi-agent. Ce cas limite mis à part, les systèmes multi-agents se composent de plusieurs agents en interaction.

Nous présentons dans cette section deux composantes d'un SMA intimement liées à la création d'un collectif d'agents : les Interactions et l'Organisation.

2.2.1 L'Interaction

Pour pouvoir considérer les agents comme un collectif, il doit exister entre eux des possibilités d'interaction, sans quoi tout se passe comme si les agents n'étaient pas même en présence. Les moyens d'interaction employés dans les systèmes multi-

²Citons Fodor : *“D'une part, toutes les capacités comportementales qui sont distinctes à première vue n'ont pas des étologies réellement différentes, et l'élaboration de théories a précisément pour but de découvrir les régularités causales qui se cachent derrière les apparences superficielles. D'autre part, certaines capacités résultent certainement de l'interaction de causes sous-jacentes [...]”*[Fodor, 1983]

³Un système multi-agent peut être vu comme un agent à un niveau supérieur.

agents vont de modèles issus de la physique (tels que des modèles à base de forces) à des types d'interaction de haut niveau (tels que des actes de langage adressés).

Cette section est volontairement brève, l'interaction entre agents logiciels étant plus particulièrement traitée dans le chapitre 4.

a) Modes de communication

Parmi les différentes méthodes de communication utilisées, deux grandes catégories peuvent se distinguer, sur la base de la destination du message. Dans la première catégorie, la communication adressée, le destinataire est un ou des agents. Dans la seconde, la communication par l'environnement, le message est déposé dans un espace commun.

Communication adressée La communication adressée est une communication directe. Certains des paramètres spécifiant le message portent l'identifiant du ou des destinataires. Ce mode de communication est marqué par les techniques classiques de la communication en informatique, typiquement le modèle de Shannon. Dans cette situation, l'agent émetteur a un rôle actif, tandis que l'agent récepteur est passif dans sa *réception* le message. Le message est déplacé *jusqu'à* son destinataire.

C'est le cas des messages basés sur KQML ou FIPA-ACL (voir section 4.2.1).

Communication par l'environnement Une communication par l'environnement est une communication indirecte. Là où la communication adressée envoie un message à un ou plusieurs agents, la communication par l'environnement dépose le message dans un ou plusieurs environnements. Le message n'atteint ensuite les autres agents que lorsque ceux-ci iront le *percevoir* dans l'environnement. Dans cette situation, l'ensemble des agents a un rôle actif, la simple réception, passive dans le cas de la communication adressée, se transformant en une consultation de l'environnement. Le message est déplacé jusqu'à une étape intermédiaire où le destinataire va le chercher.

C'est le cas de la communication par tableaux noirs ou par dépôt et consultation de *traces* dans l'environnement.

Une différence négociable Bien que, au vu de leurs caractéristiques, la différence semble assez claire entre les deux catégories de modes de communication que nous venons de présenter, il est souvent possible de confondre le premier dans le second suivant le niveau auquel le système en interaction est considéré. Ainsi, il est toujours possible de considérer qu'un message adressé est déposé dans un environnement commun à l'émetteur et aux destinataires (après tout, tout message est porté par un médium...), et que l'attitude d'attente du message (propre aux récepteurs) est une perception de cet environnement commun. La réciproque est plus délicate, tant les environnements peuvent être porteurs de simples traces⁴ alors que les messages adressés sont (*a priori*) porteurs de symboles⁵. La communication par l'environnement permet le dépôt et la perception de messages non-voulus : les traces laissées dans l'environnement ne sont pas systématiquement liées à une décision, alors que tous les messages adressés sont *décidés* par l'émetteur.

⁴Au sens de Pierce, c'est-à-dire que le signe entretient un rapport physique, direct, avec la chose considérée.

⁵Au sens de Pierce, c'est-à-dire que le signe entretient un rapport conventionnel et arbitraire avec la chose considérée (le message).

b) Interaction et Environnement

Nous avons vu dans la sous-section précédente qu'il était possible de communiquer à travers l'environnement. Beaucoup de modèles d'agents (il suffit de consulter le panorama proposé par Olivier Boissier [Boissier, 2001]) intègrent des capacités de perception et d'action *et* des capacités de communication, distinctes. Les premières s'adressent à l'environnement, les secondes aux agents⁶. Une telle distinction entraîne que les agents sont implicitement (et parfois explicitement) considérés comme ayant des relations privilégiées entre eux, et non comme de simples éléments de l'environnement.

Distinguer l'échange de messages *via* un environnement (processus classiquement dénommé perception/action) de l'échange direct de messages entre des agents (processus classiquement dénommé communication), peut toutefois entraîner quelques confusions. Ainsi, si la perception de traces dans l'environnement relève clairement de la communication par l'environnement, la perception du comportement d'un agent en relève moins clairement, et que dire si ce comportement est spécifiquement destiné à un autre agent ?

De plus, la progression de la communication adressée sur la base d'*actes* de langage (pour les actes de langages, voir la section 4.1 et pour l'ACL-FIPA, un standard de la communication entre agents qui se veut basé sur les actes de langage, voir section 4.2.1), entraîne que chaque message *est* une action. L'évolution naturelle de ce genre de communication devrait amener les messages adressés au même niveau que toute autre action de l'agent, rendant caduque l'existence d'un module de communication entrant directement en contact avec les autres agents, et limitant les entrées sorties de l'agent à ses mécanismes de perceptions et d'action.

2.2.2 L'Organisation

La partie Organisation d'un SMA correspond à la façon dont s'articulent les relations d'un agent à l'autre. A l'intérieur d'un groupe d'humains, on emploierait le terme de *relations* au sens large. L'aspect Organisation d'un SMA rassemble les rapports entre agents, des plus formalisés (hiérarchie, statut, rôle) aux plus subjectifs (respect, confiance, préférences). Cet aspect du système est intimement lié à la partie Interaction, sans laquelle les relations à "l'autre" sont impossibles.

Nous présenterons cette section en reprenant la distribution sur trois dimensions faite par Amal El Fallah [El Fallah, 2001] : les dimensions fonctionnelle, spatiale et temporelle.

a) Distribution fonctionnelle

La distribution fonctionnelle répartit aux différents agents leurs rôles (et statuts) en tenant compte de leur aptitude à tenir ces rôles. Cette distribution spécifie aussi les liens entre agents (relations de pouvoir, interactions, etc.). Dans la littérature, l'organisation d'un système est souvent réduite à ce point tant ce choix est prépondérant au niveau de la fonctionnalité du système.

⁶Souvent, les architectures d'agent dits "réactifs" se limitent à la perception/action, les compétences dédiées spécifiquement à la communication agent/agent étant plutôt réservées à des agents dits "cognitifs".

Cristof Baeijs énumère dans sa thèse [Baeijs, 1998] cinq grandes catégories d'organisations (l'organisation à membre unique, le groupe, la hiérarchie (à un ou plusieurs niveaux), l'organisation décentralisée et le marché) qui correspondent à des relations de contrôle et de communication entre agents et entre agents et ressources. Un même système multi-agent peut être composé de différents sous-groupes relevant chacun d'un type d'organisation différent (ne serait-ce qu'en vertu du principe de récursion).

D'une manière assez générale, l'augmentation de complexité d'une structure organisationnelle va de pair avec l'augmentation de son adaptabilité et avec l'augmentation de la quantité de messages échangés au sein du système. L'analyse de Cristof Baeijs montre que, dans le cadre d'agents réactifs, il existe une complémentarité entre la coordination ("*coût de maintenance des liens de communication et coût des échanges de messages*") et la vulnérabilité ("*coût nécessaire pour s'adapter à un changement de l'environnement ou de l'organisation*").

Les notions de rôle et statut seront développées et mises en relations avec leur équivalent dans les sciences sociales page 113 et suivantes.

b) Distribution spatiale

La distribution spatiale se rapporte à la place de l'agent dans l'environnement. Cet aspect de l'organisation est intimement liée aux environnements que l'agent partage avec d'autres agents. Ainsi, dans le cadre de l'application Microb-2 [Drogoul, 2000], application à la robocup, le terrain (l'environnement) est séparé en trois zones qui conditionnent trois sous-équipes aux propriétés distinctes :

- l'équipe d'attaque, composée d'agent réactifs auto-organisés ;
- l'équipe de milieu de terrain, dotée d'une organisation dynamique ;
- l'équipe de défense, disposant de capacités d'apprentissage distribué.

Le terme "spatial" peut induire en erreur, car les environnements dans lesquels les agents ont une place ne sont pas forcément de nature spatiale au sens commun du terme⁷. Ainsi, le projet COMRIS [Van de Velde, 1997] plonge l'agent dans un espace social représentant les centres d'intérêt de l'humain qu'il représente.

c) Distribution temporelle

La distribution temporelle concerne les aspects dynamiques des organisations. Dans le cadre d'organisations statiques, la dimension temporelle n'existe pas ; dans le cadre d'organisations dynamiques, elle conduit les variations le long des deux autres axes, fonctionnel et spatial.

L'intérêt d'une organisation statique est essentiellement en terme de coût de communication et de résolution de conflits. En effet, en fixant une organisation précise, le concepteur élimine *a priori* plusieurs conflits susceptibles d'apparaître dans les interactions entre les agents [El Fallah, 2001]. De même, le concepteur peut spécifier de manière particulièrement précise les méthodes d'interaction à mettre en œuvre. Ces avantages apparaissent aux dépens de l'autonomie interactionnelle de l'agent

Les organisations dynamiques permettent de leur côté de s'adapter, au sens large, c'est-à-dire réagir à des changements d'objectifs, à l'arrivée ou à la sortie d'agents,

⁷Environnement "physique" (ou représentation d'un tel environnement) de deux ou trois dimensions, hauteur, largeur, profondeur.

à la modification de l'environnement, etc. Cette capacité d'adaptation s'acquiert souvent au prix d'une plus grande activité de communication entre les agents et/ou une plus grande complexité interne de l'agent. Face à ce problème, Kelly Fernandes [Fernandes, 2001] propose un système multi-agent basé sur une hiérarchie, capable d'ajouter ou de retirer des niveaux à cette hiérarchie (en ajoutant ou retirant des agents) selon la difficulté du problème à traiter. Si le problème est trop complexe, le système s'adapte en faisant appel à d'autres agents, si le problème est "trop" simple, les agents superflus sont supprimés. Ainsi, la complexité de la structure est précisément adaptée à la complexité du problème (les pénalisations dues aux interactions sont minimales), mais, bien sûr, au prix de mécanismes (parfois coûteux) d'adaptation de la structure.

d) Liens entre ces trois distributions

Les trois axes que nous venons de présenter ne sont pas indépendants. Nous pouvons le voir, en prenant l'exemple de la robocup (simulation). Dans une équipe où les agents ont tous les mêmes capacités, leur comportement sera essentiellement conditionné par leur position sur le terrain, la position de la balle et celle des adversaires. L'organisation spatiale des agents peut entraîner l'attribution d'un statut d'attaquant (organisation fonctionnelle) à un agent positionné de manière avantageuse [Collinot *et al.*, 1996]. Inversement, si l'organisation fonctionnelle est fixe (pas de variation sur la dimension temporelle), un agent ayant un statut de défenseur restera dans une zone de terrain donnée. De manière triviale, notons aussi que le déplacement d'un agent entraîne un changement de distribution spatiale le long de l'axe temporel.

e) Où est l'organisation ?

Nous venons de voir que l'Organisation comprenait les liens de pouvoir et de devoir entre agents, distribuait des rôles, conditionnait des déplacements, etc. Mais l'Organisation n'est pas forcément une entité en soi. Elle peut certes avoir une existence en marge des agents qu'elle rassemble, mais peut aussi n'exister qu'à travers des représentations locales (et éventuellement fractionnaires) internes à chaque agent, ou encore n'être identifiable que par un observateur extérieur au système. Le premier exemple que nous venons de donner considère l'Organisation "*comme une structure externe par rapport aux agents, et [la] représente comme un objet externe ou encore comme un agent d'un méta-niveau*"⁸[Baeijs, 1998]. Le deuxième exemple "*définit l'organisation comme un objet abstrait dont la représentation est distribuée parmi les membres qui la constituent*" [Baeijs, 1998]. Le troisième et dernier exemple relève de l'organisation émergente, dans laquelle aucune représentation explicite de l'Organisation n'existe ni au niveau du système ni au niveau des agents qui le compose.

⁸Dans le domaine de la sociologie, Émile Durkehim dirait de cet aspect de l'Organisation : "*Il faut considérer les phénomènes sociaux en eux-mêmes, détachés des sujets conscients qui se les représentent.*" ("Les règles de la méthode sociologique", cité par Patrick Champagne [Champagne, 1997]).

2.3 Juxtaposition d'agents ou communauté?

Au sein de différents SMA, certains agents peuvent coopérer, d'autres s'ignorer; certains peuvent obéir à un contrôleur central, d'autres disposer d'une large autonomie; certains peuvent suivre des objectifs qui leurs sont propres, d'autres travailler à un but collectif ou aux buts d'autres agents, et ainsi de suite.

La littérature abonde de définitions du concept d'agent, mais est moins abondante en ce qui concerne les *systèmes* multi-agents. *A minima*, on peut dire que toute entité dont le comportement est influencé par une autre entité forme le germe d'un SMA. La définition laconique de Brahim Chaïb-Draa et ses collègues est même encore plus vaste puisque pour eux "*un système multi-agents est un système distribué composé d'un ensemble d'agents*" [Chaïb-Draa et al., 2001]. Les contraintes qu'ils apportent par la suite précisent que "*généralement*", chaque agent a un point de vue partiel, qu'il n'y a pas de contrôle global, que les données sont décentralisées et que le calcul est asynchrone.

Cependant, indépendamment d'une définition stricte d'un SMA, il est possible de considérer des situations d'agents en interaction dont la différence sera suffisamment marquée pour montrer la variété des situations possibles. Nous distinguerons donc trois cas : le premier comparera les systèmes (ou agents) où le comportement naît d'objectifs explicites à ceux où le comportement est issu d'un réflexe; le deuxième comparera les systèmes (ou agents) où les buts sont partagés par plusieurs agents à ceux où les buts sont propres aux agents; le troisième comparera les systèmes (ou agents) où le but est unique et fixe à ceux où les buts sont multiples et susceptibles d'évoluer. Dans ces présentations, nous considérerons le système comme tout regroupement d'agent, quelle que soit la structure de ce regroupement et quel que soit le nombre d'agents impliqués dans ce regroupement (particulièrement, il peut s'agir d'un sous-groupe, et non d'un système "dans son ensemble").

2.3.1 Comportement à base de buts explicites ou sans buts

a) Pour les agents

Selon Jacques Ferber [Ferber, 1995], les agents sont mus par des tendances. Ces tendances peuvent trouver leur source au sein de l'agent lui-même ou dans l'environnement. Dans le premier cas, où le comportement de l'agent est dirigé vers des buts explicites, Jacques Ferber parle de comportement *téléonomique*; dans le second, où le comportement est guidé par les perceptions de l'agent, il parle de comportement *réflexe*.

La distinction entre le comportement téléonomique et comportement réflexe se base en fait sur des critères *internes* à l'agent, c'est-à-dire sur l'analyse de leur principe de fonctionnement (interne) et non de leur comportement (externe). Un observateur extérieur n'est pas *a priori* et dans le cas général capable de distinguer ces deux types d'agents sur la seule base de leurs comportements.

b) Pour les systèmes

Il est plus délicat de distinguer si, d'un point de vue interne, la fonctionnalité globale (le comportement) du *système* est basé sur des buts explicites. Toutefois, un but collectif à l'échelle du système peut valoir pour but du système. Un tel but global peut préexister au système : dans le cadre de la Résolution Distribuée de

Problèmes (RDP) ce but *est* la fonction à remplir par le système et il est décomposé en sous-tâches *lors de la spécification du système* jusqu'à l'attribution de tâches à des agents. Mais pour certains systèmes la fonctionnalité globale (quand il est possible d'en distinguer une) est construite (et constatée) *lors du fonctionnement du système*. On parle alors de fonctionnalité émergente. On retrouve cette fonctionnalité collective dans les systèmes naturels, chez les insectes sociaux constructeurs par exemple (guêpes, abeilles, termites...), où le plan général de l'édifice n'est pas connu par les individus.

2.3.2 Actions personnelles ou collectives

Ce sont les actions collectives qui font l'essentiel de la spécificité du domaine des SMA : le fait que plusieurs entités agissent collectivement. Principalement, il est des actions de l'agent qui n'ont d'intérêt que mises en relation avec les actions d'autres agents.

D'un point de vue interne, pour des agents disposant de buts explicites, il est possible de voir si les buts de ces agents sont choisis en fonction d'un but collectif. On peut dans cette situation parler de *coopération* : les actions des agents sont communes et suivent l'identification et l'adoption d'un but commun [Ferber, 1995]. Dans cette situation, la coopération est issue d'une volonté à la fois individuelle et collective [Brassac et Pesty, 1996]. Les buts et actions *personnels* de l'agent sont *faits pour* créer un résultat collectif.

Mais que les agents disposent ou non de buts explicites, un observateur extérieur peut tenter de définir si les actions entreprises par les agents concourent en un effet à l'échelle d'un groupe. Si l'on n'a pas d'information sur le fonctionnement interne de l'agent, on ne peut dans cette situation que parler de *co-action* : les actions des agents ont un effet commun, mais cet effet n'est que la conséquence de la mise en commun des actions, et non d'une intention collective qui guiderait ces actions [Brassac et Pesty, 1996]. Dans cette situation, la coopération n'existe que du point de vue de l'observateur. Les actions *personnelles* de l'agent *apparaissent comme* créant un résultat collectif.

Que les agents soient coopératifs ou non, et que cette coopération soit voulue ou constatée, les agents se retrouvent parfois en situation de conflit. Diverses situations de conflits ont été répertoriées dans le domaine des SMA, comme les conflits de ressources (quand plusieurs agents cherchent à accéder à une ressource non partageable) ou les conflits d'objectifs (quand certaines actions d'un agent sont opposées aux objectifs d'un autre). Selon Ferber [Ferber, 1995], une des caractéristiques de la coopération est qu'une partie des actions des agents est destinée à éviter ou à sortir de ces conflits.

Dans certains systèmes, le conflit est la base du fonctionnement, comme dans le cas de systèmes compétitifs (appels d'offre ou marchés par exemple). Dans ces situations, l'ensemble des agents participe collectivement à la fonctionnalité du système, bien qu'ils soient (parfois systématiquement) localement en conflit.

La coordination des agents participant à des actions collectives a été traitée récemment, pour les agents cognitifs par Amal El Fallah [El Fallah, 2001], et pour les agents réactifs (à travers un point de vue organisationnel) par Christof Baeijs [Baeijs, 1998].

2.3.3 Rôle unique ou multiple, figé ou évolutif

Les objectifs des agents (ou des systèmes) ne sont pas systématiquement uniques et figés, comme dans le cas de la Résolution Distribuée de Problèmes. Ainsi, le système lui-même peut avoir à atteindre de manière séquentielle une série de sous-buts, et va donc avoir à changer ses objectifs, ce qui peut entraîner une redistribution des tâches des agents (cas des agents voyageurs qui réservent un ensemble de moyens de transports le long d'un trajet en plusieurs étapes par exemple) ; des agents peuvent décider temporairement d'un objectif commun, le temps d'une coalition (cas des agents déménageurs ayant ponctuellement à déplacer un objet trop lourd ou encombrant pour un seul d'entre-eux par exemple) ; un agent peut avoir un comportement opportuniste, avec plusieurs objectifs et passant de l'un à l'autre au gré de leur faisabilité (cas d'un agent voyageur ayant simultanément des objectifs sur des moyens de transports et des hébergements par exemple) ; etc. L'accomplissement d'un but suivi, la modification de l'environnement (principalement les ressources), l'arrivée ou le départ d'agents sont autant de raisons pouvant influencer sur le comportement de l'agent. Les Systèmes Multi-Agents Ouverts (SMAO) [Kozlak *et al.*, 1999] par exemple posent avec force le problème de la (re)répartition des tâches (si des objectifs collectifs existent) et du maintien de l'intégrité fonctionnelle du système.

Les situations sont multiples, mais beaucoup des combinaisons des trois critères (rôle unique ou multiple, figé ou évolutif, considéré au niveau de l'agent ou du système) ont déjà été explorés dans le domaine des systèmes multi-agents. Des notions assez générales sur l'adaptation ou le maintien d'un comportement existent, principalement à travers l'évaluation de fonctions d'adéquation, ou bien à travers la prise en compte explicite de l'intégrité fonctionnelle du système [Kozlak, 2000] (rappelons une fois de plus que ce que nous considérons au niveau d'un agent peut être considéré au niveau d'un système, et *vice versa*). Dans une certaine mesure, ces approches rejoignent la notion d'homéostasie en biologie du comportement [Laborit, 1994].

De la même façon que pour les organisations, la flexibilité des agents dans leurs changements de buts s'effectue souvent au prix d'une augmentation de la quantité des interactions. Mais cette adaptabilité est sensée permettre aux agents d'agir au mieux en fonction du contexte (environnemental, organisationnel, etc.) dans lequel ils sont plongés.

2.4 Conclusion

La spécificité principale du domaine des systèmes multi-agents logiciels est constitué par leur aspect collectif. Comme nous l'avons vu dans la dernière section, il existe de nombreuses façons d'associer les agents, et les choix faits pour cet assemblage influent directement sur le comportement collectif. Toutefois, bien que l'objectif à long terme dans lequel s'inscrit cette thèse est la création de communautés mixtes, nous nous concentrons, dans une première étape, sur la communication entre agents naturels et agents artificiels. C'est donc principalement la notion d'interaction et les architectures d'agent adaptées à ces interactions qui nous intéressent.

Présenter chez les agents la distinction statut/rôle (au sens de Christian Brassac et Sylvie Pesty) comme nous l'avons fait tout au long de ce chapitre nous permet de préciser que notre objectif n'est pas de créer des vrais "gens", mais des *acteurs*. Comme l'a fait remarquer Ken Perlin lors de Virtual World 2000, "*lorsque, dans un*

*film, on a besoin d'un docteur, on ne va pas chercher un vrai docteur, on prend un comédien qui joue un docteur". Cette vision des choses nous permet de nous débarrasser du problème de l'intelligence ontologique de l'agent pour nous contenter d'une intelligence apparente. Nos agents n'ont pas à être intelligents (si tant est que cela soit théoriquement, ou plutôt, philosophiquement, possible), il leur suffit de simuler l'intelligence. Simplement, comme le note Gérard Sabah, plus le fonctionnement interne de l'agent présentera une analogie avec le fonctionnement de l'intelligence, plus l'agent simulera l'intelligence avec efficacité [Sabah, 1997a]. Ainsi, plus le fonctionnement interne de l'agent présentera une analogie avec le fonctionnement interne de l'humain, plus leur interaction en sera facilitée. La recopie des principes de fonctionnement de l'intelligence découverts chez les humains est probablement une piste intéressante pour faire évoluer l'intelligence *artificielle*. Mais cette approche peut aussi montrer des limites. Ainsi, les avions ont commencé à voler efficacement à partir du moment où leurs concepteurs ont cessé de vouloir recopier un système pourtant fonctionnel (les oiseaux) et ont arrêté d'essayer de battre des ailes.*

Chapitre 3

Communautés mixtes

Les observations des anthropologues et des psychologues de l'enfance montrent l'existence chez l'homme d'une sorte de besoin fondamental : besoin social d'établir une relation avec un semblable, besoin de communiquer.

ALEX MUCCHIELLI, Cybernétique et cerveau humain.

Comme nous le montrerons dans la première section de ce chapitre (section 3.1), la frontière entre le monde réel et les mondes virtuels tend à s'estomper. Poussé par ce rapprochement des *espaces*, on constate un rapprochement des *membres* de ces différents espaces (humains pour le monde réel, agents pour les *cyberespaces*), entraînant la naissance de groupes d'agents et d'humains interagissant (section 3.2). Mais pour parvenir à une situation permettant d'exploiter au mieux les capacités spécifiques de chacun, il est nécessaire que chaque type d'interactant (humain et agent artificiel) s'adapte en partie à l'autre (sections 3.3 et 3.4).

De façon à illustrer nos propos, la section 3.5 présentera quelques agents qui nous semblent adaptés pour s'intégrer dans une communauté mixte.

3.1 Intégration monde réel / monde virtuel

Les récents progrès technologiques tendent à gommer l'opposition entre les deux parties du terme "réalité virtuelle", et la frontière qu'on croyait nette entre "réel" et "virtuel" prend de plus en plus la forme d'un *continuum* [Demazeau, 1999]. En prenant le point de vue de Paul Milgram et Fumio Kishin [Milgram et Kishin, 1994] on peut considérer un *continuum* allant du monde réel au monde virtuel (Cf. Figure 3.1), en passant par les étapes de la *réalité augmentée* (le monde réel s'enrichit de données issues de mondes virtuels) et de la *virtualité augmentée* (le monde virtuel contient des éléments issus du monde réel).

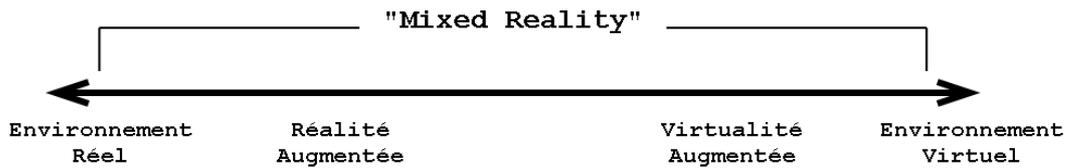


FIG. 3.1 – Représentation simplifiée d’un “continuum de la virtualité” ([Milgram et Kishin, 1994])



FIG. 3.2 – Un exemple de réalité augmentée : le *magic book*² (photo ©ATR MIC Labs)

Le long de cette représentation simplifiée (aux dires mêmes des auteurs, et nous reviendrons sur ce point plus loin), se situent les exemples de systèmes suivants en *réalité augmentée* et en *virtualité augmentée* :

- **réalité augmentée** : On peut classer dans cette catégorie les systèmes centrés sur le réel (du fait de la “réalité” de leurs environnements dominants, des agents principaux ou des objets centraux des tâches) qui sont enrichis par des systèmes centrés sur de l’information. Les applications aujourd’hui les plus avancées de la réalité augmentée se trouvent en médecine (préparation d’opérations chirurgicales et assistance lors des opérations proprement dites). Les travaux en cours dans le domaine ont un large spectre d’applications, depuis le *communityware* jusqu’aux applications ludiques. Bien que les applications les plus connues de la réalité augmentée enrichissent le réel par des données perçues visuellement se superposant au réel (voir le *magic book*, figure 3.2, qui est assez typique, ou le “tableau magique”³ développé par le CLIPS à Grenoble), il existe de nombreux travaux dans lesquels les informations sont de nature plus diverses (informations sur les interlocuteurs [Van de Velde, 1997], assistance à la navigation [Nagao, 1998], ...)

- **Virtualité augmentée** : Inversement, la virtualité augmentée enrichit des environnements virtuels, donc issus de systèmes d’information, par des éléments du réel. On retrouve ici aussi beaucoup de travaux sur le *communityware*, comme le projet FreeWalk [Nakanishi *et al.*, 1998] qui est un système de télé-conférence, basé sur un espace virtuel dans lequel les participants sont représentés par des tableaux sur lesquels sont projetés leurs images vidéos, comme dans le cadre d’une visio-conférence classique.

Il faut noter que Paul Milgram et Fumio Kishin [Milgram et Kishin, 1994] ne présentent ce *continuum* de la virtualité que comme une représentation *simplifiée*, dont les limites apparaissent rapidement dès lors que l’on cherche à ordonner les

Il faut noter que Paul Milgram et Fumio Kishin [Milgram et Kishin, 1994] ne présentent ce *continuum* de la virtualité que comme une représentation *simplifiée*, dont les limites apparaissent rapidement dès lors que l’on cherche à ordonner les

²Dans une de ses formes, le *magic book* permet de superposer à un livre des graphismes (personnages, scènes, ...) éventuellement liés au contenu du livre (Cf. <http://www.hitl.washington.edu/magicbook/>)

³qui permet à la fois la sauvegarde sous forme électronique des informations écrites par l’utilisateur et la projection d’informations sur le tableau (<http://iihm.imag.fr/demos/magicboard/>)

systèmes proches de la zone médiane. Ils proposent donc une méthode de classement plus précise, basée sur les trois axes suivants :

1. **Quantité de connaissance sur le monde** (*Extent of World Knowledge*) : la quantité de connaissance que le système de médiation a sur les informations qu’il représente, allant de “rien” pour une vidéo à “tout” pour un monde entièrement modélisé ;
2. **Fidélité de présentation** (*Reproduction Fidelity*) : la qualité, la précision, le réalisme, . . . de ce qui est présenté aux utilisateurs, allant de “faible” pour une représentation 3D en “fil de fer” à “très élevée” pour la télévision 3D haute fidélité ;
3. **Sensation de présence** (*Extent of Presence Metaphor*) : le point jusqu’auquel l’utilisateur se sent présent, impliqué, inséré dans le monde qui lui est présenté, allant de l’image unique présentée sur un écran classique jusqu’aux HMD⁴ présentant un monde en “vue subjective”.

Toutefois, il ne faut pas oublier que cette classification [Milgram et Kishin, 1994] est celle de systèmes de visualisation (comme le titre nous en prévient : “*A taxonomy of mixed reality visual displays*”), d’autres classifications existent, comme celle d’Emmanuel Dubois [Dubois *et al.*, 2000] par exemple, qui discriminent suivant un “objet de la tâche”, central. Néanmoins, elles nous permettent de bien sentir les imprécisions (et les difficultés de classement) pouvant émerger de la considération d’un axe simplifié “Réel – Virtuel”.

La communauté virtuelle sur laquelle nous avons basé notre application, le Deuxième Monde (voir section 1.2.5), est construite autour d’un cyberspace très proche de l’extrémité “virtuel” du diagramme de Milgram et Kishin représenté page 46. Il n’est toutefois pas purement virtuel car, si l’on reprend les trois axes présentés ci-dessus, on a :

- **Quantité de connaissance sur le monde** : le monde, bien qu’entièrement modélisé, peut intégrer des retransmissions vidéos ou audio, lors d’interviews ou de débats par exemple.
- **Fidélité de présentation** : plusieurs quartiers de Paris (Louvre, jardin des Tuileries ou place des Vosges, entre autres) ont été particulièrement travaillés au niveau du réalisme. Par exemple, certaines façades ont été photographiées puis digitalisées pour être ensuite intégrées dans le monde virtuel, la couleur du ciel et la luminosité changent en fonction de l’heure de la journée, on entend l’eau des fontaines couler lorsqu’on s’en approche, etc.
- **Sensation de présence** : c’est le point qui rend le Deuxième Monde le plus “virtuel”. En effet, il est un monde entièrement modélisé, avec des contraintes fortes au niveau de la complexité des scènes (les visiteurs doivent pouvoir se déplacer de manière fluide dans le monde) et de la quantité d’information (la transmission de la description de la scène ne doit pas être trop longue pour un internaute ne disposant pas d’une connexion à haut débit). Ces contraintes, ajoutées aux capacités limitées de l’interface, très classique (clavier, écran, souris et haut-parleurs), limitent la sensation de présence. La superposition d’un paysage sonore au paysage visuel dans certaines scènes (bruits de fontaine, d’oiseaux, . . .) permet toutefois d’augmenter cette sensation en respectant les contraintes précédentes.

⁴Head Mounted Display, “casque” équipé de deux écrans permettant une vision stéréoscopique, et parfois même d’écouteurs pour une immersion dans un paysage sonore tridimensionnel.

3.2 Qu'est-ce qu'une communauté mixte ?

Avec ce recouvrement des mondes réels et virtuels, se multiplient les situations d'interaction entre les "habitants" de ces différents mondes : humains et agents. On retrouve même des situations dans lesquelles l'humain et l'agent ont des rôles interchangeable⁵. Ainsi, la plate-forme présentée par Abdenour Bouzouane et ses collaborateurs [Bouzouane *et al.*, 1998] propose des jeux de rôle où les éventuels participants manquants sont remplacés par des agents. De même, les agents de type Steve (voir page 73) sont capables, indifféremment, d'interagir entre eux ou avec des humains, dans des tâches d'enseignement (un agent, plusieurs humains [Rickel et Johnson, 2000]), de jeu de rôle simple (un humain, plusieurs agents [Rickel *et al.*, 2001]) ou d'entraînement (plusieurs agents, plusieurs humains). Si les situations où seuls des humains interagissent ensemble peuvent définir une communauté virtuelle, et les situations où seuls des agents interagissent ensemble relèvent du domaine des SMA, celles mettant en interaction des agents et des humains peuvent appartenir à une troisième voie, celle des communautés mixtes (voir le tableau récapitulatif page 4).

Nous considérons donc toute situation dans laquelle des humains et des agents interagissent comme étant mixte, indépendamment des environnements dans lesquels ces interactions prennent place et indépendamment des interfaces nécessaires aux uns et aux autres pour permettre ces interactions (en effet, puisque les uns sont issus d'un monde réel et les autres d'un monde d'informations, il doit *nécessairement* exister des interfaces entre le monde réel et le monde virtuel).

Par rapport à ce qui a été dit dans la section précédente sur les réalités mixtes, où le terme *mixte* peut être employé pour qualifier des *espaces* (ou *mondes*), nous emploierons principalement cet adjectif pour qualifier des *communautés*, indépendamment d'une tâche précise ou d'un mode d'affichage donné. La mixité que nous recherchons est au niveau de *la nature* des agents, naturels ou artificiels, quels que soient les environnements dans lesquels se déroulent leurs interactions. Il est clair sur la figure 2 p.4 que notre notion de communauté mixte est transversale et indépendante de la nature de l'environnement.

Le long de l'axe de la figure 3.1, on peut associer à chaque combinaison d'environnements des rassemblements mixtes d'agents humain et artificiels, formant parfois des communautés. Cet exercice a déjà été fait en introduction de cette partie sur les communautés et nous l'avons synthétisé dans la figure 3. Parmi les combinaisons d'environnements, on peut citer :

environnement réel C'est le cas des travaux sur les interactions de groupes d'humains et de groupes de robots, comme ceux de d'Alexis Drogoul [Drogoul et Picault, 1999], expérience sur le long terme où des robots vivent dans les couloirs du laboratoire, l'étude portant sur l'aspect social des rencontres entre les groupes ; on peut aussi citer les agents d'interface "physiques" comme Muu, un agent en mousse, aux formes douces, équipé d'un énorme "œil" et capable de bouger et de produire des sons (voir la présentation au SIGGRAPH d'Okada [Okada, 2000]) pour interagir.

environnement virtuel augmenté C'est le cas, par exemple, de FreeWalk [Nakanishi *et al.*, 1998], dans lequel les utilisateurs (humains et agents) partagent un monde virtuel enrichi des vidéos des utilisateurs humains, pour permettre

⁵On remarquera d'ailleurs que selon les définitions d'*agent* (Cf. chapitre 2) il peut être possible de considérer l'*être humain* comme un agent.

par exemple des téléconférences dans lesquelles les participants retrouvent une certaine spatialité, permettant, entre autres, une matérialisation évidente des groupes de discussion.

multiples environnements Dans le cadre du projet COMRIS [Van de Velde, 1997], les agents existent simultanément dans deux mondes : le monde réel (grâce à une interface portée par l'utilisateur humain auquel il est associé) et un monde virtuel disposant d'une métrique basée sur les centres d'intérêt. L'agent REA (voir section 3.5.2) est dans un monde virtuel, tandis que l'utilisateur est dans le monde réel, et interagit avec REA par l'intermédiaire d'un écran⁶, d'un microphone et de capteurs de position.

environnement virtuel C'est le cas sur lequel nous avons travaillé : les humains et les agents se rencontrent au sein d'un espace virtuel. Très souvent, dans cette situation, l'espace est représenté en trois dimensions, et les agents (naturels ou artificiels) communiquent principalement par le biais du texte. On peut aussi citer ici des applications ludiques comme Half-Life⁷, de plus en plus nombreuses, où des équipes d'agents artificiels (appelés bots dans ce contexte) et/ou d'humains coopèrent pour atteindre un objectif tel que la prise d'une zone particulière, la protection ou libération d'otages, ou la destruction des adversaires.

Notre vision de la communauté mixte est celle d'une "symbiose" entre humains et systèmes informatiques, telle que la présentait déjà Licklider (psychologue de formation et responsable de l'Information Processing Techniques Office) en 1960 [Licklider, 1960]. Une telle relation considère les différences entre agents (naturels et artificiels) comme un avantage à exploiter⁸ plus que comme un inconvénient. Il en conclut l'intérêt d'une symbiose permettant une exploitation efficace des qualités propres de chaque agent, naturels comme artificiels.

Les travaux les plus en phase avec ces idées sont ceux du *communityware* (voir par exemple les travaux rassemblés par Ishida [Ishida, 1998a] [Ishida, 1998b]). Ils considèrent en effet des agents à la fois :

- dans une dynamique de groupe, c'est-à-dire comme faisant partie d'une communauté ;
- dans leurs interactions avec les utilisateurs humains.

Considérés sous cet angle, les agents artificiels issus du *communityware* se placent d'un côté dans la lignée des systèmes multi-agents et de l'autre dans celle des agents d'interface.

Pour aboutir à la constitution de ces communautés mixtes (et donc arriver à exploiter les différences de ses membres), le problème principal est celui de la communication entre agents. En effet, suivant la définition de Quentin Jones (voir page

⁶C'est ce qu'on appelle une interface WoW : Window on World.

⁷Half-Life est un produit de Sierra-Studio. Son principe est le suivant : l'utilisateur (le joueur) dirige un personnage dans un environnement virtuel contenant d'autres personnages (adversaires ou partenaires, joueurs humains ou bots) et dans lequel il peut récupérer et utiliser divers matériels, principalement des armes, pour mener à bien une mission qui peut être aussi simple que la mort de l'ensemble des autres joueurs ou aussi complexe que la libération d'otages.

⁸*Computing machines can do readily, well, and rapidly many things that are difficult or impossible for man, and men can do readily and well, though not rapidly, any things that are difficult or impossible for computers. That suggests that a symbiotic co-operation, if successful in integrating the positive characteristics of men and computers would be of great value.*[Licklider, 1960]

10), il ne peut y avoir communauté que s'il existe une interactivité minimale entre les différents participants. Pour franchir cet obstacle, il faut arriver à définir un mode de communication exploitable au mieux par chacune des deux parties. Actuellement, et malgré les progrès rapides des interfaces à base de langue naturelle, la plupart des interactions homme-machine utilisent soit des langages très adaptés à la machine (langages de programmation) soit des langages peu expressifs dédiés à une tâche précise (comme pointer/cliquer sur des boutons). C'est-à-dire des interactions qui :

- soit emploient une langue très proche de la machine et d'une grande capacité expressive en ce qui concerne les capacités de la machine (cas des langages de programmation) ;
- soit emploient une langue plus adaptée à l'être humain mais en général très limitée dans les choix proposés à l'utilisateur⁹.

Les interfaces à base de langue naturelle tiendraient une troisième place, celle d'un mode d'interaction très proche de l'humain, mais les systèmes actuels sont encore loin de permettre l'emploi de tout le potentiel expressif de la langue naturelle.

Outre la forme du message (instruction d'un langage de programmation, clic sur une icône ou énoncé en langue naturelle), il faut tenir compte des différences dans les informations traitées, ou plus précisément, des affinités avec tel ou tel type d'information. "*Instructions directed to computers specify courses; instructions directed to human beings specify goals*¹⁰" [Licklider, 1960]. Ainsi, quel que soit le langage employé, la *nature* des interlocuteurs conditionne une forme d'expression privilégiée : faire s'exprimer un humain en terme d'action à entreprendre facilite son interaction avec un ordinateur (c'est ce que font les programmeurs) ; être capable, pour un agent, de raisonner sur des buts simplifie l'interaction avec des humains. De plus, ces remarques laissent supposer que, tels quels, aucun des langages spécifiques aux agents artificiels ou aux humains n'est fortement adapté aux interactions mixtes. Si l'on souhaite se baser sur un langage existant (la langue naturelle, ou plus précisément, la théorie des actes de langages, dans notre cas) pour développer un langage mixte, il faudra adapter et *enrichir* ce langage pour que ni agents artificiels ni humains ne soient bridés par les capacités expressives du langage.

Nous cherchons donc un point de concours entre les capacités et les nécessités de chacun des types d'agents, pour les raisons que nous avons présentées ci-dessus (affinités avec des types d'informations différentes), mais aussi pour d'autres raisons telles que la tendance à l'anthropomorphisme (voir par exemple les travaux de Clifford Nass [Nass *et al.*, 1994]). Un agent doué de réelles capacités conversationnelles doit permettre l'emploi d'un langage plus expressif et plus adapté à l'être humain, alors que les modèles couramment utilisés en informatique sont très spécifiquement adaptés à la communication entre entités artificielles. Symétriquement, un humain cherchant à exploiter au mieux les capacités des agents avec lesquels il interagit devra adapter son langage, qui est, lui, spécifiquement adapté à la communication entre humains. Les sections qui suivent présentent les adaptations souhaitables pour chacun de ces types d'agents (naturels pour la section 3.3 et artificiels pour la section 3.4) de façon à atteindre ce point de concours.

⁹du moins dans l'interface standard : il existe généralement des options permettant d'accéder aux fonctions plus complexes, mais l'accès en est souvent d'autant plus complexe (cas des interfaces "pointer-cliquer").

¹⁰On notera ici le lien fort entre pensée (ou traitement de données) et langage.

3.3 Nécessités côté humain

Un système efficace n'est pas forcément un système simple d'emploi, et inversement. C'est-à-dire qu'il est courant que la simplicité d'interaction avec un système aille de pair avec une limitation de l'accès à ses capacités. Une des solutions évidentes pour obtenir un système simple d'emploi est de le doter d'une interface qui soit "évidente", "instinctive", "naturelle", etc ; mais plus la méthode d'interaction se rapproche de ce à quoi les humains sont habitués, par exemple une situation de dialogue en face-à-face, plus on s'expose au risque de ne pouvoir accéder *facilement* à l'ensemble des potentialités du système. Comme nous l'avons fait remarquer dans la section précédente, agents artificiels et humains n'ont pas les mêmes besoins lorsqu'ils communiquent entre-eux (agents-agents et humains-humains).

Donner à des agents des capacités de communication aussi proches que possible de celles des humains permettra facilement aux agents et humains de communiquer sur des tâches pour lesquelles les méthodes de communications employées sont adaptées, c'est-à-dire, typiquement, des problèmes destinés à être résolus par des humains. En effet, les langues naturelles ont évolué (dans leurs aspects lexicaux, syntaxiques et sémantiques) de façon à permettre la communication entre humains, sur des sujets ayant trait au monde partagé par les humains. Nous pensons que la langue naturelle (ou, dans le cadre de nos travaux, la théorie des actes de langages), peut être une bonne base pour un langage commun, à condition de la faire évoluer à nouveau, en l'enrichissant et en l'adaptant de façon à pouvoir donner un accès *facile* aux capacités spécifiques de l'agent.

Un deuxième problème lié à l'humain dans le cadre d'une coopération avec des agents artificiels, se pose en terme d'acceptation d'une nouvelle classe d'interactants avec toutes ses spécificités, et de *conscience* de ces spécificités. Par exemple, on sait que du côté de l'être humain, même s'il a conscience d'interagir avec une machine, il considère inconsciemment cette machine sous un angle social, et non purement fonctionnel¹¹. La relation entre l'être humain et l'agent artificiel n'est pas aisée à définir et les stratégies d'interaction implémentées par l'agent peuvent se trouver entre les deux points suivants :

1. méthodes d'interaction recopiant plus ou moins complètement les mécanismes d'interaction des êtres humains (typiquement, les relations de face-à-face, avec une emphase particulière portée sur la langue naturelle) ;
2. méthodes d'interaction mettant en évidence les compétences particulières de l'agent (typiquement, et actuellement, langages de programmation, interfaces spécifiques pour la résolution d'une tâche donnée, ...)

Des agents ayant des méthodes d'interaction comme celles du point 1 auront des communications simplifiées avec l'utilisateur, mais ces communications, plus simples, peuvent rendre plus délicat l'emploi des aptitudes spéciales de l'agent. Quant au point 2, il nécessite une adaptation de la part de l'humain, mais améliore l'exploitation des capacités de l'agent (il s'agit du cas le plus courant actuellement).

On peut noter que la langue naturelle (moyen de communication humain par excellence) évolue sans cesse, principalement en enrichissant son vocabulaire. On peut donc imaginer "l'adaptation" des capacités de communication de l'humain non comme une révolution, mais plutôt comme un simple enrichissement, tel que la

¹¹"[...]we demonstrate that users can be induced to elicit a wide range of social behaviors, even though users know that the machines do not actually possess feelings, "selves", genders, or human motivations" ([Nass et al., 1994]).

langue en vit quasi quotidiennement. Mais il est aussi possible que les (futures) capacités des agents soient si différentes de ce que la langue naturelle peut exprimer¹² qu'il faille employer un langage qui y soit profondément lié si l'on souhaite pouvoir profiter pleinement des avantages de l'agent.

3.4 Nécessités côté agent

Les travaux sur les agents, dans la droite ligne de l'intelligence artificielle s'intéressent en grande majorité à des agents qu'il est convenu d'appeler sincères et rationnels. De telles architectures d'agents sont insuffisantes pour intégrer nos agents dans une communauté virtuelle : elles ne sont en effet pas, en l'état, suffisamment adaptées à la génération de comportements anthropomorphes ou à la gestion d'interactions avec des humains. L'apparition de l'humain dans le monde des agents (ou de l'agent dans le monde des humains) entraîne cette nécessité de passer d'un modèle classiquement "rationnel et sincère" à un modèle "rationnel *mais* social" ou "rationnel *mais* ayant de la personnalité". En effet, le psychologue social Erving Goffman [Goffman, 1974] prétend que sans "l'hypocrisie sociale", qui peut pousser à agir d'une manière opposée à ses sentiments véritables, les sociétés perdraient toute cohésion. Une profonde intégration d'agents artificiels dans une communauté d'humains devra donc passer par la perte d'une partie de la sincérité des agents, ou l'acceptation par les humains de comportements qu'ils ne toléreraient pas de la part d'autres humains. Ces deux solutions apportent avec elles d'autres problèmes. Les humains employant déjà inconsciemment avec les ordinateurs des comportements similaires à ceux qu'ils emploient avec d'autres êtres humains, la dernière solution risque, même si l'humain à conscience de la différence de son interlocuteur, de venir perturber leur interaction. Quant à la première, elle risque, si elle est trop poussée, de faire passer l'agent pour un hypocrite, voire un menteur, ce qui est loin de correspondre à une situation idéale de communication.

Bien que dans son fonctionnement l'agent doive rester valide (au sens du maintien de l'intégrité fonctionnelle de l'agent, voir section 3.4.1), il existe une certaine liberté (autonomie) dans le choix des stratégies de résolution de la tâche qui ont été confiées à l'agent, ainsi que dans son mode de présentation à l'utilisateur (voir section 3.4.2). Ces deux degrés de liberté (sur la résolution de la tâche et sur les façons d'interagir avec l'utilisateur) permettent, pour une même fonctionnalité, d'adapter l'agent aux différences intra-utilisateurs. Ces exigences d'adaptation à l'utilisateur peuvent globalement être rassemblées sous le terme de "conscience sociale" de l'agent (dont l'hypocrisie sociale d'Erving Goffman fait partie). Bien sûr, pour permettre l'exploitation de la liberté accordée au niveau de la communication avec l'utilisateur, l'agent devra être doté de capacités de communication suffisamment riches et adaptées à cet usage (voir section 3.4.3).

¹²On peut par exemple se demander si la langue naturelle serait adaptée à la communication avec des abeilles, ou, pour prendre un exemple moins réaliste mais peut-être plus proche du cas des agents vivant dans un monde d'informations, avec d'hypothétiques habitants d'un monde quantique, pour lesquels les "simples" notions d'"ici" et "maintenant" seraient problématiques.

3.4.1 De la cohérence

La cohérence du comportement de l'agent est un point fondamental de sa plausibilité (*"belivability"*, voir particulièrement les travaux autour du projet OZ, section 3.5.1), mais elle est aussi, de manière encore plus profonde, une exigence fonctionnelle. Des aspects "sincères et rationnels", nous ne souhaitons nous affranchir (et encore, en partie seulement) que de la partie "sincérité". Conserver un fonctionnement interne rationnel est un point permettant d'obtenir un comportement perçu par les interlocuteurs comme cohérent.

Il faut, avant toute chose, que l'agent puisse remplir son rôle. L'autonomie que nous jugeons désirable d'attribuer à l'agent doit faciliter l'interaction avec les utilisateurs, mais pas au détriment de l'objectif qu'humain et agent doivent atteindre à *travers* cette interaction. En d'autres termes, l'autonomie interactionnelle de l'agent doit aider à améliorer le fonctionnement du système.

On distingue donc deux types de contraintes,

- fonctionnelles : l'agent est valide (il accomplit ce pour quoi il est prévu),
- interactionnelles : l'agent doit communiquer de manière à être compris par l'utilisateur,

marquant chacune les limites de deux degrés de liberté différents

- fonctionnel : l'agent est libre de ses stratégies de résolution de problème, qu'il peut (ou doit) adapter au contexte (utilisateur, requête, historique des demandes) ;
- interactionnel : l'agent est libre de choisir la *forme* de la communication de façon à l'adapter à son contexte.

La contrainte de validité fonctionnelle est fondamentale, puisqu'en dernier ressort, il faut tout de même que l'agent accomplisse la tâche qui lui incombe. C'est aussi la plus simple à atteindre, puisqu'il s'agit de *la* contrainte classique considérée lors du développement et de la validation de tout système.

La liberté interactionnelle est moins critique puisqu'elle ne touche qu'à la forme que prend l'interaction utilisateur/agent. Toutefois, un agent aux capacités interactionnelles déroutantes, ou, plus généralement, inadaptées à tel ou tel utilisateur, peut tout à fait atteindre l'intégrité fonctionnelle de l'ensemble agent/humain.

3.4.2 Une personnalité

La personnalité¹³ est le nom que l'on peut donner aux spécificités des comportements d'un agent donné. Elle interviendra dans le comportement de l'agent (i.e. essentiellement dans ses dialogues pour les situations qui nous intéressent) comme générateur ou modificateur : certains comportements de l'agent ne seront "que" l'expression de sa personnalité, tandis que d'autres seront influencés par la personnalité. Elle permet aussi, à la condition d'être suffisamment stable et exprimée, de générer des habitudes (régularités) comportementales et donc la création d'attentes (de la part des interlocuteurs) propres à simplifier les interactions (ceci s'intégrant d'ailleurs dans l'objectif, plus vaste, de la cohérence du comportement). Elle permet enfin l'individuation (tous les agents ne sont pas identiques), menant à la construction de relations privilégiées entre agents et humains, et renforçant un côté social déjà latent dans toute relation, fût-elle avec une machine (voir par exemple les travaux de Clifford Nass [Nass *et al.*, 1994]).

¹³"Ce qui caractérise une personne, dans son unité, sa singularité et sa permanence." (Hachette). Voir aussi la section 4.4 *La place des interprétants*

Nous verrons dans les prochains chapitres (particulièrement le chapitre 4 *Modèle de l'interaction*) que les messages échangés entre les différents participants humains d'une interaction sont chargés d'informations sur le contexte d'énonciation. La façon dont un agent (naturel ou artificiel) va réagir dans une situation donnée est une manifestation conjointe de son rôle (ce qu'il doit faire) et de sa personnalité (comment il doit le faire).

Dans les systèmes informatiques classiques, la partie fonctionnelle (que nous assimilons, de manière assez traditionnelle, au "rôle") est largement dominante, résolvant, en les éludant, les problèmes liés à l'*interprétation* et à l'*expression* des messages, c'est-à-dire aux traitements liés à l'enrichissement du message par le contexte¹⁴, respectivement en entrée et en sortie. Nous ne nous intéresserons dans cette thèse qu'à l'influence que la personnalité peut exercer sur la forme de la *communication* et non pas sur les stratégies de *traitement de la tâche* attribuée à l'agent (son rôle), qui pourraient l'amener à choisir différents algorithmes ou jeux de paramètres, pour résoudre les problèmes qui lui sont confiés, avec un certain "style" (en fait, une manifestation de la personnalité) qui lui serait propre.

Pour donner à des agents des capacités conversationnelles leur permettant d'influer sur la *forme* de leurs interactions en fonction de l'utilisateur, il faut que ces agents disposent d'une personnalité, qui, avec la prise en compte du contexte social et d'une dynamique émotionnelle représente l'essentiel des influences que le message subit. Trois types de problèmes sont liés à l'intégration d'une personnalité dans un agent : la gestion interne d'un modèle de personnalité, sa fusion avec les autres paramètres influant sur le message, et sa manifestation externe, par l'intermédiaire des messages échangés.

Le point le plus important, du point de vue de l'utilisateur, est externe : de la personnalité doit venir une amélioration. Elle doit aider à l'adaptation du message à la situation conversationnelle ; elle ne doit pas nuire à la cohérence du discours ; elle ne doit pas dénaturer le message sur lequel elle se greffe. Enfin notons, même si cela peut sembler trivial, qu'elle doit être exprimée, si l'on souhaite qu'elle ait un effet quelconque sur la conversation, mais que les modifications qu'elle induit peuvent tout à fait se satisfaire d'une perception inconsciente de la part de l'humain, comme cela est déjà le cas en situation de communication humain/humain.

À ce problème de la manifestation externe de la personnalité s'ajoutent, pour le concepteur, ceux de la modélisation interne de la personnalité et de l'influence de celle-ci sur le message, que cette influence ait lieu

- à un niveau interne : influence sur la représentation interne du message ;
- au niveau de la traduction du message : influence sur les mécanismes de traduction du message d'un format en un autre, typiquement, d'un format interne en un format compréhensible par l'utilisateur (par exemple la langue naturelle) ;
- à un niveau externe : modification du message alors qu'il est déjà sous une forme compréhensible par l'utilisateur¹⁵.

ou à une combinaison de ces trois niveaux. La proposition de modèle que nous faisons (Cf. chapitres 4 et 5) considère une influence de la personnalité sur les deux premiers niveaux. Les composantes indépendantes qui agiraient sur le troisième niveau sont

¹⁴L'ensemble de tout ce qui est au contact du message : personnalités de interactants, informations sur les groupes en présence, état du monde, historique de l'interaction, etc.

¹⁵Cette situation est possible lorsqu'un type de message (par exemple les expressions faciales, ou la langue naturelle) contient des composantes indépendantes (par exemple, rougissement de la peau et contractions musculaires faciales, ou texte et prosodie). Il est alors possible de modifier certaines de ces composantes pour refléter la personnalité.

classées comme des modalités à part entière, même si elles peuvent n'être porteuses que de messages complémentaires (voir page 145).

Inversement, il est intéressant du point de vue de l'agent de considérer la personnalité de l'utilisateur, de façon à pouvoir y adapter son mode d'interaction. Cette étape est indispensable pour faire agir l'agent avec ne serait-ce qu'un minimum de sens social. Toutefois, avoir des agents trop attentifs à la personnalité de ses interlocuteurs peut avoir un effet inverse à celui recherché, si l'humain se sent manipulé par l'agent. Nous avons dit qu'un comportement social ne pouvait pas toujours s'accommoder de sincérité (correspondance exacte entre ce qui est "pensé" et ce qui est "dit"), mais il ne faut pas avancer trop loin sur le chemin de l'hypocrisie¹⁶, au risque d'interdire toute relation de confiance.

Les chapitres 4 et 5 présentent la façon dont nous prenons en compte la personnalité dans nos modèles d'agent et de conversation.

3.4.3 Des capacités conversationnelles

De même que nous estimons les architectures "classiques" d'agents comme insuffisantes dans le cadre d'une intégration de ces agents dans une communauté mixte, nous considérons les mécanismes "classiques" d'interaction comme trop limités pour cette application [Pesty *et al.*, 1997]. Les modèles d'interaction les plus couramment employés sont trop stricts pour pouvoir être employés dans une interaction avec un humain : règles d'enchaînement figées (protocoles), sémantique fixée et commune, etc.

Cette situation est différente de celle de la communication humaine, où le sens n'est pas porté par le message seul, mais dépend du contexte d'énonciation. Lorsqu'un message est reçu, les humains ne le *décodent* pas, mais ils l'*interprètent*, particulièrement en fonction de leurs connaissances sur l'interlocuteur (personnalité, état mental courant), de leur propre personnalité, des relations qui les lient (conscience sociale) et de l'état de la conversation (focus courant, objectifs), paramètres que nous rassemblons sous le terme de "contexte conversationnel".

¹⁶Ceci est clairement présenté par les membres du projet OZ quand ils précisent que leurs agents peuvent décider d'agir de manière amicale pour atteindre leurs fins, même si l'état interne de l'agent n'est pas assimilable à de l'amitié envers son interlocuteur : "[...]it [l'agent] may decide to act friendly to get what it wants, even if the agent isn't feeling especially friendly" [Bates *et al.*, 1992].

Dialogue oral*	<i>Chat</i> †	Texte écrit‡
<p>Ginette : Monsieur B. Roger : Il checke ça lui. Ginette : C'est cela. Roger : Hein ? Ginette : J'ai dit oui il checke ça. Roger : ils ont l'air fines hein les : Ginette : Oui mais elle avait l'air moins jasante elle Roger : Elle moins jasante ? Ginette : Oui. Roger : Que l'autre ? Ginette : Oui. Roger : Elle veut pas parler. Ginette : Elle était peut-être : je veux dire : plus gênée. Roger : Elle veut pas parler d'elle</p>	<p>Bob : bonjour Fred .. Bob : surle 2m c est ta 1er visite ? Fred : ouais je viens de d/l le 2ème monde Bob : oauis chouette bon bena lros je v t expliquer qq trucs. Fred : ouais vas y ! Bob : deja pour te trouver un avatar Bob : tu vois les ongles sous la 3d sur la droite de l ecran ? Fred : ouais ça serait cool Bob : tu avs aller jusqu a celui qui s appelle options Fred : session options et tout ça ouais ! [...] Fred : Bob j'ai rien capté!!!</p>	<p>Monsieur le Président, Me permettez-vous, dans ma gratitude pour le bienveillant accueil que vous m'avez fait un jour, d'avoir le souci de votre juste gloire et de vous dire que votre étoile, si heureuse jusqu'ici, est menacée de la plus honteuse, de la plus ineffaçable des taches ? Vous êtes sorti sain et sauf des basses calomnies, vous avez conquis les cœurs. Vous apparaissez rayonnant dans l'apothéose de cette fête patriotique que l'alliance russe a été pour la France, et vous vous préparez à présider au solennel triomphe de notre Exposition Universelle, qui couronnera notre grand siècle de travail, de vérité et de liberté. Mais quelle tache de boue sur votre nom — j'allais dire sur votre règne — que cette abominable affaire Dreyfus !</p>

TAB. 3.1 – Langue orale, “tapotée” et écrite

* [Vincent *et al.*, 1995], 2 :95, activité 1.

† Extrait d'un dialogue du Deuxième Monde. Les pseudonymes ont été changés.

‡ Extrait de “J'accuse”, d'Émile Zola.

Les trois exemples de la table 3.1 montrent que les dialogues oraux spontanés ne semblent pas suivre de règles rigoureuses mais disposent de méthodes de correction très efficaces. L'exemple de *chat* montre une situation où l'enchaînement semble plus structuré, mais dans laquelle apparaissent des messages de confirmation. Quant à l'exemple de texte écrit, il présente une structure rigoureuse et un enchaînement précis. Le *chat* montre des ressemblances avec la langue parlée, dans sa spontanéité et sa robustesse face aux imperfections syntaxiques ou lexicales. Il montre aussi des ressemblances avec l'expression écrite, dans sa transmission sans dégradation du message et la faible quantité d'information transmise (absence de prosodie, pauses).

Au niveau de l'enchaînement de ces messages, un deuxième conflit apparaît entre les modèles classiques de l'interaction entre agents et la communication humaine telle qu'elle est pratiquée. Les modèles informatiques conduisent les successions de messages en fonction de protocoles généralement très contraints. À l'inverse, dans la conversation humaine, les interactants disposent d'une liberté quasi-infinie, impossible à modéliser sous la forme d'un protocole spécifiant des séries de questions/réponses. Les agents devront donc, pour pouvoir interagir efficacement avec les humains, être capables de gérer cette complexité.

Toutefois, les remarques précédentes ne doivent pas laisser penser que l'emploi de protocoles est impossible pour interagir avec des humains, puisqu'il s'agit de la situation actuelle. Simplement, pour atteindre l'objectif de simplification des interactions agents/humains, un modèle de communication *entièrement* basé sur des protocoles est inadapté.

Ces points sont développés plus avant dans le chapitre 4 "Modèle de l'interaction".

3.5 Quelques agents participant à des groupes mixtes

Dans cette section, nous allons présenter quelques projets dans lesquels agents et humains sont en interaction. Le premier projet (sous-section 3.5.1), OZ, de l'université de Carnegie Mellon, né d'une réflexion commune avec des membres du monde du théâtre, adopte un point de vue très centré sur l'humain et illustre ses théories par des applications très variées. Le deuxième projet (sous-section 3.5.2), REA, agent immobilier développé par le *Gesture and Narrative Language Group* du MIT, s'oriente plus vers les aspects communication avec l'être humain. Ce groupe présente également d'autres projets que nous aborderons dans la même sous-section. Enfin, nous présenterons de manière plus succincte (sous-section 3.5.3) (1) le projet *Amusement*, de l'université de Madrid, qui traite principalement des interactions entre humains par l'intermédiaire d'un monde virtuel; (2) STEVE, un agent versatile de l'université de Californie du Sud, participant à une réelle dynamique *mixte* dans des groupes d'humains et d'agents; (3) le projet *persona*, et les travaux connexes de Microsoft, qui proposent un modèle de raisonnement particulièrement intéressant dans le cadre de nos travaux.

3.5.1 Le projet OZ (CMU)

a) Contexte

Le projet OZ, développé à l'université de Carnegie Mellon pendant les années 90, est l'un des projets précurseurs ainsi que le plus représentatif en ce qui concerne les modèles de comportement d'agents en interaction entre eux et avec des être humains. Les objectifs (et les questions) initiales du projet OZ (voir l'article synthétique de Joseph Bates, responsable du projet [Bates *et al.*, 1991], ou celui, plus étoffé de Margaret Kelso [Kelso *et al.*, 1992]) avaient trait aux questions que peuvent se poser des metteurs en scène, des scénaristes ou des acteurs : comment impliquer le spectateur dans la pièce qui lui est présentée, et que faire pour qu'il soit "touché" par le spectacle. À ceci, le projet Oz souhaitait ajouter une dimension interactive au "spectacle", qui estomperait la frontière entre spectateur et acteur. Joseph Bates, responsable du projet, donne pour objectif à OZ d'apporter de l'aide aux artistes souhaitant créer des *pièces*¹⁷ *hautement interactives*. Il développe le choix de ces termes ("*highly interactive drama*") en trois points [Kelso *et al.*, 1992] :

- *Pièce interactive* : il s'agit de la présentation par un système informatique de mondes riches hautement interactifs, peuplés de personnages aux comportements dynamiques et complexes.
- *Hautement interactives* : le terme "interactif" marque la différence par rapport aux médias classiques, tandis que le terme "hautement" précise que les interactants du monde (humains ou agents participants à la pièce) ont une liberté plus grande que dans les médias où les choix sont fixés et peu nombreux (hypertexte par exemple).
- *Pièce* : bien que l'utilisateur soit libre de dire et de faire ce qu'il désire, il existe une forme de finalité au système (ou fil conducteur, thème, scénario, "destinée", ...), qui permet de structurer l'interaction sur le long terme.

À partir d'une expérience "grandeur nature" (c'est-à-dire avec des acteurs et un metteur en scène humains, dans le monde réel [Kelso *et al.*, 1992]), l'équipe d'OZ distingue trois entités en interaction : l'interactant (nous dirions l'agent naturel), les acteurs (nous dirions les agents artificiels) et le metteur en scène (nous dirions le planificateur central). Dans leurs travaux suivants, le planificateur central est peu abordé, cette expérience montrant que plus l'agent est autonome et que sa connaissance des objectifs de la pièce est grande (en fait, plus la connaissance est distribuée entre les acteurs et non centralisée entre les mains d'un metteur en scène), moins le planificateur a à intervenir, et plus la qualité de l'expérience vécue par l'utilisateur est bonne. Parallèlement, trois thèmes de recherche sont distingués : la construction "d'agents à large champ" (*broad agents*), une théorie permettant la gestion du déroulement de la pièce, et un système apportant des notions de *style* à la pièce. De ces objectifs présentés par Joseph Bates [Bates, 1992], à notre connaissance, seul le premier (architecture d'agent) a été étudié en profondeur.

L'architecture globale du système est présentée par la figure 3.3. On remarque particulièrement les points suivants :

- Les différents agents (les agents artificiels et l'agent humain) partagent un même environnement : les agents artificiels y sont "directement" présents (*Character*), l'agent humain y est représenté par un agent (*Interactor*) qu'il contrôle à travers une interface (*Interface*).
- Un planificateur central (*Drama Manager*) dispose de contrôle sur le monde,

¹⁷au sens de pièces de théâtre.

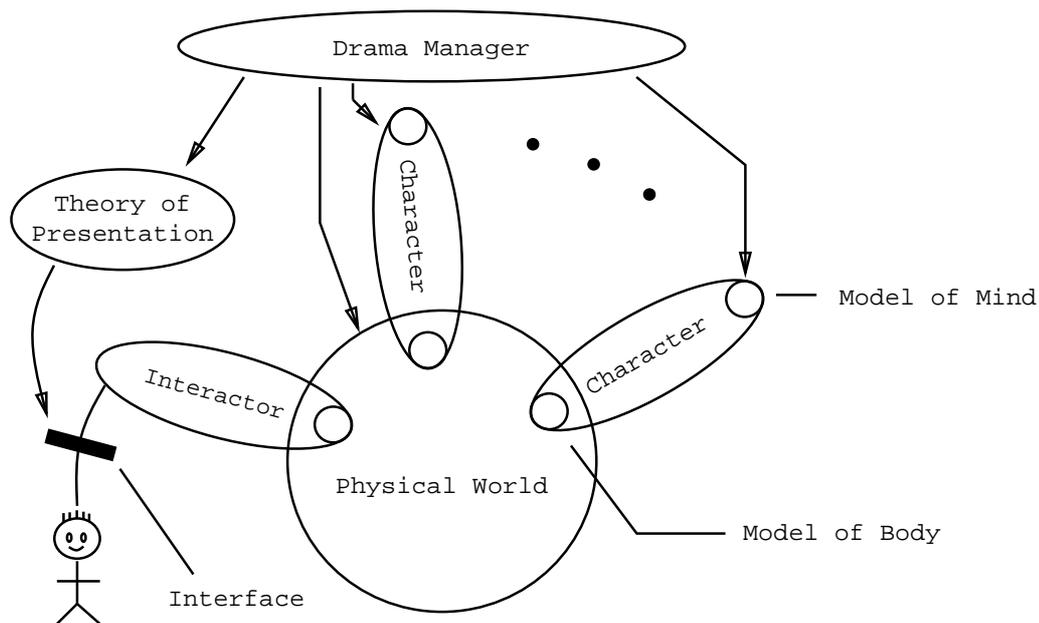


FIG. 3.3 – L’architecture du système [Kantrowitz, 1990]

les agents artificiels et les perceptions/actions de l'utilisateur.

b) "L'esprit" du projet OZ : *Broad but shallow*

Dans l'objectif global de pièce interactive, un sous-objectif fondamental est identifié : maintenir les interactants dans une situation plausible sans que les réactions de l'environnement ou les comportements des autres agents ne viennent rompre cette impression (ce qu'en Anglais on résume par *suspension of disbelief*). Pour amener les interactants dans cette position, les agents (ou *characters*, personnages, qui semble être un terme préféré) doivent présenter à l'utilisateur un comportement donnant l'impression de s'appuyer sur des buts, des émotions, des capacités en langue naturelle et des connaissances sur les agents et sur le monde¹⁸. Pour maintenir l'interactant dans cet état (de *suspension of disbelief*), il n'est pas nécessaire que l'agent soit particulièrement actif ou malin, du moment qu'il n'est pas clairement stupide¹⁹. Ces agents dont le comportement donne l'illusion de vie sont appelés plausibles (*believable agents*). Dans l'esprit du projet OZ, un agent plausible passe par une architecture à large champ (*broad agent*), permettant à l'utilisateur de prêter à l'agent des capacités supérieures à celles dont il dispose en réalité ("l'effet Eliza", en référence au chatbot "Eliza" de Weizenbaum [Weizenbaum, 1966]).

¹⁸ *We believe this means that agents must provide some signs of internal goals, reactivity, emotion, natural language ability, and knowledge of agents (self and other) as well as of the simulated physical world.* [Bates et al., 1991] "Emotion is one of the primary means to achieve this believability, this illusion of life, because it helps us know what characters really care about what happens in the world, that they truly have desires." [Reilly et Bates, 1992].

¹⁹ *In the context of Oz, instead of demanding that our agents be especially active and smart, we require only that they not be clearly stupid or unreal. An agent that keeps quiet may appear wise, while one that oversteps its abilities may destroy the suspension of disbelief.* [Bates et al., 1991]).

c) Architecture et algorithme

L'architecture d'agent est segmentée en trois modules (Figure 3.4) :

- un module de planification (module “Hap”);
- un module d'émotion (module “Em”);
- un module de perception et représentation du monde (module “Sensory routines and integrated sense model”).

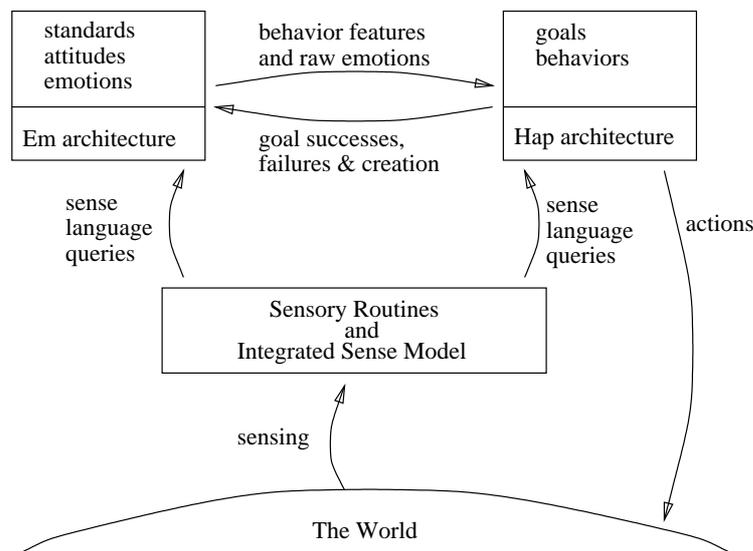


FIG. 3.4 – L'architecture de l'agent [Reilly et Bates, 1992]

Le module de planification Hap est le cœur du comportement du personnage : il gère la base de buts et les plans permettant de les atteindre ; il décide des actions ; il synthétise les informations d'ordre émotionnel et social issues de Em. L'algorithme de fonctionnement d'un agent est basé sur une boucle classique “perception → raisonnement → action” et se déroule comme suit [Bates *et al.*, 1992] :

1. Mise à jour des perceptions²⁰ et donc du modèle du monde ;
2. Evaluation de l'applicabilité des plans en cours et du succès éventuel des buts ;
3. Choix d'un but
 - Si le but est associé à une action, effectuer cette action,
 - Sinon, choix d'un plan pour accomplir le but.
4. Reboucler.

Le module émotionnel (Em) gère les émotions en fonction de la comparaison de trois *paires* de critères (inspirés des travaux d'A. Ortony [Ortony *et al.*, 1988]) :

1. Des *événements* et des *buts* : par exemple, le succès d'un but entraîne une *joie* proportionnelle à l'importance de ce but et *l'espoir* est lié à la probabilité de réussite d'un but ;
2. Des *actions* et des *standards (normes)* : par exemple, le *reproche* est dirigé vers quelqu'un dont les actes sont répréhensibles (faire quitter au personnage un lieu où il se sentait bien), tandis que si c'est le personnage lui-même qui effectue cet acte vis-à-vis d'un autre, il pourra ressentir de la *honte* ;

²⁰La perception peut être incomplète, incorrecte ou absente ; les conditions d'application des plans et les conditions de succès des buts étant vérifiées par rapport au modèle du monde et non directement aux perceptions.

3. Des *objets du monde* et des *attitudes* : certains objets (ou agents) peuvent être, par leur seule proximité, générateurs d'émotions comme la haine ou l'amitié.

Ces deux derniers points recouvrent le côté social de l'agent : (1) la prise en compte de relations personnelles (point 3) ; (2) la prise ne compte des normes sociales, intégrées en partie dans l'évaluation des actes (point 2) et en partie dans la base de plans et de buts.

d) Les capacités interactionnelles

Les capacités interactionnelles des agents sont particulièrement mises en évidence dans la deuxième partie (*Believable Social Agents*) de la thèse de Scott Reilly [Reilly, 1996]. La situation est celle d'une cour d'école où l'utilisateur peut rencontrer plusieurs personnages (nommés Melvin et Sluggo) avec lesquels négocier l'échange de cartes. Cette situation représente une version restreinte de l'architecture d'OZ, puisque :

- Elle ne propose d'interactions qu'entre un agent et un humain : les agents ne négocient pas entre eux l'échange de cartes, mais prennent toutefois en compte les autres dans la gestion de leurs émotions (Melvin a peur lorsque Sluggo est proche) ou de la négociation avec l'utilisateur (Melvin accepte de donner à l'utilisateur une carte A qui lui permettra d'obtenir de Sluggo la carte B qui était son véritable désir²¹).
- Il n'y a pas de planificateur central.

Bien que la négociation se déroule en langue naturelle, il s'agit en fait de phrases génériques complétées (en sortie) et de reconnaissance de mots clés (en entrée). Scott Reilly précise toutefois que l'intégration d'un système performant de traitement de la langue naturelle (en génération et en compréhension) avec Em, le module émotionnel aurait un intérêt indéniable, et que ce choix rudimentaire n'a été fait que pour des raisons technologiques.

L'intégration dans les agents OZ d'un système avancé de génération de langue naturelle a été proposée [Kantrowitz, 1990; Loyall et Bates, 1997]. Ce système, baptisé GLINDA — pour des raisons liées à des personnages de l'histoire du "Magicien d'Oz" —, est un système de génération de texte, employé à l'origine essentiellement pour la génération de descriptions du monde et des actions/événements y prenant place [Kantrowitz, 1990]. Plus tard, le module GLINDA a été adapté à la génération d'énoncés communicatifs (pour l'application "*Edge of intentions*" [Loyall et Bates, 1997]), après quelques modifications lui permettant de s'intégrer avec Hap. Hap gérant les *actions* de l'agent, les adaptations apportées à GLINDA visaient à lui faire générer ses messages en employant le formalisme de représentation des actes. Ainsi, pour des raisons technologiques, le module expression langagière des agents d'OZ se retrouve en accord avec les considérations des philosophes du langage pour lesquels communiquer, c'est agir (voir "les actes de langage" dans la section 4.1).

e) OZ et les mondes mixtes

Le projet OZ considère le monde comme construit pour un utilisateur unique. Bien qu'aucune objection technique ni théorique n'interdise le fonctionnement en multi-utilisateurs, il semble que cette voie n'ait pas été explorée. De plus, il pose l'existence d'un planificateur central capable d'agir à la fois sur le monde, les agents

²¹Le mécanisme permettant cette planification coopérative n'est pas précisé.

artificiels et l'interface de l'utilisateur. Ces deux points (utilisateur unique et planification centralisée) sont liés : l'être humain est particulièrement difficile à prévoir, et plus un système acceptera d'interaction avec les êtres humains plus un contrôle central sera délicat à mettre en œuvre. Contrôler le comportement de grands groupes humains est, dans le cadre de communautés virtuelles, impossible de l'aveu même de personnes qui ont essayé [Morningstar et Farmer, 1990]. En effet, un tel contrôle requiert des capacités de manipulation (au sens, disons "Machiavélique" du terme) et des connaissances sur le comportement humain bien au-delà des théories et des capacités des systèmes informatiques actuels. De plus, comme nous le faisons remarquer précédemment (p.58), une expérimentation d'Oz montre l'intérêt de distribuer au maximum les connaissances de l'agent et de lui laisser le plus d'autonomie possible.

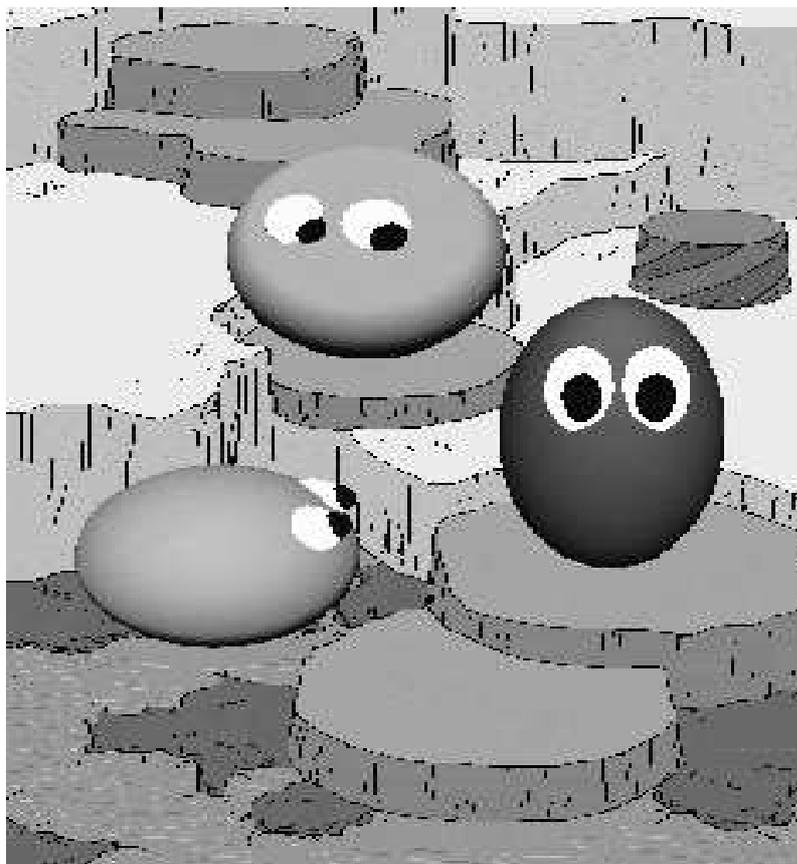


FIG. 3.5 – Les woogles dans leur monde

Par ailleurs, contrairement aux autres groupes travaillant sur des domaines équivalents, et conformément à leur philosophie "*broad but shallow*", le groupe d'OZ a expérimenté plusieurs situations différentes, sans chercher à avoir un agent synthétisant tous leurs travaux mais, par là même, restreignant le domaine d'application. On peut citer :

- *Lyotard*. Un monde textuel, dont le personnage central est Lyotard, un chat synthétique, interagissant *éventuellement* avec l'utilisateur. Les interactions sont purement "physiques" (caresses, déplacements, ...). Dans cette application, l'emphase est mise sur les émotions de Lyotard, et leurs liens avec son comportement.

- *Playground* (proche de *Robbery World* et d'*Office politics*). Un monde textuel où l'utilisateur a un objectif (échanger des cartes), qu'il peut atteindre en interagissant avec les agents présents. Les interactions se déroulent en langue naturelle.
- *Edge of intentions*. Dans un monde visuel, trois *Woogles* (personnages ovoïdes colorés, proches parents des "barbapapas", voir figure 3.5) vivent et un humain peut prendre la place d'un quatrième. Les interactions sont physiques (déplacement, direction du regard, changement de couleur, ...) et l'emphase est mise sur l'étude de la *personnalité* de chacun des *Woogles*.

Pour finir, rappelons que le champ d'étude et d'expérimentation du projet OZ est né d'un désir *artistique* : là où d'autres étudient la communication en tant que telle, il la voit comme un *moyen* d'atteindre leurs objectifs. Tous ces faits font du projet OZ un projet très riche surtout si l'on considère qu'il s'étend depuis une réflexion profonde sur un modèle d'agent jusqu'à une série de réalisations variées.

3.5.2 REA et projets connexes (MIT)

Des travaux du "*Gesture and Narrative Language group*" (GNL) au MIT, nous nous intéressons particulièrement à "*Body chat*" ([Vilhjálmsson, 1997], ou [Vilhjálmsson et Cassell, 1998] pour une présentation plus synthétique) et à REA, une application qui synthétise la plupart de leurs travaux (voir [Cassell *et al.*, 2000b] pour une vue d'ensemble).

Body chat est un système qui analyse les phrases tapées par l'utilisateur lors de sa discussion (type *chat*) avec son interlocuteur et qui se sert de cette analyse pour générer le comportement de l'avatar matérialisant l'utilisateur. Ainsi, la figure 3.7 montre un exemple d'animation du visage de l'avatar : le mot "*very*" est accentué par un hochement de tête, et le point d'exclamation retranscrit par un haussement de sourcils.

REA (Real Estate Agent) quant à elle, est "une" agent chargée de vendre de l'immobilier, apparaissant sur un écran de projection (voir figure 3.6). Elle dispose de moyens d'action visuels (expression corporelle, faciale, regard mais elle emploie aussi divers mouvements déictiques, symboliques, ...) et vocaux (communication par synthèse vocale). Elle dispose en entrée de caméras (pour juger de la présence/absence d'utilisateurs, ainsi que pour pouvoir percevoir certains de leurs mouvements, comme les gestes d'interruption) et d'un micro (reconnaissance vocale).

La spécificité la plus marquée de ces travaux est la prise en compte de l'aspect visuel de la conversation, avec l'emploi d'agents humanoïdes²² tant dans leur représentation que dans leur comportement. Ces interfaces personnifiées²³ de conversation sont plus qu'une interface revêtue d'une forme humaine ou animale ; elles sont aussi plus que le comportement réaliste de ces interfaces ("*lifelike or believable*"). La spécificité revendiquée par ces interfaces est la *conversation*, particulièrement dans les détails suivants [Cassell *et al.*, 2000b] :

- perception et interprétation multi-modale ;
- action multi-modale ;
- gestion de fonctions conversationnelles (tours de parole, ruptures et négociation, ...)
- expression de la situation de la conversation.

²²Les représentations animales ne sont pas exclues de leurs travaux, mais seuls des agents anthropomorphes ont été réalisés.

²³Au sens du Robert : "Représenté sous la forme d'un être humain".



FIG. 3.6 – Une interaction avec REA

Le GNL aborde la quasi-totalité des points pertinents des interfaces conversationnelles personnalisées, avec toutefois, comme nous l'avons introduit, une emphase sur un point rarement traité auparavant, le côté visuel de la communication.

a) Agents autonomes ou avatars contrôlés

Là où REA est un agent autonome, *Body Chat* [Vilhjálmsson, 1997] propose des avatars semi-autonomes. Ce système met en contact deux utilisateurs par l'intermédiaire d'une interface de *chat*, d'une interface vers un environnement 3D permettant la visualisation de son propre avatar et de celui de son interlocuteur, ainsi que quelques composants de contrôle permettant de spécifier des lignes directrices du comportement de l'avatar.

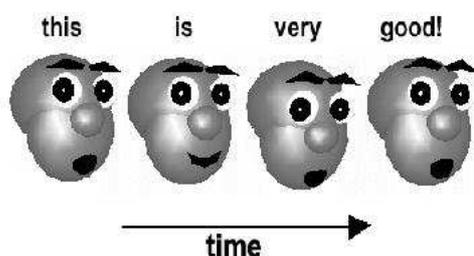


FIG. 3.7 – Exemple d'action de Body Chat [Vilhjálmsson, 1997]

Le projet *Body Chat* comble un espace entre les systèmes où l'avatar est entièrement contrôlé par l'utilisateur (avec toute la complexité due aux interfaces actuelles que cela implique) et les systèmes où un agent artificiel contrôle l'avatar. *Body Chat* permet à l'utilisateur de déléguer le contrôle de l'avatar : les utilisateurs communiquent en utilisant le *chat*, et les avatars calquent leur comportement sur les informations transmises par ce

biais (Cf. figure 3.7). Les gestes de l'avatar sont des gestes d'accompagnement²⁴ de l'information (matérialisations d'emphase sur certains mots) ou de la conversation (prise ou don de parole).

²⁴En effet ils ne peuvent qu'avoir un sens qui complète le texte échangé puisque le système se base sur ce texte pour les générer.

Le système Amusement (qui met l'emphasis sur l'interaction entre humains par l'intermédiaire d'un espace virtuel, cf. sous-section 3.5.3) reprend cette idée de la délégation de contrôle, en laissant à l'utilisateur le choix du degré d'autonomie de l'agent et des points sur lesquels il laisse cette autonomie, contrairement à *Body Chat* où l'utilisateur ne peut décider ni de prendre un contrôle direct sur son avatar, ni de laisser converser l'agent.

b) Des agents conversationnels *incarnés*

Bien que les travaux du GNL couvrent l'ensemble des domaines nécessaires à la création d'un agent conversationnel, ils se concentrent surtout sur les aspects visuels de la communication. Trois types de gestes pouvant accompagner, compléter ou préciser la parole sont distingués [Cassell, 2000] :

- *Les symboles (emblems)*. Gestes très ancrés dans un contexte culturel (le V de la victoire, pouce en l'air, gestes insultants, etc.)
- *Les mouvements propositionnels*. Mouvements conscients apportant des précisions sur le discours ou remplaçant des mots (“*mets ça là*”, “*c'était grand comme ça*”).
- *Les mouvements spontanés*. Il s'agit de la plus grande classe de mouvements accompagnant la conversation (*co-verbal gestures*). Ils caractérisent quatre sous-types de mouvements spontanés :
 - *iconique*. Le mouvement reprend une partie de la phrase pour l'accompagner (“*La pièce a la forme d'un triangle*[forme de triangle avec les mains]”, “*La caméra prend le visage entier*[mains de chaque côté de la tête]”).
 - *métaphorique*. Le mouvement sert à réifier un concept (“*Je reviendrais plus tard sur ce sujet* [les mains font un mouvement circulaire, indiquant une idée de retour et de continuité]”).
 - *déictique*. Les mouvements désignent un objet du discours, qui peut avoir une existence physique (“*Ça ne marche pas* [avec désignation du “ça”]”) ou non (“*Tu peux utiliser la méthode X ou la méthode Y* [main passant de droite, paume vers le haut, à gauche, paume vers le bas]”).
 - *rythmique (beat gesture)*. Les mouvements rythmiques ont en général une fonction pragmatique, servant à marquer les méta-performatifs, ou à vérifier l'attention des auditeurs. Ce genre de gestes sert aussi à marquer les points importants de l'énoncé (typiquement, les rhèmes, Cf. paragraphe “Agents incarnés *conversationnels*” plus bas).

L'ensemble de ces mouvements correspond à la manifestation physique (consciente ou non) d'intentions de communication, par opposition aux mouvements des lèvres lors de la parole, aux mouvements respiratoires, etc. Parmi les fonctions communicatives que portent ces gestes, on trouve particulièrement :

- la planification : gestion des tours de parole ;
- l'accentuation : emphase sur le contenu ;
- la rétroaction (*feedback*) : confirmation ou infirmation de la compréhension de la discussion par l'auditeur, ou requête de retour de la part du locuteur ;
- l'information : le geste a valeur informationnelle ([Cassell *et al.*, 2000d]).

Cette séparation entre ce qui est à exprimer et ce qui est exprimé (de la même façon que l'équipe d'OZ différencie émotion et manifestation de l'émotion) apporte de la souplesse sur plusieurs points, dont les plus évidents sont :

- une plus grande liberté au niveau de l'expression de la personnalité (un même sentiment, la *peur*, peut entraîner différents comportements selon les person-

nalités : *fuite* ou *agression*, selon un exemple d'OZ).

- une simplification du problème de la multi-modalité du comportement, qui se retrouve déporté en aval de la partie raisonnement de l'agent (l'agent peut spécifier des comportements sans savoir de quelles modalités il dispose, l'instanciation du comportement étant traitée à part).
- une indépendance (relative) entre cette partie raisonnement et les modalités dont dispose l'agent pour s'exprimer, permettant une modification des moyens d'action de l'agent qui ne nécessite pas une modification de l'ensemble du système.

On note particulièrement ici le problème de la multi-modalité de la conversation [Torres, 1997]. En effet, la personnification des agents leur donne une prise particulièrement marquée sur le monde. Le fait d'avoir une représentation et de n'être pas "un pur esprit" pose de manière évidente le problème de la multi-modalité des interactions (au moins en expression), qui n'était pas présent dans les *chatbots*. Cette augmentation de complexité en apporte une deuxième : le conflit d'expression au sein d'une même modalité. En effet, une expression faciale peut être un effet de bord (pour prononcer un mot, il faut remuer les lèvres), peut viser un effet meta-conversationnel social (sourire poli) ou pragmatique (acquiescements qui encouragent à poursuivre ou qui marquent l'intérêt) ou remplir une fonction sémantique, complète (acquiescements) ou partielle (en complément du message).

c) Des agents incarnés *conversationnels*

Bien que le point le plus systématiquement étudié par le GNL soit la communication visuelle, une emphase est mise sur l'importance d'intégrer cette étude dans le champ plus vaste de la *communication*. Comme nous l'avons dit précédemment, les mouvements (expression corporelle et faciale) ne sont que la manifestation dans une certaine modalité (ici, visuelle) d'intentions de communication. Ces intentions sont générées suivant des modèles de conversation qui s'abstraient en grande partie des modalités des messages échangés, marquant bien la différence entre l'intention communicative (appelée *conversational function*) et le message lui-même (qui est un *conversational behaviour*). Outre cette différence entre ce qui est manipulé par le système et la forme qui lui est donnée pour être communiqué, une deuxième distinction est faite entre d'une part les messages dont le but est de transmettre de l'information (*propositional function*) et d'autre part ceux qui aident à la transmission de cette information (*interactional function*), comme les actes meta-conversationnels de demande de confirmation.

La manipulation explicite d'un *sens* qui sera *par la suite* matérialisé par un message permet la considération de tout un ensemble de propriétés sémantiques, qu'il ne serait pas possible d'avoir si, comme dans le cas des *chatbots* classiques, le système ne manipulait nulle part explicitement un sens. On peut ainsi générer des messages parallèles (complémentant ou enrichissant) au message principal, comme les postures de fin de conversation, ou marquer des emphases (visuelles, prosodiques, ...) sur le *focus* de l'énoncé. En effet, une des approches proposées par le GNL pour la structuration du discours fait intervenir les notions de *thème* et *rhème*²⁵. On peut considérer le thème comme étant ce sur quoi porte l'énoncé tandis que le rhème correspond à ce qui est nouveau ou intéressant à propos du thème, et qui n'aurait *a priori* pas pu être inféré du contexte conversationnel. Cette séparation

²⁵aussi connus sous la forme des couples thème/propos, topique/commentaire ou présupposition/focus.

permet par exemple de distinguer la partie à mettre en avant — le rhème — , car, étant la partie la plus difficile à prévoir par l’auditeur, c’est elle qu’il est intéressant d’appuyer. Comme le montrent certaines expériences [Cassell *et al.*, 1999b], cette insistance peut être manifestée par le biais de la direction du regard, mais aussi de la prosodie ou de mouvements rythmiques, selon les capacités expressives de l’agent et les modalités disponibles.

Les mécanismes pris en compte pour la gestion de la conversation tranchent avec les approches radicalement monologiques et réactives des *chatbots*. Outre les points cités précédemment, on retrouve au cœur des préoccupations conversationnelles du GNL, une dimension *temporelle*, *pragmatique* ainsi que *sociale* évidente.

La dimension *temporelle* apparaît au niveau des étapes de planification. Dans REA, on peut distinguer deux types de planification :

1. une planification à l’échelle de la conversation, qui décide de l’enchaînement des messages (ou plus précisément des intentions conversationnelles) ;
2. une planification à l’échelle du message qui coordonne l’enchaînement des différents messages sur plusieurs modalités, ou qui synchronise ces messages dont l’ensemble est la matérialisation d’une intention communicative.

Pour illustrer la planification à l’échelle de la conversation, on peut citer le cas où REA commence une conversation sur le temps, pour dériver vers le temps à Boston, et après avoir amené Boston dans la conversation, fait une remarque sur le prix de l’immobilier à Boston avant, enfin, d’aborder la question de moyens financiers de l’acheteur (cf. table 3.2). De même, pour illustrer la planification à l’échelle des messages multi-modaux composant la forme d’une même intention communicative, on peut considérer les situations déictiques où l’agent désigne par un moyen visuel un des objets de l’énoncé. On pourrait encore citer la synchronisation labiale²⁶ ou le cas où la direction du regard donne des indications sur le passage de parole (fin de tour, requête de parole, ...).

Les considérations pragmatiques sont dans une certaine mesure à rapprocher de l’aspect social de la conversation, en ce sens qu’elles envisagent la conversation dans l’axe du rapport à l’autre. Dans ce cadre, les agents du GNL peuvent gérer les interruptions en milieu d’énoncé par exemple (voir la retranscription de la table 3.3), avec plus ou moins de priorité selon le type d’interruptions (simplement gestuelle, matérialisant un désir de prise de parole, ou vocale, indiquant une contestation de la situation conversationnelle en cours). REA intègre aussi des comportements de *feed-back* qu’elle utilise durant le tour de l’utilisateur.

Quant à l’aspect social de la conversation, il est intimement lié à celui de la confiance²⁷. C’est cet aspect qui est abordé dans REA (voir par exemple [Bickmore et Cassell, 2000]), où REA peut décider qu’il n’est pas opportun d’aborder le sujet du prix de la maison tant qu’elle n’est pas dans une situation de confiance suffisante avec l’utilisateur. Ce niveau de confiance augmente tout au long des tours de parole,

²⁶Il suffit de regarder (et d’écouter !) un film mal doublé pour se rendre compte du désagrément causé par la mauvaise synchronisation entre le mouvement des lèvres et les sons sensés être produits. *L’effet McGurk* représente l’extrême de la mauvaise synchronisation : on présente à un sujet un stimulus composé d’une séquence vidéo où un acteur prononce /ba/, mais dont la bande son contient le phonème /ga/, et le sujet déclare percevoir la syllabe /da/.

²⁷“*Agents [...] must be able to establish social relationship with users in order to engage their trust which, in turn, eases cooperation*” [Bickmore et Cassell, 2000]

```

[...]
REA : Are you one of our sponsors?
User : Yes
REA : Were you at our last sponsor meeting?
User : ...
REA : I got so exhausted at the last sponsor meeting. I think
I was starting to lose my voice by the end.
User : ...
REA : I have lived in Boston all my life. Come to think of
it, I have lived inside this room all my life. It is so de-
pressing.
User : ...
REA : How about this weather?
User : ...
REA : I think winters in Boston are awful
User : ...
REA : Boston is certainly more expensive than it used to be
User : ...
REA : So, what can you afford?
[...]

```

TAB. 3.2 – REA : dialogue de mise en confiance. Les réponses de l'utilisateur marquées “...” sont omises car elles n'ont aucune importance : REA les ignore. (extrait de [Bickmore et Cassell, 2000])

et particulièrement lors de discussion sans contenu informationnel (parler du temps, de la famille, ... Voir le dialogue de la table 3.2).

d) Architecture et algorithme

L'architecture de REA suit une segmentation classique en modules *perception* (*Input Manager*), *raisonnement* (*Deliberative Module*) et *action* (*Action Scheduler*), ainsi qu'une dérivation directe de la perception vers l'action pour des comportements réactifs comme le suivi de l'utilisateur par le regard de l'agent.

Perception. (Représenté sur la figure 3.8 par *Input Manager* (IM)) Cette partie transfère les messages à la partie raisonnement en instanciant certains champs (ceux dont les informations sont extractibles du percept) d'un message au format KQML (voir un exemple dans la table 3.9), dont les autres champs seront remplis au fur et à mesure du traitement. Dans le cadre des messages langagiers, le message KQML contient en plus les gestes récemment employés, permettant un traitement conjoint de messages issus de modalités différentes.

Raisonnement. (Représenté sur la figure 3.8 par *Déliberative Module* (DelM, et non pas DM qui correspond au *Decision Module*)) Cette partie commence par une phase d'interprétation : le module de compréhension (*Understanding module*) renseigne les champs `:prop` (sens propositionnel) and `:intr` (sens interactionnel) du message KQML avec les champs sémantiques reconnus. Le message KQML est ensuite transmis au module de décision (*Decision Module*) qui traite à la fois le

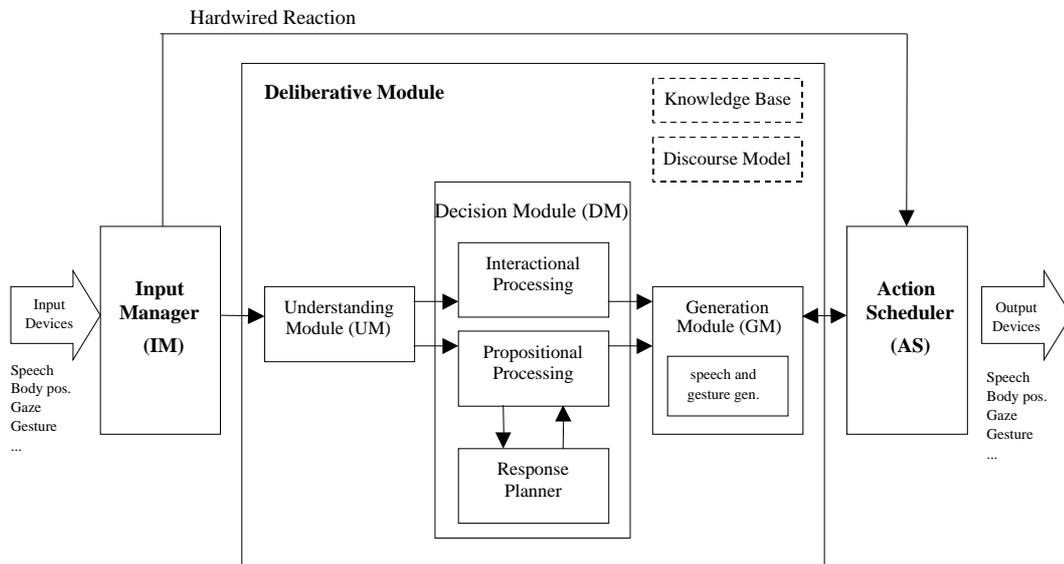


FIG. 3.8 – L'architecture de REA (extrait de [Cassell *et al.*, 2000c])

```
(tell
  :sender UM
  :recipient DM
  :content
    (commact
      :sender USER
      :recipient REA
      :input [(speaking :state TRUE)
              (gesturing :state TRUE)
              ]
      :prop NONE
      :intr [ (takingturn) ]
    )
)
```

FIG. 3.9 – Exemple de message KQML envoyé de l'*Understanding Module* (UM) vers le *Decision Module* (DM)

```

[...]
REA : This is a nice Victorian on a large lot [gestures toward
the house]. It has two bedrooms and a large kitchen with..
User : [Interruption from the user, by looking at REA and gesturing with
hands while speaking]. Wait. Telle me about the bedroom
REA : The master bedroom is (...description of the be-
droom...). Do you want to see the master bedroom?
User : [overlapping with Rea]. Sure, show me the master bedroom.
REA : I'm sorry, I didn't quite catch that, can you please
repeat what you said?
[...]

```

TAB. 3.3 – REA : exemples d'interruptions. Extrait de [Cassell *et al.*, 1999a]

message sous son angle propositionnel et interactionnel, créant éventuellement un autre message KQML contenant les intentions de réponse. Cet éventuel message est envoyé au module de génération (*Generation Module*) qui le décompose en une série de comportements qui sont confiés au module d'action.

Action. (Représenté sur la figure 3.8 par *Action Scheduler (AS)*) Cette partie planifie l'exécution des différents comportements spécifiés par le module de génération, en fonction de ses connaissances sur les capacités des effecteurs, particulièrement leurs disponibilités et de leurs influences mutuelles (une emphase marquée par un mouvement de la main empêche de se servir de cette même main pour désigner un objet).

On pourra se reporter à [Cassell *et al.*, 2000c] pour voir la décomposition d'une interaction avec un utilisateur.

e) REA et les mondes mixtes

Que ce soit Gandalf (une application visant à évaluer les effets de retours multimodaux, principalement visuels et émotionnels sur la conversation, voir [Cassell et Thórisson, 1999]), *Body Chat* (déjà présenté dans cette section), REA (déjà présenté dans cette section), ou plus récemment, SAM (un partenaire de jeu virtuel pour des enfants, voir [Cassell *et al.*, 2000a]), la situation conversationnelle se réduit à deux interlocuteurs (deux humains assistés pour *Body Chat*, un humain et un agent artificiel pour REA, Gandalf et Sam). Un système comportant plus de deux interactants (que les interactants supplémentaires soient humains ou artificiels) n'est pas même évoquée. Toutefois, comme pour OZ, les modèles internes et les architectures des agents ne présentent pas d'incompatibilité avec une situation où l'agent aurait à s'intégrer dans une communauté mixte. Les choix techniques, eux, peuvent présenter une certaine entrave à l'évolution vers un système multi-utilisateurs. En effet, l'agent artificiel et l'humain ne partagent pas un monde, ils restent chacun dans le leur, chacun de son côté de l'écran. Bien qu'elle puisse être considéré dans une première approche comme une manifestation intéressante de la mixité des mondes (réel / virtuel), cette séparation entraîne qu'il n'existe pas de "lieu" commun, mais une multitude d'interfaces entre les mondes, où aucun des participants n'a de représenta-

tion dans le monde de l'autre²⁸. Cette approche ouvre des perspectives intéressantes, mais dans le cadre qui nous intéresse, considérer un environnement partagé (qu'il s'ancre dans le monde réel, comme les communautés robotiques, ou sur une représentation informatique, comme les communautés virtuelles, voir le début de ce chapitre) permet de simplifier la situation (particulièrement en matière d'interaction) pour se concentrer sur les dynamiques communautaires. En outre, nous avons montré dans la section 1.2 que l'existence d'un lieu de rencontre commun fait partie des points importants pour la constitution d'une communauté. Tout ceci nous amène à dire avec d'autres : "*Gandalf and REA were not developed for conversation in virtual reality*" [Rickel et Johnson, 2000], mais ce n'est, semble-t-il, que la conséquence d'un choix technique, non une contrainte du modèle.

3.5.3 Autres projets

a) Amusement

L'objectif principal du projet Amusement²⁹ est la création d'un cyberspace dans lequel les utilisateurs peuvent interagir librement, ceci incluant une réflexion sur les (et un développement des) moyens d'interaction mis à la disposition des utilisateurs [Imbert *et al.*, 1999a; Imbert et de Antonio, 2000]. Le *groupe de recherche sur les environnements virtuels* de l'université polytechnique de Madrid s'intéresse particulièrement aux interactions entre utilisateurs humains par l'intermédiaire d'un monde virtuel.

Dans le cadre de ce projet, un environnement virtuel placé entre les utilisateurs et servant de support à leurs interactions est explicitement pris en compte [Imbert *et al.*, 1999a]. Pour analyser cette situation et tenter d'améliorer la qualité de l'interaction, le problème est séparé en quatre couches d'interaction : Interaction entre utilisateurs (couche 1), Interaction entre l'utilisateur et son propre avatar (couche 2), Interaction entre l'utilisateur et l'avatar d'un autre utilisateur (couche 3), Interaction entre avatars (couche 4), voir figure 3.10.

Après avoir constaté que l'essentiel des communications transitent par la couche 1 (dans la plupart des situations, l'avatar n'est considéré que comme une marionnette, destinée à montrer "l'emplacement" de l'utilisateur dans l'espace virtuel, réduisant quasiment à néant les transits d'informations sur les couches 2, 3 et 4), il est proposé de donner à l'avatar une certaine autonomie d'action (et de communication). Cette autonomie entraînera une utilisation accrue des couches 2 (indications de l'utilisateur sur le comportement à adopter, ou demande de conseils de la part de l'avatar à l'utilisateur lorsqu'une décision complexe est à prendre) et 3 (emploi d'expressions corporelles, par exemple), entraînant elle-même une amélioration de la communication entre utilisateurs (couche 1).

Le centre d'intérêt est donc l'avatar, auquel l'utilisateur peut déléguer une partie des tâches de communication, ce qui améliore en retour la qualité d'interaction entre utilisateurs, à la fois par l'enrichissement que pourra apporter l'avatar sur certaines modalités (la modalité visuelle est prise comme application), et par l'allègement de la charge de l'utilisateur, qui n'a pas à contrôler systématiquement l'avatar. L'autonomie donnée à l'avatar joue sur trois types d'action : l'envoi "spontané" de messages

²⁸Les interfaces de type moniteur, écran de projection, etc. sont d'ailleurs, dans le domaine de l'IHM, rassemblées sous le vocable de WoW : Window on World.

²⁹Amusement est un projet ESPRIT, comprenant plusieurs partenaires, dont l'*Universidad Politécnica de Madrid*.

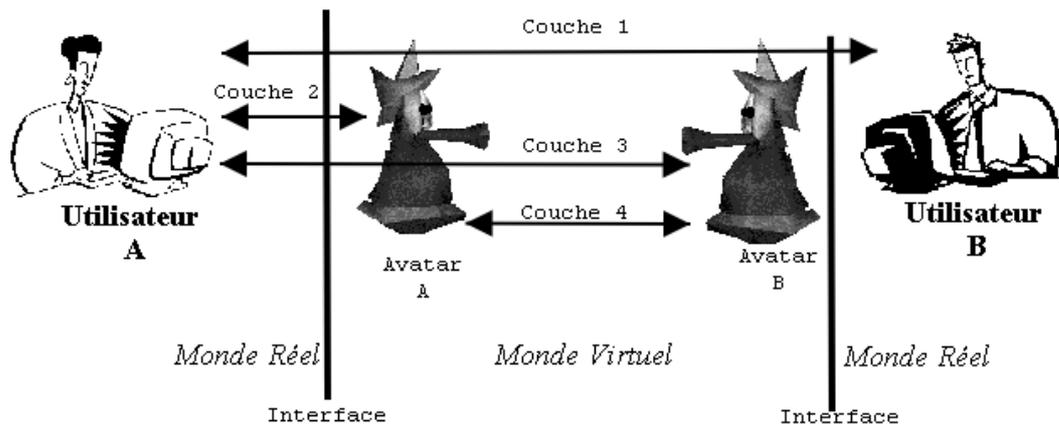


FIG. 3.10 – Niveaux d’interaction dans le projet Amusement (figure d’après [Imbert *et al.*, 1999a])

(lorsque l’avatar dispose d’une grande autonomie, il peut prendre l’initiative d’envoyer des messages), la manière d’accomplir les actions (l’agent peut influencer sur la forme d’une action requise par son utilisateur) et les messages complémentaires aux messages de l’utilisateur (expression faciale adaptée aux messages de l’utilisateur, par exemple).

L’autonomie de l’agent passe par un modèle interne basé sur les notions de *personnalité*, d’*humeurs*, d’*attitudes* et d’*intentions*. Ces quatre traits sont corrélés de façon à maintenir un comportement cohérent. Lorsque l’utilisateur décide d’en modifier un ou que l’avatar lui-même, par des mécanismes d’apprentissage, en décide la mise à jour, certains des autres paramètres peuvent être automatiquement adaptés [Imbert *et al.*, 1998].

Le projet *Amusement* est intéressant car il présente un modèle d’avatar/agent qui a pour but d’interagir avec d’autres humains. Si on laisse à l’avatar le maximum d’autonomie, il peut alors être considéré comme un agent autonome. Le comportement social de l’agent (de toutes façons inhérent à toute interaction) est abordé de manière relativement classique : les *attitudes* dépendent de chaque interactant, les rencontres avec un groupe entraînant une réaction correspondant à la moyenne des réactions que l’agent aurait eues en cas de rencontre individuelle avec chaque membre du groupe. Toutefois, de manière moins classique, l’agent a la possibilité d’exprimer un comportement différent de celui qui serait le plus en adéquation avec son état interne. Par exemple, une contrainte comme la *diplomatie* peut amener l’agent à agir amicalement même s’il ne ressent aucune amitié pour la personne avec laquelle il interagit. C’est ce que Goffman [Goffman, 1974] appelle l’*hypocrisie sociale* comportement qu’il estime indispensable à la cohésion sociale d’un groupe et dont nous avons parlé page 52.

Notons de plus qu’un travail important a été effectué sur les expressions faciales, particulièrement manifesté à travers une application de jeux de cartes [Imbert *et al.*, 1999b], où les expressions faciales servent à la fois à communiquer avec son partenaire et à exprimer l’état courant de l’agent, état dépendant directement du jeu en main. Bien entendu, ces derniers signaux, sensés être spontanés, sont eux aussi susceptibles d’être manipulés. Le même mécanisme d’*hypocrisie* permettant à l’agent d’être diplomate peut être utilisé dans cette application pour induire l’adversaire en erreur...

b) Steve et les agents enseignants

STEVE³⁰ est un agent enseignant évoluant dans un monde virtuel qu'il partage avec ses élèves (voir par exemple [Rickel et Johnson, 2000]). Une de ses utilisations est la formation à la manipulation de moteurs de bateaux, pour la marine américaine. Ses emplois plus récents mettent en place plusieurs agents de son type, pour créer pour l'utilisateur un contexte de tâche plus complexe, plus réel. Par exemple, une situation où l'utilisateur prend le rôle d'un officier américain en Bosnie ayant à gérer une situation délicate impliquant des civils, des blessés, des objectifs militaires, etc. [Rickel *et al.*, 2001] Dans cette situation, trois agents STEVE sont présents : un sous-officier Américain, un médecin militaire, et la mère d'un civil accidentellement blessé. D'autres applications de STEVE mettent en scène plusieurs utilisateurs et plusieurs agents artificiels, dans une perspective d'entraînement à des tâches collaboratives. Dans ce genre de situations, un agent STEVE peut remplacer un des utilisateurs, permettant ainsi l'entraînement même si les équipes humaines ne sont pas au complet.



FIG. 3.11 – Steve et un de ses élèves

Steve présente plusieurs particularités intéressantes, parmi lesquelles nous retiendrons les deux suivantes : premièrement, l'intégration dans un monde virtuel. L'utilisateur emploie des interfaces dites immersives comme les HMD (Head Mounted Display) et des gants de données pour s'interfacer avec l'environnement de la tâche. Deuxièmement, les mécanismes internes de contrôle de l'enseignement de la tâche, qui sont basés sur des objectifs à satisfaire. Des plans sont construits pour satisfaire ces objectifs, entraînant la mise en évidence de sous-objectifs. Cette décomposition permet de justifier les actions à entreprendre si l'élève demande des explications. De plus, pour que la tâche soit accomplie, les sous-objectifs doivent être accomplis, peu importe par qui. Ceci permet à l'élève d'intervenir au milieu de la démonstration de STEVE pour demander à finir la tâche.

STEVE présente plusieurs particularités intéressantes, parmi lesquelles nous retiendrons les deux suivantes : premièrement, l'intégration dans un monde virtuel. L'utilisateur emploie des interfaces dites immersives comme les HMD (Head Mounted Display) et des gants de données pour s'interfacer avec l'environnement de la tâche. Deuxièmement, les mécanismes internes de contrôle de l'enseignement de la tâche, qui sont basés sur des objectifs à satisfaire. Des plans sont construits pour satisfaire ces objectifs, entraînant la mise en évidence de sous-objectifs. Cette décomposition permet de justifier les actions à entreprendre si l'élève demande des explications. De plus, pour que la tâche soit accomplie, les sous-objectifs doivent être accomplis, peu importe par qui. Ceci permet à l'élève d'intervenir au milieu de la démonstration de STEVE pour demander à finir la tâche.

Pour ce qui est de son architecture, on distingue dans STEVE trois modules : perception, cognition et contrôle moteur (voir figure 3.12). Le module cognition est organisé en trois couches principales :

- un noyau central, basé sur l'architecture SOAR, qui fonctionne suivant un classique cycle perception → raisonnement → action développé dans les paragraphes suivants.
- une couche fournissant à STEVE des capacités indépendantes d'un domaine de tâche précis. Principalement, des mécanismes de démonstration, et, de manière très liée, de conversation. En effet, STEVE construit ses interactions sur la base de plans qui lui servent à enchaîner, déterminer et justifier ses interventions.
- une base de connaissances déclaratives du domaine, qui permet d'alimenter la couche précédente.

³⁰Acronyme de Soar Training Expert for Virtual Environments, un projet commun à l'université de Californie du sud, au Lockheed AI center et au CARTE (Center for Advanced Research in Technology for Education).

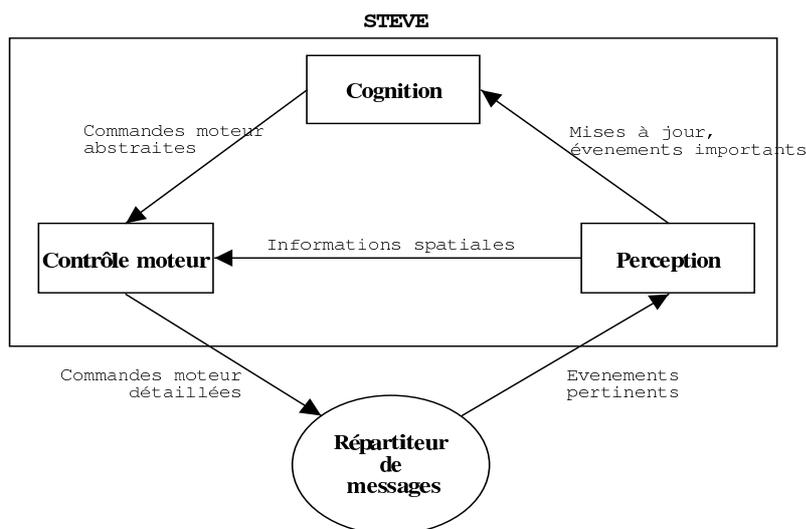


FIG. 3.12 – Les trois modules principaux de STEVE (d’après [Rickel et Johnson, 2000])

STEVE conserve aussi des informations sur le contexte de tâche et le contexte de dialogue. Le contexte de tâche contient des connaissances sur la tâche, telles que étapes, contraintes, liens de cause à effet, etc. Les plans sont construits de façon dynamique à partir de ces connaissances et remis à jour à l’arrière-plan en permanence. Le contexte de dialogue comprend l’état de la conversation (qui a la parole, quels objets sont visibles par l’étudiant, qui est en train d’agir, etc.), une liste des actions précédentes (propres et des autres), des informations sur l’action en cours (objectif ou sous-objectif, qui permet de détecter les digressions entre autres) et des informations sur le sujet courant de la discussion (est-il issu de STEVE ou de l’étudiant, a-t-il été expliqué, réalisé, les résultats ont-ils été discutés, etc.).

Le comportement de STEVE suit une boucle perception → raisonnement → action, la partie raisonnement distinguant deux étapes : une phase de sélection d’opérateurs (équivalents à de grandes catégories d’action comme répondre ou argumenter), puis une phase d’application de ces opérateurs. La séquence perception → raisonnement → action est parcourue à chaque fois que STEVE consulte son environnement (sachant qu’il est omniscient pour ce qui est de l’état de son monde) à une fréquence qui dépend des capacités du matériel sur lequel il tourne.

Du panorama que nous dressons ici, STEVE est le seul agent à évoluer dans un monde mixte, proposant des exercices d’équipe où les équipes sont composées à la fois d’agents artificiels et d’agents naturels (humains). Parmi les points qu’il est prévu d’améliorer dans STEVE [Rickel *et al.*, 2001] on remarquera le désir d’y introduire une personnalité et des émotions, ainsi qu’une amélioration des capacités communicatives de STEVE. L’ajout d’émotions permettrait par exemple à STEVE, en tant qu’enseignant, de faire ressentir aux élèves une différence entre “Consultez la jauge pour connaître le niveau d’huile” et “N’appuyez jamais sur le bouton rouge”, énoncés pour le moment produits par STEVE de manière équivalente. Dans un but de réalisme des situations d’entraînement, le comportement de STEVE doit suivre au plus près celui d’un humain, et ce, particulièrement dans les situations de stress (le STEVE actuel est qualifié de “*unrealistically rational [...] teammate*”). Enfin, pour ce qui est des capacités de communication, STEVE est particulièrement frustré en

matière de reconnaissance, la version d'enseignement pour la marine Américaine ne comprenant que quelques rares phrases très précises, telles que “*What next ?*”, “*Why ?*” et “*Let me finish.*”, contraignant grandement les interactions.

c) Microsoft : *Persona* et au delà

Le projet *Persona*, débuté dans les années 92 au laboratoire de recherche de Microsoft, s'intéressait aux agents dans une optique de communication humain/agent. Ce projet fut à la base des “Microsoft Agents” maintenant disponibles au grand public. La première réalisation de ce groupe fut *Peedy* (prononciation à l'anglaise du début de PDP1 : “Personal Digital Parrot one”) le perroquet, un agent chargé de sélectionner et de jouer des morceaux de musique selon les requêtes de l'utilisateur et une base de compact-discs [Ball *et al.*, 1997].

Cette application proposait déjà une forme d'interprétation et d'expression, dans le sens que nous leur donnons aux pages 97 et 148, c'est-à-dire la possibilité de prendre en compte le contexte d'énonciation pour associer un sens et un message. En effet, le système a à sa disposition, en production, plusieurs paraphrases permettant d'exprimer une même idée, et considère, en réception, plusieurs structures de phrases différentes comme étant associées à la même requête. Toutefois, dans les premiers travaux sur *Peedy*, en émission, le choix de la paraphrase est aléatoire, tandis qu'en réception, les différentes paraphrases renvoient toutes vers un même sens. Les travaux qui ont suivi s'intéressent aux liens qui existent entre, d'un côté un message et de l'autre une personnalité et un état émotionnel. Ces travaux plus récents portent sur un réseau Bayésien capable d'associer des mots (ou du moins des intentions communicationnelles [Breese et Ball, 1998]) et des comportements “physiques” [Ball et Breese, 1999] à des personnalités et des émotions. L'intérêt des réseaux Bayésiens étant leur réversibilité, permettant leur emploi aussi bien en interprétation (“J'ai reçu un message, quel est l'état de l'émetteur de ce message ?”) qu'en expression (“Je suis dans tel état et je cherche à exprimer tel sens, quel message employer ?”).

L'architecture d'agent présentée par Gene Ball et Jack Breese [Ball et Breese, 2000] propose une gestion du comportement à base de scripts, donnant un ensemble très contraint. Pour éviter d'avoir un agent trop répétitif (qui serait rapidement ennuyeux), ils proposent comme alternative à un apprentissage classique une mise à jour régulière de ces scripts, que ce soit de façon manuelle, ou automatique par exploitation de pages web (dans leur exemple, l'agent propose de suivre les informations). Cette partie, purement dédiée à la tâche de l'agent ne prend pas en compte la personnalité ni l'émotion (que ce soient celles de l'utilisateur ou celles de l'agent), contrairement au module chargé de la gestion des interactions avec l'utilisateur.

La prise en compte de la personnalité et de l'émotion n'apparaît qu'au niveau de la communication et non de la tâche : en parallèle à la séquence

- perception (des intentions de l'utilisateur)
- raisonnement (sur la tâche à accomplir en réponse aux désirs de l'utilisateurs)
- action (déclaration d'une intention communicative)

il s'exécute une séquence identique, mais non liée à la tâche :

- perception (de l'état mental de l'utilisateur)
- raisonnement (modification des états mentaux de l'agent en fonction d'une politique interactionnelle)
- action (spécification de la forme à apporter au message à transmettre)

La stratégie proposée de réaction aux émotions de l'utilisateur est simple : l'agent

agit avec empathie³¹, calquant ses émotions sur celles de l'interlocuteur, avec toutefois certaines restrictions dans le cas de la colère par exemple où une réaction positive se mettrait en place.

Le lien avec la langue naturelle passe par une liste de paraphrases associées avec une certaine pondération à certains nœuds du réseau Bayésien, permettant l'obtention d'une probabilité conjointe entre ces paraphrases et les états mentaux. Il existe un réseau associé à l'agent, qu'il utilise pour choisir le message dans la phase d'expression, et un réseau associé à l'interlocuteur, employé pour l'interprétation de ses énoncés. Notons que les paraphrases font partie d'une base prédéfinie, et que les connexions des paraphrases aux nœuds sont faites au cas par cas par le concepteur.

Dans ces agents, *la personnalité* est considérée comme une direction à long terme des modèles mentaux : elle fait partie des éléments qui garantissent la cohérence, la stabilité du comportement, là où les émotions induisent une variation plus rapide de ce comportement. Elle est représentée en utilisant le cercle interpersonnel [McCrae et Costa, 1989]. Les valeurs pouvant être prises suivant les deux axes (dominant/soumis et amical/hostile) autour desquels s'articule le cercle sont les trois suivantes : minimum, neutre et maximum. *L'émotion* est quant à elle considérée comme une variation rapide des états mentaux. Elle est elle aussi représentée par une position dans le plan dont les deux axes sont la valence et l'activation et dont les valeurs possibles sont aussi le neutre et les deux *extrema*.

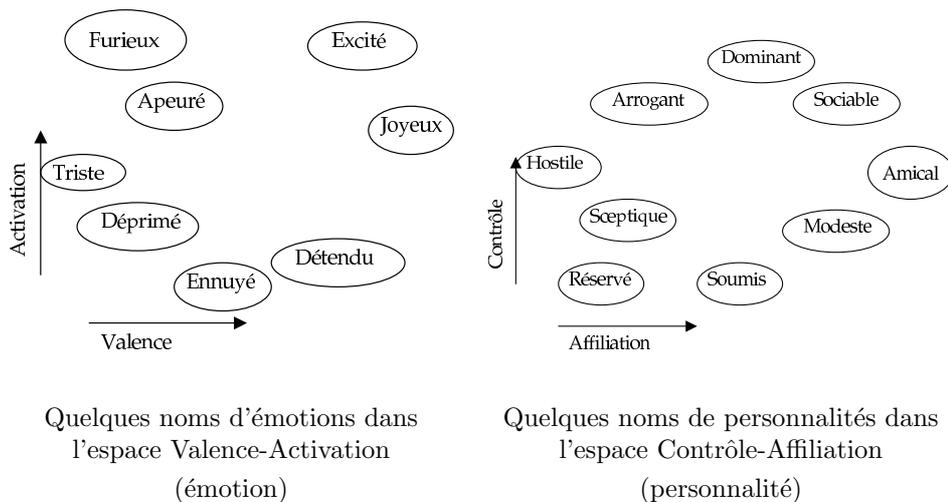


FIG. 3.13 – Modélisation de l'émotion et de la personnalité (d'après [Ball et Breese, 2000])

Un point intéressant soulevé par Gene Ball et Jack Breese est qu'ils posent une émergence du dialogue. Ils prétendent en effet que l'enchaînement des tours de parole doit venir "de lui même" à partir du moment où les ambiguïtés peuvent être levées sur la base de questions de la part de la machine. Il est possible de considérer ce point au regard de la co-construction de sens dans les dialogues, telle que nous la proposons pour gérer la dynamique conversationnelle de notre agent (voir section 4.5, et plus particulièrement 4.5.3). Si le premier énoncé satisfait les interlocuteurs,

³¹Une politique d'opposition systématique a été testée, il en ressort qu'elle n'est intéressante que lorsque l'interlocuteur s'ennuie.

la discussion se clôt d'elle-même. S'il ne satisfait pas *entièrement* les interlocuteurs, ils vont apporter des modifications (ou précisions, objections, etc.) à l'objet de la conversation, par l'intermédiaire d'un autre énoncé, amorçant ainsi un dialogue. Si par exemple un utilisateur donne à un agent de réservation de billets de train l'ensemble des informations en un seul énoncé, il n'y a pas nécessité de dialogue. Mais s'il manque des informations (aller simple ou aller-retour, par exemple) ou que de multiples réponses sont possibles (requête d'un train "aux alentours de 14h"), l'objet de la conversation n'est pas satisfaisant pour l'agent (incomplet, ambigu) et il va chercher à rendre valide cet objet en produisant d'autres énoncés ("vous voulez le billet retour aussi?", "J'ai plusieurs possibilités, laquelle préférez-vous?", etc.).

d) Entre la marionnette et l'agent autonome

Un dernier type d'agent est intéressant à considérer : il s'agit d'agents, donc doués d'une certaine autonomie, mais desquels l'utilisateur peut prendre partiellement le contrôle. Cette possibilité se retrouve dans les agents du projet *Amusement* (voir page 71), et, dans une moindre mesure, dans les agents de *Body Chat* (voir page 64). Lorsque l'utilisateur contrôle une *partie* du comportement de l'agent tout en lui laissant de l'*autonomie* dans la gestion de certaines tâches, on peut considérer l'*ensemble* humain / agent comme étant en symbiose, chacun profitant des avantages de l'autre. Dans ces cas-ci cependant, la symbiose dont il est question est bien plus "intime" que celle dont nous avons parlé dans la section 3.2. En effet, il s'agit ici de symbiose entre *agents* (artificiels ou humains) alors qu'auparavant, nous nous intéressions à la symbiose entre *groupes* d'agents.

Les travaux actuels sont peu nombreux et embryonnaires mais laissent penser que les premières générations d'agents symbiotiques devraient bientôt apparaître. On peut citer principalement le projet *Amusement*, dans lequel l'utilisateur peut intervenir à sa guise dans la gestion des interactions entre l'avatar qu'il contrôle et les autres participants : il peut laisser faire l'agent, tout prendre en charge, ou lui déléguer une partie des tâches (comme la gestion d'expressions corporelles adaptées à l'interaction en cours). De la même manière, les avatars de *BodyChat* analysent le texte tapé par l'utilisateur pour y rechercher des mots particuliers ou des clés de contrôle et dirigent l'avatar en conséquence. Dans d'autres applications, l'autonomie laissée à l'agent est si faible que le terme d'agent est probablement usurpé, et celui d'outil serait plus adapté. Néanmoins, ces applications s'inscrivent dans une évolution qui semble mener, à terme, vers des organismes symbiotiques.

On peut citer trois exemples de ces outils acquérant de plus en plus d'autonomie dans la résolution d'une tâche *commune* avec un être humain. Tout d'abord, le pilote automatique d'un avion, auquel le pilote délègue de nombreuses tâches, mais auquel il peut aussi bien reprendre le contrôle. Ensuite, les assistants de jeux "*doom-like*", comme le *Z-bot* de *Quake Arena* qui aide le joueur à viser ses adversaires en prenant en partie le contrôle de son arme, et en laissant à l'utilisateur le soin de se déplacer et de choisir les cibles. Enfin, bien que jusqu'à présent ce terme ait été quasi-exclusivement réservé à la science-fiction et à l'entomologie, les exosquelettes³², systèmes robotisés fixés à un être humain et permettant d'améliorer sa précision, sa force, sa vitesse, ...

³²Dans le cadre d'une réunion en Janvier 2001 sous l'égide de la DARPA (appel "*Exoskeletons for Human Performance Augmentation*"), les premiers prototypes de fantassins "assistés" ont été prévus pour l'année 2005.

3.6 Conclusion

En se basant sur les travaux présentés dans les chapitres précédents, traitant respectivement des communautés d’humains et des communautés d’agents, nous avons cherché à distinguer quelles étaient les contraintes pesant sur l’existence de communautés mixtes. Il ressort de manière évidente que la *communication* entre les différents membres est une condition (sinon *la condition*) *sine qua non* à la constitution de communautés mixtes. Les différences d’environnement, les différences de représentation, les différences dans les méthodes de raisonnement, dans les capacités d’action ou de perception sont, somme toute, secondaires. Comme le dit Licklider : “*The basic dissimilarity between human languages³³ and computer languages may be the most serious obstacle to true symbiosis.*” [Licklider, 1960]

Différents projets ont été présentés dans la dernière section de ce chapitre. Tous ces projets présentent des agents en interaction avec des humains et pour lesquels nous avons essayé de considérer un grand nombre de situations d’interaction (groupe ou un par un, environnements partagés ou séparés, communication orale ou écrite, représentation graphique ou non, ...) Ces projets nous ont permis de mettre en évidence un ensemble de difficultés et de points importants à prendre en compte pour faire interagir efficacement agents artificiels et humains.

³³Par le terme “*language*” Licklider comprend l’ensemble des mécanismes d’interaction.

Deuxième partie

Principes du dialogue et modèle
d'agent conversationnel pour les
communautés mixtes

Chapitre 4

Principes du dialogue

Toutes choses produites par notre propre discours et suffisance, autant vraies que fausses, sont sujettes à incertitude et à débat.

MICHEL DE MONTAIGNE

La plupart des modèles de communication actuels entre entités artificielles sont basés, explicitement ou non, sur le modèle de Shannon [Shannon, 1948] (voir Figure 4.1). Ces modèles suivent l'esprit de l'introduction même de l'article de Shannon :

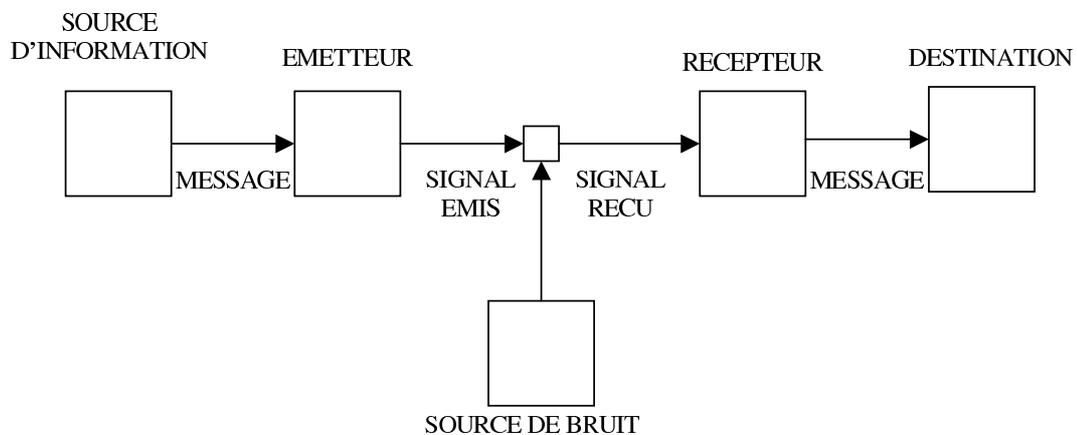


FIG. 4.1 – Diagramme d'un système général de communication (d'après [Shannon, 1948])

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point”. Dans un tel modèle, la communication est considérée comme la “simple” recopie d'un message d'un endroit à l'autre. Et, le fait est, cette approche a prouvé son efficacité dans le cadre de la communication entre entités informatiques (que ce soient des processus, un ensemble client-serveur, des agents...) ainsi que — et les applications existantes en IHM sont là pour le prouver — dans de nombreuses situations de communication entre humains et systèmes informatisés. Toutefois, dans certaines de ces situations de communication avec l'humain, ce modèle peut se montrer inadapté, comme nous le montrerons dans la section 4.2 qui traite des langages d'interaction entre agents.

Si l'on considère les études portant sur la communication humaine (et particulièrement la théorie des actes de langage présentée dans la section 4.1), on constate qu'au-delà d'une transmission de message (l'aspect *locutoire* du message, pour reprendre les concepts de la théorie des actes de langage), l'acte même de communiquer et les effets (visés ou effectifs) de l'envoi du message sont importants (ce sont les dimensions *illocutoires* et *perlocutoires*). En effet, la communication humaine ne se résume pas à un échange de symboles, mais est beaucoup plus en prise avec le monde dans lequel interagissent les humains, et est intimement liée aux notions d'*action* et de *sens*. Ces remarques rejoignent les idées de Shannon, qui, à la suite de l'extrait précédemment cité, déclare : "*Frequently the messages have meaning; that is, they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem*".

Dans le domaine de l'informatique, après un demi-siècle de travaux sur la communication comme *copie* de message d'un point à un autre, la possibilité d'interagir avec des humains pose maintenant le "problème du *sens*" que Shannon ne considérait pas comme pertinent dans le cadre auquel il s'intéressait. Ce problème du sens devient particulièrement pertinent dans le contexte auquel nous nous intéressons, celui de communautés mixtes où agents et humains ont à *communiquer*. Il a suffisamment été dit (citons par exemple Gérard Sabah [Sabah, 1997a] ou Alex Mucchielli [Mucchielli, 1995]) que communiquer, cela peut aussi être chercher à influencer les autres, convaincre, contester ou affirmer les relations qui nous lient avec nos interlocuteurs, et pas seulement à transmettre de l'information. Nous présenterons donc dans la section 4.3 ce problème du sens, qu'il est important de prendre en compte si l'on souhaite faire communiquer agents artificiels et humains.

La communication entre êtres humains n'étant pas uniquement la recopie d'un signal¹, mais nécessitant sa prise en compte *en contexte*, nous montrerons dans la section 4.4 par quels processus nous proposons de remplacer les simples *encodages* et *décodages* issus du modèle de Shannon pour passer à des mécanismes d'*interprétation* et d'*expression*. Cette section traitera de la place des interactants dans la communication et justifiera, en se basant en grande partie sur les sciences de la communication humaine, notre désir d'accorder aux agents une plus grande autonomie interactionnelle, alors qu'elle est en général limitée à leurs processus décisionnels.

Après cette section consacrée au traitement d'*un* message, nous nous intéresserons aux *enchaînements* de ces messages, en proposant un modèle basé sur la négociation de sens entre les différents interactants d'une conversation, modèle qui entraîne un suivi de conversation co-construit et non pas prédéfini. La section 4.5 s'intéressera donc à l'aspect dynamique de la conversation, tant dans les situations normales de communication (négociation de sens ou enchaînement de séquences attendues) que dans ses ruptures (ruptures sociales, incompréhension, ...).

4.1 Les actes de langage

Dans les années soixante, et à la suite des travaux fondateurs d'Austin [Austin, 1962], la philosophie du langage ordinaire s'est enrichie du concept d'acte de langage². Après ce travail d'Austin, essentiellement descriptif, Searle et Vanderveken

¹"[...] *to communicate is more than to send and to receive. Do two tape recorders communicate when they play to each other and record from each other?*" [Licklider, 1968]

²*Speech act*, parfois aussi traduit par "Acte de Discours".

[Searle, 1969; Searle et Vanderveken, 1985; Vanderveken, 1988; Vanderveken, 1990a; Vanderveken, 1990b] ont théorisé la logique illocutoire qui permet la formalisation et la manipulation des actes de langage. Nous commencerons par présenter la notion de performatif, centrale à l'étude des actes de langage (section 4.1.1), puis nous présenterons l'un des six paramètres caractérisant les performatifs, le but illocutoire, qui sert de critère discriminant pour la constitution d'une taxonomie des actes de langage (section 4.1.2). Les cinq autres composantes seront présentées dans la section suivante (section 4.1.3), puis, nous terminerons sur les notions de conditions de succès et de satisfaction des actes de langage (section 4.1.4).

4.1.1 La notion de performatif

Dans un premier temps, Austin fait un travail de description, en énonçant que "*dire, c'est faire*" [Austin, 1962]. Par cette formule, il montre que toute communication est action. Principalement, il identifie une catégorie de verbes qu'il baptise *performatifs* (de l'anglais *to perform*, accomplir, exécuter, effectuer, ...), et au sujet desquels il emploie le terme de *magie* : il suffit qu'un prêtre prononce la phrase "Je vous déclare Mari et Femme" pour que le couple soit effectivement marié. Notons ici pour lever les ambiguïtés qu'Austin emploie le terme "performatif" pour caractériser une classe particulière d'actes de discours, celle que John Searle a baptisée Déclaratifs. Dans la théorisation de la logique illocutoire de Searle, un "verbe performatif" est un verbe existant dans une langue donnée et pouvant participer à la construction d'un acte de langage. Nous emploierons pour notre part le terme performatif en lui donnant le sens de Searle.

Après une tentative de séparation des actes de discours en CONSTATIFS et ASSERTIFS, qu'il jugea lui-même insuffisante, Austin pose les bases de la logique illocutoire (qui sera réellement développée par Searle et Vanderveken) en distinguant pour un énoncé ses aspects

- locutoire (qui se rapporte à l'énonciation en elle-même),
- illocutoire (qui se rapporte à l'acte effectué par l'énoncé),
- perlocutoire (qui se rapporte aux effets de l'énoncé).

Par exemple, énoncer "Le ciel est bleu" correspond à la dimension locutoire. La dimension illocutoire de cet énoncé est une assertion. La dimension perlocutoire peut être, suivant le contexte, une simple transmission information, un refus d'aller au cinéma, une incitation à la randonnée. . .

À la suite des travaux de formalisation de Searle et Vanderveken, un acte de langage se compose d'une force illocutoire et d'un contenu propositionnel, respectivement baptisés F et P.

A	=	F	(P)
Un acte de langage	est égal à	une force illocutoire	appliquée à un contenu propositionnel

TAB. 4.1 – Forme d'un acte de langage

Le contenu propositionnel d'un énoncé est sa composante *représentationnelle*. Il correspond à une description d'un état d'un monde possible. La force illocutoire de l'énoncé est la composante *intentionnelle* de cet énoncé. Ainsi, dans les énonciations

“Est-ce qu’il pleut ?” et “Est-ce que tu vas bien ?”, la force illocutoire est *a priori* la même³ et correspond à une question, tandis que les contenus propositionnels sont différents. De même, les énonciations “Pourvu qu’il fasse beau demain” et “Il fera beau demain” ont, toujours *a priori*, des forces illocutoires différentes, la première pouvant être considérée comme un souhait, la seconde comme une affirmation, mais un contenu propositionnel identique (exemples de [Koning et Pesty, 2001]).

Le force illocutoire est définie par six composantes, dont la première, le but illocutoire, permet de faire le lien entre le monde actuel et le monde possible décrit dans le contenu propositionnel. Cette composante est la plus importante des six et la taxonomie classique des actes de langage, présentée dans la section suivante, se sert de cette composante comme critère discriminant.

4.1.2 But illocutoire et taxonomie des actes de langage

Après les premières tentatives de classification d’Austin, Searle [Searle, 1969] proposa une taxonomie plus rigoureuse (bien que contestable sur plusieurs points comme nous le verrons plus loin). Il décida de classer les forces illocutoires selon leur direction d’ajustement. Comme le dit Daniel Vanderveken : “*Lors de l’accomplissement d’un acte de discours de la forme F(P), le locuteur exprime en général le contenu propositionnel avec l’intention qu’une correspondance soit établie entre le langage et le monde suivant une certaine direction d’ajustement*” [Vanderveken, 1988]. Le but illocutoire est, parmi les composantes d’une force illocutoire, celle qui détermine la direction d’ajustement. Il existe cinq buts illocutoires distinguant cinq classes de performatifs :

- les ASSERTIFS, qui vont des mots au monde (“Il pleut”);
- les DIRECTIFS, qui vont du monde aux mots (“Ouvrez la fenêtre”);
- les ENGAGEANTS (ou COMMISSIFS), qui vont, eux aussi, du monde aux mots (“Je pense que”);
- les EXPRESSIFS, qui ne sont reliés ni au monde, ni aux mots (“merci”);
- les DÉCLARATIFS, qui vont dans les deux sens (“Je déclare la guerre”).

Notons l’existence de deux buts illocutoires distincts pour la même direction d’ajustement, les DIRECTIFS et les ENGAGEANTS. Daniel Vanderveken le justifie en faisant remarquer que “*le locuteur et l’allocutaire jouent des rôles tellement fondamentaux [...] que le langage distingue tout naturellement deux buts illocutoires [...]. Dans le cas des forces engageantes, la responsabilité du succès [...] est dévolue au locuteur; dans le cas des forces directives, par contre, elle est dévolue à l’allocutaire.*” [Vanderveken, 1988]. Denis Vernant [Vernant, 1999] a proposé une classification plus systématique, dans laquelle par exemple la distinction entre locuteur et allocutaire est appliquée à toutes les directions d’ajustement, et non uniquement à celle générant DIRECTIFS et ENGAGEANTS. Cette classification propose aussi un but illocutoire correspondant à la direction d’ajustement des mots vers eux-mêmes, les meta-performatifs, qui permettent d’agir sur le discours lui-même (“*Je conclus*” par exemple). Ces actes méta-discursifs permettent aussi de gérer les dysfonctionnements discursifs, comme l’a abordé Violaine Prince par exemple [Prince, 1996].

Vanderveken a effectué une analyse des verbes performatifs français [Vanderveken, 1988]. À la suite d’un avant-propos où il présente les difficultés intrinsèques à

³Nous verrons ultérieurement, section 4.4, que contrairement à Austin que se plaçait dans une perspective purement monologique, nous prenons en compte les interlocuteurs dans une optique de *co-construction* de la conversation, situation dans laquelle le sens d’un énoncé n’est pas figé, mais constamment négociable.

la formalisation de la langue naturelle, il y décrit l'ensemble des verbes performatifs du français. Dans chacune des catégories de performatifs, il précise une force illocutoire primitive, c'est-à-dire une force illocutoire dont toutes les composantes (qui seront présentées dans la section suivante, 4.1.3) sont neutres, ou identiques pour l'ensemble de la classe.

Pour les ASSERTIFS, il retient “*affirmer*” comme force illocutoire primitive et y inclut *suggérer, prédire, jurer, insister, confier, ...*

Pour les ENGAGEANTS, il retient “*s’engager à*” comme force illocutoire primitive et y inclut *promettre, accepter, assurer, certifier, parier, ...*

Pour les DIRECTIFS il précise qu’il n’existe pas de force illocutoire primitive en français⁴, tous les DIRECTIFS ayant des conditions spéciales (“*En général, ces forces ont un mode spécial d’accomplissement qui consiste ou bien à donner une option de refus à l’allocutaire ou bien à enlever préemptoirement une telle option.*” [Vanderveken, 1988]). Daniel Vanderveken inclut dans les *directifs demander, interdire, supplier, revendiquer, conseiller, ...*

Pour les DÉCLARATIFS il retient “*déclarer*” comme force illocutoire primitive, et il y inclut *démissionner, renoncer, bénir, définir, ...*

Pour les EXPRESSIFS il dit que la “*force illocutoire expressive primitive est une entité théorique, un cas limite des forces illocutoires*”. En effet, de même que pour les directifs il n’existe pas de performatif “neutre” parmi les EXPRESSIFS, mais cette fois-ci, pour des raisons théoriques : les expressifs sont, par nature associés à des états mentaux qui constituent les conditions de sincérité (une des composantes de la force illocutoire, voir section suivante, 4.1.3) de la force illocutoire. Chaque expressif spécifie donc, au minimum des contraintes sur ses conditions de sincérité. Daniel Vanderveken inclut dans les *expressifs remercier, s’excuser, féliciter, huer, saluer, ...*

4.1.3 Les composantes de la force illocutoire

Selon Daniel Vanderveken, toute force illocutoire est divisée en six composantes. Il a montré que ces six composantes suffisaient à elles seules pour caractériser l'ensemble des performatifs d'une langue. Ces six composantes, qui ne sont pas indépendantes les unes des autres, sont :

- le but illocutoire ;
- le mode d’accomplissement ;
- les conditions sur le contenu propositionnel ;
- les conditions préparatoires ;
- les conditions de sincérité ;
- le degré de puissance.

Le but illocutoire en tant que facteur discriminant de la classification des forces illocutoires, a déjà été présenté dans la section précédente.

Le mode d’accomplissement détermine comment “*le but illocutoire doit être accompli sur le contenu propositionnel. [...] D’un point de vue logique, le mode d’accomplissement d’une force illocutoire restreint les conditions d’accomplissement de son but en exigeant certains moyens ou façons spécifiques d’accomplir ce but.*” [Vanderveken, 1988]. Ainsi, une prière a un mode d’accomplissement qui spécifie que

⁴contrairement, par exemple, à l’anglais qui dispose du verbe *to direct*.

la requête doit être humble ou polie ; dans un ordre, le mode d’accomplissement précise que l’allocutaire ne dispose pas d’option de refus ; etc.

Les conditions sur le contenu propositionnel représentent les “*conditions sur l’ensemble des propositions qui peuvent être des contenus propositionnels*” de tel ou tel acte de langage. “*Le contenu propositionnel d’une promesse, par exemple, doit représenter une action future du locuteur.*” [Vanderveken, 1988]. Ce type de conditions temporelles évite des énoncés tels que “*Je t’ordonne d’avoir appris le portugais l’année passée*” que Vanderveken qualifie de “*linguistiquement bizarres*”.

Les conditions préparatoires déterminent les propositions que le locuteur doit présupposer dans le contexte d’énonciation de l’acte. Ainsi, les conditions préparatoires d’un blâme comprennent le fait que l’allocutaire doit avoir fait quelque chose de mauvais ou répréhensible. Un acte ne respectant pas ses conditions préparatoires (comme par exemple, un *témoignage* de la part de quelqu’un n’ayant pas été témoin) est dit *défectueux* (même si le locuteur a perlocutoirement convaincu le jury, dans l’exemple précédent).

Les conditions de sincérité vérifient que le locuteur a bien certains états mentaux. Ainsi, tous les ASSERTIFS ont comme condition de sincérité que le locuteur croit le contenu propositionnel. Selon la validation de cette condition, l’acte est dit *sincère* ou *insincère*.

Le degré de puissance détermine à quel point le locuteur s’investit dans l’énonciation. Ainsi, “*le degré de puissance d’un témoignage est plus fort que celui d’une assertion, parce qu’un locuteur qui témoigne exprime une croyance plus forte*” [Vanderveken, 1988].

Les différentes composantes de la force illocutoire peuvent servir à guider l’identification de la force illocutoire de l’énoncé, ou, une fois cette force identifiée, à compléter les informations que l’on peut obtenir d’un message (voir notre section 4.4.1 sur l’interprétation et l’expression d’un message). Ainsi, une demande ôtant explicitement toute option de refus (par l’emploi de termes tels que “impérativement” ou “sans faute”) orientera l’interprétation de l’énoncé vers un ORDRE ou un COMMANDEMENT, tandis que le même énoncé, accompagné de ce que Goffman qualifie de *softeners* (“excusez-moi”, “si je puis me permettre”, ...), sera plus considéré comme une SUGGESTION ou une RECOMMANDATION. De même, si l’on considère que l’énonciation est réussie, l’analyse de la force illocutoire employée permet de déduire des informations telles que les *attitudes propositionnelles* du locuteur (espérer, regretter, ... quelque chose) à travers les conditions de sincérité, l’existence d’une possibilité de refus à travers le mode d’accomplissement, les connaissances du locuteur sur le contexte conversationnel à travers les conditions préparatoires, etc.

4.1.4 Les conditions de succès et de satisfaction

Chaque acte de discours est décomposé en une force illocutoire et un contenu propositionnel qui servent à déterminer ses conditions de succès et de satisfaction. Les *conditions de succès* d’un acte illocutoire sont les conditions qui doivent être remplies dans un contexte d’énonciation pour que le locuteur *réussisse* à y accomplir cet acte [Vanderveken, 1999]. Les *conditions de satisfaction* d’un acte illocutoire

sont les conditions qui doivent être remplies dans un contexte pour qu’il y ait correspondance entre les mots et les choses dans le monde de ce contexte selon la direction voulue d’ajustement [Vanderveken, 1999]. Ainsi, en prenant un point de vue très schématique sur l’exemple d’un assertif, les conditions de satisfaction sont la correspondance entre le contenu propositionnel et le monde (ce contenu est il “vrai” ou “faux”), et les conditions de succès sont la sincérité du locuteur. La table 4.2 présente les différentes combinaisons de cette situation simplifiée.

Succès \ Satisfaction	Vrai	Faux
Sincère	Assertion réussie	Erreur (échec de l’assertion)
Non sincère	Échec du mensonge	Mensonge

TAB. 4.2 – Exemple de combinaison des conditions de succès et de satisfaction sur une assertion

Les *conditions de succès* ont été introduites par Austin. Pour qu’un acte soit effectivement réussi (par exemple le fait de marier ou de faire une promesse), il ne suffit pas de prononcer les mots. Le contexte d’énonciation doit aussi respecter certaines contraintes, différentes selon l’acte de langage que le locuteur cherche à réaliser. Par exemple, il peut être nécessaire que le locuteur aie un statut particulier (seul un prêtre peut marier), ou que le locuteur entreprenne certaines actions futures (dans le cas d’un engagement). Si toutes ces conditions ne sont pas remplies, l’acte est dit *malheureux*. Austin énumère une série de six règles, dont l’infraction peut mener à l’échec de l’acte, chaque règle entraînant un type particulier d’échec. Ainsi, si la procédure n’est pas accomplie correctement, il s’agit d’une *défectuosité* ; si le locuteur ne se comporte pas par la suite comme il s’y est engagé, il s’agit d’une *rupture* ; si le locuteur ne dispose pas du statut requis, ou si les circonstances ne sont pas adaptées, il s’agit d’un *emploi indu* ; etc. [Austin, 1962].

Les *conditions de satisfaction*, apportées par Searle et Vanderveken, s’appliquant à la relation entre l’acte de discours et le monde. Associées aux conditions de succès, elles ouvrent la voie à l’introduction de l’acte dans le monde et à la formalisation de situations telles que le *mensonge* (il n’y a ni succès ni satisfaction de l’acte : le locuteur n’est pas sincère et son énoncé ne correspond pas à l’état du monde) ou l’*erreur* (il y a succès, mais pas satisfaction de l’acte ; pour Austin, qui se limitait aux conditions de succès, l’acte serait réussi, bien qu’il ne corresponde pas à la réalité). Si l’on prend l’exemple des Directifs, dont le sens d’ajustement va du monde aux mots (“Ouvrez la fenêtre” par exemple) les conditions de satisfaction sont les suivantes :

1. le contenu propositionnel est vrai ;
2. il le devient du fait du Directif.

Ainsi, si à l’ordre “Fermez la fenêtre” on répond en fermant la fenêtre et en ajoutant “De toutes façons, j’allais la fermer”, la théorie dit que le DIRECTIF n’est pas satisfait.

4.2 La communication entre agents

Les modèles de communication issus des sciences humaines sont en général qualitatifs et souvent chargés d'une longue culture, rendant d'autant plus difficile leur compréhension, ne serait-ce que dans leurs grandes lignes, en vue de leur exploitation dans un système automatisé. Toutefois, depuis plusieurs années, la communication dans les systèmes multi-agents a cherché à s'inspirer des phénomènes de communication humains [Moulin *et al.*, 1999]. En se basant sur le formalisme des actes de langage (Austin [Austin, 1962], Searle [Searle, 1969] puis Vanderveken [Vanderveken, 1988]), des langages d'interaction ont été développés, parmi lesquels KQML [Finin *et al.*, 1994] et ACL-FIPA [FIPA, 1997; FIPA, 2001a].

Nous présenterons dans cette section l'ACL (*Agent Communication Language*) de la FIPA (*Foundation for Intelligent Physical Agents*) qui fait office de standard parmi les langages d'agents (section 4.2.1), et qui revendique parmi ses ascendants la théorie des actes de langage, présentée précédemment (section 4.1), puis nous expliquerons en quoi nous estimons les langages d'agents actuels (parmi lesquels l'ACL-FIPA) inadaptés à une communication impliquant un agent humain (section 4.2.3).

4.2.1 L'ACL-FIPA

L'ACL-FIPA, sur lequel nous nous concentrerons dans cette section, est issu des travaux de la FIPA et s'est fortement inspiré de KQML [Finin *et al.*, 1994]. Outre ce langage d'agents, la FIPA a de même spécifié de nombreux autres domaines nécessaires à l'élaboration de systèmes multi-agents (architectures d'agents, comportements, protocoles, ...)

Nous nous contenterons ici d'une simple description du langage à travers sa syntaxe et sa sémantique d'abord, puis à travers les différents types d'actes définis, enfin à travers sa spécification des protocoles d'interaction. Les critiques que nous portons à ce langage seront reportées dans la section suivante (section 4.2.3).

a) Syntaxe et sémantique

La figure 4.2 présente à travers un exemple les éléments principaux d'un message au format ACL-FIPA.

```
(request
  :sender (agent-identifiant :name i)
  :receiver (set (agent-identifiant :name j))
  :content ((action (agent-identifiant :name j)
    (deliver box17 (loc 12 19))))
  :protocol fipa-request
  :language FIPA-SL
  :reply-with order567)
```

FIG. 4.2 – Exemple d'un message ACL (extrait de [Koning et Pesty, 2001])

En reprenant la décomposition des messages de type KQML [Finin *et al.*, 1994] (les dernières versions de l'ACL-FIPA ont beaucoup profité des travaux autour du langage KQML), on peut distinguer trois niveaux à l'exemple de la figure 4.2 :

- Un niveau *message*, qui précise le type d’acte de langage employé, l’ontologie employée, le protocole éventuel dans lequel s’intègre le message, etc. Dans l’exemple de la figure 4.2, ce niveau rassemble le champ `:language` et le type d’acte, placé en tête du message, ici, `request`. Ce niveau, associé au niveau *contenu*, correspond à l’aspect *informationnel*, selon la distinction que nous ferons page 120.
- Un niveau *communication*, qui donne les informations nécessaires à la transmission du message, comme émetteur, destinataire, identification du message, ou le protocole dans lequel s’insère le message en cours. Dans l’exemple de la figure 4.2 ce niveau rassemble les champs `:sender`, `:receiver`, `:reply-with` et `:protocol`. Ce niveau correspond à l’aspect *interactionnel*, selon la distinction que nous ferons page 120.
- Un niveau *contenu*, qui se limite à un seul champ, `:content`. Ce niveau, associé au niveau *message*, correspond à l’aspect *informationnel*, selon la distinction que nous ferons page 120.

La sémantique des actes de l’ACL-FIPA a fait l’objet d’une représentation formelle à l’aide du langage SL (*Semantic Language*) [FIPA, 2001d]. Cette sémantique définit les agents en terme d’attitudes mentales :

- BELIEF, qui caractérise une proposition considérée comme vraie ;
- UNCERTAINTY, qui caractérise une proposition pour laquelle l’agent est incertain, mais qui serait plutôt vraie que fausse ;
- CHOICE, qui caractérise les propositions pour lesquelles l’agent souhaite qu’elles deviennent vraies (dans une certaine mesure, cette attitude rejoint les *buts*, ou les *intentions*).

De plus, trois opérateurs permettant le raisonnement sur les actions sont définis :

- FEASIBLE, qui associe une action qu’il est possible d’effectuer et ses conséquences ;
- DONE, qui associe une action effectuée et ses conséquences ;
- AGENT, qui associe un agent et une action, en cours ou prévue.

D’autres opérateurs sont ensuite définis à partir de ces éléments fondateurs (voir la synthèse de Jean-Luc Koning et Sylvie Pesty [Koning et Pesty, 2001], ou les spécifications de la FIPA [FIPA, 2001d]).

b) Les catégories d’actes

L’ACL-FIPA définit deux types d’actes — les actes primitifs et les actes composés — les seconds étant la composition des premiers. Tous les actes ont le même format général, qui comprend l’agent émetteur, l’agent receveur, le type d’acte et la proposition à laquelle cet acte s’applique (on retrouve ici une forme de distinction proche de celle faite dans la théorie des actes de langage entre performatif et contenu propositionnel, voir section 4.1). En outre, lors de leur définition, chaque acte précise un ensemble de préconditions requises pour l’emploi de cet acte (FP, *Feasibility Preconditions*), et un ensemble de conséquences attendues (RE, *Rational Effects*).

Quatre actes primitifs sont définis, à partir desquels dix-huit actes composés sont construits, pour être ensuite classés en cinq catégories définies suivant l’objectif de chaque acte. Ces catégories sont : transmettre une information (typiquement, `Inform`, ou `Confirm`), demander une information (typiquement, `Query-if`), négocier (typiquement, `Propose`), accomplir une action (typiquement, `Agree` ou `Request`) et gérer un problème (`Not Understood` et `Failure`) (voir la synthèse d’Alexandra Berger [Berger, 2001]).

c) Protocoles d'interaction

Outre un format et une sémantique pour les messages, l'ACL-FIPA offre la possibilité de les intégrer au sein d'un protocole de communication. Cette spécification de la dynamique de la conversation est ascendante : les spécifications portent sur les messages à échanger, qui se combinent pour construire la conversation. Cette approche est à l'opposé de celle des conduites de conversation, au fonctionnement descendant : les spécifications portent sur l'objectif global de l'interaction qui conditionne les messages (voir la définition de protocole page 130 et la section sur les conduites de conversation, page 90).

La FIPA propose une bibliothèque de protocoles, destinés à encadrer les interactions dans certaines situations précises⁵. Ces protocoles spécifient des actions autorisées (souvent en nombre limité) suivant l'état de la conversation, et les réactions possibles (souvent en nombre limité) à ces actions. Ainsi, lors d'une interaction "protocolaire", chaque participant à l'interaction se conforme à sa partie du protocole, permettant une convergence rapide vers la résolution du problème auquel le protocole est dédié, mais au prix d'une extrême contrainte de la conversation.

Notons toutefois que la FIPA considère aussi l'existence de plans d'interaction qui peuvent se mettre en place par la seule articulation des actes de discours et d'un comportement rationnel de l'agent. Ces plans, dont l'idée n'est qu'abordée [FIPA, 2001b], permettent une gestion de la communication ne nécessitant pas le recours à un protocole.

4.2.2 Les conduites de conversation (*conversation policy*)

Les travaux sur la communication entre agents ne se limitent pas à la spécification de langages et de protocoles d'interactions, comme on pourrait le supposer à la lecture des sections précédentes. Ils portent aussi, plus récemment, sur une spécification plus globale de la communication, là où les travaux précédents se concentraient sur une approche plus locale, au niveau des messages échangés. Ces nouvelles approches sont baptisées *conversation policies*, que nous traduirons par *conduites de conversation*. Là où les protocoles d'interaction spécifient des échanges entraînant la constitution d'une interaction (approche ascendante, ou "*bottom-up*"), les conduites de conversation spécifient les objectifs d'une interaction qui entraîne une décomposition en échanges de messages (approche descendante, ou "*top-down*").

On peut considérer que les conduites de conversation ne commencent à se constituer comme champ de recherche que depuis la fin des années 90 [Bradshaw *et al.*, 1997; Greaves et Bradshaw, 1999]. La jeunesse du domaine fait que bien des points restent encore discutés, à commencer par une définition précise de ce qu'est une conduite de conversation. Mark Greaves et ses collègues [Greaves *et al.*, 1999] ont proposé d'en faire une "*spécification déclarative qui dirige les communications entre agents logiciels utilisant un langage de communication d'agents*". Lors du développement de cette définition, les auteurs précisent certains points qui trouvent un

⁵"Ongoing conversations between agents often fall into typical patterns. In such cases, certain message sequences are expected, and, at any point in the conversation, other messages are expected to follow. These typical patterns of message exchange are called interaction protocols. [A] very pragmatic view is to pre-specify the IPs, so that a simpler agent implementation can nevertheless engage in meaningful conversation with other agents, simply by carefully following the known IP." [FIPA, 2001c]

écho particulier dans le cadre de cette thèse. Principalement, ils notent qu'un même message pouvant atteindre différents buts communicationnels (*communicative goals*, nous parlerions "d'effets perlocutoires"), il est nécessaire de considérer ce message dans son contexte⁶ et que toute conversation est sujette à des exceptions. Ce point est fondamental dans le cadre d'interactions entre agents hétérogènes, principalement s'il s'agit d'humains. Comme le font remarquer Renée Elio et ses collègues [Elio et al., 2000], "*un agent doit être développé en sachant qu'un protocole peut être enfreint*".

Alors que les actes de langage échangés entre agents bénéficient de caractérisations précises, il n'existe pas encore ne serait-ce qu'un consensus sur les spécifications d'interactions à un haut niveau [Phillips et Link, 1999]. Les conduites de conversation ont pour but de préciser ce domaine. L'apport des conduites de conversation par rapport à ce qui peut être obtenu d'un travail restreint aux actes de langage (donc au niveau du message lui-même) est triple. (1) En spécifiant des directions globales à tenir (par la spécification d'un objectif général ou de points de passages intermédiaires) les risques d'interblocages sont limités; la gestion des exceptions simplifiée (elles n'ont pas à être explicitement prévues en chaque étape de l'interaction); le choix des protocoles de communication adaptés à la situation est conditionné par ces instructions de haut niveau, permettant par exemple leur remise en cause en cas de problèmes dans leur déroulement. (2) L'explicitation du contexte conversationnel permet à l'agent de traiter des messages ne faisant pas partie du cours normal de la conversation en mettant en place un mécanisme (qui recouvre en partie ce que nous présenterons section 4.4.1 sous le nom d'*interprétation*) indépendant des protocoles suivis⁷. (3) Plusieurs auteurs considérant les conduites de conversations comme liées à une ontologie des domaines de discours, les discours eux-mêmes n'en étant que des spécialisations, il est possible de définir un nombre réduit de conduites conversationnelles couvrant un grand nombre de conversations. Ce dernier point de vue simplifie de plus la séparation entre les compétences liées à la tâche et les compétences purement conversationnelles.

Cependant, la plupart des approches des conduites de conversation requièrent que chacune d'elle soit déclarée, à la manière des protocoles où chaque agent déclare quel(s) protocole(s) il suit et s'y conforme. Il nous semble important de remarquer ce point, qui apporte un intérêt indéniable lors de la communication entre agents artificiels, mais qui risque de devenir problématique pour faire communiquer agents naturels et agents humains, si l'on cherche une interaction "confortable" n'imposant qu'un minimum de contraintes. Déclarer les conduites de conversation suivies doit-il faire partie de ce minimum? Néanmoins, la déclaration des conduites conversationnelles a tout de même des avantages. En donnant aux différents interactants des informations sur les conduites suivies, la création d'un modèle prédictif propre à simplifier l'interaction est facilitée. Et pour un agent se tenir à une (ou plusieurs) conduite(s) conversationnelle(s) apporte une garantie de cohérence de son compor-

⁶À propos de l'architecture d'agent KAoS, Jeffrey Bradshaw, un collègue de Mark Greaves dit "*Unlike most agent communication architectures, KAoS explicitly takes into account not only the individual message in isolation, but also the various sequences in which a particular message may occur.*" [Bradshaw et al., 1997]

⁷"*By following the behaviour of the conversation specification when possible and deferring to the policy to derive behaviour in exceptional circumstances, an agent is able to function predictably under normal situations and still act rationally in abnormal situations.*" [Phillips et Link, 1999]

tement, situation elle aussi propre à simplifier l'interaction. Toutefois, la question de savoir ce qu'un agent doit déclarer explicitement comme guidant son comportement, et la fidélité avec laquelle il doit se tenir à ces déclarations dépasse largement le cadre des conduites de conversations.

La jeunesse du domaine fait qu'il n'existe pas encore de modèle éprouvé, ni même reconnu ou étudié à grande échelle. Étant donnée la coexistence de diverses conceptions d'une conduite de conversation, il existe plusieurs modèles cherchant à en rendre compte. Certains [Elio *et al.*, 2000] considèrent les conduites de conversation comme des règles d'assemblage d'échanges question-réponse (à la manière de Marc-Philippe Huget assemblant des micro-protocoles [Huget, 2001]). D'autres [Phillips et Link, 1999] y voient essentiellement le moyen de gérer toutes les situations d'échec, ou inattendues, permettant la simplification du protocole d'interaction qui se limite alors à la description de la conversation telle qu'elle se déroule dans une situation idéale. D'autres enfin [Lin *et al.*, 1999] lient les conduites de conversation à des *sujets* (*topics*) de conversation, ces derniers donnant le cadre général de l'interaction tandis que les premières organisent des *schemata* générant directement des actes de langage (rejoignant à ce niveau les travaux de Renée Elio précédemment cités). On remarquera dans ce dernier exemple que pour de nombreux autres chercheurs ce sont les *sujets de conversation* qui remplissent le rôle défini pour les *conduites de conversation*, marque supplémentaire du manque actuel de consensus dans le domaine.

Sur l'ensemble des modèles existant, peu ont fait l'objet d'une spécification précise, et encore moins ont fait l'objet de réalisations concrètes menant à une interaction entre agents. Lin et ses collègues vont jusqu'à une formalisation logique de leurs *schemata* [Lin *et al.*, 1999]. Renée Elio et ses collègues présentent un agent capable d'effectuer des accès à une base de données à travers une conversation [Elio *et al.*, 2000]. Enfin, notons que les travaux de Greaves et ses collègues [Greaves et Bradshaw, 1999] sont directement issus de KAoS, une architecture d'agent ayant été, elle, à la base de plusieurs applications [Bradshaw *et al.*, 1997].

4.2.3 Les langages d'agents (artificiels) et la langue naturelle

Le langage d'agent que nous avons présenté (ACL-FIPA, section 4.2.1) est, comme son nom l'indique, un langage d'*agent artificiel*. De la même façon que les autres langages d'agents (nous ferions les mêmes remarques pour KQML, par exemple), il a été développé de façon à permettre l'interaction entre agents artificiels. Nous cherchons pour notre part à définir un modèle de conversation qui puisse permettre à des agents artificiels de communiquer avec des humains.

Cette section traite donc des problèmes que poserait l'emploi de ces langages d'agent dans le cadre d'une interaction impliquant des êtres humains. Nous commencerons par le problème de la rigueur de la communication machine-machine, puis nous parlerons des liens somme toute assez éloignés qu'entretiennent les langages d'agents et la théorie des actes de langage.

a) La rigueur de la communication machine-machine

Comme le fait remarquer Licklider "*men naturally speak redundant languages [...] employing 20 to 60 elementary symbols. Computers "naturally" speak nonredundant*

languages, usually with only two elementary symbols [...]". Cette distinction, que Licklider place au niveau de symboles très "primitifs" (lettres ou phonèmes pour l'humain, bits pour l'ordinateur) se retrouve à des niveaux plus élevés lorsque l'on considère des énoncés en langue naturelle et des messages ACL-FIPA ou KQML.

Il est simple de faire communiquer des machines en employant un nombre réduit de symboles, combinés ensuite de façon à créer des ensembles plus complexes. C'est d'ailleurs la voie qui est suivie lors de la définition des langages d'agents, que ce soit au niveau des "*communicative acts*" ACL (pour lesquels quatre actes primitifs sont définis, les autres en étant des compositions) ou de la définition de leur sémantique (qui s'effectue par combinaison de seulement trois attitudes mentales). Inversement, chez l'humain, la tendance est plus à la multiplication de symboles. Ainsi, même s'il est possible de représenter chaque force illocutoire à l'aide de six paramètres (voir page 85), la langue naturelle préfère employer des performatifs qui correspondent à un ensemble de valeurs données de ces six paramètres. Si nécessaire, les performatifs peuvent ensuite être ajustés en employant des adverbes ou des tournures de phrases particulières qui vont influencer sur telle ou telle composante de la force illocutoire.

Ainsi, dans leur façon de gérer les messages, les langages d'agent ont actuellement tendance à employer un nombre réduit de symboles, et à les combiner de manière éventuellement complexe. Dans la gestion de leurs enchaînements, la stratégie est identique, puisque, lors du déroulement de classiques protocoles de coordination, d'enchères ou de négociation, il peut arriver que les agents émettent une quantité de messages qui serait inacceptable si certains des agents participant à l'interaction devaient être des humains⁸.

De même, dans l'optique d'une communication agréable avec l'humain, il n'est pas possible d'employer des protocoles tels qu'ils sont définis pour les agents. Principalement (comme nous l'avons vu page 90), il faudrait que l'humain s'engage à suivre le protocole, ce qui nécessite, soit une préparation préalable, soit que l'interface de communication présente les actions disponibles à chaque étape. Il faudrait de plus que l'humain sache en quelle étape du protocole il se trouve, comment il y évolue suivant les actions de ses interlocuteurs, etc. Ces contraintes ne sont pas rédhibitoires, beaucoup d'interfaces homme-machine en sont pourvues, et les utilisateurs s'en accommodent, plus ou moins bien. Toutefois, dans notre objectif d'une interaction "naturelle" et "confortable" avec l'être humain, il nous faut chercher un moyen de lever au maximum ces contraintes.

b) Le lien avec la théorie des actes de langage

Bien que se réclamant de la théorie des actes de langage, l'ACL-FIPA n'en a retenu que certains aspects formels et terminologiques. Ainsi, par exemple, les "performatifs" de l'ACL-FIPA (*communicative acts*) ne correspondent pas aux performatifs de la théorie des actes de langage.

Si l'ACL-FIPA se base sur les actes de langage, il se base plutôt sur la vision qu'en avait Austin et Searle (voir les remarques de Ludovic Alidra [Alidra, 1997]), c'est-à-dire une vision purement monologique. Les performatifs sont associés à une syntaxe et une sémantique strictement définies et *communes*, permettant une communication au sens de Shannon : *encodage* du message, puis *décodage* par le récep-

⁸Sur une tâche de déplacement concerté de trois agents dans un espace d'une quinzaine de places, Aaron Armstrong et Edmund Durfee [Armstrong et Durfee, 1997] montrent que selon les algorithmes de résolution de conflit, la quantité de messages échangés peut varier de 300 à 700.

teur, suivant une procédure strictement inverse à celle d'encodage. Ainsi, en quelque sorte, l'émetteur du message conditionne la méthode de décodage que doit employer le récepteur si la communication doit réussir, interdisant toute liberté de traitement, toute *interprétation* du message. Ce type de message reste centré exclusivement sur l'intention communicative du locuteur, l'allocutaire ayant un rôle secondaire [Alidra, 1997].

Notons enfin que les "performatifs" de l'ACL-FIPA sont loin de présenter la même variété que les performatifs de la langue naturelle. Ainsi, un *request* pourrait (comme nous l'avons dit il n'y a pas de correspondance globale entre les performatifs ACL-FIPA et les performatifs de la théorie des actes de langage) correspondre, en français, à *exiger, solliciter, demander* ou *supplier*. Une certaine quantité d'information disponible en langue naturelle grâce à la variabilité qui peut exister entre plusieurs performatifs n'est pas exprimable par l'ACL-FIPA.

Les langages d'agents, et plus particulièrement l'ACL-FIPA, ont été développés de façon à servir les différentes fonctions des agents et ont été adaptés aux besoins des agents artificiels (en somme, de la même façon que la langue naturelle est adaptée aux besoins communicationnels des humains). Ils ne cherchent pas à suivre une théorie de l'interaction humaine (malgré les emprunts à la théorie des actes de langage), et donc, il est normal qu'ils n'y soient que peu adaptés. Néanmoins, les tentatives d'application de la théorie des actes de langage aux langages d'agents participent au rapprochement de ces langages et de la langue naturelle, même si comme nous l'avons montré, cette application n'a été que limitée.

4.3 Un oubli majeur : le Sens

Le paradoxe de la plupart des travaux qui exploitent le sens (tel que nous le faisons), est que pour travailler sur un sens, il est souvent extrait et transcrit sous forme de symbole(s). Or une fois que ce sens est devenu symbole, est-il encore porteur de toutes les caractéristiques qui en faisaient sa spécificité et son intérêt ? Searle se base sur son exemple, maintenant classique, de la chambre chinoise⁹ pour montrer que les ordinateurs qui manipulent des symboles sont fondamentalement différents des humains dont le cerveau manipule, en schématisant, des contenus sémantiques. La manipulation de symboles sensés retranscrire le sens nous priverait donc d'une partie (fondamentale ?) de ses propriétés. Toutefois, ce point n'est pas excessivement dérangeant pour nos travaux dans la mesure où notre objectif n'est pas la recopie de phénomènes de l'interaction humaine, ni même la conception (théorique ou pratique) d'un système au fonctionnement (ou aux résultats) humainement plausible. Notre objectif consiste en la recherche d'outils permettant d'améliorer l'interaction entre hommes et machines, ouvrant la porte à la création de communautés mixtes ayant une réelle dynamique de symbiose entre ses membres.

Pour une vue d'ensemble de l'approche du traitement automatique du sens, voir le texte synthétique de Gérard Sabah [Sabah, 1997b].

⁹Un homme ne comprenant pas le chinois est enfermé dans une chambre et reçoit des instructions en chinois. Il répond à ces instructions en suivant une table de correspondance entre les symboles chinois, et donne ensuite une réponse, elle même en chinois. Durant tout le processus, l'homme n'a fait que manipuler des signes, sans avoir accès aux sens portés, pourtant, il remplit la tâche qu'on lui demande.

4.3.1 La représentation du sens

Comme nous venons de le préciser, représenter le sens (de façon, dans notre cas, à pouvoir se livrer à des traitements automatisés) pose un problème profond. Deux grandes classes de solutions ont été proposées pour représenter le sens : celles qui font correspondre d'une manière directe un sens et un symbole et celles pour lesquelles le sens apparaît dans les relations qui existent entre différents symboles. Une fois encore, ces deux classes constituent deux extrêmes, et des solutions intermédiaires existent (association d'un sens à un ensemble de symboles par exemple).

La première classe est extrêmement simple. En faisant correspondre de manière biunivoque un sens à un symbole, il est possible de manipuler directement le sens en employant des méthodes habituelles en informatique où le traitement s'effectue sur des variables explicites. Cette classe est celle des méthodes dites *symboliques* de traitement de l'information et correspond au mouvement du *cognitivism* dont Jerry Fodor est un des initiateurs.

La deuxième classe rassemble essentiellement des représentations à base de réseaux, où le sens d'un élément est en partie¹⁰ défini par les relations qu'il entretient avec les autres éléments. On peut citer les réseaux sémantiques, les graphes conceptuels [Sowa, 1984] ou la base de connaissances du projet CYC. Cette représentation n'est pas sans rappeler le principe du procès de semiosis¹¹. On trouverait donc dans cette représentation formelle un moyen de capturer (ou plutôt d'évoquer) le sens qui ait une justification philosophique.

Des approches plus récentes en robotique abordent le problème de l'ancrage perceptif des symboles, et le sens d'un symbole peut alors être vu comme un faisceau de traits perceptifs associés entre eux. Souvent, ces approches sont couplées à des systèmes d'apprentissage (voir par exemple les travaux de Luc Steels et Frédéric Kaplan [Steels et Kaplan, 1999; Kaplan et Steels, 2000]) permettant au robot de faire progressivement un lien entre une abstraction interne (qui peut être vue comme un sens) et un ensemble de perceptions.

Nous employons pour notre part la première méthode (voir les chapitres suivants), plus simple à mettre en œuvre dans un système automatisé, pour ne pas ajouter à la difficulté des traitements du sens que nous souhaitons mettre en évidence la difficulté de manipulation d'une représentation complexe. Puisque nous employons le formalisme des actes de langage pour modéliser les messages échangés par les agents, nous employons aussi ce formalisme pour représenter les sens lors de leur manipulation par l'agent, car, comme le fait remarquer Vanderveken : *“Le langage exerce une fonction médiatrice essentielle dans l'expression de pensées. Toute pensée conçue par un sujet humain est en principe exprimable par les moyens de son langage lors de l'accomplissement d'actes de discours.”* ([Vanderveken, 1988], préface).

Toutefois, nous nous plaçons dans cette perspective symbolique du sens au niveau du traitement et non au niveau de la transmission. Nous considérons, à la différence

¹⁰Souvent, le sens est quand même défini *a priori* par les concepteurs du système, les relations entre éléments n'apportant que des enrichissements.

¹¹Parmi les approches philosophiques du sens celle de Pierce considère le sens comme une construction dynamique de l'esprit d'un interprète : dans ce qu'il définit comme le procès de sémiotique, “[le signe] s'adresse à quelqu'un, i.e. il crée dans l'esprit de cette personne un signe équivalent[...]. Le signe qu'il crée, je l'appelle interprétant du premier signe. [...] L'interprétant renvoie à un objet auquel lui-même renvoie de la même manière, l'interprétant devenant lui-même un signe, et ainsi de suite” [Pierce, 1978]. Il considère donc un processus dynamique sans fin, dans lequel des signes renvoient à des signes.

des modèles classiques d'interaction, qu'un message n'est pas porteur d'un sens, mais d'un *potentiel* de sens. Il n'est pas, comme le dit Shannon, encodé par un émetteur pour être ensuite décodé par un récepteur. Dans le modèle que nous allons présenter, chaque interactant manipule des sens précis, mais le monde ne contient pas de sens précis. Il ne contient que des signes, laissés à l'interprétation.

4.3.2 Le potentiel de sens

La plus ancienne définition de pragmatique (selon Françoise Armengaud [Armengaud, 1985]) est de Morris, qui dit : "*La pragmatique est cette partie de la sémiotique qui traite du rapport entre les signes et les usagers des signes*". Cette approche est fondamentale dans le travail présenté ici, car elle montre que l'interprétation d'un signe dépend de l'interprétant. Le signe perd donc toute possibilité d'être porteur d'un sens absolu, et se retrouve porteur d'un *potentiel de sens*. La réduction de ce potentiel (pour reprendre le terme de la physique quantique qui "réduit" un paquet d'ondes — probabiliste — en une particule à la position mesurée) dépend de la personne qui l'interprète, chargée de toutes ses expériences, attentes et connaissances, du contexte dans lequel elle est et perçoit le signe, ainsi que du contexte dans lequel est placé le signe.

L'ensemble du chapitre "Répliques et réponses" de l'ouvrage "Façons de parler" d'Erwing Goffman [Goffman, 1981] s'attache précisément à définir de quelle façon la conversation peut être abordée pour étudier ce que nous appelons potentiel de sens, et qu'il définit comme "[...] *la différence entre ce qui est dit et ce qui est signifié et la variété de ce qui peut être signifié par ce qui est dit [permettant] au locuteur de transmettre consciemment, au moyen des mêmes mots, un certain sens à tel auditeur, et un autre sens différent (ou d'autres sens supplémentaires) à tel autre.*"

Enfin, comme nous le verrons dans le dialogue présenté dans la table 4.7 p.118 dans le cadre de conversations humaines, un énoncé comme "*Tu as le téléphone ici*" (ou, "*Tu ne m'as pas dit avec qui tu as mangé à midi*") peut être l'origine de bien des dialogues, tous plausibles. De la même façon, l'exemple classique de l'énoncé "Je reviendrai", montre bien à quel point un sens peut être dépendant du contexte d'énonciation :

- d'un ami à un autre, se quittant, c'est une promesse ;
- venant du client d'un restaurant, c'est un compliment ;
- de la part d'un policier, c'est un avertissement ;
- d'un propriétaire à son locataire, ce peut-être une menace ;
- venant d'un soldat mobilisé, c'est une espérance ;
- ... et ainsi de suite.

Ces exemples montrent bien que chaque message est susceptible d'être interprété et que le contexte (relations entre les interactants ou échanges précédents par exemple) participe à l'obtention d'un sens à partir d'un message. Ces exemples ne vont toutefois pas à l'encontre de notre hypothèse selon laquelle ces interactants manipulent, de façon interne, un sens clairement défini. Ainsi, pour un locuteur donné, il existe de façon précise un sens intentionné et pour un auditeur donné, il existe de façon précise un sens interprété, mais ces sens ne sont pas "transportés" tels quels par le message.

Toutefois, un tel choix n'est pas incompatible avec un traitement de l'ambiguïté sous une forme de liste (éventuellement ordonnée, par ordre de probabilité ou de pertinence) d'interprétations concurrentes d'un même message, permettant d'explorer

d'autres pistes si le choix fait n'était pas validé par l'autre interlocuteur. De manière cohérente avec nos remarques précédentes, nous ne prétendons pas qu'il existe une relation directe et non équivoque entre un message (externe) et un sens (interne). Nous posons simplement qu'un message est porteur d'un potentiel de sens, réductible à un ensemble discret de sens, chacun étant représenté, de manière interne à l'interlocuteur, par un symbole (ou ensemble de symboles) précis.

4.4 La place des interactants

Un signe ne se révèle porteur d'un sens ("le signe *fait sens*") qu'à travers un processus d'interprétation, et ce processus d'interprétation est le fait d'un *agent* (une entité qui *agit*, qu'elle soit naturelle ou artificielle) extérieur au message. La communication entre agents nécessite la participation active des différents interactants, sans quoi on retombe dans la situation d'un simple transfert d'information¹². Il s'ensuit que dans une communication entre individus différents les uns des autres le message (signe) sera le même pour chacun, mais le sens extrait sera personnel.

Nous montrerons dans cette section comment les interactants s'intègrent dans la conversation comme vrais *acteurs* et non comme simples émetteurs et récepteurs. Principalement, nous traiterons des différents aspects des processus d'*interprétation* et d'*expression* qui constituent le cœur de cette thèse. Après une sous-section qui synthétisera l'essentiel de ces mécanismes, nous verrons quelles théories des sciences humaines ou de l'informatique ont déjà été proposées pour effectuer ces traitements¹³. Nous présenterons ensuite quels paramètres il nous semble pertinent de prendre en compte chez chacun des interactants pour introduire les phénomènes d'interprétation et d'expression.

4.4.1 Interprétation et expression

Searle et Shannon s'accordent sur une différence claire entre les humains et les systèmes automatisés : les premiers manipulent des sens tandis que les seconds manipulent des symboles. Nous avons nommé *expression* et *interprétation* les mécanismes qui permettent à un humain de passer d'un contenu sémantique¹⁴ à un signe et *vice versa*. Nous avons dit dans la section précédente que les messages étaient porteurs d'un potentiel de sens et que le passage entre ce message porteur d'un potentiel de sens et un sens nécessitait la confrontation à un contexte. L'interprétation d'un message dépend, entre autres, de l'interprétant, tel interprétant obtenant tel sens d'un message, tel autre obtenant un autre sens. Inversement, l'expression d'un sens donné produit des messages différents selon l'individu qui réalise cette expression. De telles considérations sur le message s'accroissent fort bien d'une vision du message comme une entité à part entière, telle qu'elle a été décrite par Alexandre Ribeiro [Ribeiro, 2000], par exemple.

¹²Citons à nouveau Licklider : "[...] *to communicate is more than to send and to receive. Do two tape recorders communicate when they play to each other and record from each other?*" [Licklider, 1968].

¹³Les agents conversationnels décrits dans la section 3.5.3 ne seront pas redétaillés dans cette sous section, mais les mécanismes qu'ils mettent en œuvre, particulièrement dans le cadre des projets *Persona* et *Amusement*, y seraient particulièrement pertinents.

¹⁴Rappelons que ce contenu sémantique "présent à l'esprit" de l'humain n'est qu'une hypothèse que nous avons posée et dont la justification, linguistique, philosophique et/ou psychologique, reste à déterminer.

Nous développons dans cette section les différents points ayant trait à la relation entre les messages et les interactants, faisant d’eux plus que des magnétophones, pour reprendre l’analogie de Licklider. Nous rassemblons tous ces phénomènes sous le nom d’enrichissement contextuel (ou d’adaptation contextuelle). Comme le fait remarquer Gérard Sabah, “*un système de dialogue intelligent doit être capable de comprendre différemment des énoncés identiques apparaissant dans des contextes distincts.*” [Sabah, 1989]. Nous commencerons donc par présenter l’*interprétation* (sous-section a)), puis l’*expression* (sous-section b)), d’un message, processus que nous proposons pour faire le lien entre message et sens, en ancrant ce message dans un contexte conversationnel. Nous verrons ensuite (sous-section c)) que des interactants ayant des mécanismes conversationnels simples, comme les systèmes artificiels limitant l’expression à un simple encodage et l’interprétation à un simple décodage, limitent les informations qu’il est possible d’obtenir d’un message. Nous aborderons ensuite le problème de la multi-modalité (sous-section d)), qui correspond aussi à un enrichissement contextuel dans la mesure où des messages de différentes modalités peuvent agir de concert. Nous présenterons enfin les problèmes pouvant émerger des adaptations contextuelles et des connaissances mutuelles des différents interactants (sous-section e)), avant de préciser les restrictions de notre modèle (sous-section f)).

a) Interprétation

Le processus d’interprétation est une fonction ayant comme paramètres le message reçu et le contexte d’énonciation du message, ce contexte comprenant principalement des connaissances sur le monde, des connaissances sur soi et des connaissances sur l’interlocuteur. Pour être plus précis, les paramètres sont le message reçu et la *représentation* du contexte d’énonciation du point de vue de l’agent (donc, le *modèle* que l’agent a de lui-même et de son interlocuteur, et les connaissances *subjectives* de l’agent sur le monde). Le résultat de cette fonction est le *sens* (ou *les sens*, en cas d’ambiguïtés) de l’énoncé, aux réserves près que nous avons faites sur la représentation du sens dans la section 4.3.

Si l’on se contentait d’une simple *traduction* littérale d’un énoncé pour en obtenir un acte de langage, c’est-à-dire, si l’on se contentait d’une analyse lexico-syntaxique, on obtiendrait les résultats présentés dans la table 4.3. Dans ces deux exemples, la forme de surface de l’énoncé est identique pour ce qui concerne la force illocutoire employée, seul le contenu propositionnel diffère. Une simple traduction associerait à chacun de ces énoncés le performatif “suggérer”, qui ne porte pas de contraintes particulières au niveau de son mode d’accomplissement.

Énoncé	Force illocutoire	Mode d’accomplissement
<i>J’aimerais bien qu’on aille au cinéma</i>	SUGGÉRER	Pas de contraintes particulières
<i>J’aimerais bien que tu ranges ta chambre</i>	SUGGÉRER	Pas de contraintes particulières

TAB. 4.3 – Traduction littérale de deux énoncés vers une même force illocutoire.

Le passage d’un énoncé à une force illocutoire que nous venons de montrer, par

simple *traduction*, est insensible au contexte de l'énoncé. La table suivante (4.4) présente une *interprétation* des énoncés, tenant compte du contexte. Nous y montrons comment une différence de relation entre les interlocuteurs peut entraîner une différence d'interprétation au niveau du mode d'accomplissement de la force illocutoire *littérale*. L'ajout de contraintes dans le mode d'accomplissement entraîne, à son tour, la modification du performatif associé à l'énoncé.

Sens perçu ^a		+	contexte		→	sens interprété ^b	
Énoncé			Relation au locuteur			Force illoc.	Mode d'accomplissement
<i>J'aimerais bien qu'on aille au cinéma</i>			Ami-ami			SUGGÉRER	Pas de contraintes particulières
<i>J'aimerais bien que tu ranges ta chambre</i>			Mère-enfant			ORDONNER	Pas d'option de refus

^aSens représenté sous la forme d'un acte de langage. Cet exemple ne montre que l'énoncé en langue naturelle, la force illocutoire perçue est SUGGÉRER, comme dans l'exemple de la table 4.3

^bSens représenté sous la forme d'un acte de langage.

TAB. 4.4 – Interprétation de deux énoncés vers deux forces illocutoires.

Dans l'exemple présenté, le *message reçu* correspond à la colonne "Force illocutoire" du tableau 4.3, et correspond à la simple traduction "de surface" du contenu de la colonne "Énoncé", dans laquelle se trouve un message en langue naturelle. Le *contexte* est pris en compte à travers les *connaissances sur l'interlocuteur*, ici, la relation sociale. Essentiellement pour des raisons de simplicité les *connaissances sur le monde* et la personnalité du locuteur (les *connaissances sur soi*) ne sont pas prises en compte dans cet exemple. Une partie du *sens* est représentée par la force illocutoire interprétée, le mode d'accomplissement n'étant qu'un des paramètres de cette force illocutoire.

Cet exemple montre bien que le sens d'un énoncé n'en dépend pas exclusivement. Les informations lexicales et syntaxiques d'un énoncé en langue naturelle permettent de donner une indication sur son sens, mais des informations extérieures au message —son contexte— peuvent elles aussi participer à la constitution du sens. Notons enfin que "ce" sens est purement subjectif et personnel à l'agent interprétant, contrairement aux situations classiques où le message *encodé* à son émission est *décodé* à sa réception. Nous verrons dans la section 4.5 comment la différence entre ce sens interprété et le sens intentionné du locuteur participe à la dynamique de la conversation.

b) Expression

Le processus d'expression est une fonction qui prend comme paramètres un *sens* (avec les mêmes réserves que ci-dessus) et la *représentation* du contexte d'énonciation du point de vue de l'agent (donc, de la même façon que ci-dessus, le *modèle* que l'agent a de lui-même et de son interlocuteur, et les *connaissances subjectives* de

l'agent sur le monde). Le résultat de cette fonction est un message, potentiellement porteur du sens intentionné, destiné à un allocutaire (ou un groupe d'allocutaires) précis, dans un contexte conversationnel donné.

L'exemple de la table 4.5 montre comment une différence de confiance en la source de données peut entraîner le choix d'un énoncé différent. Dans cet exemple, nous nous plaçons dans le cadre de la recommandation de film, où les évaluations de films sont données par les utilisateurs. Dans un tel contexte, un film bien noté par une centaine d'utilisateurs, sera plus susceptible d'être un "bon" film qu'un film bien noté par quelques utilisateurs seulement. Dans une situation donnée, un agent peut chercher à énoncer un acte de langage dont le performatif est l'ASSERTIF primitif (*affirmer*). Cet acte intentionné peut être enrichi d'informations contextuelles permettant de refléter la confiance qu'a l'agent en les données sur lesquelles il se base. Dans l'exemple présenté, une confiance élevée se traduit par une augmentation du degré de puissance de la force illocutoire intentionnée, entraînant la transformation d'*Affirmer* en *Soutenir* en cas de grande confiance et en *Penser* en cas de faible confiance.

Contexte + Sens intentionné ^a		→ Sens exprimé ^b		
Validité de la note	Force illoc.	Force illoc.	Degré de puissance	Énoncé possible
Bonne	AFFIRMER	SOUTENIR	+2	Je suis certain que ...
Moyenne	AFFIRMER	PENSER	-1	À mon avis, il est possible que ...

^aSens représenté sous la forme d'un acte de langage.

^bSens représenté sous la forme d'un acte de langage et accompagné d'une *traduction* possible en langue naturelle pour la lecture de l'exemple.

TAB. 4.5 – Deux expressions possibles d'une même force illocutoire

Dans cet exemple, une partie du *sens* est représentée par la force illocutoire intentionnée. Le *contexte* considéré est une *connaissance sur le monde* et correspond à la confiance en la note. Les *connaissances sur l'interlocuteur* et *connaissances sur soi* ne sont pas prises en compte, une fois encore pour conserver à l'exemple sa simplicité. Le *résultat* de l'expression est la force illocutoire exprimée, dont une traduction en langue naturelle est présentée dans la dernière colonne.

Une conséquence du principe d'expression est que le message va être porteur de nombreuses informations sur le locuteur. Quelles sont les politiques de conversation employées par le locuteur? Quelle relation estime-t-il avoir avec ses allocutaires? Quelle vision d'eux a-t-il? Quels rôles et statuts le locuteur se donne-t-il et donne-t-il aux allocutaires? Enfin, comment tous ces paramètres changent-ils dans des situations conversationnelles différentes? Les réponses à ces différentes questions permettent à chaque participant de construire un modèle de ses interlocuteurs, modèle sur lequel il pourra s'appuyer pour la suite de ses conversations. De manière très générale, on peut dire que la phase d'expression est sensée adapter le message

de façon à ce qu'il remplisse au mieux son objectif. Il est intéressant de noter que cet objectif n'est pas forcément celui d'une conversation sans heurts : on peut exploiter les connaissances que l'on a sur ses interlocuteurs pour les insulter plus efficacement, et la rupture de la conversation qui suivrait une insulte efficace peut constituer, du point de vue du locuteur, la situation conversationnelle souhaitée.

c) Connaissances sur le monde et communication

Parmi les fondements de la théorie des actes de langage, on trouve l'application d'une force illocutoire à un contenu propositionnel. Le contenu propositionnel représente un *état de choses* (vrai ou non) d'un monde (monde réel ou monde possible) ([Vanderveken, 1988], p. 83). Pour être exact, cet état de chose est une description d'un monde du point de vue subjectif du locuteur. Cette description du monde (correspondant à la composante *représentationnelle* de l'acte de langage), une fois reçue, peut servir à enrichir les connaissances du récepteur, la force illocutoire (correspondant à la composante *intentionnelle* de l'acte de langage) qui lui est appliquée permettant de savoir de quelle manière cette connaissance est reliée au monde. En effet, les composantes de la force illocutoire permettent par exemple de savoir :

- que le locuteur est capable d'accomplir un acte donné (condition préparatoire des ENGAGEANTS) ;
- quel est le point de vue du locuteur sur les connaissances de l'allocutaire (les conditions préparatoires de RAPPELER présupposent l'oubli, supposé ou effectif, de l'information) ;
- que le monde décrit est un monde possible souhaité (le but illocutoire des DIRECTIFS va du monde aux mots)
- quelle est l'attitude du locuteur vis à vis de l'allocutaire (le mode d'accomplissement de SUPPLIER requiert une insistance humble et soumise).

Dans un système simple où, particulièrement, tous les interactants sont sincères et rationnels, il suffit à chaque récepteur d'un message d'en consulter la force illocutoire (après une éventuelle *interprétation* de l'énoncé) pour savoir de quelle manière il faut considérer l'état du monde décrit par son contenu propositionnel. Dans le cadre d'un *assertif* ou d'un *expressif*, le contenu propositionnel peut être ajouté directement à la base de connaissances (connaissances sur le monde pour l'un et connaissance sur le locuteur pour l'autre). Dans le cadre d'un *Directif* ou d'un *Engageant*, le contenu propositionnel représente un état du monde *futur*, obtenu par l'action du locuteur (*Engageant*) ou de l'allocutaire (*Directif*).

Le modèle de communication de Shannon présente l'extrême de cette situation, puisque le mécanisme d'interprétation est l'exact symétrique de celui d'expression (l'objectif de la communication selon ce modèle est en effet la *recopie* d'un message d'un point à un autre). La composante intentionnelle est alors réduite uniquement au but illocutoire. On voit bien ce phénomène dans le cadre des langages d'agents, comme l'ACL-FIPA (voir section 4.2.1) qui se réclame de la théorie des actes de langage, mais ne spécifie pour chaque classe de performatifs qu'une force illocutoire primitive. Si l'on ne considère que des situations où l'émetteur souhaite transmettre ou demander de l'information à des récepteurs, qui eux intègrent cette information ou répondent aux questions, on retrouve la vision classique en informatique de communication entre deux systèmes s'échangeant des messages de type **ask** (ou **request**) ou **tell** (ou **inform**).

Dans un système plus complexe (interlocuteurs humains par exemple), toutes les situations sont imaginables : le doute sur les informations transmises, le mensonge,

les faux engagements...et surtout la non littéralité des énoncés. Certaines de ces situations sont d'ailleurs parfaitement descriptibles par la logique illocutoire.

d) Multi-modalité

Certains signes de l'environnement peuvent être considérés comme des messages, mais peuvent aussi venir en complément d'un message, pour participer aux processus d'interprétation ou d'expression. Parmi ces signes les plus courants et les plus susceptibles d'effets sont, chez les humains, les expressions faciales et corporelles, et, dans la cadre de la parole, la prosodie. Par exemple, un énoncé accompagné d'un sourire pourra entraîner, lors de la phase d'interprétation, une décrémentation du degré de puissance de la force illocutoire littérale de l'énoncé. Ainsi, un énoncé langagier correspondant à une stricte interdiction mais accompagné d'un sourire peut perdre son statut d'*ordre* pour devenir une *recommandation*, éventuellement négociable (voir figure 4.3). De la même façon, pour exprimer un *Directif* sans laisser d'option de refus (comme *Ordonner* par exemple), il est possible d'employer un énoncé dont la force illocutoire littérale laisse cette option de refus, mais accompagné d'un regard agressif. Dans cette situation, c'est le mode d'accomplissement de l'énoncé langagier qui est modifié lors de l'intégration du message visuel (voir figure 4.3).



“Il est interdit de fumer ici.”

“J’aimerais bien que ce travail soit fini pour la semaine prochaine.”

FIG. 4.3 – Un exemple d'influence multi-modale.

La prise en compte de la multi-modalité, couplée à l'expression et l'interprétation des messages, autorise ainsi une plus grande variabilité dans les messages exprimés, mais aussi une plus grande subtilité dans les différentes composantes de la force illocutoire d'un énoncé. En effet, en français du moins et *a priori* dans l'ensemble des langues naturelles, chaque point de l'espace généré par l'ensemble des valeurs possibles des composantes d'une force illocutoire ne correspond pas forcément à un performatif. Vanderveken signale que l'emploi de modificateurs linguistiques (adverbes par exemple) permet d'atteindre d'autres points de cet espace, occupés ou

non par des performatifs (voir [Vanderveken, 1988], particulièrement le chapitre v). La prise en compte d'éléments extérieurs au message, comme, entre autres, la prosodie, les expressions faciales ou corporelles, les vêtements¹⁵, l'attitude... permet elle aussi d'atteindre d'autres points de cet espace.

e) Adaptation à l'autre et croyances mutuelles

Prendre en compte des interactants qui, à la fois adaptent leurs messages à leurs allocutaires et se basent sur les messages reçus pour se créer un modèle de ces mêmes allocutaires de façon à pouvoir y adapter leurs messages peut facilement entraîner des situations surprenantes d'un point de vue humain, voire paradoxales ou tautologiques. Nous évoquerons dans cette sous-section trois écueils principaux : l'altruisme excessif, dans lequel le locuteur ne centre ses actions *que* sur son interlocuteur ; les références circulaires, où chaque participant à l'interaction intègre dans son modèle des autres la représentation que ces autres se font de lui-même ; et l'exploitation de règles de communication de plus en plus précises, qui peut amener à donner à un agent des capacités de communication "surhumaines" qui peuvent perturber l'interaction.

L'altruisme excessif. On pourrait imaginer approcher de la situation interactionnelle idéale en adaptant entièrement le message à son destinataire, faisant fi de toute la personnalité du locuteur. Un tel comportement, qualifiable de purement altruiste, entraîne pourtant des difficultés. En choisissant une stratégie de communication uniquement centrée sur ses allocutaires (et donc, à première vue, idéale de leur point de vue), l'agent prive ces mêmes allocutaires des informations *personnelles* qui les aideraient à construire eux-mêmes des messages adaptés à l'agent. De plus, à vouloir ainsi s'adapter absolument à chacun de ses interlocuteurs, l'agent risque de paraître incohérent lors d'une discussion impliquant un groupe, car il est susceptible d'*exprimer* ses messages de façon radicalement différente selon ses interlocuteurs. Cette situation peut être évitée en chargeant chaque message de la personnalité de l'agent (et non en l'adaptant *uniquement* aux interlocuteurs). Pour remplir ce rôle "stabilisateur", la personnalité se doit d'être —au moins en partie— stable à l'échelle d'une interaction.

Il n'existe pas, à notre connaissance, de travaux portant sur ce paradoxe : un agent s'adaptant exclusivement à l'autre, au mépris de sa propre personnalité perturbe la conversation plus qu'il ne la facilite. La raison la plus probable est que, dans le domaine des interactions incluant des entités artificielles, il n'existe pas encore d'agents exploitant complètement ces mécanismes somme toute complexes et subtils ; tandis que dans le domaine de la communication humaine, il n'existe que peu de politiques de conversation occultant complètement la personnalité du locuteur (comme dans les cas particulièrement ritualisés d'interaction avec un haut dignitaire religieux par exemple).

Les références circulaires. En toute rigueur, le modèle qu'un agent se fait d'un de ses interlocuteurs doit contenir le modèle que l'interlocuteur se fait de l'agent, lui-même contenant le modèle que l'agent a de son interlocuteur, etc. Plus les niveaux de récursion s'accumulent, plus les effets sur le comportement deviennent marginaux, mais obtenir un ou deux niveaux de récursion permet une anticipation plus

¹⁵Insignes de grade par exemple.

précise des réactions des allocutaires, et donc une adaptation des messages plus en adéquation avec la politique conversationnelle du locuteur.

Ce problème des croyances mutuelles (le fameux : “je sais qu’il sait que je sais. . .”) a été largement étudié en intelligence artificielle, par exemple à travers le paradoxe de Conway¹⁶ et en sciences humaines, principalement à travers la théorie de l’esprit¹⁷.

Jusqu’où contrôler la conversation ? Un dernier problème plus subtil peut venir à l’esprit si l’on considère une interaction entre agents humains et agents artificiels. Considérons un agent qui emploierait toutes les ressources des sciences humaines et de l’intelligence artificielle pour modéliser ses interlocuteurs, anticiper leurs réactions, adapter son style de conversation de façon à ce que ses messages reçoivent l’accueil le plus favorable possible, chercher à gagner leur confiance, etc. Un humain communiquant avec un tel agent, et connaissant les capacités de cet agent, pourrait tout à fait juger l’agent comme manipulateur et hypocrite, entraînant une situation conversationnelle tendue, état complètement opposé à l’objectif de l’agent. Comme le fait remarquer Goffman, il est des règles de comportement qui ne remplissent leur rôle que tant qu’elles ne sont pas explicitées.

Dans le cadre de relations humaines, la prise en compte de l’autre est indispensable, mais les écueils que nous venons de présenter (altruisme excessif, références circulaires et contrôle de la conversation) y sont rarement présents. Le premier car il est rare d’effacer complètement son identité dans une relation (excepté par exemple dans des cas de différences de statuts extrêmes : un domestique, un soldat, ou bien vis-à-vis d’un haut dignitaire religieux). Le deuxième pour de simples raisons de limitations cognitives de l’humain (excepté par exemple lors de la phase préparatoire d’un débat, où les différents protagonistes cherchent à préparer leurs arguments et contre-arguments). La troisième pour les raisons que nous avons déjà citées : certaines règles gagnent en efficacité à ne pas être explicites.

En résumé, “*si l’on veut parler avec quelqu’un, mieux vaut se mettre à sa portée et parler son langage. [...] Lorsque la communication passe bien entre deux individus, on s’aperçoit que ces deux personnes ont des attitudes similaires, des postures et des gestes en harmonie. Leurs échanges verbaux et paraverbaux sont synchronisés.*” [Mucchielli, 1995]

¹⁶Dans sa thèse, Humbert Fiorino [Fiorino, 1998] présente ainsi ce paradoxe : “*Imaginons n enfants jouant ensemble. [...] Supposons que, pendant le jeu, k enfants se salissent le front. Chaque enfant peut voir la tache des autres, mais pas l’état de son propre front. Peu après, le père dit « au moins l’un d’entre vous a le front taché » , exprimant donc un fait que tous connaissent avant qu’il ne parle (si $k > 1$).*” Et pourtant, cette information, en définissant une connaissance commune, permet aux enfants de répondre à la question « *L’un d’entre-vous peut-il prouver qu’il a le front taché ?* », en raisonnant sur les connaissances des autres enfants.

¹⁷“*L’expression « théorie de l’esprit » désigne l’aptitude à expliquer et à prédire ses propres actions et celles d’autres agents intelligents. [L’un des deux grands courants] postule la mise en œuvre d’une méthodologie simulationniste pour expliquer l’aptitude à comprendre et prédire les conduites intentionnelles. [...] Les états mentaux prédits et inférés pour expliquer les conduites intentionnelles peuvent être perceptifs (notamment l’attention [...]), volitionnels (désir) ou épistémiques (savoir que, croire que, penser que, etc.)*” [Nadel et Melot, 1998]

f) Restrictions

Nous ne travaillons que sur la composante intentionnelle des énoncés (la force illocutoire, voir section 4.1). Plus précisément, nous considérons la composante représentationnelle des énoncés (le contenu propositionnel, correspondant à la description d'un état du monde) comme n'étant pas sujette à négociation. Les composantes de la force illocutoire (particulièrement le mode d'accomplissement et le degré de puissance) sont, dans notre modèle, susceptibles d'être modifiées lors de l'interprétation ou de l'expression, à l'exception du but illocutoire qui, par définition, assure le lien entre la force illocutoire et son contenu propositionnel.

De fait, nous laissons volontairement de côté ici le problème du sens implicite des messages. Si à la question "Est-ce que Paul est marié?" on obtient comme réponse "Paul est homosexuel" (exemple cité par Daniel Vanderveken [Vanderveken, 1988], p.74), c'est parce que le locuteur estime comme faisant partie de l'arrière-fonds conversationnel que la plupart des homosexuels ne se marient pas. Il devient délicat dans cette situation de faire la distinction entre ce qui relève de l'interprétation et ce qui relève d'un mécanisme plus classique de raisonnement. On peut en effet imaginer qu'une "boîte noire" chargée de l'interprétation puisse transformer cet énoncé en un simple "non". Notre choix de limiter la variation à la force illocutoire des énoncés nous amène à déléguer à la partie raisonnement de l'agent ce qui, pourtant, peut être considéré comme une *interprétation* (puisque'il s'agit d'un enrichissement contextuel) de cet énoncé en une réponse négative.

4.4.2 Du message au(x) sens et *vice versa*

Le passage du message au sens et du sens au message se synthétise en deux questions : comment le récepteur d'un message en extrait un (ou des) sens? Comment l'émetteur d'un message le crée-t-il à partir de son sens intentionné? Plus généralement, ces deux questions se fondent en une : pourquoi les choses ont été dites ainsi et pas autrement? Ou encore : qu'est-ce qui justifie le choix de tel énoncé, plutôt que de tel autre? Ces problèmes rejoignent celui du lien entre un état interne et un comportement.

Nous allons présenter dans cette sous-section plusieurs *éléments de réponse* à ces questions. Nous commencerons par présenter certains des travaux de Grice, philosophe du langage, qui justifient les choix d'énoncés (et guident l'interprétation de ces énoncés) à travers quatre maximes. Puis, nous présenterons la théorie des faces (et du territoire) d'Erving Goffman, psychologue de l'interaction, pour qui une partie de la conversation entre humains est conditionnée par une recherche d'équilibre entre la mise en avant et la mise en retrait de la face de chaque interlocuteur (dans une première approche simplificatrice, la face est assimilable à l'*ego*, ou à l'amour-propre). A la suite de cette section, nous présenterons brièvement la "métaphore théâtrale" du même Erving Goffman, qui montre comment les interactions humaines peuvent aussi être conditionnées selon le *rôle* (au sens théâtral) que chaque interactant endosse à un moment donné et devant un public donné. La logique illocutoire, de Searle et Vanderveken, philosophes, proposant elle aussi des règles de choix, nous les synthétiserons dans la dernière sous-section.

Rappelons que la section 3.5.3 a présenté des agents conversationnels mettant chacun en œuvre des stratégies différentes pour faire le lien entre sens et message. A cet égard, les projets *Persona* et *Amusement* sont particulièrement intéressants.

a) Les maximes conversationnelles de Herbert Paul Grice

Selon Herbert Paul Grice [Grice, 1975], lors d'une conversation, les locuteurs s'efforcent de respecter quatre maximes, tandis que les allocutaires supposent que leur interlocuteur s'efforce de les respecter. Ces maximes sont les suivantes :

Maximes de quantité *Transmettez une quantité d'information à la fois suffisante et minimale.* Quelqu'un qui demande dans la rue la direction d'une pharmacie, peut s'attendre à la description d'un itinéraire, pas à une simple direction indiquée d'un vague geste de la main ni à ce que son interlocuteur commence la rédaction d'une liste d'instructions, accompagnée d'un plan et de dessins explicatifs.

Maxime de qualité *Ne communiquez pas une information que vous savez fausse.* Dans la même situation que précédemment, la personne interrogée cherchera à donner des instructions correctes.

Maxime de relation *Soyez pertinent; parlez à propos.* Toujours dans la même situation, une diatribe sur l'incompétence du pharmacien ne serait pas une réponse appropriée.

Maximes de manière *Évitez l'ambiguïté, soyez clair, bref et méthodique.* À nouveau dans le cas de la pharmacie, une description de l'itinéraire approximative ("prenez l'une des rues sur votre gauche"), confuse ("... sur la gauche, mais un peu en face d'un immeuble qui ressemble à..."), prolixe ("... vous vous souviendrez bien qu'après avoir tourné à gauche... je vous rappelle que...") ou déstructurée ("... vous passez devant la poste, vous remontez la rue, mais pour arriver à la poste, il vous faut passer par...") est à éviter.

Dans le cadre d'une communication ces quatre maximes permettent d'apporter, du point de vue du locuteur, des contraintes guidant le choix des énoncés et de leur contenu sémantique, et du point de vue de l'allocutaire des informations supplémentaires pour le guider dans sa compréhension de l'énoncé. Il peut par exemple supposer que l'énoncé contient toutes les informations dont il a besoin (maxime de quantité) ou que l'énoncé, bien que complexe, est aussi simple que possible dans la situation en cours (maxime de manière). Elles ont de plus l'intérêt de porter à la fois sur des contenus sémantiques (maximes de quantité, de qualité et de relation) et sur la forme même du message (maximes de manière).

Notons enfin que pour Grice, les quatre maximes sont, dans les cas généraux, "inviolables". La violation apparente de l'une d'entre elles à un certain niveau est juste l'indication que le message doit être interprété à un autre niveau. Grice fait explicitement apparaître ici un moyen d'atteindre le(s) sens implicite(s) d'un message. Ceci dit, la violation d'un de ces principes peut aussi tout simplement signifier que l'interlocuteur n'est pas coopératif. Comme le dit Grice, avant de chercher une interprétation d'une proposition, "*il faut d'abord qu'il n'y ai pas lieu de supposer qu'il [le locuteur] n'observe pas les règles de la conversation, ou, au moins, le principe de coopération*" [Grice, 1975].

Intérêt. Les maximes de Grice font appel à des traitements trop complexes et/ou trop peu formalisés pour pouvoir être employées telles quelles dans une application informatisée, au moins pour ce qui est du traitement des messages entrant. Pour ce qui est de la génération de messages, l'intégration de ces maximes semble plus simple, mais il ne faut pas oublier que les phénomènes décrits par Grice sont à la fois généraux et de très haut niveau. Par exemple, la maxime de qualité ("ne communiquez

pas une information que vous savez fausse”), qui est pour le moment quasi implicite dans tout système informatique, doit être tempérée par la notion d’“hypocrisie sociale” (voir pages 52 et 55). Néanmoins, les maximes de Grice trouvent leur place dans cette section car elles représentent une synthèse et donnent un cadre hors duquel il n’est ni intéressant ni souhaitable de se trouver, même si ce cadre n’est pas explicitement représenté en un endroit du modèle de l’agent, mais qu’il est plutôt diffus.

b) La théorie des faces d’Erwing Goffman

La théorie des faces d’Erving Goffman présente un modèle minimal, qui considère deux paramètres pour chaque interactant : la face (comme dans l’expression “sauver la face”) et le territoire (qui rassemble l’espace revendiqué, mais aussi le temps de parole, l’ordre dans une file d’attente, la vie privée, le corps, ...¹⁸). Au territoire est ensuite associé deux principes : le principe de *ménagement* (éviter de menacer) et le principe de *modestie* (éviter de trop mettre en valeur). Catherine Kerbrat-Orecchioni [Kerbrat-Orecchioni, 1989] a clairement exprimé que ces principes sont également applicables à la face. Une relation sera socialement acceptable si elle respecte les faces et territoires de tous les participants à l’interaction, soi-même inclus (le tableau 4.6 synthétise pour la face les comportements à éviter). Une telle situation est très difficile à obtenir, c’est pourquoi Goffman précise que l’accord est bien souvent consensuel, constamment remis en cause par les interactants, et résultant de “*l’effet combiné [de ces] règles d’amour propre et de considération*” [Goffman, 1974].

Principe de ménagement (empiètement). Pour qu’une communication se déroule au mieux, il faut ménager à la fois sa face et celles de nos interlocuteurs. Cette règle condamne les comportements agressifs, comme la supplication (qui agresse sa propre face) ou l’insulte (qui agresse la face de l’interlocuteur). Appliquée au territoire, elle recommande de ne pas se tenir trop près de ses interlocuteurs (territoire spatial), de ne pas leur couper la parole (territoire (spatio-)temporel) ou de ne pas être trop curieux ou de fouiller dans leurs affaires (vie privée).

Principe de modestie (renfermement). Le principe de modestie prévient de l’excès inverse : il ne faut pas mettre exagérément en avant sa face ou celle de ses interlocuteurs. Cette règle condamne des comportements comme la flatterie ou le narcissisme. Appliquée au territoire, elle recommande de ne pas se tenir trop éloigné de ses interlocuteurs ou de meubler les “silences inconfortables”.

Ces principes, de ménagement et de modestie sont graduels et variables selon les interlocuteurs et les situations. Tel locuteur pourra fort bien accepter des infractions à ces principes venant d’un membre de sa famille et trouverait déplacé un même

¹⁸Goffman spécifie huit “territoires du moi” [Goffman, 1973] (Volume 2, p44), parmi lesquels on peut noter *l’espace personnel* (la portion d’espace qui entoure une personne et dans laquelle toute pénétration est ressentie comme une intrusion), *le tour* (qui relève du territoire temporel : il s’agit de l’ordre dans lequel un sujet peut prétendre avoir droit à un bien), *les domaines d’information* (l’esprit de l’individu, sa vie privée, son courrier, ...), *les territoires réservés de conversation* (“*Le droit qu’a l’individu d’exercer un certain contrôle sur qui peut lui adresser la parole et quand; et encore le droit qu’a un groupe d’individus qui se parlent de protéger leur cercle contre l’intrusion et l’indiscrétion d’autrui*”).

Sujet subissant l'infraction	Principe enfreint	Ménagement	Modestie
	Soi	servilité	narcissisme
L'Autre		insulte	flatterie

TAB. 4.6 – Types de comportements à éviter selon la théorie des faces

comportement qui serait le fait d'un subordonné par exemple. De même, comme le précise Goffman, il n'existe pas forcément de situation agréant à tous les participants : pour des raisons culturelles, une Finnoise discutant avec une Espagnole se sentira agressée par sa trop grande proximité spatiale et cherchera à retrouver une distance confortable en reculant, tandis que sont interlocutrice se sentira rejetée et persistera à s'en rapprocher. C'est pourquoi Goffman note que chaque participant s'efforce de maintenir une situation acceptable, estimant qu'en général il y a "*plus à perdre qu'à gagner*" [Goffman, 1973] à bouleverser les règles établies ou à expliciter un malaise (qui reste en général inconscient).

Goffman pose des règles absolues, dont l'infraction doit entraîner réparation (suivant un rituel codifié [Goffman, 1974]) et n'indique pas, contrairement à Grice, la nécessité d'une lecture à un niveau différent. Toutefois, même si la règle est absolue, son seuil de déclenchement est très relatif : les interlocuteurs adaptent leur seuil de tolérance à la situation et peuvent décider de "laisser passer" certaines choses qu'ils estimeraient inacceptable dans une situation différente.

Intérêt. Erwing Goffman résume ainsi sa théorie (qu'il qualifie lui-même de minimale) : "*C'est en se demandant sans cesse [...] « Est-ce que, en faisant ou en ne faisant pas cela, je risque de perdre la face ou de la faire perdre aux autres ? » que [l'interactant socialisé] décide à chaque moment, consciemment ou non, de sa conduite.*" Pour Erwing Goffman, les faces représentées par une personne dérivent de faces prototypiques (ou stéréotypiques), dont les subtiles modifications donnent la personnalité de chacun. Sachant de quelles "faces stéréotypiques" hérite la face de l'interlocuteur, il est possible de s'en servir pour avoir un modèle, grossier mais robuste, permettant une évaluation des comportements. Erwing Goffman nous fournit ici des simplifications justifiées.

La forme même qu'Erwing Goffman a donné à sa théorie la rend particulièrement adaptée aux systèmes automatisés : deux principes appliqués à deux domaines pour chaque interlocuteur et pour soi donnant un ensemble de huit dimensions (dans le cadre d'une relation à deux interlocuteurs) à travers lesquelles il est possible de juger un comportement. Bien entendu, la projection d'un comportement donné sur ces différentes dimensions d'évaluation est loin d'être triviale, sauf en ce qui concerne le territoire, pour lequel il existe une métrique (en terme de distance et de durée) aisément manipulable. Nous n'avons pas connaissance de travaux s'étant penchés sur le sujet pour ce qui concerne la face. Comme souvent en sciences humaines, Erwing Goffman semble utiliser à cette étape une évaluation purement subjective. Néanmoins, des travaux en psycholinguistique sur la charge interpersonnelle de certaines entrées lexicales [Wiggins *et al.*, 1988] — ou le degré de puissance d'un performatif en philosophie du langage [Vanderveken, 1988] — semblent pouvoir mener à un lien

entre énoncés et théorie des faces.

c) La métaphore théâtrale d'Erving Goffman

Goffman s'appuie beaucoup sur la "métaphore théâtrale". Il considère que la vie est une suite de "représentations", où chaque participant joue un rôle (dans le sens théâtral du terme et non pas dans le sens qui lui est couramment attribué pour les agents artificiels), différent suivant le public pour lequel il "joue". Pour Goffman, dans toute relation, chaque participant endosse le rôle adapté à la situation, aux participants et à leurs attentes. Nous ne sommes donc qu'une suite de rôles, bien marqués par la dissociation entre l'acteur ("*artisan infatigable des impressions d'autrui, engagé dans d'innombrables mises en scène quotidiennes*") et le personnage ("*silhouette habituellement avantageuse, destinée à mettre en évidence l'esprit, la force, et d'autres solides qualités*"). Comme le fait remarquer Robert Ezra Park (cité par [Goffman, 1973]) : "*Ce n'est probablement pas par un pur hasard historique que le mot personne, dans son sens premier, signifie un masque. C'est plutôt la reconnaissance du fait que tout le monde toujours et partout, joue un rôle, plus ou moins consciemment. [...] C'est dans ces rôles que nous nous connaissons les uns les autres, et que nous nous connaissons nous mêmes.*"

Nous sommes plus ou moins conscients d'endosser des rôles lors de nos rencontres, ou représentations. A un extrême, nous avons l'acteur qui est "*sincèrement convaincu que l'impression de réalité qu'il produit est la réalité même. Lorsque son public partage cette conviction — ce qui semble être le plus souvent le cas —, alors, momentanément du moins, seul le sociologue ou le misanthrope, peut avoir des doutes sur la "réalité" de ce que l'acteur présente.*" A l'autre bout de cette échelle, on trouve l'acteur qui "*ne croit pas en son propre jeu*", que Goffman qualifie de cynique. Notons que ce terme de cynisme (et celui d'hypocrisie, qui apparaît tout au long des ouvrages de Goffman) n'a pas ici les connotations négatives qui lui sont souvent attribuées ; par exemple, un médecin qui prescrit un placebo à un malade sera amené à jouer le rôle d'un médecin, tout en enfreignant consciemment le fait que toute personne endossant le rôle de médecin est sensée être profondément digne de confiance et ne pas mentir. Face à ces remarques, on est en droit de se demander dans quelle mesure on atteint l'acteur à travers ses personnages, c'est-à-dire, s'il existe une personnalité profonde, fortement indépendante de la situation, une forme de fonds commun qui servirait de base à l'ensemble des rôles endossés, ou bien si l'acteur est parfait, et que tout se passe comme s'il n'avait comme personnalité que celles de ses personnages. Goffman n'apporte pas de réponse à cette question¹⁹.

Intérêt. La notion de rôle est de plus en plus employée dans le domaine des systèmes multi-agents, et la notion de statut y est souvent associée. Ces deux notions permettent de structurer des organisations d'agents, en marquant de façon explicite les capacités et responsabilités des agents. À chaque rôle et statut peut être associé des comportements prototypiques (comme le remarque Goffman), qui apportent des contraintes sur les actions que l'agent (naturel ou artificiel) est sensé entreprendre. En tant que fondateur du comportement, le rôle (que ce soit au sens de l'IAD ou au

¹⁹Dans un autre contexte Albert Camus ("le mythe de Sysiphe") dit : "*Il est certain qu'apparemment, pour avoir vu cent fois le même acteur, je ne l'en connaîtrais personnellement pas mieux. Pourtant, si je fais la somme des héros qu'il a incarnés et si je dis que je le connais un peu plus au centième personnage recensé, on sent qu'il y aura là une part de vérité. [...] un homme se définit aussi bien par ses comédies que par ses élans sincères*".

sens théâtral) permet, d'un côté d'avoir une ligne de conduite qui oriente les actions et de l'autre, d'anticiper les grandes lignes de comportement d'un interlocuteur. Ainsi, endosser un rôle aide à la décision des actions à entreprendre ; reconnaître un rôle chez ses interlocuteurs aide à la structuration de l'interaction en permettant certaines anticipations.

d) Le cercle interpersonnel de McCrae et Costa

La théorie du cercle interpersonnel se base sur une représentation d'un cercle ayant deux axes : l'axe domination/soumission (appelé axe de *contrôle*) et l'axe hostilité/amitié (appelé l'axe d'*affiliation*) [McCrae et Costa, 1989]. Il est possible de définir une position sur ce cercle qui corresponde au comportement de l'individu dans ses relations aux autres.

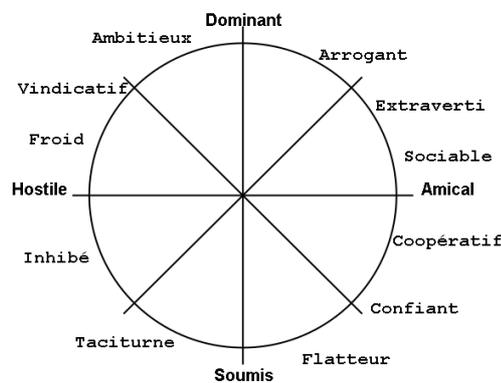


FIG. 4.4 – Types de personnalités dans l'espace contrôle / affiliation

Tout le long de ce cercle sont placés des points de repère correspondant aux comportements typiques des individus tenant cette place sur le cercle (confiant, inhibé, social), opposés deux à deux avec le comportement diamétralement situé (méfiant, extraverti, asocial). Trois principes sont associés à cette représentation :

- Complémentarité : en situation d'interaction, les relations se construisent sur des bases de complémentarité suivant l'axe de contrôle : un comportement hostile-soumis invite à un comportement hostile-dominant (et *vice versa*), tandis qu'un comportement amical-dominant invite à un comportement amical-soumis (et *vice versa*).
- Variabilité : la capacité pour un individu de se positionner sur une large plage de valeurs ou en un point précis et fixe donne une indication sur la stabilité de l'individu, ses capacités d'adaptation, et, en conséquence, sa facilité de relation aux autres²⁰.
- Enchaînement circulaire : la représentation sous la forme d'un cercle apporte des relations de voisinage entre différents comportements, permettant de considérer que tel ou tel comportement est plus proche que tel ou tel autre.

Intérêt. À travers un ensemble de données et de processus très simples, ce modèle permet, à l'instar de la théorie des faces de Goffman, de prendre en compte

²⁰Lors des tests visant à déterminer ces paramètres chez les humains, cette valeur est liée à la déviance statistique des mesures.

les relations interpersonnelles. Ce modèle est employé dans plusieurs systèmes en interaction avec l'humain pour modéliser l'état interne de l'agent, souvent associé à une représentation de l'émotion sous un format proche (espace à deux dimensions : valence et activation). Citons par exemple les travaux de Gene Ball et Jack Breese [Ball et Breese, 2000], déjà présentés dans cette thèse (voir page 75 et suivantes). De plus, des études, malheureusement limités aux langues anglaise et espagnole, mettent en évidence un lien direct entre le lexique employé par un individu et son profil interpersonnel. La réalisation de telles études pour la langue française (1) permettrait la constitution aisée du profil interpersonnel des interlocuteurs de l'agent sur la base de la simple étude statistique du vocabulaire employé et (2) donnerait une méthode simple de choix d'un vocabulaire adapté à l'interlocuteur.

e) La logique illocutoire de Searle et Vanderveken

La logique illocutoire telle qu'elle est présentée par Vanderveken [Vanderveken, 1988] apporte un ensemble de contraintes sur le choix des forces illocutoires qu'il est possible d'employer à un moment donné d'une conversation. Le chapitre V ("Lois fondamentales de la sémantique générale") décrit vingt lois, desquelles nous retiendrons particulièrement que chaque acte de discours dépend de conditions préparatoires et que les interlocuteurs doivent respecter une rationalité minimale. Tout accomplissement d'acte de discours respectant ces lois, elles ont leur intérêt dans les situations d'interprétation et d'expression. En interprétation, un auditeur peut supposer que l'ensemble des conditions préparatoires est respecté par le locuteur et ainsi obtenir des informations supplémentaires à ce qui est dit dans le message. En expression, un locuteur se doit de ne pas affirmer une chose et son contraire, en vertu du principe de rationalité. Toutefois, plusieurs des lois présentées dans ce chapitre (ou tout au long du livre) ne sont pas particulièrement surprenantes dans le cadre de la communication entre entités artificielles. Des lois comme les deux que nous venons de citer sont quasiment implicites dans l'ensemble des mécanismes de communication entre entités artificielles et c'est leur non-respect qui serait surprenant.

Par ailleurs, Vanderveken signale l'existence de *marqueurs de force illocutoire*, qui, dans un énoncé, prennent la forme d'un mot ou d'un trait syntaxique, et qui introduisent des conditions de sincérité supplémentaires ou un mode d'accomplissement particulier. Ces marqueurs permettent eux aussi d'apporter des contraintes guidant l'interprétation ou des modifications affinant un message à émettre. Ainsi, "*Malheureusement pour toi, ...*" sert à affirmer le contenu propositionnel en pré-supposant que l'état de choses qu'il représente est malheureux pour l'allocutaire [Vanderveken, 1988].

Intérêt. La logique illocutoire formalise un ensemble de contraintes sur le langage, comme les conditions de succès et de satisfaction (que nous avons vues page 86 et suivantes) ou les lois sur les conditions préparatoires, sur les mécanismes d'altérations des forces illocutoires (présentées par exemple par Daniel Vanderveken [Vanderveken, 1988]), etc. Ces contraintes sont des informations parallèles à chaque message qui permettent d'en guider l'interprétation. La formalisation de la force illocutoire en six composantes permet de plus des manipulations de forces illocutoires dans un espace à six dimensions, certains des points de cet espace correspondant à des performatifs (i.e. des forces illocutoires existantes dans la langue naturelle considérée), les autres pouvant (éventuellement) être atteints par l'emploi de marqueurs de force illocutoire. Une telle formalisation apporte une grande liberté dans les mécanismes d'expression.

4.4.3 Quels paramètres pour les interactants ?

Les interlocuteurs que nous considérons sont des individus, différents les uns des autres (voir notre définition d'agent section 2.1). Nous allons préciser dans cette section les paramètres que nous considérons pour marquer cette différence.

Tout d'abord, les *connaissances* de l'agent sont prépondérantes. C'est dans ces connaissances que l'agent va puiser pour constituer ses messages. La richesse (variété des sujets traitables, finesse du traitement) d'une conversation est profondément liée à la quantité (et la qualité) des connaissances à la disposition de l'agent. Puis nous parlerons de la notion de *rôle* et de *statut*, dans le domaine des sciences humaines et, en parallèle, dans le domaine de l'IAD. Enfin, de l'analyse conjointe des travaux en psychologie de l'interaction et en interface homme-machine, nous avons synthétisé une notion de *personnalité*, elle-même constituée de trois composantes principales, un modèle "*psychologique*", une *conscience sociale*, et enfin, à la suite de travaux plus récents, une *dynamique émotionnelle*. Ces différents points sont développés de manière générale ci-dessous, et présentés d'une manière plus approfondie dans le domaine des agents conversationnels dans la section 5.1.

a) Connaissances sur le monde

L'importance des connaissances de l'agent dans le cadre d'une communication non triviale est si évidente que nous nous contenterons dans cette sous-section de quelques généralités. Les connaissances importantes tombent dans deux catégories, celles permettant d'apporter une forme aux messages, et celles permettant d'apporter un contenu à la conversation.

Les connaissances linguistiques de formation des énoncés sont indispensables. Elles correspondent de manière minimale aux règles de construction d'énoncés bien formés (règles syntaxiques), mais peuvent s'enrichir de règles sur les enchaînements des énoncés, sur la prise en compte des interlocuteurs (règles pragmatiques), etc. Même dans le cas de la gestion d'une conversation extrêmement limitée (une fois qu'on en connaît les mécanismes internes) comme celle d'ELIZA [Weizenbaum, 1966], on ne peut se passer de certaines connaissances sur la construction des énoncés. Ainsi, ELIZA est capable, lors de ses reformulations des énoncés de son interlocuteur d'adapter les pronoms, permettant une réponse de la forme "*Do you believe you are sick ?*" à la question "*Am I sick ?*".

Les connaissances apportant un contenu à la conversation ne sont pas indispensables, si l'on considère des systèmes écholaliques comme ELIZA, dans lesquels le contenu des messages de l'agent est soit vide (du type "*In what way ?*") soit directement emprunté à l'énoncé reçu (du type "*What makes you think that . . .*"). Toutefois une interaction de ce type ne présente, en général, qu'un intérêt limité.

Si l'on consulte la typologie des dialogues argumentatifs de Douglas Walton [Walton, 1990], on constate que sur huit types de dialogues, trois ont directement à voir avec les connaissances de chaque interlocuteur (*Inquiry*, *Pedagogical* et *Expert consultation*) et que quatre autres y sont fortement liées (*Critical Discussion*, *Debate*, *Negotiation* et *Planning Committee*), le dernier type (*Quarrel*) étant plus lié à une confrontation d'émotions que de connaissances. Ainsi, dans le cas des dialogues argumentatifs (nous nous référons souvent dans cette thèse aux situations de *consultation d'expert*) l'essentiel de l'interaction est basé sur la confrontation des connaissances des différents interlocuteurs.

b) Rôle et statut

Les notions de rôle et de statut existent dans les domaines de la psychologie et de la psychologie sociale, et l'IAD en a principalement retenu celle de rôle.

Dans le domaine des sciences humaines, Anne-Marie Rocheblave-Spenlé [Rocheblave, 1994] pose le rôle et le statut comme lien entre les groupes et les individus :

Deux concepts [...] possèdent cette fonction de chaînon entre la structure sociale et l'individu. Il s'agit des concepts de statut et de rôle. Le premier renvoie davantage à la structure sociale, puisque les statuts désignent les différentes positions, liées les unes aux autres, qui ponctuent cette structure sociale et y définissent des systèmes relativement autonomes (par exemple, le système familial, caractérisé par les positions : père, mère, enfant). Le concept de rôle est plus orienté vers les individus, puisqu'il se réfère à des conduites, ou plutôt à des modèles de conduites, rattachés au statut.

Roger Mucchielli, lui, propose les définitions suivantes [Mucchielli, 1983] :

Rôle : Au sens théâtral, personnage d'une pièce jouée par un acteur. Le même mot a trois sens en psychologie sociale : 1) L'attitude d'un individu dans un groupe ; 2) La fonction dans une organisation sociale ; 3) tout comportement caractéristique attendu par le groupe de la part de l'un de ses membres.

Statut (social) : Position reconnue ou accordée à une personne dans un groupe, par rapport aux autres membres du groupe. Il s'agit d'une position "morale", à laquelle est associé un degré de considération, d'estime sociale, de respect.

La notion de rôle est très importante dans l'œuvre de Goffman puisqu'il considère que chaque individu endosse un rôle (au sens théâtral) dans chaque situation interactionnelle, rôle qui dépend certes de l'individu, mais aussi du public, de la situation, des objectifs de la communication, ... (voir p.109)

De ces deux points de vue concourants, les points qui nous intéressent sont que le statut est un *état*, qui conditionne en général un ensemble de rôles à tenir, tandis que le rôle est plus directement lié au comportement ("*Le rôle est défini par Linton tantôt comme l'aspect dynamique du statut, tantôt comme « la somme totale des modèles culturels associés à un statut particulier » [...] Pour Linton, alors que le statut constitue un concept statique et structural, le rôle représente un point de vue dynamique et fonctionnel.*" [Rocheblave, 1994].

Du côté de l'IAD, la notion de rôle, selon Ferber ([Ferber, 1995]) correspond à "[...] l'ensemble des activités qu'un agent est supposé accomplir dans une organisation considérée." Cette définition, qui fait explicitement référence à une organisation, montre qu'en IAD aussi (voir aussi Weiss [Weiss, 1999]) le rôle est vu comme une fonction de la position sociale, et, de ce fait, rejoint en partie les définitions présentées auparavant, où le statut conditionne le rôle, et donc les conduites à tenir. Cependant, dans les faits, le rôle d'un agent est souvent confondu avec sa fonction, et donc sans lien avec une quelconque notion de statut.

Notons aussi la notion de statut et de rôle telle qu'elle est présentée par Sylvie Pesty et Christian Brassac [Brassac et Pesty, 1996], pour qui le statut n'est pas social mais ontologique, le rôle étant lié au comportement. Ainsi, des agents humains, ayant un statut cognitif, peuvent agir de manière réactive, endossant ainsi un rôle réactif.

La connaissance des rôles et statuts de chaque agent permet donc d'en connaître certaines capacités et certaines des ressources auxquelles il a accès. Revendiquer tel ou tel rôle permet (s'il n'y a pas abus, évidemment) à un agent de déclarer une liste de compétences. Revendiquer un statut permet à l'agent de se positionner sur une échelle sociale, ce qui apporte des indications sur les manières d'interagir qui peuvent être employées et parfois sur les ressources auxquelles l'agent peut avoir accès. Goffman considère des ensembles de rôles et statuts (et les comportements qui y sont associés) *prototypiques* (comme médecin ou prêtre), qui permettent de donner "en un bloc" un ensemble d'indications aux interlocuteurs sur le comportement à adopter en présence les uns des autres.

Notons enfin que les rôles et statuts sont dynamiques, même s'il est parfois difficile de faire la distinction, chez le même individu, des statuts et rôles dont il peut se réclamer à différents moments. Un même individu peut rassembler les rôles et statuts associés à maire, médecin, parent d'élève et mari, par exemple, passant de l'un à l'autre suivant les situations. Toutefois, il est communément admis dans les sciences humaines, que dans les situations les plus classiques, un seul statut est revendiqué à la fois. De plus, comme les fait remarquer Anne-Marie Rocheblave-Spenlé les rôles sont des "*modèles de conduite*" et non pas des conduites. Ainsi, "*le rôle joué constitue toujours un compromis entre le modèle social prescrit afférent au statut et la personne qui, cherchant à se conformer à ce modèle, l'interprète évidemment d'une façon unique*". Ainsi, le rôle laisse à l'acteur une marge de liberté lui permettant de *s'exprimer* (au sens que nous lui donnons dans le mécanisme d'expression, section 4.4.1).

c) La personnalité

Selon les disciplines que l'on consulte (et parfois même au sein de ces disciplines), il est difficile de trouver une définition unique de la *personnalité*. Les tentatives d'analyse de la personnalité (au sens large) ne sont pas un champ d'analyse récent, puisque Hippocrate (IV^e siècle avant J.C.) proposait déjà une classification de tempéraments, et qu'on trouve dans l'encyclopédie de Diderot et D'Alembert (*ca.*1750) des planches sur la morphopsychologie, associant des formes de visages à des animaux, et prêtant aux humains des traits comportementaux typiques des animaux auxquels ils ressemblent.

Plus récemment (début du XX^e siècle), une première vague de travaux se base sur la définition de types caractérogiques ou de "traits de caractère" (éventuellement héréditaires) dont la composition ou juxtaposition permettrait de définir la personnalité. C'est aussi la période de la psychométrie, avec Binet et le quotient intellectuel (rapport entre l'âge réel et l'âge mental déterminé par une série de tests réalisables à partir d'un certain âge seulement) et Spearman et le "facteur g" ("*adaptabilité en général, niveau global de l'efficacité dans l'adaptation au réel et aux situations nouvelles.*" [Mucchielli, 1971]).

Durant la première moitié du XX^e siècle, la psychanalyse considéra la personnalité comme étant le produit unique de l'histoire personnelle et des "traces" laissées par les événements vécus. Cette théorie "*des profondeurs bouleverse complètement la conception antérieure des aptitudes liées à l'existence de fonctions psychiques élémentaires, ou plus exactement d'opérations mentales*" ([Mucchielli, 1971]).

Enfin, de façon plus marquée dans la deuxième moitié du XX^e siècle, les psychologues "culturalistes" introduisirent l'idée que les comportements individuels étaient conditionnés par les influences culturelles et les interactions sociales, aussi bien à un

niveau “immédiat” (le comportement dépend du groupe avec lequel on interagit et serait différent si l’interaction avait lieu avec un autre groupe ; voir à ce sujet la métaphore théâtrale de Goffman, p.109) qu’à un niveau plus profond (la “personnalité” d’un individu se construisant par ses échanges avec les autres membres des groupes qu’il fréquente).

Nous retiendrons de ces différentes approches deux aspects essentiels pour la modélisation d’un agent :

- La personnalité est un paramètre interne qui conditionne le comportement ;
- La personnalité rassemble les paramètres comportementaux qui font d’un individu qu’il est distinct des autres.

Lorsque l’on parcourt tous ces “modèles du comportement humain” on se rend compte de leur grande hétérogénéité. En confrontant ces modèles aux théories de l’interaction, qu’elle soit naturelle ou artificielle, et en consultant aussi ce qui s’est déjà fait dans le domaine des agents artificiels revendiquant une personnalité, nous avons retenu trois paramètres principaux. Le premier concerne ce qui est souvent appelé le *profil psychologique*, et est couramment confondu avec l’ensemble de la personnalité. Il rassemble les paramètres influant sur l’ensemble du comportement, et ayant une variation plutôt lente (i.e., il reste stable à l’échelle d’une interaction). Nous avons nommé le deuxième *conscience sociale*. Il concerne tous les comportements qui sont influencés par “les autres”, et plus particulièrement les rôles et statuts de ces “autres”. Le dernier paramètre que nous considérons comme constituant la personnalité d’un individu est sa *dynamique émotionnelle*. Nous y considérons tout ce qui amène une variation brusque du comportement, souvent déclenchée par un objet ou une situation.

Bien que chacun de ces trois paramètres ait une influence sur le comportement lors d’une interaction, la conscience sociale y tient une place prépondérante²¹. C’est pourquoi nous développerons essentiellement ce point, restreignant au minimum le profil psychologique et la dynamique émotionnelle.

Parmi les connaissances exploitées lors des processus d’interprétation et d’expression, on trouve les relations qui nous lient aux autres interlocuteurs. Les relations sociales influent sur la structure du discours, ainsi que sur son contenu. Sur sa structure car il peut exister des règles²² (implicites ou explicites) qui gèrent la conversation avec des personnes remplissant des rôles particuliers ou ayant un statut particulier. Sur son contenu car il est des sujets (ou un vocabulaire) qu’il n’est pas séant d’aborder (ou d’employer) avec telle ou telle personne.

On retrouve ici le concept central de Goffman et de sa métaphore théâtrale, le rôle (voir page 109). Chaque situation interactionnelle se passe sur la scène d’un théâtre où chaque interactant endosse un rôle²³ adapté à son public. Dans le cadre de conversations humaines, une grande partie des échanges sert à se positionner

²¹Comme le dit Roland Barthes, “*la langue, selon une intuition de Benveniste, c’est le social même*”. (Leçon inaugurale de la chaire de sémiologie littéraire du Collège de France, 1977)

²²Goffman emploierait le terme de rituels.

²³Le terme *rôle* est employé ici dans son sens théâtral, et non dans le sens qui lui est couramment associé en IAD (voir les définitions page 113) : en effet, ce dernier, s’il correspond aussi à un comportement à tenir, est bien plus lié à une tâche à accomplir qu’à une directive d’ordre social.

vis-à-vis du rôle des autres et de son propre rôle (confirmation ou contestation) : “*Chaque comportement face à une personne, quelle qu’elle soit, est finalement une communication de la façon dont on voit sa relation à la personne*” estime Watzlawick, cité par Alex Mucchielli [Mucchielli, 1995].

Un autre aspect intéressant de la conscience sociale est la notion de confiance. La confiance peut servir à obtenir de la part de ses interlocuteurs des informations, actions ou engagements. Par exemple, l’agent REA (chargé de la vente d’immobilier, voir section 3.5.2) peut décider de ne pas aborder le problème du salaire de son interlocuteur (qui est typiquement un sujet sensible) avant de l’avoir mis en confiance, et ce par l’intermédiaire d’une série d’énoncés n’ayant pas directement à voir avec la transaction en cours, mais montrant qu’elle peut “s’intéresser” à son interlocuteur. Vendant un appartement à Boston, elle pourra parler du temps à Boston, ou appeler des commentaires sur certaines parties de la maison qu’elle fait virtuellement visiter. Le caractère clairement non-informatif de ces énoncés apparaît lorsque l’on sait que d’un point de vue fonctionnel, REA ignore les réponses qui peuvent lui être faites (voir le dialogue page 68). L’essentiel pour elle (ou plutôt, pour ses concepteurs) est de montrer par ces énoncés qu’elle est capable de tenir une discussion (ou plutôt, d’en faire illusion) qui rendra son interlocuteur plus à l’aise, plus confiant et l’amènera à prêter une certaine humanité à REA, pour, au final, améliorer la qualité de la conversation.

4.5 La dynamique de la communication

Jusqu’à l’arrivée de la logique illocutoire et particulièrement l’introduction des conditions de satisfaction (voir les travaux de Searle et Vanderveken [Searle et Vanderveken, 1985] [Vanderveken, 1988]) la théorie des actes de discours voulait aborder l’étude de la conversation à travers un énoncé unique.

Erwing Goffman [Goffman, 1981] doute de l’existence d’un sens littéral des énoncés conversationnels, ceux-ci étant trop dépendants du contexte d’élocution : “*De même qu’une déclaration immédiatement antérieure est souvent nécessaire pour donner un sens à la réponse qui suit, de même, on a souvent besoin de la réponse qui suit pour donner un sens à la déclaration devant laquelle on se trouve (si l’on n’est pas celui à qui elle s’adressait)*”. Goffman s’arrête donc sur un “*tour long de deux énonciations*” comme unité dialogique minimale. Il introduit la possibilité d’interpréter et de montrer au premier locuteur les effets de son message.

Christian Brassac [Brassac, 1994] quant à lui, affirme la nécessité de travailler sur une unité encore un peu plus longue qui permette l’introduction des concepts de négociation et de co-construction du sens. “*Le minimum consiste alors à travailler non pas sur un tour de parole auquel répond un auditeur, [...] mais à envisager une réflexion sur trois tours de parole.*” Il précise par ailleurs que “[...] le sens [d’un énoncé] n’appartient pas au premier locuteur, mais [qu’il] est construit dans l’échange, mieux, [qu’il] est co-construit par les agents qui interagissent”.

Nous partageons ce dernier avis, et les travaux de Brassac et ses collègues peuvent donner des indications sur les enchaînements des énoncés. Principalement, ils montrent qu’il est utopique, dans le cadre de conversations impliquant l’être humain, de traiter²⁴ un énoncé en tant que tel et indépendamment des énoncés précédents.

²⁴Par “traiter” nous considérons l’ensemble des processus déclenchés par la perception d’un message : interprétation, raisonnement sur le message (quels que soient les moyens mis en œuvre à ce niveau), éventuellement préparation, expression, et présentation dans

Nous allons donc traiter dans cette partie des aspects dynamiques de la communication. Nous commencerons par présenter les mécanismes de *co-construction* (ou de *négociation*) de sens (section 4.5.1), qui proposent une solution à la régulation de l'intercompréhension dans une situation où chaque interactant est libre d'interpréter les messages comme il l'entend, et non pas comme un protocole le lui impose. Une fois cette base posée, nous parlerons de l'*enchaînement* des messages (section 4.5.2), en distinguant l'enchaînement des messages liés au thème de la conversation (*enchaînement informationnel*) et l'enchaînement des messages liés à la régulation de la conversation (*enchaînement conversationnel*). Après quoi, nous proposerons un moyen de modéliser les mécanismes d'enchaînement en liant la notion de co-construction à un *focus conversationnel* (section 4.5.3), ce qui formera le cœur de notre modèle de l'interaction. Pour compléter ce modèle, nous rajouterons des *attentes* à nos messages (section 4.5.4). Enfin, nous rassemblerons dans la dernière section (4.5.6) des explications sur les *ruptures conversationnelles*, qui participent pleinement à la dynamique d'une conversation.

4.5.1 Co-construction de sens / négociation de sens

Afin de donner une idée générale de la notion de co-construction de sens, nous présentons dans la table 4.7 huit extraits de dialogues. Ces extraits semblent tous "naturels" et "possibles". Ils débutent tous par le même énoncé. Pourtant, on voit dans ces enchaînements que le sens de ce premier énoncé est largement négociable par les interactants (et d'ailleurs, est négocié).

Chacun de ces extraits présente une séquence où les interlocuteurs déclarent (implicitement) leurs interprétations des énoncés, et confirment ou infirment les interprétations faites. "En t_2 [B1], l'auditeur de l'acte initial propose une interprétation du premier acte exprimé en t_1 [A1]. En t_3 [A2], le locuteur initial ratifie ou non cette proposition d'interprétation. t_1 n'acquiert son statut conversationnel qu'après ce double jeu." [Brassac, 1995]. On a ainsi,

- pour l'extrait n° 1 : interprétation assertive suivie de la validation de cette interprétation ;
- pour l'extrait n° 2 : interprétation directive suivie de la validation de cette interprétation ;
- pour l'extrait n° 3 : interprétation assertive suivie de l'invalidation de cette interprétation ;
- pour l'extrait n° 4 : interprétation directive suivie de l'invalidation de cette interprétation.

Ces dialogues montrent comment un interlocuteur détecte et résout un échec de la conversation (extraits n° 3 & n° 4), ou, au contraire, avalise la réaction de l'autre interlocuteur (extraits n° 1 & n° 2). Dans ces extraits, les interactants échangent des énoncés qui sont soumis au jugement de l'autre, qui les accepte ou les conteste, montrant que la conversation est une activité collective, nécessitant une *évaluation* et un *retour* de la part des différents interactants. De cette activité collective, Brassac dit que le sens d'un énoncé est *négocié* et *co-construit* par les interactants.

On pourrait tout simplement considérer ce mécanisme de co-construction comme une boucle de rétroaction, qui s'assurerait d'une "bonne" interprétation du message, et qui tendrait (éventuellement) vers une stabilisation du système, c'est-à-dire une situation où les deux interlocuteurs aient négocié un sens qui les satisfasse tous deux. La situation est plus complexe, car il n'est pas possible de "rétro-agir" : la

l'environnement d'un ou de plusieurs messages en "réponse" ou "réaction" au message initial.

Quatre dialogues “téléphone”	Quatre dialogues “repas”
<p>Extrait n° 1</p> <p>A1 : Tu as le téléphone ici</p> <p>B1 : Oui, c’est moderne</p> <p>A2 : Ah... Je n’aurais pas cru</p>	<p>Extrait n° 1</p> <p>A1 : Tu ne m’as pas dit avec qui tu as mangé à midi</p> <p>B1 : C’est vrai on n’a pas eu le temps d’en parler</p> <p>A2 : Oui, c’est vrai, et puis, il faut se dépêcher de partir</p>
<p>Extrait n° 2</p> <p>A1 : Tu as le téléphone ici</p> <p>B1 : Oui, c’est le numéro 83-35-36-09</p> <p>A2 : Ah ben je pourrais t’appeler comme ça</p>	<p>Extrait n° 2</p> <p>A1 : Tu ne m’as pas dit avec qui tu as mangé à midi</p> <p>B1 : Avec Michel</p> <p>A2 : Et il va bien ?</p>
<p>Extrait n° 3</p> <p>A1 : Tu as le téléphone ici</p> <p>B1 : Oui, c’est moderne</p> <p>A2 : (rires) Ah ben j’aurais bien aimé que tu me donnes le numéro</p>	<p>Extrait n° 3</p> <p>A1 : Tu ne m’as pas dit avec qui tu as mangé à midi</p> <p>B1 : C’est vrai on n’a pas eu le temps d’en parler</p> <p>A2 : Ben dis le moi !</p>
<p>Extrait n° 4</p> <p>A1 : Tu as le téléphone ici</p> <p>B1 : Oui, c’est le numéro 83-35-36-09</p> <p>A2 : Euh, mais je ne te demandais pas le numéro</p>	<p>Extrait n° 4</p> <p>A1 : Tu ne m’as pas dit avec qui tu as mangé à midi</p> <p>B1 : Avec Michel</p> <p>A2 : Mais je ne te le demandais pas !</p>

TAB. 4.7 – Négociation de sens dans un dialogue (exemples donnés par Christian Brassac : [Brassac et Trognon, 1992] pour le téléphone, et [Brassac, 1994] pour le repas.)

conversation s'inscrit dans le temps, et on ne peut influencer sur ce qui a été dit qu'en rajoutant un message sur la pile de messages déjà transmis. Cette situation est bien matérialisée par le concept de “*spiral pyramids*” de [Leigh, 1995] (voir page 119), qui synthétise : “*you can't decommunicate*”.

On voit bien, à la lumière de ces exemples que le sens absolu d'une énonciation est une utopie. Le sens d'un énoncé ne peut se concevoir qu'à travers son contexte d'énonciation (dans le cas présent, les énoncés précédents, *mais aussi les énoncés suivants* puisque le sens d'un énoncé est sans cesse susceptible d'être remis en question), et, comme le dit Brassac, ce sens “*tend vers une fixation effectuée conjointement par les deux conversants.*” [Brassac, 1995]. Le sens intentionné par les locuteurs en devient même secondaire : “[...] *il n'est pas possible d'être certain que le premier locuteur voulait, intentionnait de poser une question. L'un exprime avoir compris cet énoncé comme porteur d'une question, l'autre accepte que l'un aie pu le comprendre comme tel.*” [Brassac, 1995] Il existe toutefois un sens que l'on peut qualifier de générique, standard, littéral, canonique, ... et qui est en quelque sorte un référent commun permettant l'intercompréhension. Ce sens littéral, pour un mot, peut être celui que l'on trouve dans un dictionnaire. Il peut servir de référence pour la compréhension d'un énoncé, mais il reste négociable par les interlocuteurs.

Notons enfin que laisser une liberté d'interprétation sur les symboles entraîne qu'à chacun d'entre eux est associé une sémantique qui ne peut être stricte, mais au contraire, indicative et négociable. Mais même lorsque les interlocuteurs négocient ce potentiel de sens jusqu'à penser être d'accord sur un sens consensuel commun, on ne s'affranchit jamais de ce que Goffman appelle “*l'ambiguïté résiduelle*”, qui marque toute communication.

Une représentation géométrique : une pyramide et une spirale Pour clore cette sous-section, nous présentons un travail issu du monde du *management* et traitant de la co-construction, montrant ainsi que ce concept n'est pas limité au domaine de la linguistique.

Alan Leigh [Leigh, 1995] présente la communication sous la forme, tout d'abord, d'un triangle dont les trois sommets sont le message, l'auditoire et le contexte, et le sens du message comme étant quelque chose de commun à ces trois pôles (voir figure 4.5). Il considère ensuite un processus dynamique, dans lequel l'information transite du locuteur vers l'auditoire, à travers le contexte. Le message apparaît de manière explicite car Leigh dit que l'information ne peut être transmise qu'à travers un objet physique (le message) ; le contexte est là pour marquer que le message s'enrichit de la personnalité du locuteur, des connaissances réciproques, etc. Leigh donne à ce processus dynamique la forme d'une spirale, car à chaque passage sur l'un des pôles, le contexte s'agrandit d'un message, de l'interprétation de ce message, des inférences tirées de l'interprétation de ce message, ... Il pose alors un point fondamental : “*You can't decommunicate*”. La spirale de la communication s'élargit sans cesse, et tout acte conversationnel vient s'ajouter aux précédents, prenant ainsi à contre-pied le modèle de Shannon dans lequel on peut intégrer une rétroaction.

La figure 4.5 est ensuite plongée dans le temps, le triangle inscrit dans la spirale devenant pyramide inscrite dans un ressort conique.

Il retire plusieurs enseignements de ce modèle :

- Le monde est une construction de textes et de signes, qui sont interprétés par l'auditoire et que nous influençons (et non pas contrôlons totalement) : “*you can't decommunicate*”.

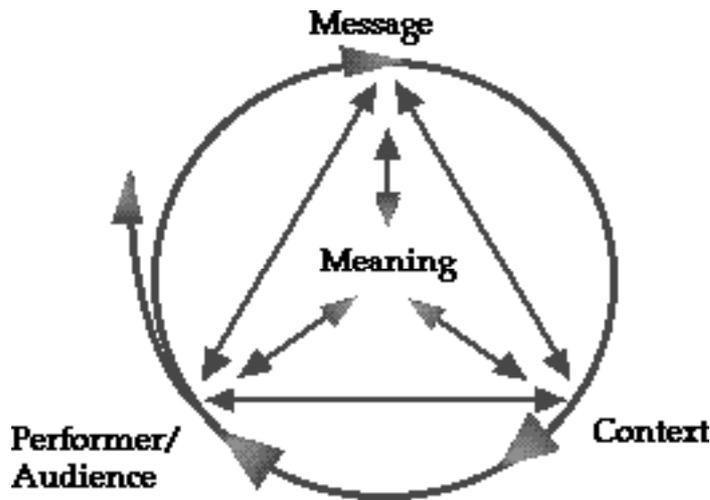


FIG. 4.5 – La spirale de la communication

- Tout est signe et fait partie de la communication. Même une absence de réponse est un message. On ne peut pas sortir de la spirale : “*you can't not communicate*”.
- On ne peut pas “rater une communication”; on ne peut qu'échouer à transmettre le sens que l'on voulait faire passer (ce qui peut entraîner une phase de négociation).
- Ni le locuteur ni l'auditoire ne peuvent prétendre posséder la réalité de la transaction (le message n'est pas le sens, mais n'est que porteur d'un potentiel).
- Enfin, il préconise l'AOT (*audience-oriented thinking*) comme un moyen d'améliorer les interactions, de la même façon que nous introduisons un processus d'expression qui adapte le message à ses destinataires (Cf. section 4.4).

4.5.2 Enchaînements

Les dialogues auxquels nous nous intéressons sont constitués de plusieurs tours de parole. Les situations dans lesquelles l'interaction se réduit à un seul échange (comme dans le cas d'une réponse jugée suffisante à une question qui était le seul motif de l'interaction) pourront donc être considérées comme des cas particuliers. Hors de ces cas particuliers, chacun des interlocuteurs sera amené à intervenir plusieurs fois dans la conversation, pour énoncer des messages différents, à des moments différents. Les problèmes qui se posent sont donc de savoir quand placer les différentes interventions, et quelles sont les formes qu'elles doivent prendre.

Pour répondre à ces questions, il est confortable de séparer les messages échangés en deux grandes catégories, comme le fait Justine Cassell (voir par exemple [Cassell *et al.*, 2000b]) qui travaille dans le domaine des agents conversationnels. Nous pouvons voir d'un côté les messages qui relèvent de l'*interactionnel* (ou *conversationnel*), et de l'autre ceux qui relèvent du *propositionnel* (ou *informationnel*). Justine Cassell distingue ces deux catégories en considérant comme *interactionnel* tout comportement visant à réguler la conversation, tandis que les comportements *propositionnels* sont liés au contenu de la conversation elle-même²⁵.

²⁵Justine Cassell n'aborde pas les attitudes méta-conversationnelles, où la conversation en cours devient explicitement son propre sujet, dans le cas de l'explicitation d'une incompréhension par exemple. Il semble qu'elle considère que cette situation ne nécessite pas un

a) Enchaînement des tours de paroles (interactionnel)

Les messages interactionnels (pour être plus précis, les messages à sens *intentionné* interactionnel) servent à réguler la conversation. Il s'agit essentiellement pour un agent de faire sentir à ses interlocuteurs dans quel état conversationnel il se trouve. Les états les plus importants à manifester dans la conversation étant :

- interaction en cours ou non. Par exemple, les interlocuteurs cherchent à rester en contact visuel et à des distances respectant leurs règles territoriales (voir le territoire selon Goffman, page 107) pour montrer qu'ils considèrent être en interaction.
- désir de prendre ou de laisser la parole au locuteur en cours. Par l'intermédiaire de gestes d'interruption l'allocutaire peut manifester son désir de prendre la parole, ou, par l'emploi d'acquiescements, montrer qu'il suit la conversation et qu'il laisse la parole au locuteur.
- désir de passer la parole ou de la conserver. Par l'intermédiaire d'indices prosodiques (pauses) ou visuels (direction du regard), par exemple, le locuteur peut manifester son désir de laisser la parole.
- certaines incompréhensions. Ce dernier point est un peu particulier puisque nous ne considérons dans cette section que les messages interactionnels. Les incompréhensions interactionnelles comprennent par exemple les problèmes de transmission du message (message bruité), les infractions sociales, et donc, très généralement, l'ensemble des problèmes n'ayant pas un lien avec le sujet de l'interaction.

Les messages interactionnels permettent de réguler la conversation là où les systèmes informatiques classiques emploieraient un protocole et des messages du type "prêt à émettre" ou "acquiescement" par exemple.

b) Enchaînement des informations transmises (propositionnel)

Les messages propositionnels (pour être plus précis, les messages à sens *intentionné* propositionnel) sont des messages traitant du contenu de la conversation. Ce sont ces messages qui correspondent à la partie la plus évidente de la conversation et qui participent principalement à la co-construction de sens dont nous avons parlé précédemment.

C'est l'enchaînement de ces informations qui forme la cohérence sémantique du discours. Les règles justifiant les séquences de messages peuvent être spécifiques à un type d'interaction (consultation d'expert, débat, ... voir par exemple la typologie des dialogues argumentatifs de Walton [Walton, 1990]) ou à une politique conversationnelle donnée. Mais, si l'on se concentre sur les idées de Christian Brassac, ce sont ces informations (celles issues des messages propositionnels) qui participent directement à la construction / négociation de l'objet du discours. Le contenu des messages propositionnels est donc construit en relation avec cet objet, que ce soit pour le confirmer, le contester ou le compléter.

On peut noter que la théorie des actes de discours [Vanderveken, 1988] fournit une série de contraintes sur les contenus propositionnels (comme par exemple le principe de rationalité des locuteurs) et sur les forces illocutoires (comme des conflits sur les modes d'accomplissement). De la même façon que les maximes conversationnelles de Grice, ces contraintes sont supposées être respectées par tout locuteur. Leur infraction apparente indique que le sens littéral n'est pas le sens intentionné par le

locuteur et qu'il est nécessaire de chercher un autre sens au message. Les maximes de Grice et les contraintes d'utilisation des forces illocutoires et des contenus propositionnels peuvent être vues en interprétation comme des conditions d'arrêt : tant que l'interprétation trouvée ne les respecte pas toutes, l'interprétation n'est pas correcte, et il faut en chercher une autre qui soit cohérente. Cette recherche d'une autre interprétation peut passer par l'obtention d'informations supplémentaires, entre autres par le biais de la conversation, entraînant ainsi l'enchaînement des tours de parole.

c) Synchronisation

Traiter des enchaînements de messages et du côté dynamique de la conversation pose la question de la synchronisation des messages. Dans une conversation humaine, les pauses entre messages ou aux changements d'interlocuteurs, de même que les interruptions et le débit de la parole sont des informations pertinentes, très souvent de nature interactionnelle (voir les deux sous-sections précédentes). La notion temporelle est aussi particulièrement présente dans le cadre de messages multi-modaux, où une mauvaise synchronisation entre les messages peut détruire complètement leur synergie, voire produire des effets opposés à ceux souhaités (voir la remarque faite sur l'effet McGurck, note 26 page 67).

Nous avons décidé de simplifier ce problème en nous limitant à l'étude du séquençement des messages (tel message arrivant avant, après ou pendant tel autre) plutôt que d'envisager un aspect temporel précis qui nous aurait posé ces problèmes de synchronisation.

4.5.3 Le focus conversationnel

Comme le fait remarquer Brassac ([Brassac et Trognon, 1992]), "*Le déroulement de la conversation est imprévisible. En revanche, c'est un objet accessible aux conversants et constructible par eux.*" Parler de construction (ou co-construction) suggère un travail commun des interlocuteurs sur un objet commun. Nous avons décidé de construire une partie de la dynamique de la conversation autour de cet objet commun, que nous désignons par le terme de *focus conversationnel*. Ce focus contient les informations échangées par les interlocuteurs et sert de support à la conversation. La notion de focus seule étant insuffisante pour traiter simplement des situations de dialogues qui nous intéressent, nous y ajoutons la notion d'*attentes*, présentée dans la section suivante.

a) Contenu du focus conversationnel

Pour rester fidèles à la subjectivité de la conversation, nous ne considérons pas un focus conversationnel qui soit véritablement commun aux interlocuteurs, mais nous attribuons à chacun d'entre eux un point de vue propre sur la conversation. Chaque participant à la conversation possède un focus local, qu'il peut espérer aussi proche que possible des visions locales des autres interactants²⁶. Ce focus contient l'ensemble des informations non-contestées (voir sous-section suivante) qui ont été échangées. Ces informations elles-mêmes, issues de la conversation, sont obtenues par l'intermédiaire d'un processus d'interprétation, et donc déjà marquées par le récepteur du message, argument supplémentaire en faveur de la subjectivité du focus.

²⁶Toutefois, par commodité de lecture, nous parlerons souvent *du* focus conversationnel.

Le focus conversationnel doit être initialisé avec l'ensemble des connaissances supposées communes (ce que John Searle appelle "connaissances d'arrière plan" ou Vanderveken l'"arrière-fonds (*sic*) conversationnel" [Vanderveken, 1999]). En effet, des conflits peuvent apparaître du fait de certaines de ces connaissances, supposées communes à tort. L'intégration de ces connaissances dans le focus permet le traitement commun des informations issues de la conversation et celles issues des connaissances de l'interactant. L'inconvénient de cette méthode étant la difficulté à déterminer l'ensemble de ces connaissances qui serait considéré comme *implicite* dans le cadre d'une conversation entre humains. Le risque étant de passer un temps démesuré à se mettre d'accord sur ces connaissances supposées communes au détriment de l'objectif de l'interaction.

b) Lien avec la dynamique de la conversation

Nous nous appuyerons sur le focus conversationnel pour donner une dynamique à la conversation. Si l'on considère que les différents interactants co-construisent leur conversation, cette conversation s'arrêtera (dans l'idéal) lorsque la construction réalisée les satisfera tous. Le deuxième cas d'arrêt que nous considérons est celui de l'échec de la conversation, que cet échec soit reconnu par les deux interlocuteurs (déclaration d'incompétence de la part d'un des interlocuteurs, conflit de connaissances trop profond, ...), ou bien que la rupture soit unilatérale (l'un des interlocuteurs décide de mettre fin à la conversation pour des raisons exprimées ou non, sans qu'il y ait consensus sur la décision de cette fin). Nous laissons par exemple de côté les situations où les interlocuteurs s'interrompent et reprennent plus tard une conversation sur un même sujet.

L'objectif (idéal) général étant donc pour les interactants de construire un focus les satisfaisant tous, la dynamique de la conversation est une négociation du contenu de ce focus. De la même façon que Brassac montre des négociations sur les énoncés eux mêmes, à l'échelle de quelques tours de parole (voir les extraits de la table 4.7, p.118), on peut considérer des négociations à l'échelle de la conversation.

Le cœur de notre modèle d'interaction est cette négociation autour du focus. Les différents interactants participent à la conversation en complétant le focus ou en négociant les informations qui y sont contenues, ce qui correspond aux deux comportements suivants : *tentative de modification du focus*, et *commentaire sur le focus*.

A chaque instant de la conversation, les interactants doivent se demander si le contenu du focus les satisfait²⁷. Plusieurs situations sont alors possibles :

- le contenu est satisfaisant : l'interactant le déclare (*commentaire sur le focus*), explicitement ou non (il peut par exemple décider de ne plus intervenir dans la conversation). Lorsque tous les interactants sont dans cet état, on peut considérer que la négociation s'est terminée sur un succès.
- le contenu est insuffisamment satisfaisant ou non satisfaisant. L'interactant peut soit :
 - exprimer l'inadéquation du focus (*commentaire sur le focus*) de manière explicite.

²⁷La fonction de calcul de satisfaction peut être rendue aussi complexe que souhaité, intégrant par exemple des contraintes sociales ou temporelles : tel interlocuteur aura tendance à acquiescer à tout message émanant d'une autorité, tandis que tel autre aura tendance à accepter plus facilement certaines choses si la discussion s'éternise.

- proposer une modification du focus, que ce soit par l’ajout d’informations supplémentaires ou le retrait d’une information déjà présente (*modification du focus*), ce qui est aussi une expression implicite de l’inadéquation du focus, vue au point précédent.

Typiquement, la *modification du focus* s’obtient par l’ajout, le retrait ou le remplacement d’une ou plusieurs informations du focus. Contrairement à la conversation en elle-même, pour laquelle il n’est pas possible de modifier les énoncés antérieurs (voir p.119), le focus conversationnel contient les informations sur lesquelles les interactants sont d’accord (ou plutôt, pensent être d’accord) à un moment donné, il est donc tout à fait possible d’en ôter des parties si l’évolution de la conversation les rend caduques — voire contestables — aux yeux de certains d’entre eux.

De leur côté, *les commentaires sur le focus* ont plusieurs rôles. Ils peuvent avoir un rôle de confrontation entre les différentes visions locales du focus ; ils peuvent aussi servir à expliciter l’incompatibilité entre le message perçu et le contenu courant du focus ; enfin, ils permettent d’émettre un avis sur le focus : incomplet, incohérent, ...

4.5.4 Les attentes

Nous avons pris les dialogues de la table 4.7 (page 118) comme exemple d’une négociation de sens entre différents interlocuteurs. Ces dialogues montrent aussi que lors de la production d’un énoncé, le locuteur a des attentes sur les réactions de son interlocuteur. Selon les différents extraits de ces dialogues, le locuteur réagit en faisant remarquer que l’énoncé reçu est, ou n’est pas, celui qu’il attendait en réaction à son propre énoncé. Avec le focus, ces anticipations constituent le deuxième mécanisme participant directement à la dynamique de la conversation.

Les attentes représentent certains des comportements possibles (idéalement, les plus probables ou les plus courants) de la part de l’interlocuteur dans une situation donnée et sont associées à des actions à entreprendre si l’interlocuteur venait à se comporter de cette façon.

L’exemple le plus parlant que l’on puisse présenter pour montrer l’intérêt des attentes est celui de la réponse à une question. Ainsi, des messages tels que “Clint Eastwood”, “Pourquoi ?” ou “Oui” ne peuvent pas être traités indépendamment du ou des échanges précédents, qui, en l’occurrence, pourraient être ceux du dialogue de la table 4.8.

Tour	Locuteur	Énoncé
1	Agent	Veux-tu un acteur précis dans ton film ?
	Client	Clint Eastwood.
2	Agent	Je peux te proposer « Impitoyable »
	Client	Pourquoi ?
3	Agent	Le film que tu cherches est avec Clint Eastwood.
	Agent	Souhaites-tu plus d’informations sur ce film ?
	Client	Oui

TAB. 4.8 – Un dialogue nécessitant une prise en compte du contexte.

Dans l’ensemble de cette section, nous nous appuyerons sur ce dialogue comme exemple.

Le principal avantage à l'emploi d'attentes est que la réaction de l'agent est fortement contextualisée et est sensée être adaptée de manière très précise à la situation en cours. En ce sens, elle remplit un rôle similaire à celui des étapes d'interprétation et d'expression en mettant en évidence l'importance du contexte. Un deuxième avantage en terme de complexité du traitement est apporté par les attentes : l'action à entreprendre est prédéfinie et associée à un événement ; la perception de cet événement entraîne directement l'accomplissement de cette action.

Nous commencerons cette section en présentant les principes de fonctionnement des attentes, puis nous étudierons un exemple de dialogue. Les attentes, en offrant la possibilité d'associer à un événement (en général une action de la part des interlocuteurs) une action à entreprendre, puis à se mettre en attente d'autres événements offrent à première vue des similarités avec les systèmes utilisant des protocoles d'interaction. C'est pourquoi nous montrerons dans une troisième section en quoi nos attentes se démarquent des protocoles. Enfin, nous détaillerons une spécificité des attentes, leur durée de vie.

a) Principes de fonctionnement des attentes

Une attente est un ensemble {événement, action, priorité, liste d'attentes, durée de vie}. Après avoir décidé d'une action à entreprendre, le locuteur peut prévoir une série de réactions possibles et prédéfinir ses propres futures actions, selon les réactions de son interlocuteur. Pour cela, il définit une (ou plusieurs) attente(s), dans laquelle (lesquelles) il précise

- à quel percept de l'environnement (principalement un comportement de l'utilisateur) cette attente est destinée à réagir (c'est l'événement) ;
- l'orientation du comportement à adopter si l'attente est validée (c'est l'action) ;
- l'importance de l'action à entreprendre, relativement aux autres attentes susceptibles d'être déclenchées (c'est la priorité) ;
- si d'autres comportements (et lesquels) sont à prévoir de la part de l'interlocuteur, une fois l'action effectuée (c'est la liste d'attentes) ;
- un temps pendant lequel l'attente est pertinente et au delà duquel elle doit être oubliée (c'est la durée de vie).

À chacune de ses interventions, l'agent peut décider de créer des attentes. Les origines de ces attentes sont diverses, par exemple, certaines sont valables en permanence, tandis que d'autres sont intégrées au cas par cas, suite à des énoncés spécifiques. Il s'ensuit qu'il est possible de trouver simultanément actives des attentes sensibles aux mêmes événements. Pour éviter les conflits, nous proposons trois méthodes, les deux premières étant intégrées dans la définition même des attentes, la suivante étant liée à des groupes d'attentes plutôt qu'aux attentes elles-mêmes :

1. Donner une *priorité* à chaque attente. Des heuristiques telles que donner une priorité maximale aux attentes les plus récentes, à celles ayant les événements déclenchant les plus spécifiques ou à celles qui cadrent au plus près du rôle de l'agent semblent raisonnables²⁸.
2. Associer une *durée de vie* à chaque attente (ce point est développé page 131). Typiquement, des réactions comme "refus" ou "acceptation" ne peuvent être comprises qu'à condition de faire référence à un énoncé extrêmement récent

²⁸Ce sont celles qui ont été employées dans notre application, voir chapitre 6.

(un seul tour de parole dans la très grande majorité des cas). Ainsi, les attentes deviennent caduques dès que leur temps de vie est dépassé, et ne restent pas à perturber les attentes plus récentes (ou valides un temps plus long).

3. Former des *groupes d'attentes*, rassemblant les attentes nées d'une même action. Par exemple, dans la table 4.9, chaque énoncé de l'agent amène ses propres attentes et l'on peut considérer trois groupes d'attentes, un par tour. Dès lors que l'une d'entre elles est validée, les autres attentes cessent d'être considérées. Ainsi, une seule des réactions possibles est prise en compte et nous négligeons les situations, rares et complexes, où l'interlocuteur décide d'avoir plusieurs réactions à une action unique de l'agent.

Nous verrons dans le chapitre traitant de l'application que la priorité (et, par effet de bord, une partie de la “décrépitude”) est mise en œuvre à travers l'empilement des attentes. Leur consultation est séquentielle et s'arrête dès qu'une attente valide est rencontrée. L'ordre d'empilement définit donc implicitement un ordre de priorité entre les attentes.

Les attentes ne se limitent pas à la construction de séquences action/réaction purement réactives. L'action associée est plus une *orientation* à donner au comportement. Il peut tout à fait s'agir d'une action uniquement interne à l'agent et non un comportement visible. Cette action interne pouvant éventuellement déboucher sur un comportement visible, mais dont la génération peut être aussi complexe que souhaitée, et faire appel à toutes les connaissances disponibles à l'agent.

b) Exemple de fonctionnement

Nous allons dans cette sous-section montrer les mécanismes internes entraînant le dialogue de la page 124 (qui correspond à un comportement externe, observable par les interlocuteurs). Ce dialogue est extrait d'une interaction entre un client humain et notre agent artificiel et l'ensemble des actions de l'agent est sous-tendu par l'emploi d'attentes.

Tour 1. Lors du premier énoncé, l'agent propose des contraintes supplémentaires pour la sélection de films²⁹. Sur un énoncé de ce type, trois réponses sont particulièrement attendues :

1. Un nom d'acteur. Le client, par ellipse, peut répondre une phrase tronquée, comme “Clint Eastwood”, plutôt qu'une phrase complète comme “Je veux un film avec Clint Eastwood”.
 2. Un refus. Le client peut décliner la proposition de l'agent.
 3. Une acceptation. Le client peut répondre au sens littéral de la phrase.
- Dans l'exemple que nous donnons, la réponse de l'utilisateur correspond au point 1.

Ces trois réponses attendues correspondent aux événements déclenchants de trois attentes, pour lesquelles les actions à entreprendre seront respectivement :

1. Intégration de la nouvelle information (le nom de l'acteur) au sein du focus conversationnel, puis consultation de ce même focus pour décider de l'éventuelle action suivante. Dans cette situation, l'attente ne spécifie pas directement un comportement, mais permet de contextualiser le message de façon à

²⁹La fonction de l'agent —la recommandation/recherche de films— est développée dans le chapitre 6 traitant de l'application.

pouvoir l'intégrer dans le focus. Ainsi, l'ellipse est complétée, qui plus est en levant l'ambiguïté sur la fonction de Clint Eastwood, acteur ou réalisateur.

2. Consultation du focus. Dans cette situation, l'attente se contente de passer la main au focus. Ce comportement quasi transparent de l'attente (dans notre modèle, le focus aurait, de toutes façons, proposé une réponse) peut se justifier par trois points : premièrement, spécifier ce genre d'attente avec une priorité élevée peut permettre d'ignorer d'autres attentes, et, d'une certaine façon, de *garantir* que la réaction de l'agent sera conditionnée par le focus ; deuxièmement, si cette attente est validée (et selon la stratégie de gestion des attentes décidée), les autres attentes issues du même tour (basées sur un nom d'acteur ou une acceptation dans notre exemple) peuvent être considérées comme caduques et supprimées ; troisièmement, mais cet intérêt n'est pas fonctionnel, cette attente permet de marquer explicitement le refus comme une conséquence probable.
3. Demande explicite. Dans cette situation où l'interlocuteur ne fait que répondre à une interprétation littérale de l'énoncé de l'agent (à la manière du classique exemple "Peux-tu me passer le sel ?" – "Oui."), l'action à entreprendre est une demande d'explicitation de la réponse faite.

Tour 2. Lors du deuxième énoncé, l'agent propose un film au client. Sur un énoncé de ce type, trois réponses sont particulièrement attendues :

1. Refus. Le client refuse explicitement le film proposé.
 2. Demande d'explicitation. Le client demande pourquoi ce film lui a été proposé.
 3. Acceptation. Le client accepte la proposition de l'agent.
- Dans l'exemple que nous donnons, la réponse de l'utilisateur correspond au point 2.

Ces trois réponses attendues correspondent aux événements déclenchants de trois attentes, pour lesquelles les actions à entreprendre seront respectivement :

1. Retrait du film et consultation du focus. Dans cette situation, le client refuse le film proposé. Ce film est donc intégré dans le focus comme une nouvelle contrainte négative et la décision de l'action à entreprendre est laissée au focus, une fois remis à jour.
2. Transmission du contenu du focus et proposition d'informations supplémentaires. Dans cette situation, l'agent déclare les informations contenues dans le focus et qui correspondent à l'ensemble des critères sur lesquels il s'est basé pour déterminer le film à proposer³⁰. Après quoi, il transmet des informations sur le film lui-même.
3. Continuer sur une procédure de conclusion de vente. Dans cette situation, l'agent abandonne son mode de fonctionnement relativement "libre", basé sur le focus et les attentes, pour employer des méthodes plus contraintes, plus adaptées à la conclusion d'une vente.

³⁰Notons que le dialogue de la page 124 est généré par notre agent et qu'il n'implémente qu'une partie restreinte de ce que nous présentons dans ce chapitre théorique. Par exemple, l'agent n'intègre pas dans sa justification les *préférences* du client, bien que cela soit techniquement possible et que ces préférences soient prises en compte dans la recherche d'un film à proposer.

Tour 3. Lors du troisième énoncé, l'agent déclare les informations ayant guidé le choix du film qu'il a proposé, puis offre au client des informations supplémentaires. Pour ce dernier énoncé, deux réponses sont particulièrement attendues :

1. Acceptation.

2. Refus.

→ Dans l'exemple que nous donnons, la réponse de l'utilisateur correspond au point 1.

Ces deux réponses attendues correspondent aux événements déclenchant de deux attentes, pour lesquelles les actions à entreprendre seront respectivement :

1. Transmission d'information sur le film. Dans cette situation l'agent va rechercher les informations dont il dispose sur le film, puis les déclare au client.

2. Consultation du focus. De la même façon que nous l'avons vu précédemment, cette attente se contente de passer la main au focus.

L'ensemble des énoncés de l'agent, ses attentes et les énoncés du client sont synthétisées dans la table 4.9.

Tour	Énoncé de l'agent	Extraits des paramètres de l'attente		Réponse de l'utilisateur	Énoncé résultant
		Réaction attendue	Actions associées		
1	Veux-tu un acteur précis dans ton film?	Nom d'acteur	intégration au focus & consultation du focus	Clint Eastwood.	Je peux te proposer « Impitoyable » .
		Refus	Consultation du focus		
		Acceptation	Demande explicite de la forme : "Lequel?"		
2	Je peux te proposer « Impitoyable » .	Refus	Retrait du film du focus & consultation du focus	Pourquoi?	Souhaites-tu plus d'informations sur ce film?
		Demande d'explicitation	Transmission du contenu du focus & proposition d'informations supplémentaires		
		Acceptation	Continuer sur une procédure de conclusion de vente		
3	Souhaites-tu plus d'informations sur ce film?	Acceptation	Transmission d'informations.	Oui.	Tourné en 1992, Gene HACKMAN joue dedans, [...]
		Refus	Consultation du focus		

TAB. 4.9 – Liens entre message et contexte : une solution exploitant les attentes.

c) Liens avec les protocoles.

Le système d'attentes permet une régulation de la conversation *ressemblant* au contrôle de l'enchaînement conversationnel obtenu par l'emploi de protocoles d'interaction. En effet, les *événements* sont assimilables aux conditions de passage d'une étape à l'autre d'un protocole, étapes auxquelles sont associées des *actions*, ainsi qu'une *liste* de conditions de passage vers une autre étape, etc. Ainsi, lorsque l'on analyse dans son ensemble les attentes possibles de l'agent, on peut en rassembler certaines en des séquences *pouvant* être suivies à la manière d'un classique protocole d'interaction.

Toutefois, attentes et protocoles présentent des différences qui nous permettent de justifier les choix des premières par rapport aux seconds. La thèse de Marc-Philippe Huget synthétise un ensemble de définitions au terme "protocole", pour en obtenir la suivante [Huget, 2001] :

Un protocole est un ensemble de règles qui guident l'interaction entre plusieurs agents. Pour un état donné du protocole, il n'existe qu'un nombre fini de messages en émission et en réception. Si un agent accepte d'utiliser un protocole, il accepte de se conformer à ce protocole et à en respecter les règles. De plus, il approuve la sémantique du protocole.

Une règle est soit syntaxique, soit sémantique. Une règle syntaxique porte sur l'architecture du protocole, i.e. la construction des transitions reliant les états du protocole. [...] Les règles sémantiques définissent les actions que les agents doivent effectuer lors de l'émission et de la réception d'un message. [...] En effet, les agents connaissent quels sont les messages qu'ils peuvent recevoir pour un état donné de l'interaction, [et] quels sont les messages qu'ils peuvent envoyer [...]

Le système à base d'attentes que nous avons mis en place se distingue principalement des protocoles sur le point des *engagements*. Dans le cas des protocoles, un agent s'engage à suivre le protocole ; il garantit qu'il n'emploiera que des actions attendues par le protocole ; il certifie que les messages reçus seront traités de la manière prévue par le protocole ; il a l'assurance que, du côté de ses interlocuteurs, un protocole adapté au sien guide leurs interventions, et que ces interlocuteurs respectent les mêmes engagements que lui-même.

En revanche, un système à base d'attentes tel que nous le proposons laisse une plus grande liberté à chacun des interactants. L'agent est libre d'employer ou non des attentes et de changer ce choix sans avoir à en informer ses interlocuteurs. Il peut disposer d'autres moyens de suivi de conversation (et c'est le cas avec notre focus conversationnel), ce qui entraîne que les actions entreprises par l'agent peuvent ne pas être dictées par le système d'attentes, et que l'agent est capable de traiter des événements non prévus par ses attentes. Enfin, l'agent peut décider d'employer un système d'attentes indépendamment des choix de ses interlocuteurs quant à leurs méthodes de gestion de l'enchaînement conversationnel.

De plus, en général, un protocole est explicite et nécessite l'accord des deux partenaires. Lorsque l'un des partenaires est humain, ces deux points sont plus délicats à garantir, à moins de donner à l'humain des contraintes (obligation de suivre rigoureusement telle ou telle procédure, interfaces contraignantes, etc.).

Outre ces "engagements", plus limités dans le cadre d'attentes, employer le concept d'attentes plutôt que de protocole nous permet d'éviter plusieurs problèmes qui

seraient apparus si nous avions souhaité employer *exactement* des protocoles. Tout d’abord, le modèle que nous proposons a pour but d’améliorer l’interaction avec l’être humain. Les situations d’interaction classiques entre entités artificielles sont souvent guidées par des protocoles, et donc, de fait, strictes, extrêmement sensibles aux détails, aux séquences parfois enchevêtrées, susceptibles de parcourir de nombreux niveaux de récursion, exigeantes et restrictives quant aux actions possibles en un moment précis, etc. Ces contraintes peuvent rapidement devenir trop complexes pour être suivies par des humains. Les attentes nous libèrent du côté strict des protocoles. Elles nous permettent d’avoir un mécanisme de gestion de la dynamique conversationnelle qui ne soit pas bloquant, qui puisse être choisi indépendamment des choix des interlocuteurs et qui puisse coexister avec d’autres moyens de gestion de cette dynamique.

Bien sûr, la notion d’attentes gagnerait à être enrichie de nombreux concepts classiques dans le domaine des protocoles d’interaction, mais nous nous sommes limités à définir pour les attentes les points ayant une importance dans le cadre de cette thèse et ne nous entraînant pas trop loin de nos objectifs initiaux.

d) Durée de vie.

La notion de durée de vie de l’attente permet de moduler la capacité de l’agent à garder le fil de la conversation malgré des digressions, tout en permettant d’“oublier” un fil de discussion précédent, si la digression s’avérait finalement être une réorientation de la conversation.

ttl ^a de l’attente d’une acceptation = 1	ttl de l’attente d’une acceptation = 2
A1 : Je te propose “à bout de souffle”	A1 : Je te propose “à bout de souffle”
H2 : donne moi des infos sur à bout de souffle.	H2 : donne moi des infos sur à bout de souffle.
A3 : Tourné en 1959 , réalisé par Jean-Luc GODARD [...]	A3 : Tourné en 1959 , réalisé par Jean-Luc GODARD [...]
H4 : d’accord.	H4 : d’accord.
A5a : “d’accord” quoi ?	A5b : Tu veux prendre “à bout de souffle”, c’est bien ça ?

^aTime To Live : temps de vie de l’attente.

TAB. 4.10 – Influence du temps de vie (ttl) des attentes sur la dynamique de la conversation (extrait d’un dialogue entre un humain et l’agent présenté dans le chapitre 6)

Dans l’exemple de dialogue présenté dans la table 4.10, l’agent propose un film à son interlocuteur (A1), et, parmi les réactions attendues, il y a l’**acceptation** de ce film. La seule différence entre les dialogues de gauche et de droite est que cette attente sera valide pendant un seul tour de parole pour le dialogue de gauche et pendant deux pour le dialogue de droite. Ainsi, dans le premier cas lorsque l’interlocuteur (H2) demande des informations sur le film proposé, l’agent (A3) répond à sa requête, mais “oublie” l’attente liée à l’acceptation. Il s’en suit que lorsque l’interlocuteur accepte

(H4) le film proposé deux tours de parole auparavant (A1), l'agent ne peut plus compter sur cette attente pour réagir et doit faire appel à des stratégies différentes³¹. Dans le deuxième cas, l'attente d'une acceptation est toujours valide au moment de l'énoncé H4, et l'agent peut donc réagir suivant l'action associée à cette attente (demande de confirmation et déclenchement d'une procédure de conclusion de vente en l'occurrence, suivant le même principe que l'on peut voir dans les attentes du deuxième tour du tableau 4.9).

L'intérêt principal du temps de vie est de permettre d'abandonner certaines directions de la conversation dès lors que l'interlocuteur n'y fait plus référence. Les règles de variation du temps de vie peuvent être aussi complexes que voulues, l'approche la plus simple consistant à décrémenter le temps de vie de l'attente à chaque tour de parole. Mais d'autres stratégies, basées sur le temps (durées de vie exprimée en secondes par exemple), ou même en permettant l'augmentation³² sont tout à fait possible.

4.5.5 Liens entre focus et attentes

Le focus conversationnel et les listes d'attentes jouent des rôles complémentaires dans la gestion de la dynamique de la conversation. Les attentes sont particulièrement adaptées à la prise en charge de situations précises, mais, pour cette même raison et du fait de leur extrême contextualisation, ne peuvent être employées que dans des situations très spécifiques et définies. Le focus, de son côté, ne peut traiter que les messages ayant trait aux objets de la discussion, mais n'aboutit jamais à une situation bloquée puisqu'il peut en permanence faire des propositions de modification ou des commentaires pour faire "avancer" la discussion. De manière très générale, on peut dire que le focus présente une spécification de la conversation descendante : il donne un cadre à l'interaction, sans spécifier quoi que ce soit à l'échelle des échanges eux-mêmes. En ce sens, il se rapproche des conduites de conversation présentées dans la section 4.2.2. Les attentes, elles, spécifient l'interaction de manière ascendante, la traitant à l'échelle de l'énoncé, et indépendamment d'un but global. En ce sens, elles se rapprochent des protocoles d'interaction.

La frontière qui sépare les situations idéales à gérer par l'une ou l'autre de ces méthodes n'est pas clairement définissable. D'un côté, multiplier les attentes permet de traiter des situations qui pourraient être gérées par le focus. Si l'on pousse à l'extrême le fonctionnement en "tout attentes", on obtient un système fonctionnant sur un principe proche des protocoles d'interaction. D'un autre côté, ajouter des traitements dans l'étape d'interprétation permet de déléguer au focus des traitements qui pourraient revenir à des attentes. On peut en effet mettre en place au niveau de l'étape d'interprétation (c'est son rôle) une prise en compte de l'historique qui effectuerait les conversions³³ présentées dans la table 4.11.

³¹Nous verrons dans le chapitre lié à l'application que, dans le cas présent, c'est un système purement réactif (de type *chatbot*) qui a réagit.

³²Si l'on considère l'augmentation du temps de vie de l'attente en fonction de son importance (potentielle) dans le contexte conversationnel courant, la valeur "temps de vie" acquiert finalement le statut plus générique de "pertinence" de l'attente. Nous avons préféré conserver le terme de "temps de vie" car nous ne considérons que les variations de pertinence liées à l'"oubli", donc au temps. Le terme est donc non seulement plus clair et plus aisé à manipuler, mais aussi plus proche de la réalité de nos travaux, l'étude de la *pertinence* des attentes, dans le cas général, dépassant largement le cadre de cette thèse.

³³Ce genre de mécanisme serait à peine plus complexe à mettre en œuvre que le traitement des anaphores et des déictiques, déjà pris en charge par le module d'interprétation. Notre

Tour	Énoncé de l'agent	Énoncé de l'interlocuteur	Résultat possible en sortie d'interprétation ^a
1	Veux-tu un acteur précis dans ton film ?	Clint Eastwood	Je veux un film avec Clint Eastwood.
2	Je peux te proposer « Impitoyable ».	Pourquoi ?	Pourquoi me proposes-tu « Impitoyable » ?
3	Souhaites-tu plus d'informations sur ce film ?	Oui.	Donnes moi des informations sur le film « Impitoyable ».

^aReprésentation en langue naturelle du message en sortie de l'étape d'interprétation.

TAB. 4.11 – Liens entre message et contexte : une solution exploitant l'interprétation

Ainsi, un observateur extérieur du dialogue de la table 4.8 ne peut pas savoir si celui-ci est généré par des attentes, comme nous l'avons montré page 126, par le focus, comme nous venons de le montrer, ou par une méthode mixte.

De plus, comme nous le verrons dans le chapitre “Application”, et comme cela apparaît déjà sur certains des exemples précédemment cités, il est tout à fait possible que l'action associée à une attente, après avoir effectué quelques traitements internes, décide de “passer la main” au focus, qui se retrouve en charge de la suite à donner à la conversation. De la même façon, certaines actions, décidées suite à la consultation du focus conversationnel entraînent la mise en place d'attentes.

4.5.6 Confrontations, ruptures, négociations

Nous considérons qu'une partie de la dynamique de la conversation émerge de la co-construction d'un focus commun à l'ensemble des interlocuteurs (rappelons qu'il n'est pas exact de parler d'un focus, car chacun des interlocuteurs possède sa propre vision de la conversation, vision locale qu'il espère aussi proche que possible de celle des autres), mais il faut noter que cette dynamique peut naître autant des situations de conflit que des situations où les interlocuteurs approuvent chaque nouvel énoncé. En effet, même s'il existe des situations conversationnelles dans lesquelles les énoncés s'enchaînent sans que les interlocuteurs n'ait à en contester aucun, les *ruptures* (malgré leur nom) ne sont pas des situations d'exceptions (au sens informatique du terme) et font tout autant partie de situations “normales” de la conversation.

Une fois de plus (comme nous l'avons fait pour les enchaînements, dans la partie 4.5.2, et avec les mêmes précautions d'emploi), il est confortable de considérer les ruptures suivant deux catégories : les ruptures interactionnelles (ou conversationnelles) et les ruptures propositionnelles (ou informationnelles). Les raisons pouvant entraîner ces ruptures peuvent tout à fait être transversales (i.e., une même raison pouvant entraîner des ruptures des deux types), comme celles liées à la personnalité, les éventuelles caractéristiques psychologiques, sociales et émotionnelles des interlo-

application, présentée chapitre 6, se limite au traitement de certaines anaphores.

cuteurs étant tout à fait susceptible d'entraîner des ruptures. Ce point ne sera pas développé dans cette thèse.

a) Ruptures conversationnelles

Les ruptures conversationnelles (ou interactionnelles) rassemblent les conflits qui portent sur la forme des énoncés. En tant que telles, elles n'ont que peu de lien avec le focus conversationnel, car elles sont issues d'un traitement amont, mais participent directement à la dynamique de la conversation. Elles font état des conflits syntaxiques ou lexicaux par exemple, signalant à l'auditoire que la forme de l'énoncé pose problème. Ces ruptures interviennent aussi dans le cas où le message a pu être bruité lors de son transfert, s'il existe des ambiguïtés lexicales ou syntaxiques, bref, dans toute situation où le problème à la source de la rupture interdit de pouvoir passer à un traitement sémantique.

b) Ruptures informationnelles

Les ruptures informationnelles (ou propositionnelles) interviennent lors de problèmes de traitement du contenu du message. L'incohérence (dont le quiproquo serait un cas particulier) est une situation typique pouvant induire des ruptures propositionnelles. Elle marque une incompatibilité entre le contenu d'un message et les connaissances de l'agent, que ces connaissances lui soient propres ou issues des étapes précédentes de la conversation. Par exemple, lors de la co-construction du focus, il est possible d'avoir à y introduire des informations en conflit avec celles déjà présentes. La levée de l'incohérence peut être effectuée de manière individuelle, l'allocutaire prenant personnellement une décision. Il est ainsi possible de faire des suppositions à propos de l'information ambiguë ou manquante, et voir dans la suite de la conversation si cette supposition se confirme, ou n'est pas contestée. C'est, par exemple, l'approche de Leuhen et du système Coala [Lehuen, 1997]. Mais la résolution de l'incohérence peut aussi devenir à son tour objet de négociation, l'allocutaire signalant (par une rupture de la conversation) le conflit.

On peut distinguer trois grandes classes de solutions à ce genre de ruptures : soit les interlocuteurs se mettent d'accord sur ce qu'il faut accepter, refuser ou supprimer pour lever le conflit ; soit les interlocuteurs réalisent qu'il y a eu incompréhension au niveau du message lui-même (le sens interprété est trop éloigné du sens intentionné) ; soit ils réalisent que le problème se situe à un niveau encore plus élevé, et qu'ils ne communiquaient pas en référence au même objet (focus) de discussion (c'est le cas du quiproquo).

Enfin, il est possible que le conflit détecté n'ait pas de solution. Soit que les interlocuteurs ne souhaitent (ou ne peuvent) pas modifier leurs connaissances, soit que les informations transmises soient incompréhensibles, soit que la construction d'un focus commun soit considérée comme impossible (dans le cas de la négociation de films, l'agent peut avoir à signaler qu'à sa connaissance, aucun film ne peut rassembler tous les critères intégrés dans le focus).

Un cas limite de la rupture propositionnelle serait la levée d'ambiguïté : un léger décalage du focus conversationnel pour se concentrer sur un point particulier (du message ou du focus lui-même) le temps de confirmer un détail, ou d'obtenir des précisions.

4.6 Conclusion

La communication entre humains et la communication entre entités informatiques ont toutes deux leurs domaines d'étude dédiés. Les théories s'appliquant à ces différents domaines sont parfois compatibles, parfois contradictoires. Nous avons cherché à rassembler dans ce chapitre différents points de ces deux domaines pour former un ensemble permettant à des humains et à des agents artificiels de communiquer en respectant (et en exploitant) les spécificités de chacun.

Nous avons principalement insisté sur trois points. Le premier point traite de l'importance de considérer les messages *en contexte*, et donc de les interpréter et de les exprimer plutôt que de les coder et de les décoder comme c'est le cas classiquement en informatique. Les deux derniers points portent sur la gestion de la dynamique de la conversation. Le deuxième point, donc, montre comment la conversation, considérée comme une négociation autour d'un objet commun aux interlocuteurs, peut être structurée par la co-construction d'un *focus conversationnel*, représentant cet objet commun. Ce mode de contrôle de l'interaction précise un objectif à atteindre, sans donner de contraintes au niveau des échanges qui peuvent permettre d'atteindre cet objectif. Le troisième point, enfin, propose une méthode de structuration de la communication basée sur les *attentes* qu'un locuteur peut avoir quant aux effets (perlocutoires) des ses propres actions. Ce mode de contrôle de l'interaction précise, lui, des échanges, indépendamment d'un objectif global à atteindre.

Dans leur ensemble, les principes que nous avons présentés dans ce chapitre sont plus issus des sciences humaines (donc de l'interaction humain-humain) que des sciences de l'information. En effet, nous avons montré que plusieurs choix faits pour la communication entre agents (principalement l'emploi de protocoles) sont inadaptés à l'humain. Mais nous avons aussi dit dans un chapitre précédent qu'une reproduction des mécanismes conversationnels de l'humain n'était pas souhaitable, car ces mécanismes sont eux aussi inadaptés à l'exploitation de certaines compétences des agents artificiels. Nous avons donc cherché, tout au long de ce chapitre, à suivre l'idéal de symbiose de Licklider, c'est-à-dire, à définir des principes permettant de retrouver humains et agents artificiels comme de réels *partenaires* dans l'interaction.

Chapitre 5

Modèle et architecture d'agent

Car il est besoin de remarquer que le principal effet de toutes les passions dans les hommes est qu'elles incitent et disposent leur âme à vouloir les choses auxquelles elles préparent leur corps.
RENÉ DESCARTES, “*Les passions de l'âme*”

Ce chapitre constitue une transition entre les principes de dialogue que nous venons de présenter (chapitre 4) et l'agent que nous avons réalisé (chapitre 6).

En partant des principes d'interaction que nous souhaitons voir présents dans le comportement de notre agent, nous développerons dans ce chapitre non pas un modèle complet d'agent, mais uniquement les quelques éléments à ajouter ou à inclure à un modèle quelconque d'agent¹. Par exemple, aucun modèle *complet* de raisonnement ne sera présenté ici. À la suite de ces *éléments* de modèles, donc une perspective fonctionnelle mais fractionnaire, nous proposerons une architecture possible, dans une perspective plus structurelle, mais complète.

Nous commencerons donc ce chapitre en rappelant qu'un agent conversationnel ne se définit pas uniquement comme un programme capable d'employer la langue naturelle, mais qu'il s'agit d'un *agent* différencié — donc un individu, doté d'une personnalité — doué de capacités sociales (section 5.1). Les deux sections suivantes reprendront les deux points principaux de cette thèse, tout d'abord la prise en compte contextuelle des messages (interprétation et expression, section 5.2) puis leur enchaînement (focus et attentes, section 5.3). Après cette présentation des éléments de ce modèle d'agent conversationnel, nous synthétiserons dans la dernière section (5.4) une architecture possible, dans une transition vers la dernière partie de cette thèse qui présente l'agent tel qu'il a été réalisé.

¹Les modèles “cognitifs” sont bien sûr privilégiés, mais rien n'interdit à un agent “réactif” de disposer d'étapes d'interprétation et d'expression, si ce n'est que sa catégorisation “réactive” risque d'en devenir caduque.

5.1 L'agent conversationnel : un individu dans un groupe

Comme nous l'avons dit tout au long du chapitre 4, nous considérons des agents ayant une individualité, c'est-à-dire distincts les uns des autres dans leurs comportements.

Pour s'intégrer dans une communauté humaine et former une communauté mixte, les agents doivent disposer d'une forme de *conscience sociale* que nous présenterons au début de cette section. Outre cette conscience sociale, nous présenterons également les connaissances dont doit disposer un agent conversationnel destiné à intégrer une communauté mixte. Trois ensembles de connaissances de l'agent sont particulièrement pertinents : (1) l'*historique* de l'interaction, trace des messages échangés et des sens interprétés ou intentionnés qui y sont liés ; (2) les *connaissances sur les autres*, qui comprennent à la fois les informations que l'on peut avoir sur les interlocuteurs et les façons de se comporter vis à vis d'eux ; (3) les *paramètres d'individuation*², rassemblant tout ce qui est spécifique à l'agent en tant qu'individu (tant dans les informations —par exemple, la personnalité— que dans ses conduites —par exemple, sa politique conversationnelle).

Nous n'aborderons pas dans cette section les connaissances et compétences spécifiques à une tâche, car nous traitons ici des capacités *conversationnelles* de l'agent. Puisque nous souhaitons un modèle aussi générique que possible, nous devons pouvoir nous affranchir de toute description de la tâche de l'agent lorsque nous présentons ses capacités conversationnelles. Cette généralité ne vaut que pour les *principes* que nous proposons, car il est bien évident que la tâche à accomplir marquera l'ensemble de la chaîne de traitement du message. Ainsi, pour les traitements en lien avec la langue naturelle, le lexique de l'agent est profondément conditionné par la tâche qu'il doit accomplir.

5.1.1 Conscience sociale

La conscience sociale de l'agent rassemble ce qui a trait à son comportement en présence d'autres agents. Elle comprend principalement (1) une série de règles (protocoles ou rituels d'interaction, normes, etc.) qui conditionnent le comportement de l'agent suivant les agents en présence et (2) des informations sur les interlocuteurs. La conscience sociale a pour but d'adapter les messages de façon à les rendre socialement acceptables (en émission) et de vérifier l'adéquation sociale de messages reçus (en réception). Elle acquiert et met aussi à jour certaines parties du profil des interlocuteurs.

En phase d'émission du message (plus précisément, dans l'étape d'expression), on peut faire le parallèle avec le paradigme structuro-expressif des psychologues, tel qu'il est présenté par Alex Mucchielli [Mucchielli, 1995]. Dans ce paradigme, des structures profondes (désirs, valeurs, certitudes) passent par une suite de transformations (particulièrement sociales) pour générer des *phénomènes de surface* (nous dirions *comportements*) : “*Les phénomènes de surface [étant], dans notre cas, les*

²Bien que ce dernier point offre une certaine symétrie avec le précédent, le terme de “paramètres d'individuation” a été préféré à celui de “connaissances de l'agent sur lui-même”, qui poserait le problème de l'introspection et de l'accès direct à des paramètres devant alors être explicites.

conduites et les expressions verbales des sujets; les structures sous-jacentes [étant] leurs psychismes.” [Mucchielli, 1995]. Une fois le contenu du message décidé par la partie raisonnement de l'agent (le “psychisme”), pour avoir des agents *socialement corrects* [Chicoisne, 1998; Chicoisne et Pesty, 1999], il faut prendre en compte ce qu'Alex Mucchielli appelle le *principe de réalité*. À ce niveau-ci une intention est confrontée aux connaissances de l'agent sur le monde pour savoir si oui ou non (et si oui, comment) cette intention peut donner naissance à une action. “*La communication est alors l'expression d'un désir sous une forme socialisée*”[Mucchielli, 1995].

Symétriquement, en phase de réception d'un message (dans l'étape d'interprétation) la conscience sociale a deux fonctions : mettre à jour les connaissances sur l'interlocuteur et confronter ses actions aux normes.

Premièrement, donc, les messages échangés sont susceptibles d'être porteurs d'informations sur la politique conversationnelle suivie par le locuteur, ses connaissances, ses compétences linguistiques (au sens le plus large) ou sa personnalité. C'est dans la phase d'interprétation que ces paramètres sont extraits de façon à constituer (ou affiner) un modèle de l'interlocuteur. Pour être plus exact, les informations obtenues ne portent pas directement sur l'interlocuteur, mais sur le “personnage” joué par notre interlocuteur dans la situation interactionnelle du moment³. Mais, au fur et à mesure des interactions, le modèle qu'on se fait de nos interlocuteurs peut s'affiner, et l'on peut espérer toucher l'acteur à travers ses personnages (voir note numéro 19 page 109). Cependant, tant que les différents interlocuteurs restent cohérents, et qu'il est possible de savoir qu'ils incarnent tel personnage à tel moment, le fait de savoir si les informations obtenues appartiennent plus au personnage ou à l'acteur est finalement secondaire. Il suffit alors de se baser sur les informations précédemment obtenues sur le personnage en cours.

La deuxième fonction de la conscience sociale en phase d'interprétation est la vérification de l'adéquation du comportement des interlocuteurs aux standards sociaux : telle action (tel message) est-elle (il) ou non acceptable ? La conscience sociale de l'agent permet la détection de ces infractions, et peut mener jusqu'à une rupture de la conversation en cours pour initier un rituel de réparation [Goffman, 1974].

5.1.2 Historique de l'interaction

L'historique de l'interaction est une trace des messages que l'agent a perçus dans son environnement ainsi que de ceux qu'il y a déposés. Cette trace, contient aussi, en regard de chaque message perçu le (ou les) sens extraits de ce message, et en regard de chaque message déposé dans l'environnement, son sens intentionné (voir table 5.1). Un historique particulièrement complet pourrait aussi conserver la trace des mécanismes mis en œuvre dans les traitements des messages.

L'usage d'un historique permet de garder une trace du passé, ce qui a un intérêt direct sur le suivi de la conversation. L'une des caractéristiques principales d'une conversation “confortable” entre humains est la cohérence du comportement des interlocuteurs. La cohérence implique une continuité entre les actions passées et l'action présente, voire les actions futures (la cohérence du discours peut n'apparaître qu'une fois achevé l'envoi d'un ensemble de messages). Une phase de prise en compte des événements passés peut donc aider à assurer la cohérence du discours⁴.

³L'interlocuteur n'étant, finalement, qu'un *acteur*, “*artisan infatigable des impressions d'autrui*” pour reprendre la métaphore théâtrale de Goffman, voir page 109.

⁴Notons que toute considération d'un événement passé pour décider d'un comportement

Messages perçus	Messages formalisés	Messages interprétés	Contenu du focus & attentes	(éventuelles) informations liées à la tâche	Contenu du focus & attentes	Message intentionné	Message exprimé	Message énoncé
-----------------	---------------------	----------------------	-----------------------------	--	-----------------------------	---------------------	-----------------	----------------

TAB. 5.1 – Format des enregistrements de l'historique

D'un pur point de vue conversationnel (et donc, une fois de plus, indépendamment des besoins d'une tâche donnée), nous identifions trois fonctionnalités à l'historique :

1. au niveau du message en langue naturelle, l'historique permet de résoudre les anaphores. Ainsi, dans l'énoncé "Donne *le* moi", le pronom anaphorique *le* fait référence à un objet cité précédemment et que l'on peut identifier par consultation de l'historique.
2. au niveau de l'enchaînement conversationnel, le maintien d'un historique permet de revenir sur les échanges précédents de façon à les justifier, ou tout simplement les ré-exprimer. Ainsi, l'agent peut déclarer explicitement les attentes qu'il avait et montrer en quoi les actions de ses interlocuteurs les ont déclenchées ; il peut de même déclarer le contenu de son focus, montrer à ses interlocuteurs comment celui-ci a évolué en fonction des différents échanges et ainsi justifier son comportement, ou permettre de détecter dans l'interaction l'instant où a commencé l'incompréhension.
3. au niveau de la gestion des incompréhensions, conserver les différentes étapes du traitement du message permet de les détailler. Ainsi, l'agent peut revenir sur l'ensemble des interprétations possibles d'un message ambigu, ou bien énoncer explicitement les résultats de l'interprétation des messages de l'interlocuteur (ce qui, en langue naturelle correspond à : "Tu m'as dit <Message perçu>, et j'en ai compris <Message interprété>").

5.1.3 Connaissances sur les autres

Les connaissances sur les autres sont fondamentales si l'on souhaite communiquer. Si l'agent n'a pour but que d'émettre de l'information, sans en espérer aucun retour, ou d'agir indépendamment d'autres agents, cette connaissance serait effectivement superflue. Mais à partir du moment où l'agent désire que ses messages aient un effet particulier sur ses allocutaires, il ne peut se passer de connaissances sur leurs moyens de perception (et d'action), leurs compétences à comprendre et traiter les messages, ou tout simplement leurs capacités à faire ce qu'il désire.

Les connaissances sur les autres dont dispose l'agent lui permettent donc de communiquer, mais elles lui permettent aussi de communiquer mieux. Ces connaissances est finalement une forme d'apprentissage.

portent principalement sur deux domaines (non indépendants) : l'interlocuteur en tant que partenaire dans la conversation et l'interlocuteur en tant que partenaire sur la tâche à accomplir. Ce dernier point rassemble les connaissances sur le monde (*savoirs*) et les capacités d'action (*savoir-faire*) de l'interlocuteur. Ces connaissances sont spécifiques aux tâches de l'agent. Le premier point, lui, rassemble les paramètres constituant l'individualité de l'interlocuteur et ceux conditionnant son comportement, ainsi que ses capacités interactionnelles et les conduites de conversation indiquant les actions possibles vis-à-vis de l'interlocuteur.

Nous n'énumérerons pas ici l'ensemble des connaissances sur les autres susceptibles d'influer sur le cours d'une conversation, l'élaboration d'une telle liste nécessiterait en effet des recherches en psycholinguistique qui dépasseraient largement le cadre de cette thèse. De plus, ces paramètres dépendent grandement des modèles choisis pour représenter telle ou telle connaissance sur les autres. Nous nous sommes limités à la considération de certains paramètres sociaux et nous verrons simplement dans la partie traitant de l'implémentation (chapitre 6) que, principalement pour des raisons de finesse des traitements de la langue naturelle, nous nous sommes limités à l'exploitation de la théorie des faces et du territoire de Goffman.

Notons pour finir que, dans le cadre d'interactions plus classiques entre agents artificiels, une connaissance essentielle sur les autres agents concerne les protocoles d'interaction qu'ils peuvent utiliser.

5.1.4 Paramètres d'individuation

Nous avons dit que nous considérons des agents ayant une individualité. Chaque agent dispose donc de divers paramètres lui permettant d'"être" différent des autres.

Le paramètre fondamental est constitué par les compétences de l'agent. C'est le paramètre qui permet le plus facilement de distinguer les agents les uns des autres (et il s'agit du plus couramment employé), puisqu'il permet d'apporter assez rapidement une réponse aux questions telles que "*Cet agent est-il susceptible de m'aider à atteindre mes objectifs ?*" ou "*Quelle est la fonction de cet agent ?*" Parallèlement à ces compétences, il est intéressant de considérer l'ensemble des paramètres qui influent non pas sur les actions de l'agent mais sur sa manière d'agir⁵. Nous retrouvons ici une distinction plusieurs fois faite entre ce qu'il faut faire et la manière de la faire ou, dans l'autre sens, ce qui a été fait et l'objectif que voulait atteindre celui qui a fait.

Pour les mêmes raisons que pour les "connaissances sur les autres", nous ne ferons pas ici de liste des paramètres d'individuation possibles. Nous avons principalement considéré la conscience sociale (partie de la personnalité que nous considérons comme prépondérante dans les situations interactionnelles), mais, la façon de gérer les attentes ou le focus a également sa place ici. Ainsi, un agent demandant confirmation

⁵On peut faire une analogie avec les sociétés humaines en considérant deux fonctionnaires d'une administration, ayant même rôle et statut. Ces deux fonctionnaires sont a priori interchangeables (les *fonctionnaires* sont sensés remplir une *fonction*). Pourtant, en situation réelle, on constate que l'un sera plus efficace dans telle ou telle situation, que l'autre sera généralement plus diligent, ou que l'on "préfère" avoir à traiter avec l'un plutôt que l'autre. Une fois encore, puisque nous nous concentrons ici sur l'aspect interactionnel de l'agent, nous passerons sous silence les altérations qu'il est possible de faire subir aux fonctionnalités de l'agent (efficacité ou diligence de l'agent dans l'exemple précédent), pour nous concentrer sur ce qui tient à la façon d'agir (aspect "relationnel" de l'exemple précédent) plus qu'à l'intention de l'action.

de chaque information avant de l'intégrer à son focus paraîtra excessivement peu sûr de lui ou mal-comprenant.

5.1.5 Une interaction socialisée mais fonctionnelle

Il ne faut pas oublier que les aspects présentés dans les sections précédentes — issus de l'étude des relations humaines — sont sensés *améliorer* les interactions entre humains et agents artificiels : la fonctionnalité de l'agent ne doit pas s'en retrouver entravée. D'après les définitions de Gerhard Weiss [Weiss, 1999], les agents ayant à s'intégrer dans des groupes mixtes entrent dans la catégorie des agents assistants, ou, plus généralement, des agents d'interface. Selon Gene Ball et Jack Breese [Ball et Breese, 2000] (du projet *Persona*, chez Microsoft) une interface doit être :

- utile (capable de rendre un service et d'être compétente) ;
- utilisable (robuste et efficace dans la communication) ;
- confortable (tenir compte des attentes profondes en matière de conversation).

C'est dans cette optique que nous présenterons ici notre modèle d'agent : l'agent doit avant tout remplir sa fonction, et les principes que nous posons sur l'interaction ne doivent pas entraver son fonctionnement, mais au contraire l'améliorer en améliorant l'interaction. C'est en partie pour ces raisons que notre modèle d'interaction se veut le plus générique possible, le plus indépendant d'une tâche.

Le comportement de l'agent est un consensus entre sa fonctionnalité (principe fondamental du “quoi faire”) et sa personnalité (principe fondamental du “comment le faire”).

5.2 Intention conversationnelle et forme de surface

Nous avons dit que les messages n'étaient pas porteurs d'*un* sens, mais plutôt d'un *potentiel de sens* négociable et participant à la co-construction du focus conversationnel. Mais nous avons aussi posé comme hypothèse que les agents manipulent des sens précis. Il existe donc une étape de traitement qui va extraire d'un message perçu un potentiel de sens, puis le transformer en sens manipulable. Symétriquement, il existe une étape qui transformera *le* sens intentionné en un potentiel de sens, puis en message.

Notons que notre modèle d'agent (chapitres 5) n'intègre pas de capacités de communication autres que l'action et la perception, contrairement à de nombreux autres modèles qui disposent de capacités dédiées de communication d'agent à agent (voir par exemple le panorama d'Olivier Boissier [Boissier, 2001]). Ce choix est justifié par les constats que nous avons fait page 38.

Nous commencerons dans cette section par définir, du point de vue d'un modèle d'agent et non plus sous l'angle de la description d'une interaction, les mécanismes permettant le passage entre un potentiel de sens et un message : la formalisation et l'énonciation, en abordant le cas particulier de la multimodalité. La deuxième partie de cette section traitera, toujours du point de vue d'un modèle d'agent, de l'enrichissement contextuel des messages : l'interprétation et l'expression.

5.2.1 Potentiel de sens et message

Lorsqu'un message est perçu par l'agent, le premier traitement qui lui est appliqué est une *formalisation*, une traduction du message sous une forme qui soit manipulable par l'agent. Symétriquement, lorsque l'agent désire déposer un message dans l'environnement, le message doit passer par une phase d'*énonciation* qui le met sous une forme adaptée à l'environnement.

Nous laissons volontairement de côté le problème de la reconnaissance des messages dans l'environnement en considérant que les agents partagent un lexique (au moins en partie et pour chacune des modalités sur lesquelles ils souhaitent interagir) : mots, bien sûr, mais aussi, expressions faciales et corporelles, marqueurs prosodiques, ou même éléments de modalités n'ayant aucune relation avec des modalités physiques⁶. Préciser des lexiques (et donc, implicitement, des modalités) permet aussi d'éviter le problème de la reconnaissance des messages au sein de l'environnement. En effet, plus l'environnement est complexe et plus les capacités de perception et d'action des agents sont évoluées, plus il devient délicat distinguer dans l'environnement les traces susceptibles de participer à une interaction. Ces lexiques (et modalités) communs nous permettent ainsi de nous affranchir de ce que Jean-Paul Sansonnet appelle le SAP (Strong Alien Problem) [Sansonnet et Valencia, 2000].

De plus, nous supposons dans notre application que les sémantiques propres à chaque agents et associées aux éléments du lexique sont proches. Cette restriction n'est pas d'ordre théorique et n'a pour but que de nous affranchir de complexes mécanismes d'acquisition de la sémantique des symboles (voir par exemple les travaux de Luc Steels et de Frédéric Kaplan sur l'ancrage perceptif des symboles [Kaplan et Steels, 2000]). Elle nous permet de plus de limiter les effets de ce que Jean-Paul Sansonnet appelle le WAP (Weak Alien Problem).

a) Formalisation

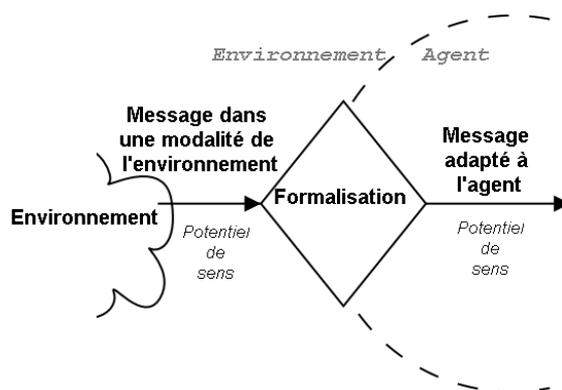


FIG. 5.1 – Le module de formalisation

Nous appelons *formalisation* l'étape de traduction d'un message de l'environnement en un formalisme adapté à l'agent (voir figure 5.1). Cette étape doit être la plus neutre possible, puisque c'est à partir du message obtenu (ou des messages obtenus, en cas d'ambiguïté) que commence l'étape d'interprétation. C'est le sens *littéral* qui

⁶Ainsi, des agents peuvent communiquer en s'échangeant des ACL.

doit apparaître ici. Cette étape, associée à celle d'intégration multimodale, constitue l'étape de *perception* des modèles perception → raisonnement → action.

Ce mécanisme n'est pas limité aux agents, puisque, intégré à une interface, il permet à un humain d'interagir à travers un environnement dans lequel les modalités ne lui sont pas directement compréhensibles (par exemple, un langage d'agent comme FIPA-ACL).

Les mécanismes de formalisation sont particulièrement dépendants des modalités à traiter et du format dans lequel elles doivent être formalisées, aussi nous ne proposons pas ici de méthode générale. Nous verrons dans le chapitre traitant de l'application (section 6.3.1) les méthodes que nous employons pour formaliser de la langue naturelle, des expressions corporelles et des déplacements.

Notons que dans certaines situations, la fonction de formalisation peut être extrêmement proche de la fonction identité. Par exemple, deux agents peuvent décider de s'échanger des messages sous une forme particulièrement voisine de celle qu'ils manipulent. Ainsi, des agents qui emploient KQML pour formaliser puis manipuler des messages reçus d'un humain (à la manière de REA, voir section 3.5.2) pourraient fort bien employer ce même langage pour communiquer d'agent à agent (il s'agit même de l'emploi principal de KQML). Dans une telle situation, la fonction de *formalisation* serait la fonction identité.

b) Énonciation

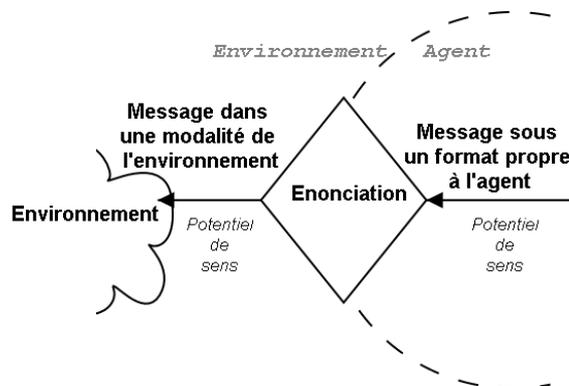


FIG. 5.2 – Le module d'énonciation

De manière symétrique à l'étape de formalisation, nous appelons *énonciation* (figure 5.2) l'étape de mise en forme d'un message pour le déposer dans l'environnement. De la même façon, associée à la répartition multimodale, cette étape constitue l'étape d'action des modèles perception → raisonnement → action. Toujours de manière symétrique à l'étape de formalisation, nous remarquons que nous ne pouvons pas donner ici de mécanisme général, mais que nous présenterons, dans la partie application, des mécanismes d'énonciation liés à la langue naturelle, aux expressions corporelles et aux déplacements. Enfin, nous faisons les mêmes remarques que précédemment sur le fait que, dans certains cas de communication, la fonction d'énonciation puisse être la fonction identité.

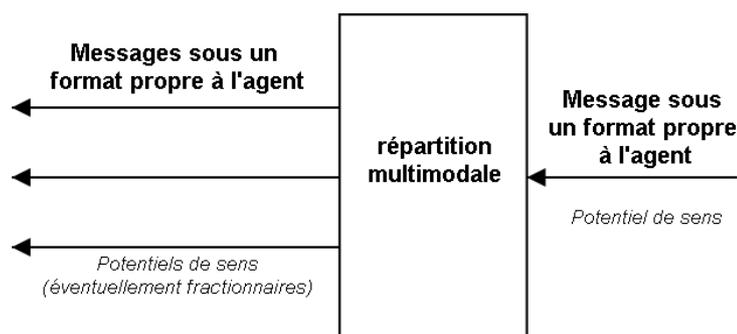


FIG. 5.3 – Le module de répartition multimodale

c) Répartition multi-modale

L'énonciation des messages est décomposée en sous-modules chacun dédié à une modalité donnée. L'étape de répartition multimodale (figure 5.3) consiste, comme son nom l'indique, à sélectionner les modalités dans lesquelles le message sera déposé dans l'environnement.

Pour choisir la ou les modalités de transmission du message nous nous intéressons seulement aux aspects *sémantiques*, laissant de côté, pour des raisons de simplicité, des critères pourtant fondamentaux comme la disponibilité des ressources (par exemple, la tentative d'emploi d'expressions corporelles alors que le corps est dédié à une autre tâche, comme porter un objet). Sur les aspects sémantiques de la répartition multimodale, nous notons deux points : il existe des modalités plus adaptées que d'autres à porter telle ou telle intention conversationnelle ; il existe des combinaisons de messages sur plusieurs modalités capables de porter une intention conversationnelle impossible à "porter" par une modalité unique.

Le choix de la modalité se base principalement sur le sens à exprimer et les capacités expressives des différentes modalités. Ces critères sont simples car ils permettent de se limiter à la question "*Telle modalité est-elle suffisamment expressive pour porter tel potentiel de sens ?*"

Lorsque aucune modalité n'est adaptée à porter un potentiel de sens donné, il est possible de combiner plusieurs modalités pour atteindre l'effet souhaité. Par exemple, si une modalité textuelle ne permet pas d'énoncer un performatif au degré de puissance trop élevé, il est possible d'accompagner le message textuel d'un message visuel, comme un froncement de sourcils ou un mouvement rythmique des mains. Pour ce faire, nous posons les concepts de *modalité dominante* et de *modalités auxiliaires*. La modalité dominante est celle qui portera l'essentiel du potentiel de sens, les modalités auxiliaires n'intervenant qu'en tant qu'*altération* de ce qui est porté par la modalité dominante.

Par ailleurs, puisque nous ne traitons que les variations sur une partie du performatif et non sur le contenu propositionnel, nous ne nous intéressons pas aux situations où le contenu propositionnel peut être réparti sur plusieurs messages, comme dans le cas du message oral "*Met ça là*" accompagné de gestes de désignation.

Le problème de la répartition multimodale revient à faire correspondre un potentiel de sens à *émittre* par l'agent à un potentiel de sens *qui peut être émis* (et

qui sera transformé en un message déposable dans l'environnement par le module d'énonciation).

Nous considérons donc un espace dans lequel sont positionnés les sens exprimables (qui peuvent être émis) et le sens que l'on cherche à exprimer (à émettre). Il suffit ensuite de chercher le sens exprimable le plus proche de celui que l'on cherche à exprimer. Pour simplifier cette étape, nous supposons que ces deux catégories de potentiel de sens ont la même représentation (nous proposons une représentation fortement inspirée des actes de langage, voir page 153). Cette hypothèse permet facilement de les projeter toutes deux dans le même espace.

L'espace dans lequel nous projetons le potentiel de sens à exprimer — et duquel nous extrayons le potentiel de sens exprimable — contient au moins autant de dimensions qu'il existe de degrés de liberté dans la spécification des actes de langage⁷. Mais selon les besoins, des dimensions ayant des significations sociales, émotionnelles ou même liées à une catégorie de tâches peuvent être ajoutées. Le calcul de la distance peut lui aussi être aussi complexe que le comportement de l'agent l'exige. Par exemple, on peut ne tolérer aucune variation suivant un axe donné, comme nous le faisons pour le but illocutoire tout en laissant une totale liberté sur un autre axe comme le degré de puissance.

De façon à rapprocher encore le potentiel de sens à exprimer et l'exprimable, nous introduisons des possibilités d'*altérations*, comme “poli”, “hésitant” ou “enthousiaste”⁸. Ces altérations servent à ajuster un performatif exprimable de façon à le rapprocher autant que possible de celui que l'agent souhaitait exprimer. Elles correspondent à des actions spécifiques à certaines modalités (expression faciale pour “enthousiaste”, choix lexical pour “poli”, prosodie pour “hésitation”, par exemple) et agissent en tant que message parallèle, sur une modalité auxiliaire (comme une expression faciale, parallèle à un message textuel), ou sur la modalité principale (comme les choix lexicaux et syntaxiques de la forme de politesse).

En plus de la répartition des messages sur différentes modalités, ce module est chargé de renseigner chacun des messages sur les contraintes de synchronicité qui les lient.

d) Intégration multimodale

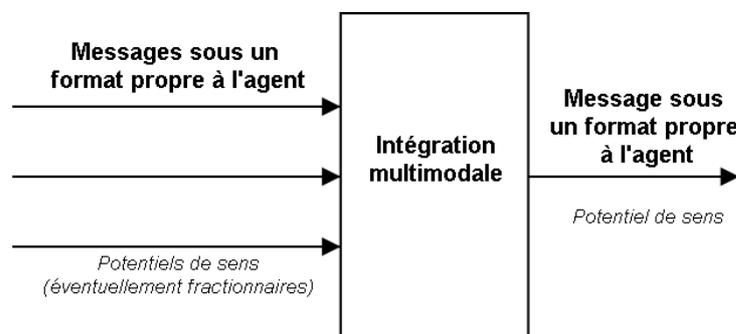


FIG. 5.4 – Le module d'intégration multimodale

⁷Vandervecken en spécifie six, voir page 85.

⁸Ces altérations ressemblent en partie à la notion de mode d'accomplissement de Daniel Vandervecken, mais sont plus vastes car elles agissent sur toutes les composantes de la force illocutoire, mode d'accomplissement compris.

Le module de formalisation des messages est décomposée en sous-modules, chacun dédié à une modalité donnée. L'étape d'intégration multimodale (figure 5.4) consiste, comme son nom l'indique, à synthétiser en un seul message (interne) plusieurs messages issus de modalités différentes (externes) et susceptibles de participer à un même potentiel de sens.

Le problème principal de cette étape est de savoir quand plusieurs messages participent à un sens commun et quand ils doivent être considérés indépendamment les uns des autres. Et une fois ce problème résolu, le problème suivant est de savoir comment fusionner des messages considérés comme participant à un message unique⁹.

Nous considérons deux situations de complémentarité de messages. Dans le premier cas, ce sont essentiellement les performatifs qui sont touchés, tandis que dans le second les contenus propositionnels des messages se complètent d'une modalité à l'autre. Puisque nous ne nous intéressons qu'aux variations sur les performatifs, nous laisserons de côté le second point. Nous noterons juste que cette situation peut être détectée en cas de manque patent d'une information dans un message d'une modalité donnée (comme pour les mouvements déictiques ou propositionnels, voir page 65), ou bien lorsque l'on constate qu'une modalité contient une information complémentaire (comme pour les mouvements iconiques ou, à nouveau, propositionnels, voir page 65).

Le critère fondamental pour décider de l'influence d'une modalité sur une autre est le *timing* de ces messages les uns par rapport aux autres. Malheureusement, les contraintes temporelles sont très variables selon les modalités, selon les altérations apportées par les messages, etc. Ainsi, nous reprenons ici la notion de *modalité dominante* introduite lors de la présentation de la phase de *répartition multimodale*. À chaque nouveau message, le système doit répondre à la question "*Si ce message est un message dominant, quelle est la fenêtre temporelle (aussi bien dans le passé que dans le futur) pendant laquelle son sens est susceptible d'être altéré par d'autres messages et dans quelles modalités ?*". Parfois, la réponse est simple : le message reçu ne peut pas être considéré comme un message dominant. En effet, certaines expressions corporelles ou bien la prosodie ne peuvent (en général) être considérées que comme des messages complémentaires. Dans le cas où *il est possible* que ce message soit dominant, son envoi pour traitement dans le module suivant (*interprétation*, en l'occurrence) apparaît lorsque :

- soit ce message se retrouve hors du champ de la question (sa propre fenêtre temporelle est dépassée et il ne peut plus être pris comme altération par les autres messages), il est donc transmis aux modules suivants en tant que tel ;
- soit ce message a trouvé une séquence de messages complémentaires cohérente et il est altéré en conséquence avant d'être transmis aux modules suivants.

Il faut toutefois voir ces règles comme générales et susceptibles d'être modifiées suivant les situations conversationnelles. Par exemple, chez l'humain certaines modalités sont souvent prioritaires par rapport à d'autres, comme le langage parlé face aux expressions corporelles.

⁹Chez les humains, l'exemple le plus flagrant d'intégration multimodale est rendu par l'effet "McGurk", voir note 26 page 67.

5.2.2 Enrichissement contextuel

L'enrichissement contextuel correspond à l'un des deux points centraux de cette thèse : le fait que les messages sont porteurs d'information sur leur contexte d'énonciation. Cette partie du modèle d'agent que nous proposons correspond directement, dans le modèle d'interaction, aux principes d'interprétation et d'expression (voir section 4.4.1). Cette section montre comment ces principes, présentés précédemment du point de vue de l'interaction, sont adaptés de façon à s'intégrer dans un modèle d'agent.

a) Interprétation

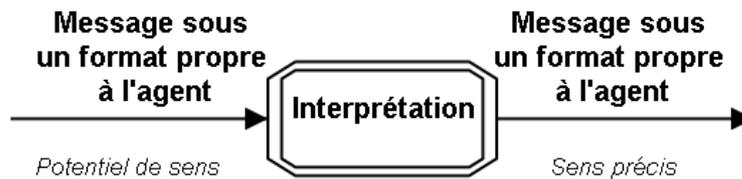


FIG. 5.5 – Le module d'interprétation

Le processus d'interprétation (figure 5.5) remplit deux rôles : premièrement, *modifier* le message en fonction du contexte d'énonciation ; deuxièmement, *extraire* d'un message reçu (ou plus généralement, perçu) des informations sur le producteur de ce message¹⁰.

L'isolement explicite de l'étape d'interprétation entre celle de perception et celle de raisonnement permet à ces dernières de conserver leurs fonctionnalités classiques.

L'interprétation peut être vue comme la confrontation du message aux connaissances qu'a l'agent sur le contexte d'énonciation du message. L'exemple donné par la table 4.4 précédemment vue, montre comment les connaissances sur les relations sociales peuvent altérer le degré de puissance, et, par là même, la force illocutoire de l'énoncé. Mais l'interprétation peut aussi modifier en profondeur le message : dans le cadre de l'application présentée dans le chapitre 6, l'agent possède un corps et est plongé dans un environnement spatialisé. Lors des déplacements des autres agents le module d'interprétation convertit les coordonnées *absolues* (par rapport à un repère lié à l'environnement) en coordonnées *relatives* (centrées sur l'agent). En l'occurrence, il s'agit de coordonnées polaires dont l'origine du repère est l'agent et l'axe de référence la direction de son regard, donnant directement accès à des informations telles que : "*l'interlocuteur est-il proche de moi ?*" ou "*l'interlocuteur est-il en face de moi ?*". Dans cette situation, l'interprétation du message permet surtout d'extraire du message les informations pertinentes pour l'agent et de les mettre sous une forme qui lui soit plus personnelle. Là où la position d'un interlocuteur n'est qu'une information de l'environnement, la distance entre cet interlocuteur et l'agent commence à faire sens car elle existe *en référence* à l'agent. Ainsi, la mise en contexte transforme un simple *percept* de l'environnement en une *représentation* en terme

¹⁰La coexistence de ces deux fonctionnalités au sein d'un même module doit entraîner une grande précaution : le système peut rapidement devenir autoréférentiel si les paramètres extraits du message servent à modifier ce même message de façon à ce qu'il soit plus conforme aux paramètres qui viennent à peine d'en être extrait !

d'objets, de monde, d'événements, etc., qui plus est *enrichie* des connaissances de l'agent.

Notons que l'étape d'interprétation peut contenir des éléments extrêmement génériques et propres à un grand nombre de rôles possibles, mais qu'il *faut* laisser la possibilité au rôle de proposer ses propres traitements dans cette étape. En anticipant à nouveau sur la description de l'application du chapitre 6, nous pouvons dire que lors du déplacement d'un interlocuteur, ses coordonnées absolues sont transformées en coordonnées relatives, mais aussi que d'autres informations sont extraites de sa position, comme par exemple, sa position relativement à des zones spécifiques (rayon des films, des disques, sortie du magasin, etc.) ou relativement au territoire (au sens de Goffman) de l'agent.

b) Expression



FIG. 5.6 – Le module d'expression

Le processus d'expression (figure 5.6) a pour rôle principal de contextualiser un message à émettre. Lors de son expression, l'intention communicationnelle de l'agent (issue de l'étape de raisonnement) est mise en forme de façon à être adaptée aux interlocuteurs à qui elle est destinée.

L'isolement explicite de l'étape d'expression entre celle de raisonnement et celle d'action permet à ces dernières de conserver leurs fonctionnalités classiques.

L'expression peut avoir comme effet l'enrichissement ou l'appauvrissement des informations initialement prévues dans le message. L'exemple de la table 4.5 page 100 montre un enrichissement du message : la force illocutoire initialement choisie (AFFIRMER) est modifiée de façon à refléter une information supplémentaire, là la confiance en l'information transmise. Inversement, on peut imaginer une situation où la relation sociale liant les interlocuteurs est si évidente qu'il n'est nul besoin d'employer le performatif ORDONNER, car tout directif serait compris comme un ordre. Du point de vue de l'auditeur, c'est ce qui se passe dans l'exemple du tableau 4.4 : on peut *supposer* que l'intention de la mère est d'ORDONNER, mais qu'elle sait que la relation qui la lie à son fils est suffisante pour qu'une force illocutoire ayant un degré de puissance inférieur (comme SUGGÉRER) soit *interprétée* comme un ordre.

5.3 Enchaînements

Les sections précédentes présentaient différents modules du modèle d'agent que nous proposons. Cette section s'en démarque un peu car elle ne présente pas un module en soi, mais un principe de fonctionnement. De plus, ce principe n'est pas intégré dans une partie générique de l'agent, mais dans la partie dédiée à la tâche. En effet, nous présentons ici une manière pour l'agent de générer une dynamique

conversationnelle telle qu'elle a été présentée dans la section 4.5 du chapitre traitant des principes du dialogue.

Deux processus participent à la dynamique de la conversation : la (co-)construction du focus conversationnel et la prise en compte d'attentes.

Contrairement aux autres points du modèle de l'interaction que nous avons précédemment développés, ces deux processus peuvent être importés tels quels dans le modèle de l'agent. Les deux sous-sections suivantes sont donc particulièrement succinctes.

5.3.1 Focus conversationnel

Le focus conversationnel contient l'ensemble des données non-contestées qui ont été échangées, ajoutées aux connaissances supposées communes. Deux catégories de messages y ont directement trait : la tentative de modification du focus et le commentaire sur le focus.

Au niveau de l'agent, le focus est situé dans la partie raisonnement. Les mécanismes qui y sont associés sont ceux décrits dans la section 4.5.3. L'agent consulte son focus et définit si son contenu est satisfaisant ; cette satisfaction peut être calculée en fonction de critères intrinsèques (critère de consistance du focus par exemple) ou extrinsèques (conflit entre les connaissances de l'agent et le contenu du focus par exemple). Selon le résultat de cette évaluation, l'agent peut décider de réagir en utilisant l'un des deux types de messages déjà introduits : tentative de modification du focus ou commentaire sur le focus.

On peut considérer le focus comme étant en partie la mémoire de travail de l'agent car il contient les données pertinentes à la gestion de l'instant précis de la conversation en cours.

Le focus en lui-même n'apporte pas de contrainte sur les *échanges*, mais uniquement sur un objectif à atteindre.

5.3.2 Prise en compte d'attentes

Les attentes correspondent à des événements possibles et précisent les actions à entreprendre si ces événements surviennent.

Au niveau de l'agent, les attentes sont gérées dans la partie raisonnement. Les mécanismes qui y sont associés sont ceux décrits dans la section 4.5.4. Lors de chacune de ses actions, l'agent peut anticiper sur les conséquences de son action et mettre en place des attentes susceptibles de prendre en charge ces conséquences. Symétriquement, à chaque événement perçu l'agent consulte la liste d'attentes valides au moment de la perception et si l'une d'entre elle peut être déclenchée, s'engage dans l'accomplissement de l'action spécifiée par l'attente.

On peut considérer les attentes comme étant des éléments de comportement réactif de l'agent.

Une attente n'apporte pas de contraintes sur un objectif général à la conversation (encore qu'un ensemble d'attentes puisse être considéré comme tel), mais uniquement à l'échelle d'un échange.

5.4 Synthèse : une architecture possible

Nous venons de présenter les différents éléments d'un modèle d'agent. Dans cette section, nous précisons maintenant comment il est possible d'articuler ces différents

éléments pour former une architecture *complète* d'agent telle qu'elle est représentée page 152. Nous commencerons par présenter le fonctionnement d'ensemble d'un agent qui serait basé sur une telle architecture (comme l'agent que nous présenterons dans le chapitre 6). Nous décrirons alors les modules qui composent cette architecture, de manière généralement sommaire, car ces modules ne font, dans l'ensemble, que matérialiser des concepts soit déjà développés dans les pages précédentes, soit courants dans le domaine des architectures d'agents. La dernière section fera une présentation croisée des données partagées et des modules qui les partagent.

5.4.1 Fonctionnement d'ensemble

L'architecture que nous proposons est une architecture classique perception \rightarrow raisonnement \rightarrow action. Nous y ajoutons des étapes d'*interprétation* et d'*expression* au statut ambivalent, car relevant à la fois de la perception et du raisonnement pour la première et de l'action et du raisonnement pour la seconde. Le traitement du message suit donc une séquence perception \rightarrow interprétation \rightarrow raisonnement \rightarrow expression \rightarrow action.

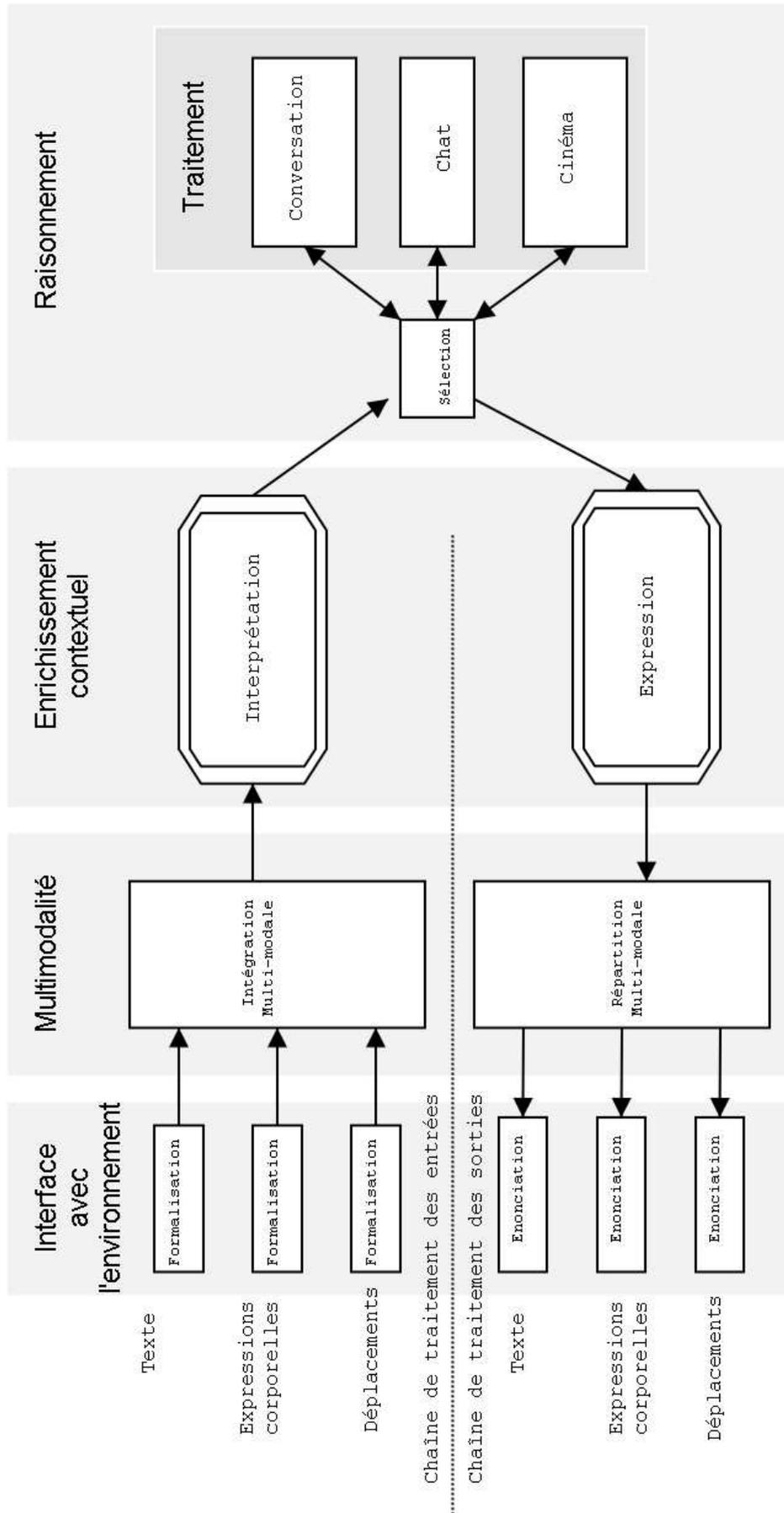


FIG. 5.7 – Architecture générale

a) Format des messages internes

Pendant les différentes étapes de traitement, des données sont échangées entre les modules. Ces données, reflets du message d'entrée ou de l'intention communicative de l'agent, sont les *messages internes*.

Puisque la théorie des actes de langages est particulièrement adaptée à la représentation d'énoncés d'origine langagière, et que Daniel Vanderveken affirme que "*Toute pensée conçue par un sujet humain est en principe exprimable par les moyens de son langage lors de l'accomplissement d'actes de discours.*" ([Vanderveken, 1988], préface), nous nous baserons sur ce formalisme pour nos messages internes. Toutefois, la modularisation de notre modèle ne nous permet pas de faire de suppositions sur les modalités employées. Nous avons donc donné à ces messages internes un statut d'*acte de communication* plus que d'*acte de langage*.

Étape	Entrée/Sortie ?	Nature du message	Format du message
Formalisation	e	Message de l'environnement	"naturel"
	s	Potentiel de sens éventuellement réparti sur plusieurs messages	interne
Intégration	e	Potentiel de sens (interprétable)	interne
	s		
Interprétation	e	Sens précis	interne
	s		
Sélection & Traitement	e	Sens précis	interne
	s		
Expression	e	Potentiel de sens (interprétable)	interne
	s		
Répartition	e	Potentiel de sens éventuellement réparti sur plusieurs messages	interne
	s		
Énonciation	e	Message de l'environnement	"naturel"
	s		

TAB. 5.2 – Nature des messages échangés

Il nous semble intéressant de conserver le même format de représentation du message tout au long de son traitement, bien qu'il ne représente pas exactement la même chose en tout point de la chaîne de traitement (voir le tableau 5.2) :

- en sortie des modules de formalisation et en entrée des modules d'énonciation, il s'agit d'un sens potentiel, éventuellement fractionnaire (car à considérer conjointement aux messages d'autres modalités) ;

- en sortie du module d'intégration multi-modale et en entrée du module de répartition multi-modale, il s'agit d'un potentiel de sens ;
- en sortie du module d'interprétation et en entrée du module d'expression, il s'agit d'un sens défini de manière unique.

Chaque module peut lever ou apporter des ambiguïtés. Pour prendre l'exemple de la langue naturelle, le module de perception peut se retrouver confronté à des ambiguïtés lexicales qui pourront être levées en phase d'interprétation ou de raisonnement. Mais l'étape d'interprétation peut elle-même apporter ses propres ambiguïtés, la résolution d'anaphores y étant par exemple très sujette. Si des ambiguïtés apparaissent, les modules produiront un message interne pour chaque ambiguïté. Ainsi, les messages internes ne sont pas porteurs d'ambiguïté.

b) Traitement d'un message

Cette section ne présente que le fonctionnement d'ensemble ; la description des différentes étapes se veut donc volontairement brève.

“Entrées” La *perception* traduit un message de l'environnement pour le mettre sous un format manipulable par l'agent. L'étape suivante, d'*interprétation*, enrichit les messages avec l'ensemble du contexte conversationnel. Les messages qui arrivent en entrée de l'étape de raisonnement ont donc déjà subi plusieurs traitements.

Nous proposons de segmenter la partie raisonnement en deux sous-ensembles : une *sélection* et des modules de *traitement* proprement dits, spécifiques.

Sélection La partie sélection est en charge de deux sélections conjointes :

- la sélection du message traité. Un seul message en entrée, s'il est ambigu, peut entraîner l'arrivée de plusieurs messages au niveau du raisonnement, chacun correspondant à un sens possible de l'énoncé initial.
- La sélection du module de traitement, obtenue par la mise en compétition des modules.

Pour effectuer ces sélections, le module lance un appel d'offre à destination des modules de traitement. À la suite de cet appel d'offre, le module sélection décide du couple message-module¹¹.

Traitements Le cœur de la partie raisonnement, qui, finalement, prend les décisions d'action, est séparé en modules spécialisés (l'agent que nous avons réalisé en comprend trois : conversation, cinéma et *chat* ; pour une présentation détaillée, voir page 193 et suivantes). Chacun de ces modules, que nous avons baptisé modules de *traitement*, a deux fonctions.

La première permet de répondre à l'appel d'offre du module de sélection. Lorsqu'on lui présente un message, le module estime (1) sa compétence à traiter le message (un message peut être complètement hors du domaine de compétences du

¹¹Nous verrons dans le chapitre application (section 6.4.2) que nous avons mis en place un mécanisme particulièrement simple : lors de l'appel d'offre, chaque module renvoie *une* valeur correspondant à la fois à sa compétence à traiter le message et à son intérêt à être chargé du traitement de ce message. La sélection se limite alors à désigner le module ayant déclaré la valeur la plus haute. Un mécanisme d'attribution plus complexe (par exemple basé sur la négociation entre les modules) serait certainement nécessaire pour le traitement de situations conversationnelles moins contraintes.

module, par exemple, un message de demande d'information sur un film vis-à-vis d'un module chargé de la gestion des infractions sociales) et (2) à quel point il est important que le message lui soit confié (par exemple, le message peut être une réponse à une question déjà issue de ce même module).

La deuxième fonction du module concerne le traitement proprement dit. Pour ce traitement, il n'y a aucune contrainte quant aux méthodes employées. Nous verrons par exemple dans le chapitre décrivant notre application que le module traitant du cinéma est construit autour d'un focus conversationnel, d'attentes et d'une base de données, que le module dédié à la gestion de la conversation se limite à une série de règles et que le module *chat* ne fait que des associations lexicales.

Notons que l'architecture que nous proposons permet aussi un *comportement* (donc, selon un point de vue extérieur) proactif de l'agent. Les modules peuvent tout à fait décider d'une action sans qu'elle ne soit une *réaction* à un événement. De plus, l'événement déclenchant peut tout à fait être un événement interne (et donc imperceptible pour un observateur extérieur), principalement un changement d'état de l'agent ou le déclenchement d'une temporisation. Enfin, notons qu'une réaction de l'agent à un événement non remarqué par ailleurs peut donner l'illusion que la *décision d'action* a été proactive. Ainsi, notre agent de recommandation de films initie "spontanément" une interaction lorsque l'utilisateur s'approche du rayon des films.

De même, un module, même s'il se déclare compétent dans le traitement de tel ou tel message, et même s'il est important pour lui de le traiter, peut tout à fait ne pas décider d'action. Pour faire un retour sur les sciences de la communication, ne rien dire *est* un acte de communication en soi (revoir page 119, où Alan Leigh synthétise : "*You can't not communicate*").

“Sorties” Le message intentionné (c'est-à-dire le message sortant de l'étape de raisonnement) est confié au module d'*expression* qui adapte ce message au contexte conversationnel. Ce message, alors porteur d'un potentiel de sens, est transmis à l'étape d'*action*, qui le traduit en un format adapté à l'environnement.

5.4.2 Les modules et leurs traitements

Dans cette sous-section, nous rassemblons les traitements fondamentaux qui sont successivement appliqués aux messages perçus. L'ordre dans lequel nous allons présenter ces différents modules correspond à l'ordre dans lequel les messages les traversent lors de leur traitement. Tout d'abord, comme nous l'avons vu sur la figure 5.7 le message passe par un module de *formalisation* (un pour chaque modalité de perception), puis, l'ensemble des messages issus des différentes modalités perceptibles passent dans le modules d'*intégration multi-modale*. Le message unique qui en est issu est ensuite mis en contexte dans le module d'*interprétation*. Le message déclenche ensuite la *sélection* du module de traitement le plus adapté, puis le *traitement* en soi. En cas de message à émettre, après avoir quitté le module de *traitement*, le message est confronté au contexte dans le module d'*expression*, puis le module de *répartition multi-modale* distribue le message suivant les modalités les plus appropriées, enfin, les modules d'*énonciation* (un pour chaque modalité d'action) mettent les messages sous une forme adaptée à l'environnement.

a) Formalisation

Les étapes de formalisation consistent en la transformation d'un message d'une modalité donnée vers un ou plusieurs (si des ambiguïtés apparaissent) messages en un formalisme interne. Chaque modalité disponible en perception de l'environnement dispose de son propre module de formalisation (figure 5.1).

Ce module est très fortement lié au module d'intégration multi-modale car les phénomènes de synergie entre différentes modalités sont complexes. Le simple ajout d'un module permettant la prise en compte d'une nouvelle modalité ne pourra se faire sans modification du module d'intégration multimodale que si l'on souhaite ne traiter que les cas les plus simples d'influence entre les modalités déjà présentes et la nouvelle intégrée.

b) Intégration multi-modale

Le module d'intégration multi-modale permet de transformer une série de messages issus de modalités différentes en un seul message représentant un sens potentiel (figure 5.4). Il est possible d'obtenir en sortie de ce module une série de messages, si l'association des messages est possible de plusieurs manières ou s'il reste des ambiguïtés qui n'ont pu être levées à cette étape. Toutefois, chacun de ces messages représente un sens potentiel complet et n'est plus lié à une modalité.

Nous considérons, pour l'intégration de messages issus de différentes modalités, qu'il existe un message dominant, qui contient l'essentiel du sens. Les messages issus des autres modalités ne sont là que pour modifier ce message (sourire atténuant une phrase légèrement agressive par exemple) ou le compléter (déictiques d'un énoncé langagier à instancier par un objet pointé par exemple).

c) Interprétation

Le module d'interprétation remplit deux rôles : adaptation contextuelle du message et acquisition des données sur les interlocuteurs (figure 5.5).

Dans un premier temps, le module d'interprétation confronte le message à traiter avec les connaissances dont il dispose sur les messages précédents (pour résoudre les anaphores par exemple), sur l'émetteur du message (pour expliciter une relation sociale par exemple, comme page 99), sur le monde (pour estimer la plausibilité d'une information par exemple, comme page 100), etc. Il tire de ces confrontations un message porteur d'un sens unique. C'est en effet en cette étape que se passe la réduction du potentiel de sens.

Dans un second temps, le module d'interprétation identifie dans chaque message de l'environnement les éléments susceptibles d'influer sur le profil de l'émetteur du message et met à jour ce profil.

Si le message initial est porteur d'ambiguïtés que le module ne peut résoudre¹², il transmettra tous les sens possibles (ou probables), de façon à ce que, en sortie de ce module, aucun message ne soit porteur d'ambiguïté.

¹²Même en restreignant le contexte à la recommandation de films, un énoncé comme "Clint Eastwood" est ambigu : il est à la fois acteur et réalisateur. Dans certains cas, cette ambiguïté peut n'être levée que par la prise en compte de l'enchaînement de la conversation (typiquement, il s'agit de la réponse à une question), qui est du ressort des modules de traitement.

d) Sélection

Le module de sélection est chargé de transmettre le message au module de traitement approprié. Il est donc en charge de deux sélections :

- la sélection du message parmi une (éventuelle) liste d’interprétations possibles ;
- la sélection du module qui sera chargé du traitement du message.

e) Traitement (modules de raisonnement)

Si l’on excepte les modules d’interprétation et d’expression —dont le statut est intermédiaire, entre perception et raisonnement— le raisonnement de l’agent est accompli par un ensemble de modules de traitement spécialisés. Certains de ces modules s’occupent des situations à contenu purement conversationnel (gestion des tours de parole ou de l’incompréhension), situations très dissociables des tâches de l’agent, donc particulièrement génériques et susceptibles d’être intégrées dans différents agents avec peu ou pas d’adaptation. Les autres situations liées aux tâches de l’agent sont gérées par des modules plus spécifiques.

f) Expression

Le module d’expression remplit une fonctionnalité symétrique du module d’interprétation, excepté la mise à jour des connaissances sur l’interlocuteur. Cette étape convertit un sens intentionné en un potentiel de sens plus adapté au destinataire du message (figure 5.6).

g) Répartition multi-modale

Ce module décide, en fonction du message qui lui est présenté, des modalités à employer (figure 5.3). Principalement, le module décide d’une modalité dominante pour le message et emploie des modalités complémentaires pour altérer ou compléter le message. Ainsi, pour exprimer une salutation, l’agent pourra simplement dire “bonjour”, accompagné d’un sourire et d’un mouvement de la tête s’il souhaite être particulièrement amical, ou au contraire, d’un regard “froid” s’il souhaite marquer l’aspect purement conventionnel de la salutation. De la même façon, ce module gère les déictiques multi-modaux, comme lors de la désignation d’un objet qui est aussi référencé dans un message textuel.

h) Enonciation

De manière symétrique au module de formalisation, le module d’énonciation transforme un message dans une modalité de l’environnement (figure 5.2). Chaque modalité de l’environnement dans laquelle l’agent désire produire des messages (c’est-à-dire agir) nécessite l’existence d’un module d’énonciation dédié.

De la même façon que nous l’avons fait remarquer pour les modules de formalisation et d’intégration multi-modale, le problème de la synchronisation entre les différentes modalités est loin d’être trivial. L’architecture que nous proposons ne peut le résoudre qu’en admettant une forte interconnexion entre les modules d’énonciation et celui de répartition ainsi qu’entre les modules d’énonciations entre eux.

5.4.3 Les données partagées

Dans cette sous-section, nous mettons en relation les données fondamentales que requiert notre modèle d'agent avec les processus dans lesquels elles interviennent (voir la table 5.3). Les modules *traitement* tiennent une place un peu particulière, car certains, pouvant être génériques, sont liés à la gestion de la conversation (gestion des ruptures par exemple), tandis que d'autres sont spécifiques et liés à la tâche. Nous ne pouvons donc nous permettre de supposer qu'ils nécessiteront ou non d'avoir accès à telle ou telle donnée. Nous les considérons donc comme ayant un accès complet à ces données, et nous n'en parlerons donc dans les sous-sections qui suivent que lorsque la catégorie de données et les modules de traitement présentent une relation singulière.

	Formalisation	Intégration multi-modale	Interprétation	Sélection	Traitement	Expression	Segmentation multi-modale	Enonciation	Voir page...
Focus	0	0	r	r	r/w	0	0	0	158
Attentes	(r)	0	r	r	r/w	0	0	0	159
Personnalité	0	0	r	(r)	r/w	r	0	(r)	159
Conn. sur les autres	0	0	r/w	(r)	r/w	r	0	(r)	160
Historique	w	r	w	w	r	w	0	w	160

Légende :

0 : Aucun lien entre le module et la donnée

r : Le module a besoin de consulter la donnée (lecture)

w : Le module renseigne la donnée (écriture)

r/w : Le module lit et écrit la donnée

(r) : Le module pourrait "raisonnablement" avoir besoin de consulter la donnée

TAB. 5.3 – Accès des modules aux données

a) Focus conversationnel

Le focus conversationnel va contenir les divers éléments qui permettent de caractériser l'objet de négociation. Ces éléments peuvent être issus directement de la conversation (éléments clairement exprimés par les interlocuteurs), en être déduits (selon d'éventuelles corrélations entre éléments) ou être supposés (paramètres instanciés avec une valeur par défaut).

Le focus conversationnel contient aussi des informations sur l'état de la conversation, en partie à la manière de REA [Cassell, 2000; Cassell et Bickmore, 2000]. Il conserve des informations sur la conversation en tant qu'objet partagé : qui a la parole ?, souhaite-t-il la garder ou la passer ?, l'un des interlocuteurs l'a-t-il réclamée ?,

etc. Il contient aussi des informations sur la gestion interne de la conversation par l'agent : quel module a-t-il eu la main lors de la dernière énonciation ?, l'agent est-il en phase contestation du focus (gestion d'une incompréhension) ?, etc.

Le focus intervient dans trois modules :

- Interprétation : lors de la phase d'interprétation, la connaissance du focus permet de lever des ambiguïtés.
- Sélection : le contenu du focus participe à la sélection du module le plus adapté au traitement, car c'est lui qui tient trace des changements d'interlocuteurs ou de l'arrivée d'informations en conflit avec celles déjà connues, situations nécessitant l'aiguillage du message vers des modules spécifiques.
- Traitement : ce sont les modules de traitement qui gèrent la dynamique de la conversation, et, en tant que tels, ils nécessitent un contrôle total sur les informations contenues dans le focus conversationnel.

b) Attentes

Les attentes correspondent à certains événements prévus par le système, en fonction essentiellement des dernières actions de celui-ci. Aux attentes sont associées des indications sur les actions à entreprendre au cas où l'attente se verrait confirmée par les événements.

Ces attentes influent (ou peuvent influencer) sur le comportement de l'agent au sein de quatre modules :

- (Formalisation) : puisqu'elles spécifient les messages auxquels l'agent doit s'attendre, les attentes peuvent agir sur la méthode de formalisation.
- Interprétation : dans ce module, les attentes peuvent guider l'interprétation en levant certaines ambiguïtés : une interprétation d'un message correspondant à une attente sera considéré comme plus "plausible".
- Sélection : lorsqu'une attente se confirme, le module ayant spécifié l'attente doit être favorisée.
- Traitement : c'est la partie traitement de l'agent qui, lorsqu'elle décide d'une action, précise simultanément les (éventuelles) attentes à mettre en place. Inversement, la partie traitement tient compte des attentes précédentes (et de leurs actions associées) lorsqu'elle a à décider d'une action.

c) Personnalité

De la personnalité, nous nous limitons à la *conscience sociale*, dont les effets sont liés (et limités) aux interlocuteurs participant à l'interaction.

La personnalité est (ou peut être) présente en quatre modules (plus les modules de traitement) :

- Interprétation : à cette étape, la personnalité (et d'autant plus la conscience sociale) vient enrichir le message.
- (Sélection) : on pourrait imaginer que la personnalité de l'agent influe sur le choix des actions à entreprendre, et donc, sur la sélection du module de traitement, mais ce point de vue va à l'encontre de notre distinction entre la fonctionnalité de l'agent (son rôle, ce qu'il fait) et sa personnalité (la façon dont il agit).
- Expression : presque par définition, la phase d'expression, chargée d'enrichir le message, se doit de prendre en compte la personnalité de l'agent.
- (Énonciation) : le choix de telle ou telle paraphrase (ou de tel ou tel geste ayant même valeur sémantique) peut prendre en compte la personnalité, bien

qu'en toute rigueur, le message à énoncer en ait déjà été chargé lors de l'étape d'expression. On peut rendre compte ici de fonctions spécifiquement liées à la modalité du message, comme des tics verbaux, ou des contraintes motrices.

d) Historique

La manière la plus simple de voir l'historique est de voir en lui une "simple" liste contenant les messages issus des modules *formalisation* et *interprétation* en entrée, et *expression* et *énonciation* en sortie. Il est en fait plus que cela puisque nous y intégrons tout ce qui peut avoir trait à la conservation d'informations.

Nous parlerons par commodité de l'historique comme d'une donnée unique alors qu'il serait bien plus pertinent de le considérer comme réparti parmi les différents modules, chaque module conservant les données nécessaires aux traitements dont il a la charge.

L'historique est une donnée susceptible d'être manipulée par la quasi-totalité des modules :

- Formalisation : à cette étape, il est nécessaire de stocker le message perçu et sa (ou ses) formalisation(s).
- Intégration multi-modale : lors de cette étape, des messages appartenant à plusieurs modalités sont fusionnés. L'historique sert de tampon pour permettre la prise en compte conjointe de messages dont la perception est étalée dans le temps.
- Interprétation : le module d'interprétation doit stocker dans l'historique les interprétations faites des messages perçus. Le module peut aussi consulter le focus pour résoudre certaines anaphores.
- Sélection : lors de la phase de sélection, on note le message qui, parmi toutes les interprétations encore éventuellement possibles à ce stade, a été sélectionné pour être traité.
- Traitement : bien que notre modèle d'agent ne suppose rien sur la tâche de l'agent (et donc sur le contenu des modules de traitement), il semble raisonnable que certains d'entre eux aient à consulter les interactions précédentes pour décider de leurs actions. De plus, parmi ces modules de traitement, il en existe certains spécifiquement dédiés à la gestion de la conversation, comme la gestion des incompréhensions. L'un des exemples de la page 140 montre comment ce genre de module peut s'appuyer sur l'historique.
- Expression, énonciation : pour les mêmes raisons que les modules de formalisation et d'interprétation, ces modules doivent pouvoir écrire dans l'historique. De la même façon, ces modules nécessitent un accès en lecture de façon à pouvoir générer des anaphores qui, au moins dans le cadre de la langue naturelle, peuvent aider à "alléger" un énoncé.

e) Connaissances sur les autres

Par *connaissances sur les autres* nous considérons l'ensemble des informations sur ses interlocuteurs dont l'agent dispose. Entre autres, ces informations comprennent un modèle des interlocuteurs, leur rôle et statut, ainsi que leurs stratégies conversationnelles. Dans le cadre de la gestion de la communication, ces informations servent principalement lors des phases d'enrichissement contextuel, mais aussi lors de la gestion de la dynamique de la conversation, si l'agent reconnaît les stratégies mises en

œuvre par ses interlocuteurs et qu'il décide de collaborer¹³.

Outre la phase de traitement, les connaissances sur les autres apparaissent (ou peuvent apparaître) dans quatre modules :

- Interprétation : à cette étape, les connaissances sur l'utilisateur sont mises à jour en fonction de connaissances extraites (ou inférées) du message. Cette manipulation se fait à la lumière des connaissances précédemment disponibles.
- (Sélection) : la sélection de l'action à entreprendre pourrait dépendre de l'interlocuteur autant que du message, mais comme nous en avons déjà fait la remarque pour la personnalité, une telle dépendance se ferait aux dépens de la distinction que nous souhaitons entre fonctionnalité et personnalité.
- Expression : c'est principalement dans cette étape que les connaissances sur les autres ont leur intérêt, puisque c'est là que le message est adapté à ses destinataires.
- (Énonciation) : enfin, le choix du message lui-même peut être influencé par ces connaissances, même si nous faisons ici la même remarque que pour la personnalité : en toute rigueur, le message a déjà été chargé de ces connaissances en phase d'expression.

5.5 Conclusion

Pour finir ce chapitre, nous revenons sur l'importance de la cohérence dans le comportement de l'agent, particulièrement sur ses compétences en langues naturelles.

Les interfaces conversationnelles ont ceci de particulier que plus le système est efficace en production, plus l'humain est exigeant, car il croit à l'humanité du système. Des expériences montrent par exemple que plus le système est "fruste" dans ses énoncés, plus les utilisateurs ont tendance à parler "petit-nègre", alors qu'inversement, un système plus volubile amènera des réponses complètes, accompagnées de formules de politesse (bonjour, s'il vous plaît, etc.) Malheureusement, il est très simple de faire produire à un agent des énoncés complexes, dans la mesure où ces énoncés peuvent tout simplement être préenregistrés. Il faut alors être attentif à ce qu'il n'existe pas un déséquilibre trop grand entre les compétences en production et en analyse de l'agent, ce qui compliquerait la mise en place, chez les interlocuteurs d'un modèle des compétences de l'agent.

Cependant, les capacités d'adaptation de l'être humain sont telles que l'on peut en général compter sur lui pour s'adapter aux compétences de l'agent. "*Certaines expériences ergonomiques (Morel 85)*¹⁴ tendent à montrer que l'interlocuteur humain s'adapte assez naturellement au niveau de langue de la machine (lexique, structure linguistique des énoncés)." [Sabah, 1989]. De plus, on remarque couramment l'« effet Eliza » dans les interactions entre humains et systèmes automatiques : les humains ont tendance attribuer aux machines plus d'intelligence qu'elles n'en ont, et à chercher dans le *comportement* de l'agent des processus typiquement humains (*intentions* particulièrement) qui n'ont aucune correspondance dans son *modèle interne*.

¹³Poussée à son extrême, cette situation est celle de l'interaction à base de protocoles, où les connaissances sur l'autre comprennent les protocoles dont cet autre dispose.

¹⁴Marie-Annick Morel, "Analyse linguistique d'un corpus d'oral finalisé", rapport d'un Gréco "Communication parlée".

Troisième partie

Application

Chapitre 6

Application

*Theory may inform, but practice
convinces.*

ANONYME

À partir des travaux conduits par Canal+ dans le cadre du Deuxième Monde (communauté virtuelle représentant la ville de Paris en trois dimensions et décrite dans la section 1.2.5), une série de prototypes d’environnements 3D à but de commerce électronique a été réalisée. Autour de ce concept d’une boutique spécialisée dans les contenus culturels (livres, musique, films par exemple), Canal+ souhaitait pouvoir enrichir son offre en y intégrant une interface de type “agent conversationnel incarné” (*embodied conversational agent*), l’objectif étant de parvenir à un système qui puisse assister les clients lors de leurs visites sur des sites de commerce en ligne. Cette assistance peut prendre la forme d’aide à la recherche d’un produit particulier ou de recommandation d’un produit (en se basant sur l’outil MMA¹), le tout en employant une interface de type conversationnelle et non pas un moteur de recherche à base de mots clés comme cela est souvent le cas sur les sites de commerce en ligne. Plus précisément, cette application se situe en amont des moteurs de recherche, l’agent participant, de manière conversationnelle, avec l’utilisateur à l’élaboration de la requête.

Pour tester les principes que nous avons présentés dans les deux chapitres précédents, nous avons développé cet agent dont le rôle est de recommander des films. Cette application nous permet d’une part d’exploiter les moyens mis à notre disposition par Canal+ et d’autre part de tester notre modèle de conversation dans un contexte contraint mais pour lequel l’emploi de la langue naturelle laisse l’utilisateur assez libre dans ses stratégies de communication.

La représentation du modèle de l’agent ayant été faite de manière fonctionnelle, nous avons suivi la même approche pour développer notre agent. Après une première section où nous développerons les *spécifications* de l’agent (section 6.1), nous présenterons les *données* manipulées par l’agent (section 6.2) et les *modules* de traitement de l’agent. Ces derniers sont répartis en deux sections, tout d’abord les *modules liés*

¹Mes Meilleurs Amis, un système de recherche de proximité entre films et utilisateurs qui peut être utilisé pour la recommandation collaborative. L’outil de recommandation “Mes Meilleurs Amis” s’appuie sur un système d’aide à la décision conçu et développé par une équipe de chercheurs en Intelligence Artificielle du LIP6. Les chercheurs à l’origine de ce projet sont Patrice Perny, Jean-Daniel Zucker et Nicolas Bredèche.

à la perception et à l'action (section 6.3), puis les *modules liés au raisonnement* (section 6.4), bien que la distinction entre les deux puisse être sujette à discussion. Après avoir ainsi présenté les spécifications du programme puis le programme lui-même, nous étudierons dans la section *analyses* (section 6.5) plusieurs extraits d'interactions entre un agent et un humain. Enfin, nous ferons quelques *remarques techniques* (section 6.6) sur l'application.

6.1 Spécifications

Le rôle de l'agent présenté dans ce chapitre est la recommandation et l'assistance à la recherche de films. Cette situation d'interaction correspond à un dialogue argumentatif de type *consultation d'expert* selon la typologie de Douglas Walton [Walton, 1990]. Cette recommandation est en fait considérée comme une phase préliminaire à un achat éventuel, situant notre agent dans un contexte de commerce électronique. Cet agent est représenté par un avatar et interagit avec l'utilisateur sur des modalités visuelles (expressions corporelles), spatiales (déplacements) et surtout langagières (texte *chat*).



FIG. 6.1 – L'environnement de l'agent

Les différents interactants partagent un *cyberespace* tridimensionnel représentant un magasin (voir figure 6.1). Cet espace virtuel, décrit en VRML, est disponible par l'intermédiaire d'une communauté de test Blaxxun installée localement². Il est possible de se connecter à cette communauté par l'intermédiaire d'un navigateur disposant d'une extension d'application spécifique (*plug-in*) disponible gratuitement. Les agents peuvent eux aussi se connecter à cette communauté, avec les mêmes possibilités que les humains, en employant une bibliothèque de fonctions dédiées.

²Il s'agit en fait d'une version gratuite du serveur de monde et de communauté virtuelle dont le nombre de connectés simultanés est limité à trois. La compatibilité entre le serveur de test et celui en exploitation, qui supporte le Deuxième Monde est totale, et le basculement d'une application d'un serveur à l'autre ne pose aucun problème technique.

Pour accomplir son rôle, notre agent est basé sur une architecture telle que celle qui a été présentée dans le chapitre 5, architecture elle-même dérivée du modèle d'interaction présenté dans le chapitre 4. Il dispose d'un module de raisonnement spécifique lié à son rôle qui sera présenté dans la section 6.4.3. Comme nous le verrons dans cette section, ses recommandations doivent se baser principalement sur deux critères : le *profil* de l'utilisateur et ses *critères de choix*. Le *profil* de l'utilisateur représente ses *préférences* et permet d'ordonner les propositions de films de façon à savoir quels sont les films les plus susceptibles de lui plaire. Les *critères de choix* correspondent aux caractéristiques que doit posséder le film. Bien évidemment, ces deux catégories d'information doivent pouvoir être obtenues par l'agent par le biais de sa conversation avec l'utilisateur.

Le *profil* de l'utilisateur est pris en compte dans un contexte collaboratif, puisqu'il est confronté aux préférences des autres utilisateurs³. Le système de recommandation se base sur l'outil MMA (dont l'interface *web* est présentée dans la figure 6.2) qui rassemble à la fois les préférences d'un ensemble d'utilisateurs, les similarités entre ces utilisateurs et les similarités entre films. Les recommandations faites au client se basent donc aussi sur les préférences des autres membres de la communauté.



FIG. 6.2 – L'interface *web* de MMA

Les *critères de choix* de l'utilisateur servent à cibler un film. Ils doivent pouvoir être négociés par l'utilisateur et le vendeur. Les mécanismes de conversation mis en œuvre par l'agent doivent donc permettre une dynamique qui amène à une focalisation commune sur un film.

6.2 Les Données

Nous présentons ici les données manipulées par l'agent et les modifications conceptuelles que nous leur avons fait subir lors de leur passage du modèle d'agent à l'instance de ce modèle. Après avoir commencé cette section par la présentation des *actes de communication* (section 6.2.1), nous présenterons le *focus conversationnel* (section 6.2.2) et les *attentes* (section 6.2.3), qui n'apparaissent qu'au niveau des modules de traitement. Nous traiterons ensuite de l'*historique* (section 6.2.4) des messages échangés et traités. Enfin, nous finirons par les *connaissances sur les autres*

³Ce genre d'application est typique du *communityware*, voir par exemple la recommandation collaborative de programmes télévisuels [Uchyigit et Clark, 2002] ou tout simplement le moteur de recherche google (www.google.fr).

(section 6.2.6) et la *personnalité* (section 6.2.5) de l'agent. Parmi les données que nous présentons dans cette section, certaines sont accessibles par plusieurs modules et appartiennent à l'ensemble de l'agent (historique par exemple) tandis que d'autres sont plus spécifiques d'une fonctionnalité (films préférés ou possédés par exemple) et sont contenues uniquement dans les modules concernés.

6.2.1 Actes de communication

Une fois passée l'étape de *formalisation* (première phase de perception), et jusqu'à ce qu'il soit *énoncé* (dernière phase de réalisation de l'action), un message est échangé entre les différents modules sous une forme unique, celle d'une variable d'un type que nous avons baptisé (page 153) "acte de communication". En effet, seul un message issu d'une modalité à base langagière puis formalisé pourrait prétendre au statut d'"acte de langage". L'emploi du terme *acte de communication* montre que cette variable est susceptible de contenir des informations extraites de modalités autres que langagières, permet d'unifier la représentation des messages et nous autorise quelques libertés que nous n'aurions pas eues si nous avions prétendu respecter à la lettre la théorie des actes de langage.

Nos actes de communication sont conceptuellement très proche des actes de langage. Entre autres, ils contiennent une description du monde (assimilable à un contenu propositionnel) ainsi que la relation entre cette description et le monde (assimilable à une force illocutoire). Nous nous permettons d'ailleurs d'employer la terminologie des actes de langage lorsque nous parlons des actes de communication. Toutefois, comme nous l'avons dit, ne pas se réclamer intégralement des actes de langage dans la spécification de ces actes de communication nous accorde une certaine liberté conceptuelle. Les actes de communication que nous utilisons dans ce chapitre désignent des variables informatiques confortables mais dont la justification et les spécifications théoriques *précises* seraient encore à réaliser.

Nom de la variable	Type	Page...
Contenu propositionnel	Liste	169
Force illocutoire	Objet composé	170
Modalité dominante	Énumération	171
Message dominant	Événement Blaxxun	171
Classe sémantique	Énumération	172

TAB. 6.1 – Structure d'un acte de communication

À la manière de REA (voir particulièrement page 69) qui fait communiquer ses différents modules en employant des messages au format KQML, nous remplissons et/ou modifions les différents champs de notre acte de communication au fur et à mesure de son passage au travers des modules de traitement⁴. Les champs principaux de la structure de données "Acte de communication" sont : un contenu propositionnel, une force illocutoire, la modalité dominante du message, le message dominant, et la classe sémantique (voir tableau 6.1).

⁴Nous n'avons pas employé KQML pour faire communiquer nos modules car cette solution nous semblait excessive au vu du déploiement des différents modules de notre agent.

a) Contenu propositionnel.

Le contenu propositionnel est la description d'un état du monde. Dans le cadre de notre application, cette variable contient essentiellement une description des objets du discours. Cette description est faite par l'intermédiaire de variables associant un *nom* à une *liste d'éléments*. Ainsi, `titreFilm(« Matrix »)` est une variable qui désigne un film dont le titre est "Matrix" et `area(areaFilms, out, 1234)` est une variable qui spécifie que l'individu dont l'identifiant est 1234 n'est pas dans la zone des films. Le tableau 6.2 rassemble les variables les plus utilisées de notre agent. Notons que le même type de variable est utilisé pour remplir le focus conversationnel.

Exemples	Description
<code>bodyXp(2)</code>	Expression corporelle numéro n° 2
<code>move(beamTo, 1234)</code>	Position : face à l'individu 1234
<code>position(ox, [...] z, id)</code>	Position de l'individu <i>id</i> en coordonnées cartésiennes
<code>dist(255, 1234, 9876)</code>	La distance entre les individus 1234 et 9876 est de 255
<code>area(rayonFilm, in, 1234)</code>	L'individu 1234 est dans le rayon des films
<code>territory(close, 1234, 9876)</code>	L'individu 1234 est trop près (selon la théorie des faces) de l'individu 9876
<code>date(1975)</code>	Année 1975
<code>date(1975, >)</code>	Après 1975
<code>note(nomActeur, clint eastwood, 80%)</code>	l'acteur Eastwood est gratifié d'une note de 80%
<code>critere(date, 1990, <)</code>	Représentation d'une date antérieure à 1980
<code>nomActeur(Clint Eastwood)</code>	Un acteur
<code>nomReal(Sergio Leone)</code>	Un réalisateur
<code>movieCat(comedie)</code>	Un genre de film
<code>titreFilm(princesse Mononoke)</code>	Un titre de film (par extension, un film)
<code>ordering(DBField, DBSimF.AnneeProd, ASC)</code>	Les films doivent être ordonnés par années, les plus anciens d'abord
<code>modifier(politesse)</code>	Altération de type "politesse"
<code>rheme(titreFilm, Hatari)</code>	Le film Hatari est le point central de l'énoncé

TAB. 6.2 – Variables employées dans le contenu propositionnel et le focus.

Ainsi, un contenu propositionnel peut par exemple contenir des informations du type :

- `titreFilm(« 1001 pattes »)` : le discours fait référence à un film dont le titre est "1001 Pattes" (comme dans l'énoncé : "Donne moi des infos sur 1001 pattes").
- `date (1995, ><)` : le discours fait référence à une date aux alentours de 1995 (comme dans l'énoncé : "Je cherche un film des années 90").

- `note(titreFilm, « Le bon, la brute et le truand », 75%)` : le discours fait référence à une évaluation sur un film (`titreFilm`) dont le titre est “Le bon, la brute et le truand”, avec une valeur de 75%⁵ (comme dans l’énoncé “J’aime bien le bon, la brute et le truand”).
- `territory(close, 13526)` : l’individu dont la référence est 13526 est trop proche de l’agent (suite à un déplacement au cours duquel l’individu en question s’est rapproché de l’agent, par exemple, mais ce contenu propositionnel peut aussi servir à qualifier un déplacement de l’agent, pour lui demander de se positionner “trop près” de l’individu 13526).

b) Force illocutoire.

La force illocutoire est le lien entre le contenu propositionnel et le monde. Bien que cela soit en théorie incorrect (nous manipulons des actes de communication et non de langage), nous employons pour remplir cette variable les forces illocutoires définies par Daniel Vanderveken [Vanderveken, 1988]. Ces forces illocutoires elles-mêmes sont des structures contenant plusieurs variables, dont les plus pertinentes dans le cadre de cette présentation sont :

- But illocutoire : direction d’ajustement entre le contenu propositionnel et le monde. *Cette variable fait partie des composantes de la force illocutoire de Vanderveken et est employée telle qu’il l’a définie.*
- Contrôle : désir de la personne à contrôler (ou à se laisser guider par) les situations, personnes ou relations. Cette valeur correspond à l’axe dominant-soumis du cercle interpersonnel (voir page 110).
- Affiliation : tendance à rechercher la compagnie des autres, indépendamment des sentiments que l’on peut éprouver envers ces personnes. Cette valeur correspond à l’axe bienveillant-hostile du cercle interpersonnel (voir page 110).
- Degré de puissance : force avec laquelle l’énoncé est exprimé. *Cette variable fait partie des composantes de la force illocutoire de Vanderveken et est employée telle qu’il l’a définie.*
- Nom : le nom de la force illocutoire. Selon la formalisation de Vanderveken, à une force illocutoire donnée correspond un ensemble de valeurs pour ses différentes composantes. Puisque nous manipulons directement certaines des composantes de la force illocutoire, et non la force illocutoire en elle-même, nous n’avons aucune garantie qu’à notre ensemble de valeurs corresponde une force illocutoire répertoriée pour le Français par Vanderveken (c’est-à-dire un performatif). Cette variable “nom” correspond à la force illocutoire la plus proche.
- Altération : les altérations apportent des modifications aux autres paramètres, permettant ainsi de rapprocher la force illocutoire d’une force illocutoire existante en français (voir page 146). Les trois altérations que nous avons considérées sont : politesse, hésitation et enthousiasme.

Parmi les composantes de la force illocutoire spécifiées par Daniel Vanderveken, nous avons ignoré pour notre application le mode d’accomplissement, les conditions sur le contenu propositionnel, les conditions préparatoires et les conditions de sincérité. Des exemples de traduction d’un énoncé en langue naturelle sont proposés dans

⁵Ce genre d’élément du contenu propositionnel est lié à la tâche de manière évidente, et montre bien que des traitements liés à une tâche se retrouvent répartis dans différents modules de l’agent (ne serait-ce que le lexique dans la partie formalisation liée à la langue naturelle).

Énoncé	Performatif(Contenu propositionnel)
j'aime bien les westerns	Inform(<code>note(movieCat,western,70)</code>)
j'adore les westerns	Inform(<code>note(movieCat,western,95)</code>)
non	Refuser()
je veux un film de Leone	Affirmer(<code>critere(nomReal, sergio leone, =)</code>)

TAB. 6.3 – Traduction langue naturelle → Acte de langage

le tableau 6.3.

Le mode d'accomplissement tel que l'a défini Vanderveken nécessite un contrôle pragmatique et sémantique qui dépasse ce que le format de notre message peut permettre⁶. Nos *altérations* jouent un rôle semblable.

Les conditions sur le contenu propositionnel et les conditions de sincérité ne concernent pas le type de dialogue auquel nous nous intéressons, étant donné qu'il n'y a pas lieu de douter des intentions de l'utilisateur et que son absence de sincérité n'entrave pas la négociation : l'agent n'est là que pour lui rendre service.

Pour les énoncés de l'agent, les conditions préparatoires sont implicites (comme cela est souvent le cas dans les systèmes automatisés). Pour les énoncés de l'humain il aurait été un peu artificiel de forcer l'existence d'une variable dédiée aux conditions préparatoires, étant donné qu'aucun mécanisme de l'agent ne l'aurait exploitée. Nous avons en effet estimé que la considération explicite des conditions préparatoires nécessitait la mise en place d'une chaîne de traitement complète et complexe (extraction, gestion de leur influence et manifestation d'un comportement en tenant compte) alors qu'une considération implicite était suffisante dans le cadre de notre application.

c) Modalité dominante.

Si l'acte de communication est issu de la perception d'un (ou plusieurs) message(s) dans l'environnement, cette variable précise de quelle modalité est issue ce message. Si l'acte de communication est destiné à être exprimé, cette variable précise la modalité principale d'énonciation.

d) Message dominant

L'acte de communication contient aussi le message dont il est issu. Cette variable existe essentiellement pour deux raisons : elle peut servir au principe d'explicitation des différents traitements qu'a subi le message dont nous avons parlé page 140 et elle permet des traitements liés au message lui-même et non à sa version formalisée et modifiée/adaptée par les différents modules (par exemple, un traitement de type *chat* nécessite de travailler sur les *mots* du message et ne pourrait rien tirer d'une formalisation de type acte de langage). Cette variable correspond directement à la structure de données qui contient l'événement perçu dans l'environnement (donc sous le format Blaxxun).

⁶Des travaux ultérieurs d'Émilie Benoit [Benoit, 2002] et de Sylvie Brincourt [Brincourt, 2003] lèvent en partie ce blocage, principalement au niveau sémantique.

e) **Classe sémantique.**

La classe sémantique indique, parmi une série de catégories d'actions liées à la tâche, à laquelle appartient le message. Cette information est redondante car elle peut être inférée du contenu propositionnel et de la force illocutoire, mais elle est explicitement représentée pour des raisons de simplicité de traitement. Dans le cadre de notre application 16 classes sémantiques ont été définies suite à une série d'expérimentations en "magicien d'OZ" [Berger, 2001]. Ces classes sont énumérées et brièvement décrites dans le tableau 6.4.

Classe	Description
Unknown	Inconnu ou indéfini
Admin	Message d'administration
PropInfos	Proposer des informations
InfosFocus	Transmissions d'information sur le focus
AskConstraintsProfile	Demander contraintes (aimes-tu X?)
AskConstraintsFocus	Demander contraintes (veux-tu X?)
Failure	Échec
GiveInfos	Transmissions d'information sur un film
Accept	Acceptation
Refuse	Refus
PropList	Proposer une liste de films
RequestList	Requête d'une liste de films
RequestInfo	Demande d'informations
GeneralInfo	Informations sur le profil ou sur le focus
Break	Clôture de conversation
Leave	Départ

TAB. 6.4 – Les classes sémantiques

Les classes sémantiques ont été définies après étude de dialogues sous la forme d'un magicien d'OZ, où un humain tenait le rôle de l'agent vendeur tandis qu'un autre jouait le rôle d'un client, l'interaction ayant lieu par l'intermédiaire de l'interface Blaxxun de façon à être dans la situation la plus écologique possible [Berger, 2001]. Ces classes sémantiques représentent un niveau d'abstraction intermédiaire pour la catégorisation des messages : plus spécifique que les forces illocutoires, mais restant toutefois assez générique pour permettre l'emploi des mêmes classes dans d'autres situations de consultation d'expert [Walton, 1990]. Les classes sémantiques adaptées à ce type d'interaction ont été étudiées plus finement par Sylvie Brincourt [Brincourt, 2003] et Émilie Benoit [Benoit, 2002], pour finalement disparaître au profit d'une redéfinition plus complète du langage.

6.2.2 Focus conversationnel

Le focus conversationnel est une liste de variables représentant certaines caractéristiques du film dont les interlocuteurs discutent. Ces variables sont un ensemble {type de critère, critère, relation}. Elles correspondent dans le tableau 6.2 à la variable nommée "critère". Les critères peuvent porter sur cinq types d'élément :

- Les acteurs. `critere(NomActeur, Clint Eastwood, <>)` : le film ne doit pas être avec Clint Eastwood.
- Les réalisateurs. `critere(NomReal, Sergio Leone, =)` : le film doit être de Sergio Leone.
- Les dates. `critere(date, 1975, ><)` : le film doit être des années 1970. `critere(date, 1990, >)` : le film doit avoir été tourné après 1990.
- Les genres. `critere(movieCat, western, =)` : le film doit être un western.
- Les films. `critere(titreFilm, « 1001 pattes », <>)` : le film ne doit pas être « 1001 pattes ».

Le focus contient par ailleurs le critère d’ordonnancement des réponses (variable nommée `ordering` dans le tableau 6.2), qui peut être :

- par ordre de préférence probable aux yeux du client (préférence calculée à partir des préférences déjà exprimées et des préférences des autres utilisateurs), ce qui constitue le comportement par défaut ;
- par ordre croissant ou décroissant de date, de façon à présenter les films les plus anciens ou les plus récents correspondant aux critères spécifiés.

Il n’existe qu’un focus conversationnel, contenu dans le module “rôle vendeur”. De plus, bien que le programme permette l’emploi du focus pour une quelconque négociation (hors film), nous n’avons pas mis en place de mécanisme permettant le passage d’un contenu de focus à un autre (en d’autres termes : un changement de sujet de conversation). Il s’ensuit que notre agent est très spécifique dans ses conversations, capable de tenir une conversation structurée par la négociation de son focus *dans le domaine du conseil de films*, mais déléguant ses réactions au module *chatbot* (ou employant d’autres mécanismes, comme les attentes, par exemple) lorsque le sujet de conversation s’en écarte.

La technologie Balxxun, sur laquelle nous nous appuyons, ne permet pas les interruptions sur les messages textuels, car ceux-ci sont transmis instantanément, et seulement lorsque le locuteur (ou plutôt, le scripteur) le désire. Ainsi, les informations telles que l’interlocuteur ayant la parole ou ayant manifesté le désir de la prendre, ne sont pas prises en compte dans notre application, elles n’ont en effet aucun sens. En théorie, ces informations sont gérées par le focus conversationnel, comme nous l’avons précisé page 158. Mais, n’étant pas disponible, le focus se trouve réduit à sa fonction fondamentale : la synthèse des informations échangées par les interlocuteurs à propos de l’objet de négociation.

Suite à cette description du fonctionnement du focus, le fonctionnement de l’ensemble {focus, attentes} sera développé en détail lors de la description du fonctionnement du module rôle, section 6.4.3.

6.2.3 Attentes

Nous nommons *attentes* un ensemble {événement, action, paramètres, importance, liste d’attentes, durée de vie}. Cet ensemble présente quelques spécificités par rapport à la description théorique qui en a été faite page 124. L’*événement* correspond au percept issu de l’environnement (ou, pour être plus précis, à l’*interprétation* qui en a été faite) auquel l’agent *peut* réagir en entreprenant l’*action* associée. Les *paramètres* sont là pour permettre de spécialiser l’attente, en apportant (éventuellement) des précisions à l’action associée ou à l’événement déclenchant (ils permettent de spécifier des variables). L’*importance* donne une indication de la priorité pour

l'agent à agir de cette façon (i.e., accomplir l'*action*) dans cette situation (*événement*) : c'est cette valeur qui est renvoyée lors de l'appel d'offre du module de sélection (voir page 194). Cette valeur est différente de la "priorité" (voir page 125) de l'attente, qui est représentée implicitement par la position de l'attente au sein d'une liste ordonnée, comme nous le verrons plus loin. La *liste d'attentes* rassemble les attentes qui deviendront valides si l'attente en cours est déclenchée, et enfin, la *durée de vie* de l'attente précise le temps (calculé en nombre de tours de parole) pendant lequel l'attente est susceptible d'être déclenchée.

Notre application n'emploie des attentes qu'au sein du module "rôle vendeur". Deux types d'attentes peuvent être distinguées, suivant leur *durée de vie* : certaines de ces attentes sont valables en permanence tandis que d'autres sont ajoutées et retirées en fonction du contexte conversationnel. Les premières représentent les situations pour lesquelles l'agent est prêt à réagir en permanence et les secondes des situations participant au suivi d'une conversation plus spécifique (le tableau 6.13 de la section 6.4.3 qui développe le fonctionnement des attentes, répertorie les attentes toujours valides).

La table 6.5 donne deux exemples d'attentes. Le premier est celui d'une attente active en permanence : si un interlocuteur avec lequel l'agent n'a pas d'interaction pénètre dans le rayon des films, l'agent cherche à débiter une nouvelle interaction avec cet interlocuteur. Cette attente est valable sans limite de durée, mais doit être supprimée une fois qu'une conversation est initiée.

Événement	Action	Import.	TTL ^a	Cond. de retrait
Pénétration dans le rayon des films	Initiation d'interaction	0.75	∞^b	Conversation initiée
Refus	Retrait de certains films du focus & action en fonction du focus	0.85	1	Expiration de la durée de vie

^aTime To Live, durée de vie.

^bUn temps de vie infini est représenté par une valeur égale à -1.

TAB. 6.5 – Deux exemples d'attentes (certaines variables spécifiques du traitement informatique sont omises)

Le deuxième exemple présente une attente dont la durée de vie est d'un tour de parole. Cette attente ne peut se comprendre qu'à travers le contexte d'où elle est issue : au tour précédent l'agent a proposé à son interlocuteur un ou plusieurs films (il s'agit d'une attente déjà vue dans les tableaux 4.9 et 4.8, pages 129 et 124, et analysée à cette occasion). Il a ensuite précisé les réactions possibles à cette proposition, et entre autres, qu'un refus devait entraîner : (1) le retrait des films proposés⁷ et (2) la consultation du focus conversationnel mis à jour pour déterminer l'action suivante. La durée de vie d'une valeur de "un" précise qu'un refus ne doit être traité par cette fonction que s'il arrive immédiatement dans la suite de la conversation. En effet, une réponse de la forme "non" ne peut se comprendre (dans les cas les plus simples) qu'en fonction du tour de parole précédent.

⁷Concrètement, le programme insère les films dans le focus, comme critères négatifs, de la forme `critere(titreFilm, 1001 pattes, <>)`.

De plus, les attentes sont rangées dans une liste ordonnée, et le parcours de la liste s'arrête à la première qui se trouve validée. L'ordre de ces attentes reflète la priorité (qui est un des paramètres des attentes, comme nous l'avons présenté dans le modèle de l'interaction, page 125) : à chaque nouveau tour de parole, les attentes sont empilées sur le haut de la pile, donnant ainsi priorité aux attentes les plus récentes. Parmi ces attentes les plus récentes, un sous-ordre peut être défini de façon à favoriser les attentes les plus spécifiques. La "priorité" n'est donc pas explicitement présente dans la structure de données représentant les attentes, mais dans la manière dont elles sont rangées.

L'*importance* de l'attente est liée au fonctionnement du module de sélection (voir section 6.4.2), elle est donc un choix technique et n'a pas de lien avec la *priorité*. Lors de l'appel d'offre du module sélection, c'est cette valeur qui est renvoyée par le module rôle (qui, rappelons le, est le seul à avoir un fonctionnement à base d'attentes) au module de sélection, et qui correspond à une estimation de la capacité du module rôle à traiter l'événement considéré, ou, sous un autre angle, de l'importance qu'a ce message pour le module rôle, ce qui conditionne son "désir" d'avoir à le traiter.

Dans les tests que nous avons effectués, les valeurs de temps de vie des attentes étaient en général infinies (situation auxquelles l'agent est toujours susceptible de réagir) ou égales à un (réaction immédiate). Quelques autres valeurs ont été testées (comme nous l'avons montré dans l'exemple de la page 131), mais une analyse fine de l'influence de la variation des temps de vie de certaines attentes sur le déroulement de la conversation aurait nécessité la mise en place d'expériences dépassant le cadre de cette thèse.

Suite à cette description du fonctionnement des attentes seules, le fonctionnement de l'ensemble {focus, attentes} sera développé en détail lors de la description du fonctionnement du module rôle, section 6.4.3.

6.2.4 Historique

Comme il a été spécifié page 160, l'historique contient les messages perçus et déposés dans l'environnement, ainsi que les actes de communication issus de ces messages — ou les générant — en différents stades de leur traitement. Toutefois, seuls sont effectivement stockés les messages ayant suscité l'intérêt de l'agent, cet intérêt étant mesuré par la réponse à l'appel d'offre effectué par le module de sélection (voir section 6.4.2). Si aucun des modules ne veut traiter le message (qu'il ne le puisse pas ou qu'il ne le désire pas), alors il n'est pas stocké dans l'historique. Cette stratégie permet par exemple d'éviter de stocker l'ensemble des déplacements de tous les présents (ce qui peut représenter jusqu'à un message toutes les 200 ms pour chaque avatar présent), pour ne se concentrer que sur les événements significatifs pour l'agent.

6.2.5 Personnalité propre

Dans l'application réalisée, nous nous sommes limités à traiter d'une partie de l'aspect social de la personnalité. Pour cela, nous avons retenu deux théories issues des sciences humaines : la théorie des faces d'Erving Goffman (voir page 107) et, dans une moindre mesure, le cercle interpersonnel de McCrae et Costa (voir page 110).

Théorie des faces. Suivant la théorie des faces, notre agent dispose de quatre variables, correspondant aux limites minimales et maximales de la zone acceptable en ce qui concerne le territoire et la face de l'agent. Pour ce qui est du territoire, les limites sont converties en distances et l'agent cherchera à se maintenir à une distance de son interlocuteur qui respecte au mieux ces contraintes. Du respect de ces contraintes émerge un comportement de suivi de l'interlocuteur lorsque l'agent est en discussion. Pour ce qui est de la face, une agressivité est associée à chaque acte de communication. Cette agressivité est principalement basée sur le degré de puissance du performatif.

Cercle interpersonnel. La théorie du cercle interpersonnel se base sur une représentation d'un cercle ayant deux axes : l'axe domination/soumission (appelé axe de *contrôle*) et l'axe hostilité/amitié (appelé l'axe d'*affiliation*). Ce cercle représente la relation revendiquée par l'agent dans le cadre de l'interaction en cours, et une règle de complémentarité entre cette revendication de l'agent et celle de son interlocuteur permet de caractériser les interactions "confortables".

Conformément aux liens qui existent entre données et processus (voir le tableau 5.3 page 158), l'usage des ces variables sera développé lorsque nous précisons le fonctionnement des modules d'interprétation (page 184) et d'expression (page 188).

6.2.6 Connaissances sur les autres

Les connaissances dont l'agent dispose sur les autres sont de deux types : celles directement liées à l'interaction et celles directement liées au rôle.

Pour ce qui est des connaissances liées à l'interaction, l'agent dispose pour chaque interlocuteur des mêmes paramètres que ceux que nous avons présentés dans la section précédente, dans le cadre de la modélisation de la personnalité de l'agent. Le premier de ces paramètres est la *face* (au sens de Goffman, voir page 107) de l'interlocuteur. Elle conditionne principalement le choix des forces illocutoires. Le deuxième de ces paramètres est le *territoire* (toujours au sens de Goffman, voir page 107). Le dernier enfin est un positionnement sur le cercle interpersonnel de McCrae et Costa, associé à une valeur représentant la variabilité.

Pour ce qui est des connaissances liées au rôle, nous considérons principalement le *profil* de l'utilisateur. Ce profil contient les informations sur les préférences de l'utilisateur. Il est renseigné de trois manières :

- Tout d'abord, l'utilisateur peut utiliser une interface web pour juger les films, dans le cadre de l'application MMA (voir figure 6.2). Les informations fournies dans ces conditions constituent le profil *permanent* de l'utilisateur.
- Ensuite, pendant ses discussions avec l'agent, l'interlocuteur va être amené à exprimer de nouvelles préférences ("Je n'aime pas les comédies", "je préfère les films avec Clint Eastwood", ...). Ces informations sont intégrées à un profil *temporaire*, qui dure le temps de l'interaction. Certaines d'entre elles (les plus spécifiques, celles qui visent un film, par opposition à celles qui visent l'ensemble des films d'une année, l'œuvre d'un acteur ou d'un réalisateur) peuvent, à la fin de l'interaction, être intégrées au profil permanent.
- Enfin, l'ensemble de la base de MMA est employé pour compléter le profil. Lorsqu'une opinion au sujet d'un film est exprimée, l'agent recherche les films proches et les intègre aussi au profil. Ainsi, le profil contient aussi des films

pour lesquels l'interlocuteur de l'agent n'a pas explicitement exprimé une opinion, mais pour lesquels il est tout de même possible d'évaluer une opinion raisonnable.

Bien entendu, l'agent conserve aussi sur son interlocuteur de nombreuses informations de fonctionnement (comme sa position) ou statistiques (comme la durée totale d'interaction), qu'il n'y a pas intérêt à développer ici.

6.3 Modules liés à la perception et à l'action

Nous avons présenté dans la section précédente les *données* manipulées par l'agent, donc suivant un point de vue particulièrement *statique*. Nous allons maintenant traiter des étapes qui composent la chaîne perception → raisonnement → action de l'agent.

Nous nous limitons dans cette section aux modules dont les traitements sont liés directement aux étapes de perception et d'action et en y en incluant les deux étapes d'enrichissement contextuel : l'interprétation et l'expression. Nous laissons donc pour la section suivante les modules liés au raisonnement, plus spécifiques d'une tâche, et surtout aux fonctionnements internes plus variables.

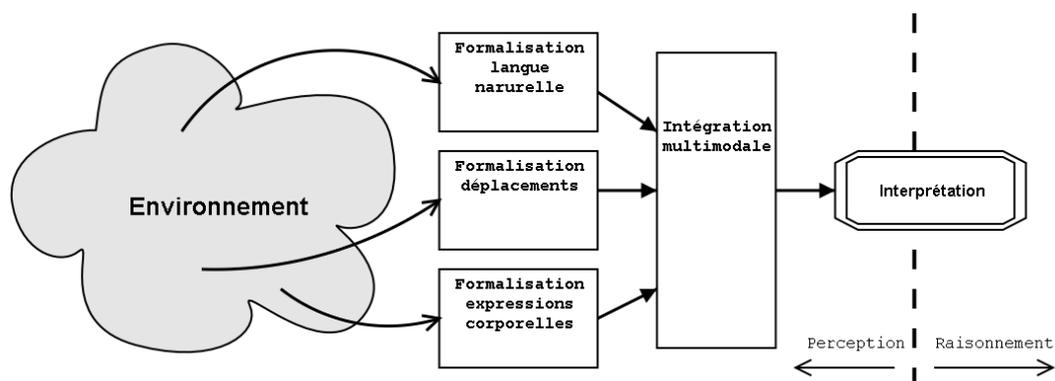


FIG. 6.3 – La chaîne « perception »

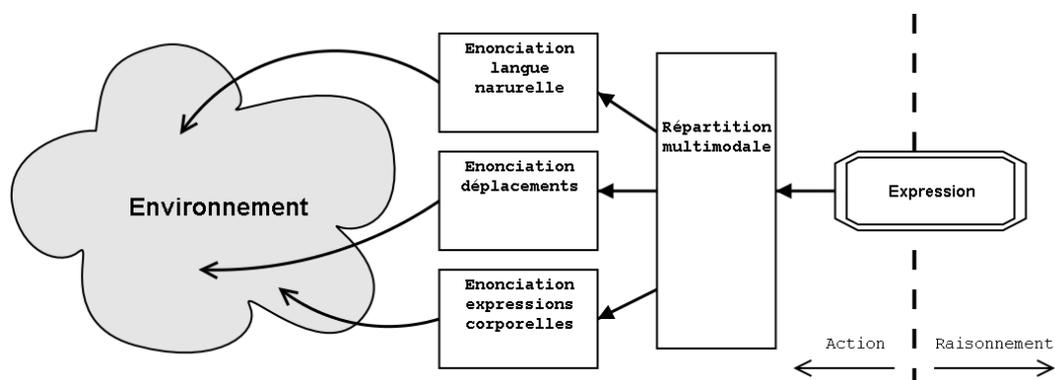


FIG. 6.4 – La chaîne « action »

Cette section présentera les modules dans l'ordre chronologique de leur activation lors des traitements de messages : *formalisation* d'un percept d'une modalité donnée vers un formalisme adapté à la manipulation par l'agent ; *intégration multi-modale* des différents messages formalisés pour en tirer un message porteur d'un sens potentiel ; *interprétation* de ce message pour en tirer un message porteur d'un sens unique (voir figure 6.3). À partir de cette étape, le message est confié aux modules de raisonnement. Si ces modules (dont le fonctionnement sera développé dans la prochaine section) décident d'une action à entreprendre, le message suivra une chaîne de traitement symétrique de celle présentée précédemment : *expression* du message pour l'enrichir du contexte conversationnel ; *répartition multi-modale* de façon à sélectionner les modalités de l'environnement les plus à même de porter le message ; *expression* des différents messages, c'est-à-dire, dépôt des messages dans l'environnement, après une traduction dans la modalité à laquelle ils sont dédiés (voir figure 6.4).

6.3.1 Formalisation

Comme nous l'avons indiqué page 148, l'étape de formalisation sert à convertir un message perçu dans l'environnement de façon à le mettre sous une forme manipulable par l'ordinateur. Notre application ne peut percevoir que suivant trois modalités : position des avatars (canal des *déplacements*), animations déclenchées au niveau de l'avatar (canal des *expressions corporelles*) et messages textuels (canal du *chat*). Les deux premières modalités sont traitées de manière très directe, par rapport à la dernière pour laquelle nous avons mis en place un traitement plus complexe.

a) Déplacements

En ce qui concerne les déplacements, les informations issues du serveur Blaxxun, sont transmises quasiment telles quelles, puisqu'il ne s'agit que des coordonnées absolues de l'avatar en déplacement. Pour des raisons de performances, les messages de position ne sont pas systématiquement traités : lors du déplacement d'un avatar, le serveur peut réactualiser sa position jusqu'à une fois toutes les 200ms. Nous avons donc décidé de bloquer la transmission d'un message de position tant qu'il ne correspondait pas à un déplacement minimum depuis le dernier déplacement traité. Cette stratégie pourrait être affinée mais est en l'état adaptée à nos besoins. De plus, le serveur Blaxxun permet la définition d'un horizon au-delà duquel les événements ne sont plus perçus par l'agent et nous exploitons cette fonctionnalité pour limiter la quantité d'événements perçus.

b) Expressions corporelles

La technologie Blaxxun n'autorise qu'un emploi très limité des expressions corporelles. Chaque avatar est défini en VRML et peut contenir une série d'animations pouvant être déclenchées par le client Blaxxun. Ces animations sont en nombre limité (une vingtaine au mieux), non (ou peu) paramétrables lors de l'exécution (i.e. elles sont définies lors de la création de l'avatar), et leur "*timing*" n'est pas maîtrisable. De plus, la forme même de l'interface sur le monde virtuel permet difficilement de distinguer de subtiles modifications, entraînant une utilisation limitée des expressions faciales. Enfin, et ceci est probablement le problème le plus dérangeant, le serveur Blaxxun se contente d'envoyer le numéro de l'expression corporelle employée (les expressions par défaut sont présentées dans le tableau 6.6). Or, toutes les expressions

corporelles peuvent être (re)définies par l'utilisateur. Ainsi, parmi les expressions par défaut la numéro 1 correspond à un geste de salutation, mais rien n'empêche un utilisateur de créer ses propres expressions et d'employer l'emplacement numéro 1 pour une expression n'ayant plus rien à voir avec une salutation. Cette absence de sémantique associée aux expressions corporelles ne nous semble pas avoir de solution simple dans le cadre de la technologie Blaxxun.

Le nombre limité et le non paramétrage des expressions corporelles nous a amené à choisir une méthode de formalisation à base de dictionnaire : à tel numéro d'expression corporelle correspond tel message formalisé. Ce choix a été fait malgré la possibilité de redéfinition des expressions corporelles. Si cette solution venait à se montrer problématique, il resterait la possibilité d'un "débrayage" de la perception des expressions corporelles dès lors que l'avatar n'est pas un avatar standard ou répertorié (l'avatar étant hébergé sur un site web, il est possible de limiter le traitement des expressions corporelles aux avatars issus de sites pour lesquels la correspondance entre une expression corporelle et sa référence est connue). La possibilité de reconnaître l'expression en percevant les mouvements de l'avatar nous semble largement irréaliste, en l'état actuel de la technologie et des recherches dans le domaine des expressions corporelles.

n°	Quelques mots-clefs déclenchant ^a	Force illocutoire associée	Altération associée
1	hello, hi	Saluer	
2	hey, watch, like		
3	yes, sure, agree	Accepter	
4	smile		Enthousiaste
5	frown, hate		
6	no, disagree	Refuser	
7	not, dislike, reject	Refuser	
8	bye, goodbye, wave	Saluer	
9	cool		Enthousiaste
10	laugh		Enthousiaste

^aCes mots sont les mots de l'interface standard Blaxxun. Canal+ distribuait une version francisée.

TAB. 6.6 – Expressions corporelles standard

Les expressions corporelles ne sont que rarement déclenchées explicitement par les utilisateurs humains : elles sont le plus souvent associées à une liste de mots-clefs, et automatiquement déclenchées sur occurrence de l'un d'entre eux. Dans cette situation, il est délicat de considérer une expression corporelle comme ayant valeur sémantique, car elle ne peut pas être considérée comme venant en *complément* du message textuel et le sens dont elle *pourrait* être porteuse n'est finalement que redondant par rapport au message textuel.

c) Texte

La conversion des messages de modalité textuelle est la plus complexe des trois. Le module de formalisation est construit autour d'un analyseur lexico-syntaxique

nommé PC-PATR, développé par le SIL⁸. Les travaux théoriques sur ce module, ainsi que la préparation du lexique et de la grammaire ont été réalisés par Alexandra Berger [Berger, 2001]. L'adaptation de PC-PATR à l'agent a été en grande partie réalisée par Yannick Chappellaz [Chappellaz, 2001].

L'analyseur lexico-syntaxique PC-PATR. PC-PATR a été intégré à l'agent suite aux travaux d'Alexandra Berger. La présentation suivante est extraite de son mémoire [Berger, 2001] :

PC-PATR est un analyseur syntaxique qui fonctionne sur la base de l'unification grammaticale, chaque programme d'analyse de syntagmes nécessite :

- un fichier lexique où l'on va expliciter chaque mot que l'on a l'intention d'utiliser ;
- un fichier grammaire qui contiendra tous les traits morphosyntaxiques nécessaires à la compréhension du lexique, la déclaration de chacune des catégories de mot, ainsi que les règles de grammaire qui régiront la bonne formation des syntagmes.

Dans le lexique, on donnera une interprétation de chaque expression (une expression pouvant être un ou plusieurs morphèmes) qui soit "compréhensible" et interprétable par l'agent.

L'analyseur vérifie tout d'abord si tous les mots (ou expressions) de l'énoncé font partie de son lexique, sans quoi, il ne peut donner d'analyse. Il "récupère" dans l'entrée du mot la catégorie, la glose et les traits morphosyntaxiques qui lui permettront ensuite de définir de quelle structure grammaticale il s'agit et d'en construire la f-structure (*feature structure*, structure de traits) et l'arbre syntaxique. Nous n'utiliserons ici que l'arbre syntaxique puisque nous n'avons besoin que de l'expression et de son interprétation (glose).

La grammaire de PC-PATR a été développée de façon à tenir compte des particularités du *chat*. Principalement, la plupart des règles d'accord sont ignorées, étant donné que, très régulièrement, les internautes les négligent. Les accords peuvent aider au traitement des messages en permettant de lever certaines ambiguïtés, mais les énoncés sur lesquels nous travaillons sont suffisamment simples pour ne pas contenir⁹ d'ambiguïtés de référence entre divers éléments. La contrainte de la définition d'une grammaire personnelle (ajoutée au problème des licences d'exploitation) a été principale dans le choix de PC-PATR, car la plupart des analyseurs lexico-syntaxiques ne permettent pas de définir la grammaire du langage : ils intègrent en général une grammaire du français correct, ce qui n'est pas adapté à nos besoins !

Étape de pré-traitement. Outre PC-PATR, le module de formalisation de la langue naturelle contient une étape de pré-traitement, qui sert principalement à normaliser les énoncés entrants. Notamment, dans cette étape

- des espaces sont insérées derrière chaque élision. En effet, les seuls séparateurs reconnus par PC-PATR sont les espaces, et donc "j'aime" ou "d'avant" doivent être transformés en "j' aime" et "d' avant" pour être considérés comme deux morphèmes et non un seul (ce qui imposerait de les avoir tels quels dans le lexique).

⁸Summer Institute of Linguistics, <http://www.sil.org/pcpatr/>

⁹ou trop rarement pour mériter un traitement spécifique.

- les signes de ponctuation sont supprimés. Comme nous l'avons dit dans le point précédent, seules les espaces sont des séparateurs, les signes de ponctuation sont considérés comme des lettres. Puisque nous ne traitons pas d'énoncés complexes dans lesquels la ponctuation aurait une importance sémantique, nous pouvons nous permettre de la supprimer. Dans le cas contraire, il aurait fallu intégrer la ponctuation dans la grammaire.
- les caractères accentués sont remplacés par les caractères sans l'accentuation. Par exemple, "itinéraire d'un enfant gâté" est transformé en "itineraire d'un enfant gate". La dernière version de PC-PATR ne traite en effet pas les caractères accentués (elle semble se limiter aux caractères de la table ascii non-étendue), mais ce remplacement nous permet aussi de traiter les mots dans lesquels l'utilisateur aurait oublié les accents (situation très courante).
- le texte est passé en minuscule. PC-PATR est sensible à la casse, et nous avons décidé de n'avoir que des lettres minuscules, sans quoi, par exemple, un titre orthographié "matrix" ne serait pas reconnu si le lexique contenait "Matrix".
- les *smileys* (émoticônes) sont extraits. Ils sont en effet de plusieurs types, comme : :o> ;-)))) (-8 :-D etc. et leur intégration dans le lexique serait une solution lourde et inadaptée¹⁰. Ils sont détectés lors du pré-traitement, supprimés de l'énoncé, puis, une fois l'énoncé traité, ils sont utilisés pour adapter la force illocutoire, l'affiliation et le contrôle (voir page 170) du performatif.

Lexique et grammaire La constitution du lexique et de la grammaire est expliquée par Alexandra Berger dans son mémoire [Berger, 2001]. Nous en reprenons ici les points essentiels.

Le lexique et la grammaire ont été réalisés à partir de l'analyse de conversation entre un humain jouant le rôle d'un acheteur et un humain jouant le rôle de l'agent, tous deux communiquant en utilisant les interfaces Blaxxun. En reprenant les travaux d'Alexandra Berger [Berger, 2001], on peut distinguer parmi les entrées lexicales une catégorie particulière, qui rassemble les morphèmes implicitement ou explicitement performatifs. Certaines de ces entrées sont rassemblées dans la table 6.7, tandis que la table 6.8 rassemble des entrées plus classiques.

Entrée	Entrée
1 \w prefere \c V \g perf, informer, (aimer, 70%, focus) \f tsing trans indpresent	4 \w peux \c V \g perf, demander \f tsing trans indpresent
2 \w en quelle annee \c PROR \g perf, questionner, (date, focus) \f rel	5 \w veux bien \c V \g perf, affirmer, (desirer) \f tsing trans indpresent
3 \w donne le moi \c ENG \g perf, accepter, (), (enthousiaste)	

TAB. 6.7 – Extraits du fichier lexique (performatifs)

¹⁰Les fonctionnalités de pré-traitement morpho-syntaxiques de PC-PATR n'étaient pas non plus facilement adaptables à cette situation.

Dans la table 6.7, nous rassemblons un ensemble représentatif des entrées lexicales à valeur performative. Les spécificités de ces entrées sont les suivantes :

- Pour l’entrée n° 1 : la glose (repérée par le préfixe \g) de l’entrée lexicale précise que l’énoncé aura pour force illocutoire “Informer”, mais elle donne aussi des informations sur le contenu propositionnel. (*aimer*, 70%, *focus*) signifie que le focus¹¹ de l’énoncé (un film, un acteur, un genre de film, ...) est l’objet d’une préférence, dont la valeur est de 70%.
- Pour l’entrée n° 2 : le mot (repéré par le préfixe \w) considéré est en fait une séquence de mots, permettant simplement (i.e. sans avoir à modifier la grammaire par exemple) la prise en compte d’une situation assez spécifique. Comme nous l’avons fait remarquer pour l’entrée n° 1, la glose comprend des informations visant à aider la construction du contenu propositionnel. En l’occurrence, le contenu propositionnel contiendra une date en référence directe avec le focus¹¹ de l’énoncé.
- Pour l’entrée n° 3 : la glose de cette entrée lexicale introduit un élément de variation sur la force illocutoire (4^e élément), précisant que cette force est porteuse d’*enthousiasme*. D’autres altérations possibles sont *politesse* et *hésitation*.
- Les entrées n° 4 et n° 5 ne sont là que pour présenter un panorama varié d’entrées lexicales à valeur performative.

Entrée	Entrée
1 \w petersen \c N \g nomReal, wolfgang petersen \f real	4 \w alain chabat \c N \g nomActeur, alain chabat \f act
2 \w cartoon \c N \g movieCat, dessin anime	5 \w dessin anime \c N \g movieCat, dessin anime
3 \w l' ete de kikujiro \c N \g titreFilm, l'été de kikujiro	6 \w bon \c ADJ \g note, +70%

TAB. 6.8 – Extraits du fichier lexique (divers)

Dans la table 6.8, nous présentons un ensemble représentatif des entrées lexicales à valeur non-performative. Les spécificités de ces entrées sont les suivantes :

- Pour les entrées n° 1 et n° 4 : on peut désigner un acteur ou réalisateur à partir de son nom (ce qui peut mener à des ambiguïtés), comme dans le cas n° 1 ou de son nom et de son prénom, comme dans le cas n° 4. La glose, quant à elle contient le nom et le prénom, ce qui permet d’unifier les traitements ultérieurs. On remarquera dans ces entrées le renseignement de certains traits (repéré par le préfixe \f, comme *feature*). Ces informations sont exploitées dans la grammaire de façon à marquer la différence entre “un film **de** Eastwood” où Eastwood est réalisateur, et “un film **avec** Eastwood”, où il est acteur¹².

¹¹Le terme “focus” est ici ambigu car il ne fait pas référence au focus conversationnel, mais au sujet central de l’énoncé. Le terme “rhème” serait plus approprié, mais “focus” est conservé dans le lexique pour des raisons historiques.

- Pour les entrées n° 2 et n° 5 : ces entrées font référence à une catégorie de films. Il y a dans le lexique autant d'entrées qu'il y a de synonymes à un mot, leurs autres paramètres (principalement la glose) étant identiques. Cette fonctionnalité pourrait être remplie par un thesaurus.
- Pour l'entrée n° 3 : cette entrée représente un film, la séquence recherchée étant au format requis par PC-PATR (pas d'accents et espace derrière l'apostrophe) et la glose étant au format nécessaire pour faire le lien avec la base de données. Une fois de plus, un identifiant unique serait préférable, mais employer directement le titre permet une analyse plus simple du fonctionnement.
- Pour l'entrée n° 6 : cette entrée représente un adjectif qualificatif, qui apporte un complément d'information à une note (éventuellement) présente dans l'énoncé.

La construction et la validation du lexique et de la grammaire se sont faite de manière incrémentale jusqu'à ce que l'agent soit capable de traiter une proportion satisfaisante des échanges du corpus issus des dialogues entre deux humains.

6.3.2 Intégration multi-modale

Le module d'intégration multi-modale est chargé de transformer une série de messages issus de différentes modalités en un seul message porteur d'un sens potentiel. Sachant que la multi-modalité n'est pas le sujet central de cette thèse (mais aussi pour des raisons plus pratiques comme le temps de réponse), nous avons simplifié à l'extrême les traitements multi-modaux.

Comme nous l'avons dit, l'ensemble des mécanismes d'intégration multi-modale se structure autour d'un message dominant, qui est complété ou adapté par une série de messages complémentaires. Nos modalités étant en nombre limité, nous avons choisi des heuristiques particulièrement simples :

- mouvement : le déplacement est un message dominant et ne peut être influencé par les autres modalités (ni les influencer). Il constitue une modalité entièrement découplée des autres.
- texte : un message textuel est toujours dominant. Il ne peut être complété par des expressions corporelles que si celles-ci ont déjà été perçues (au plus 50ms avant la perception du message textuel) ou qu'elles interviennent rapidement (au plus 50ms) de façon à ne pas bloquer inutilement la chaîne de traitement.
- expressions corporelles : les expressions corporelles peuvent à la fois être des messages dominants et des messages complémentaires. Pour être considérée comme un message dominant une expression corporelle doit satisfaire à plusieurs conditions. Tout d'abord, elle ne doit pas avoir été utilisée en tant que modificatrice d'un message textuel. Ensuite, elle ne doit pas être porteuse d'un message proche du dernier¹³ message textuel traité (salutation, acquiescement, remerciement, refus, ...). Enfin, il faut différer le traitement de 50ms de façon à attendre un éventuel message textuel susceptible de s'appuyer sur elle.

¹²Ceci est particulièrement visible dans l'extrait de dialogue suivant :
humain> donne moi un film avec sergio leone
Bot> ???
humain> donne moi un film de sergio leone
Bot> J'ai IL ETAIT UNE FOIS EN AMERIQUE [...]

¹³Dans les limites de quelques secondes toutefois, pour autoriser les répétitions. L'objectif n'est que d'éviter les expressions corporelles automatiquement générées, et donc, par principe, redondantes par rapport au texte.

L'acte de communication résultant de la confrontation d'un message textuel et d'une expression corporelle aura toutes les caractéristiques du message textuel, à l'exception du degré de puissance, du contrôle et de l'affiliation qui seront la moyenne de ceux des deux messages. Le but illocutoire est trop précisément lié au message dominant pour être influencé par les messages complémentaires. La technologie Blaxxun n'autorisant pas un emploi aisé de déictiques ("*cet objet*" accompagné d'un geste de pointage), les expressions corporelles n'influent pas sur le contenu propositionnel.

6.3.3 Interprétation

Lors de l'étape d'interprétation, l'agent acquiert des données sur ses interlocuteurs et modifie le message en fonction de son contexte.

a) Acquisition des données

Les données acquises pendant l'étape d'interprétation sont celles permettant la mise à jour du profil de l'utilisateur. Trois catégories de connaissances sur l'utilisateur sont mises à jour à ce niveau : le territoire, la face et le profil interpersonnel.

Territoire. Le territoire correspond à une zone en forme de couronne¹⁴, centrée sur un individu. Les interlocuteurs de cet individu doivent s'y trouver pour que l'interaction se déroule au mieux, sans que l'individu ne se sente oppressé (interlocuteur trop proche) ni rejeté (interlocuteur trop éloigné).

Les deux frontières du territoire sont mises à jour lors des déplacements de l'interlocuteur, suivant quelques heuristiques assez simples : si l'interlocuteur communique, c'est que l'agent est dans la zone acceptable de son territoire ; si l'interlocuteur s'approche, alors la frontière extérieure est rapprochée ; si l'interlocuteur s'éloigne, alors la frontière intérieure est éloignée.

On peut noter que ces règles peuvent facilement entrer en conflit avec les règles de respect du territoire *de l'agent*. En effet, l'agent lui-même se déplace pour faire respecter son propre territoire, ce qui, évidemment, perturbe le calcul des limites du territoire de l'interlocuteur. La situation se compliquerait encore si l'agent (et/ou ses interlocuteurs) se déplaçait au sein d'un groupe. Des règles plus complexes impliquant les connaissances mutuelles de l'agent et de ses interlocuteurs seraient à prendre en compte ici (avec toutefois les limites présentées page 103).

Des expérimentations spécifiques devraient être faites pour définir les limites des règles que nous avons implémentées, en gardant toutefois à l'esprit que, au moins dans un espace virtuel de type Blaxxun, la sensation d'espace, bien qu'existante, est fortement limitée. Les canaux de communication étant en partie en conflit (le *chat* mobilise lui aussi la vision des interactants humains, pour regarder le clavier et pour lire les messages), la modalité textuelle bénéficie d'une attention supérieure à la modalité visuelle. Une fois de plus, pour des questions d'interface, les utilisateurs se soucient peu (mais ils s'en soucient toutefois) de leur intégration dans l'espace, plus précisément, dans le cas qui nous intéresse, de la distance qui les sépare de leurs

¹⁴En toute rigueur, la zone ne devrait pas être une couronne parfaite, mais plus étendue sur l'avant, qui correspond à la zone utile de l'individu, d'autant que l'interface Blaxxun limite énormément (pour ne pas dire interdit) les perceptions hors de la zone frontale.

interlocuteurs¹⁵. L'arrivée d'interfaces orales, puis de systèmes de vision stéréoscopique devrait renforcer l'importance du territoire.

La vision que Goffman a du territoire est bien plus étendue que celle implémentée. Il spécifie en effet huit "territoires du moi" [Goffman, 1973] (voir la note n° 18 en bas de page 107) qui incluent des domaines aussi variés que le territoire temporel (temps de parole), la vie privée, les sujets de conversation, ... La partie que nous traitons correspond essentiellement aux points que Goffman baptise l'*espace personnel* et l'*espace utile*. D'autres points comme les *domaines d'information* et les *territoires réservés d'information* trouvent dans notre architecture leur place dans les modules traitant des aspects sociaux.

Face. De la même façon que pour le territoire, nous implémentons pour la face une version extrêmement simplifiée de la face de Goffman. Là où Goffman voit un ensemble de critères sociaux constituant quelque chose de très proche de notre notion de personnalité ("On peut définir le terme de face comme étant la valeur sociale positive qu'une personne revendique effectivement à travers la ligne d'action que les autres supposent qu'elle a adoptée au cours d'un contact particulier. La face est une image du moi, délimitée selon certains attributs sociaux approuvés." [Goffman, 1974]), nous n'en retiendrons que la notion d'*ego*, de sensibilité personnelle, de sensibilité aux agressions verbales.

La face, tout comme le territoire, se limite à un axe ménagement/modestie, le long duquel est délimitée une zone "acceptable". Dans l'étape d'interprétation, dont nous traitons dans cette section, l'agent est sensé définir les limites de cette zone. Nous employons là encore une stratégie relativement simple, qui consiste à prendre comme référence un profil typique pour chaque interlocuteur.

Sachant que la modalité privilégiée de manifestation de la face est langagière et que nos modules de formalisation et d'énonciation en langue naturelle sont trop basiques pour en transmettre les effets, elle n'a aucune incidence sur le comportement. Toutefois, bien que n'ayant pas de manifestation externe, elle participe aux phénomènes d'interprétation et d'expression, ce qui justifie sa présence dans cette section.

Profil interpersonnel. Le profil social correspond à une position et une valeur de *variabilité* sur la périphérie du cercle interpersonnel de McCrae et Costa.

Nous avons associé à chaque force illocutoire (et aux altérations) une valeur de contrôle et d'affiliation (nom des axes soumis/dominant et amical/hostile autour desquels est construit le cercle), permettant de les positionner sur le cercle. Nous considérons que le choix de l'acte de langage dépend, chez notre interlocuteur, d'une étape d'expression et que l'acte de langage est donc porteur d'informations sur l'interlocuteur lui-même, sur l'agent, et, de manière plus générale, sur le contexte d'énonciation. Pour simplifier le traitement, nous négligeons (cette simplification est acceptable vu le domaine de discours restreint de l'application) les problèmes pouvant émerger des croyances mutuelles de l'interlocuteur et de l'agent (voir page 103). Si l'interlocuteur, de son côté, suit un comportement particulièrement altruiste (voir page 188), le choix du performatif sera plus influencé par l'image qu'il a de l'agent que par sa propre personnalité. Dans cette situation, des stratégies plus subtiles,

¹⁵Principalement, ils se considèrent comme trop proches lorsque le visage de leur interlocuteur remplit tout leur écran; ils se considèrent à la "bonne distance" lorsqu'ils y sont amenés en utilisant la commande "faire face à" (*beam to*) de l'interface Blaxxun.

intégrant la prise en compte d'autres paramètres — principalement la personnalité de l'agent telle qu'elle est vue par son interlocuteur — seraient nécessaires.

L'agent exploite directement les valeurs de contrôle et d'affiliation associées à la force illocutoire pour les intégrer dans le profil interpersonnel de son interlocuteur. Cette intégration passe par une moyenne pondérée, permettant de limiter les brusques variations. La distance existant entre le profil au tour n et le profil au tour $n + 1$ permet la mise à jour de la valeur de *variabilité*. Une démonstration spécifique de ce processus a été faite [Chicoisne, 2000].

Les remarques apportées sur la face (modalité privilégiée de manifestation et côté élémentaire des modules dédiés à la langue), peuvent être reprises pour le profil interpersonnel. Le profil interpersonnel est présenté pour montrer de quelle manière une forme de personnalité peut être représentée et comment elle peut influencer sur le traitement d'un message. Cette influence n'apparaît qu'au niveau des traitements internes du message, les capacités de perception et d'action limitées de notre agent ne permettant pas à son comportement d'en être influencé.

b) Adaptation contextuelle

C'est lors de l'adaptation contextuelle que le message est enrichi des informations sur le contexte dont l'agent dispose. Dans notre application, nous avons restreint la variation à des domaines tenant plus du *ton* de l'énoncé — de sa présentation — que de son *sens*. Nous voulions montrer l'intérêt d'étapes telles que l'interprétation et l'expression, mais, en l'absence d'une représentation sémantique complète¹⁶, nous ne pouvions, dans cette application, prétendre à des manipulations en profondeur de la sémantique de l'énoncé. Dans l'implémentation qui a été faite de l'agent, l'enrichissement contextuel peut porter sur les deux parties de l'acte de langage : la *force illocutoire* et le *contenu propositionnel*. La *classe sémantique*, dépendant à la fois de la force illocutoire et du contenu propositionnel, est elle aussi susceptible d'être modifiée en cette étape, suivant des heuristiques propres au rôle, puisque, par définition, la classe sémantique est spécifiquement liée au rôle.

Seul le dernier des dialogues que nous présentons dans la section 6.5 met en évidence cette adaptation contextuelle. Les capacités de formalisation de la langue naturelle de l'agent ne sont pas assez subtiles, mais le respect du territoire est une adaptation contextuelle qui dépend d'une modalité spatiale, et ce phénomène est mis en évidence dans le dernier dialogue analysé.

Force illocutoire. Comme nous l'avons dit lors de la présentation des actes de communication (page 168), la force illocutoire se compose principalement des variables suivantes : But illocutoire, Contrôle, Affiliation, Degré de puissance, Nom et Altérations (les quatre premières variables conditionnant les deux dernières).

Le *but illocutoire* n'est pas interprétable. Comme nous l'avons dit en introduction, nous ne pouvons, dans cette application, effectuer des modifications sémantiques "profondes". Le but illocutoire, en tant que lien entre le contenu propositionnel (description du monde) et le monde lui-même, est tout simplement trop sensible à manipuler.

Le *degré de puissance* exprime l'intensité avec laquelle le locuteur s'implique en accomplissant l'acte de langage. En tant que tel, nous l'utilisons pour pondérer les modifications apportées à l'acte : plus le degré de puissance est élevé, moins la force

¹⁶voie sur laquelle nous avons progressé depuis, principalement grâce aux travaux de Sylvie Brincourt [Brincourt, 2003] et d'Émilie Benoit [Benoit, 2002]

illocutoire sera modifiée lors de l'interprétation. Toutefois, nous avons instauré un lien entre degré de puissance et contrôle : si l'interlocuteur a un profil interpersonnel dominant, nous augmentons le degré de puissance, et inversement, si ce profil est soumis, nous le diminuons. Cette manipulation permet de rendre compte de l'effet présenté dans le tableau 4.4 (page 99), où la "suggestion" de la mère est comprise comme un "ordre". Une fois de plus, ce mécanisme est excessivement simplificateur, car le profil interpersonnel de l'interlocuteur correspond plus à un comportement (et donc à un rôle) qu'à un statut. Or, pour reprendre l'exemple précédent de la mère et de son enfant, l'adaptation de la force illocutoire est plus liée à un statut qu'à un comportement.

Le *contrôle* et l'*affiliation* subissent les mêmes traitements. Ils sont adaptés de façon à se rapprocher du profil qui est attribué à l'utilisateur. C'est une manière un peu simpliste d'agir de façon à ce que chaque énoncé "ressemble" à ce que l'interlocuteur a l'habitude de dire. Une telle stratégie permet d'obtenir des interprétations plus conformes à l'interlocuteur mais estompe les brusques variations de ton qui seraient *voulues*. Une fois encore, une stratégie plus réaliste devrait tenir compte du profil interpersonnel de l'agent, car l'interlocuteur est susceptible de prendre en compte son propre interlocuteur (l'agent en l'occurrence) dans la préparation de ses énoncés (voir, à nouveau, les remarques de la page 103).

Le *nom* et l'*altération* ne sont pas directement modifiés dans cette étape, mais, étant liés aux paramètres précédents, ils y sont évidemment susceptibles de variation.

Contenu propositionnel. Les adaptations faites sur le contenu propositionnel sont relativement réduites. Elles se limitent à la résolution des anaphores et à la transformation d'informations de nature spatiale de façon à en obtenir une représentation qui implique plus l'agent.

Lorsque, dans le contenu propositionnel apparaît un objet désigné par "focus¹⁷", il est remplacé soit par le dernier objet qui a été central à la discussion, soit par l'objet central de l'énoncé en cours de traitement. En toute rigueur, l'information devrait être cherchée dans l'historique, mais, pour des raisons de simplification, l'agent conserve le rhème des énoncés précédents (le sien et celui de son interlocuteur) et se sert directement de cette information pour remplacer l'anaphore.

Lors du déplacement d'un individu dans le champ de perception de l'agent, le message perçu est un ensemble de sept coordonnées (trois pour la position et quatre pour l'orientation) correspondant à la nouvelle position de l'individu. Ces informations, absolues, sont converties de façon à devenir relatives à l'agent et à être chargées d'une partie des connaissances de l'agent. De ces coordonnées initiales, l'agent produit trois informations :

- il transforme la position sous forme de coordonnées polaires (distance et angle) centrées sur l'agent ;
- il renseigne la position en terme de territoire (au sens de Goffman) : l'individu est trop proche, à la bonne distance ou trop loin ;
- il précise si l'agent est dans le rayon des films.

Ces différents traitements correspondent à un enrichissement contextuel, et donc à une adaptation du message à l'agent.

¹⁷Voir note 11 page 182.

6.3.4 Expression

Comme nous l'avons dit page 157, le module d'expression remplit une fonctionnalité symétrique du module d'interprétation, pour ce qui est de l'adaptation contextuelle. Tout comme dans le module d'interprétation, l'adaptation porte sur la force illocutoire et le contenu propositionnel. Les adaptations portent sur la recherche d'un équilibre entre ce qui est acceptable (ou souhaitable) pour l'agent et ce qui est acceptable (ou souhaitable) pour son interlocuteur. Nous emploierons dans cette étape une valeur qui correspond à l'*altruisme*¹⁸, et qui précise si l'emphase doit être portée sur l'adaptation à l'agent ou à son interlocuteur.

Aucun des dialogues que nous présentons dans la section 6.5 ne met en évidence cette adaptation contextuelle, les capacités de génération en langue naturelle de l'agent n'étant pas assez subtiles.

Force illocutoire Les modifications apportées à la force illocutoire sont entraînées par la recherche du respect (1) de la *face* et (2) des règles de complémentarité du *profil interpersonnel*.

En ce qui concerne la face, la procédure est la suivante : on évalue pour chaque acte une "agressivité", principalement basée sur le degré de puissance de la force illocutoire employée et sur la présence ou non d'une altération de type "politesse". L'agent confronte ensuite cette valeur à la face de son modèle de l'utilisateur : dans l'idéal, l'agent cherchera à construire un acte de langage qui rentre dans la zone "acceptable" de la face de l'interlocuteur. Pour ramener dans cette zone la force illocutoire, l'agent peut décider d'agir directement sur le degré de puissance ou bien ajouter une altération.

En ce qui concerne les profils interpersonnels des interlocuteurs, ils sont associés, comme nous l'avons précisé lors de leurs présentations (p. 176), à une contrainte de complémentarité : une interaction se déroulera pour le mieux si les interlocuteurs ont des profils identiques suivant l'axe d'affiliation (hostile appelle hostile, amical appelle amical) et opposés suivant l'axe de contrôle (dominant appelle soumis et soumis appelle dominant). Le message à exprimer peut déjà avoir, à cette étape du traitement, ses champs contrôle et affiliation renseignés. Ces valeurs sont adaptées en fonction du profil interpersonnel de l'interlocuteur (de la manière qui vient d'être rappelée) et de celui de l'agent (de façon à être marquées par la personnalité de l'agent). La proportion respective de l'influence des profils de l'agent et de l'interlocuteur dépend de la valeur d'altruisme de l'agent. Plus l'agent aura une valeur élevée (comportement altruiste), plus le profil de l'interlocuteur sera pris en compte, aux dépens du propre profil de l'agent ; plus cette valeur sera basse (comportement égoïste) plus l'influence du profil de l'agent sera marquée, aux dépens de l'adaptation à l'interlocuteur.

On pourrait imaginer qu'un comportement purement altruiste est idéal du point de vue de l'interlocuteur. Les remarques de la page 103 montrent certains des risques d'une telle approche.

Contenu propositionnel Contrairement à l'étape d'interprétation, l'adaptation du contenu propositionnel n'est pas liée au traitement des anaphores. Le module d'énonciation en langue naturelle actuellement en place ne permettant pas la gestion des anaphores, il était inutile d'en prévoir l'intégration à ce niveau du traitement.

¹⁸Zhang et ses collaborateurs emploieraient ici leur notion (plus complète) de *Motivational Quality* [Zhang *et al.*, 2002].

Le contenu propositionnel est modifié en phase d'expression pour les messages de type *déplacement*. C'est en effet en cette étape qu'est géré le respect des *territoires*. Trois types de déplacements sont prévus : l'agent peut se déplacer jusqu'à une zone donnée (le rayon des films dont il a la charge), il peut se placer face à son interlocuteur ou bien se positionner de manière correcte vis-à-vis des territoires¹⁹. Dans le premier de ces deux derniers cas, nous ne prenons pas en compte le territoire. Dans le second, le message précise simplement de se mettre face à l'interlocuteur et le module d'expression décide des coordonnées exactes du déplacement, principalement de la distance à prendre vis-à-vis de l'interlocuteur. Le calcul de la distance est simple, puisque l'agent cherche à se placer à la fois dans le territoire acceptable de son interlocuteur, et dans le sien. Si les deux territoires ne se recouvrent pas, il cherche à se placer au plus près du territoire vers lequel l'entraîne sa valeur d'altruisme : le sien si elle est faible, celui de son interlocuteur si elle est forte. Le troisième cas est une généralisation du second, puisqu'il amène l'agent en la position la plus proche qui soit compatible avec les territoires, sans pour autant se placer face à son interlocuteur (ceci pouvant servir à suivre l'interlocuteur par exemple).

Le territoire maintient à distance : il représente des limites à ne pas franchir sous peine d'offenser le propriétaire du territoire. Mais il faut noter qu'il attire aussi : se coller à quelqu'un (tout comme le fixer du regard) est une violation, mais ne pas être proche de lui (tout comme ne pas le regarder) quand il nous parle est un affront. Tant qu'il y a compatibilité des territoires, la relation se déroule avec douceur. En quelque sorte, la liberté d'un individu s'arrête là où commence celle des autres.

6.3.5 Répartition multi-modale

Le module de répartition multi-modale sélectionne les modalités dans lesquelles le message va être exprimé. La stratégie mise en œuvre est simple et dépend de la modalité principale du message. La modalité principale est renseignée par le module de raisonnement qui a généré le message. Il pourrait ne pas renseigner cette variable et laisser le module de répartition en décider, mais nous avons choisi de laisser cette tâche aux modules de raisonnement.

Si la modalité principale est "déplacement", le module de répartition transmet le message tel quel au module d'énonciation dédié au déplacement.

Si la modalité principale est "expression corporelle", le module de répartition transmet le message tel quel au module d'énonciation dédié aux expressions corporelles.

Si la modalité principale est "texte", le traitement peut prendre deux voies, dont une triviale, empruntée lorsque le message est déjà sous une forme textuelle (et non F(p)), comme lorsqu'il arrive du module *chatbot*. Dans cette situation, le module de répartition n'a qu'à transmettre le message tel quel au module d'énonciation, dont la tâche sera elle aussi minimale, puisque le texte de l'énoncé existe déjà.

Dans le cas général d'un message de modalité principale textuelle, le module de répartition se concentre sur le nom de la force illocutoire et sur l'altération. Les forces illocutoires "saluer", "accepter" et "refuser" sont accompagnées d'expressions corporelles sémantiquement équivalentes ; les déclarations d'échec déclenchent une expression identique à celle du refus ; les altérations "enthousiaste" et "hésitant" sont respectivement accompagnées d'une manifestation d'exubérance ou de repli sur soi

¹⁹Notons que ces indications correspondent à un sens intentionné, et qu'elles seront, une fois *exprimées*, transformées en coordonnées absolues dans l'environnement de l'agent, marquant bien ainsi la fonction du module d'expression : mettre en contexte un sens.

de la part de l'agent ; enfin, certains mouvements d'accompagnement peuvent (de manière aléatoire) être déclenchés sur tout autre énoncé, de façon à briser la rigidité de l'agent.

Nous ne traitons pas les déictiques (“ce film”, avec un geste en direction d'une affiche, cassette ou DVD) pour plusieurs raisons : du point de vue de l'utilisateur, l'inadéquation de l'interface Blaxxun à ce genre d'action (la définition du monde et l'interface visuelle entraînent une difficulté à repérer l'objet pointé), et, du point de vue du développeur, le manque de contrôle sur les mouvements de l'avatar (rappelons que les expressions corporelles sont au maximum au nombre de vingt et qu'elles sont pré-enregistrées) et l'impossibilité de garantir la synchronie entre plusieurs messages (texte et mouvement, en l'occurrence).

6.3.6 Énonciation

L'étape d'énonciation est, fonctionnellement, le symétrique de celle de formalisation. Notre agent comprend trois sous-modules distincts, chacun spécifique à une modalité de l'environnement. Nous ne développerons pas les sous-modules dédiés aux déplacements et aux expressions corporelles, car leur fonctionnement est particulièrement simple : le sous-module chargé des déplacements se contente de transmettre au serveur Blaxxun la nouvelle position souhaitée, dont les coordonnées ont été calculées lors de l'expression ; le sous-module chargé des expressions corporelles se contente d'associer (s'il trouve une association possible dans son dictionnaire de mouvements) le message au numéro de l'expression corporelle, puis transmet ce dernier au serveur.

La génération de langue naturelle est le sous-module le plus délicat de l'ensemble d'énonciation. Trois grandes catégories de traitements y sont mis en œuvre pour générer un énoncé en langue naturelle :

- le message entrant peut déjà être porteur du message en langue naturelle (**phrases portées par le message**) ;
- le message entrant peut être directement traduit en langue naturelle en employant une phrase complètement définie lors de la conception de l'agent (**phrases en boîte**) ;
- le message entrant peut entraîner la sélection d'un patron de phrase, qui sera complété en fonction des autres informations portées par ce message (**phrases à trous**).

Des méthodes plus avancées de génération profonde ou stochastique n'ont pas été employées, tout d'abord pour des raisons de coût de développement, mais aussi car les premières, au fort pouvoir expressif, nécessitent des représentations sémantiques complexes dont nous ne disposons pas, tandis que les secondes, dont l'usage commence à prouver l'efficacité, nécessitent des corpus d'interaction de grande taille, dont nous ne disposons pas non plus.

La sélection de telle ou telle méthode (phrases portées par le message, à trous, en boîte) est essentiellement dépendante de la classe sémantique (voir page 172) de l'acte de communication. En effet, les classes sémantiques représentant des catégories d'actions proches, des actes de communication appartenant à la même classe sémantique vont naturellement se traduire en langue naturelle par des énoncés proches. Notons aussi un argument *a posteriori*, puisqu'une fois l'agent terminé, la discrimination des énoncés sur la base de la classe sémantique s'est révélée être un choix efficace pour sa maintenance et son extension.

Les phrases portées par le message. Dans ce cas trivial, l'acte de communication est porteur du texte à énoncer. Cette situation se rencontre principalement lorsque le message est issu du module *chatbot*, mais d'autres modules peuvent agir de la sorte, quand, par exemple, nous avons un message très précis à énoncer et que nous savons que ce message dépasserait les capacités du module d'énonciation langue naturelle. Cette solution peut sembler intéressante car elle permet de se dispenser d'une représentation sémantique (sous la forme d'un acte de communication complet) pour une représentation simplifiée contenant explicitement un message d'une modalité de l'environnement. Les dangers d'une telle approche sont multiples :

- tout d'abord, renseigner le message avec un contenu déjà spécifique à une modalité empêche son traitement par le module d'expression et de répartition multimodale (ce qui est moins grave pour ce dernier module, puisque les choix ont déjà été implicitement faits par le module ayant généré le message) ;
- ensuite, les modules rôle risquent de nécessiter des processus de construction de messages complexes, de façon à pouvoir générer eux-mêmes leurs messages, processus risquant de faire rapidement double emploi avec ceux du module d'énonciation ;
- enfin, intégrer des capacités spécifiquement dédiées à une modalité au sein d'un module rôle peut effectivement apporter un gain en matière de complexité des modules en aval, mais finirait par charger de fonctionnalités parasites, éloignées du but premier de ce module : le raisonnement.

En conséquence, nous n'avons employé cette méthode que dans deux cas : pour le module *chatbot*, qui, par principe, ne manipule que des chaînes de caractères et dans les quelques situations où le message à exprimer nécessitait un traitement particulièrement spécifique, pour lequel l'adaptation du module d'énonciation aurait été démesurée. Par exemple, certains messages d'initialisation ("*Voulez vous que je vous aide à choisir un film ?*"), de clôture ("*Clôture de négociation*"), ou de gestion de ruptures et/ou d'échec ("*Tu ne m'as pas encore donné de contraintes sur la film que tu voulais*", "*Un peu de tenue !*").

Les phrases en boîte Les phrases en boîte sont des messages intégralement pré-enregistrés. Ces phrases diffèrent des précédentes car elles sont stockées au niveau du module d'énonciation, tandis que dans le cas précédent, elles le sont au niveau du module de raisonnement. C'est une fois arrivé dans le module d'énonciation que le message (sous la forme d'un ensemble force illocutoire et contenu propositionnel, et non d'un message déjà mis en forme pour une modalité donnée) sera associé au message pré-enregistré. L'intérêt principal de cette méthode est de permettre une totale indépendance de forme entre le message tel qu'il est manipulé par l'agent et le message qui sera énoncé. Ainsi, les phrases en boîte respectent plus le découpage (et découplage) fonctionnel de l'agent, et surtout, les messages déclenchant une de ces phrases sont, jusqu'à cette étape d'énonciation, représentés sous une forme manipulable par le module d'expression et de répartition.

Les phrases à trous Les phrases à trous (ou à base de patrons, ou *template based*) ressemblent aux phrases en boîte, à ceci près qu'elles sont par endroits incomplètes et que leur complétion permet d'obtenir une variété de phrases inaccessible à un système employant uniquement des phrases pré-enregistrées. Le module d'énonciation contient des informations sur la structure de la phrase, sur les mots qui y sont déjà présents et sur les données nécessaires à la complétion de ce patron.

Une fois la catégorie d'énoncé déterminée (détermination essentiellement faite sur la base de la classe sémantique), le remplissage du patron s'effectue en employant les informations présentes dans le contenu propositionnel. Certains patrons prennent aussi en compte d'autres paramètres de la force illocutoire, principalement les altérations, de façon à pouvoir rendre la *politesse* en ajoutant un "s'il vous plaît" en début d'énoncé ou un "peut-être" en fin pour marquer l'*hésitation*.

Dans le cas où l'agent donne des informations sur un film, le patron a la forme présentée dans le tableau 6.9. Chaque élément du contenu propositionnel susceptible de prendre place dans ce patron subit un traitement particulier : une information de type `acteur(c Clint eastwood)` sera traduite en "clint eastwood joue dedans", et une de type `date(1965)` par "tourné en 1965". Une fois les informations mises en forme, elles sont intégrées dans le patron pour compléter l'énoncé.

Deux patrons pour donner des informations sur un film
<liste éléments>.
Voilà tes infos : <liste éléments>.

Élément intégrable au patron	Forme "langue naturelle"
<code>titreFilm(1001 pattes)</code>	le titre est 1001 pattes
<code>nomActeur(Robert deNiro)</code>	Robert deNiro joue dedans
<code>nomReal(Steven Spielberg)</code>	réalisé par Steven Spielberg
<code>date(1975, <)</code>	tourné avant 1975
<code>date(1975, >)</code>	tourné après 1975
<code>date(1975, ><)</code>	tourné aux alentours de 1975
<code>date(1975, =)</code>	tourné en 1975
<code>movieCat(western)</code>	genre : western

TAB. 6.9 – Un exemple de patron d'énoncé

L'emploi de patrons permet la création "à la volée" d'un grand nombre de phrases différentes ayant une forme de surface similaire. Les patrons permettent aussi de faire évoluer certaines des données sans avoir à modifier le module d'énonciation : ajouter un acteur dans la base permettra toujours la création d'énoncés du type "...<tel acteur> joue dedans."

Paraphrasage Les deux méthodes précédentes autorisent une certaine variation sur la façon d'exprimer deux messages ayant la même classe sémantique. En effet, rien n'interdit que plusieurs phrases préenregistrées (ou patrons) correspondent à une même classe sémantique, la sélection de l'une ou l'autre dépendant du hasard. Le hasard permet d'obtenir un comportement varié, même si les situations conversationnelles qui se succèdent sont rigoureusement les mêmes.

Le tableau 6.10 présente l'ensemble des paraphrases possibles pour présenter à l'utilisateur une listes de film.

Limites Le module d'énonciation en langue naturelle que nous avons mis en place est excessivement *ad hoc*. Ce problème est général dans le domaine de la génération en langue naturelle : actuellement, la quasi-totalité des générateurs en langue naturelle

1	Je peux te proposer <liste de film>
2	J'ai <liste de film>
3	J'ai <liste de film>, qu'en penses-tu ?
4	Je vois <liste de film> qui pourraient aller
5	J'ai <liste de film>, [lequel ça] te conviendrait ?

TAB. 6.10 – Un exemple de paraphrases

est dédiée à *un* domaine de génération et ils se retrouvent donc, de fait, fortement marqués par leur domaine.

Les travaux d'Emilie Benoit [Benoit, 2002] ont porté sur le développement d'un module de génération en langue naturelle qui n'a pas encore été intégré à notre agent. Ce module se base sur un logiciel libre, *charabia*²⁰, générant de la langue naturelle à partir de graphes. Cette méthode permet de générer des phrases dans lesquelles certains mots sont variables (suivant le même principe que les "phrases à trous"), mais aussi la structure même de la phrase. Ce module permet donc bien plus de flexibilité dans la génération que celui actuellement en place, il permet une plus grande puissance expressive (en terme de "subtilité" d'expression des altérations par exemple) ainsi que la génération de messages plus proches d'une langue *naturelle*.

6.4 Modules liés au raisonnement

Nous présentons dans cette section les modules directement liés au raisonnement de l'agent. Nous considérons les modules précédents et suivants (dans l'ordre de traitement de l'agent), comme étant principalement liés aux phénomènes de perception et d'action²¹. Après avoir présenté le fonctionnement général de cet ensemble de modules, nous détaillerons ces modules, en commençant par le module de *sélection*, qui joue un rôle central en décidant du module qui sera chargé du traitement du message entrant. Nous enchaînerons ensuite avec les trois modules spécifiques de traitement, celui lié au *cinéma*, celui lié à la *conversation*, et celui du *chatbot*.

6.4.1 Architecture générale

La partie raisonnement de l'agent est composée de quatre modules, dont trois jouent un rôle similaire (voir figure 6.5). L'étape de raisonnement proprement dite est effectuée par l'un des trois modules *cinéma*, *conversation* ou *chatbot*, mais un traitement préliminaire est requis, de façon à choisir *quel* module devra traiter le message entrant.

Lorsqu'un message sort de la chaîne de perception (après les étapes de formalisation, intégration multimodale et interprétation), il est confié au module de sélection. Ce module va consulter chacun des trois modules de raisonnement, et décider duquel est le plus approprié au traitement du message entrant (section 6.4.2).

²⁰<http://www.charabia.net/>

²¹Tels qu'ils sont implémentés, les modules d'interprétation et d'expression relèvent effectivement plus de la perception et de l'action que du raisonnement. Les modèles théoriques présentés dans les chapitres 4 et 5 en font des modules véritablement à mi-chemin, entre perception et raisonnement d'une part, et raisonnement et action d'autre part.

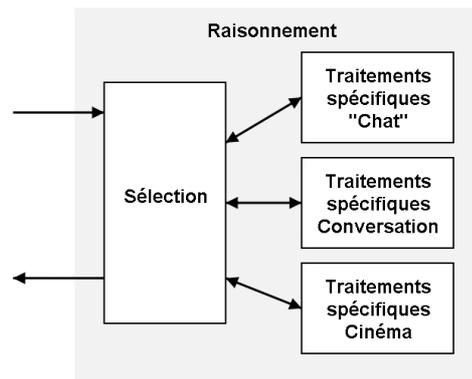


FIG. 6.5 – Raisonnement

Chaque module dispose d'une interface unifiée, permettant dans un premier temps de lui demander d'évaluer sa compétence, et dans un éventuel second temps de lui déléguer le traitement du message entrant (voir figure 6.6). Au terme de son raisonnement (traitement), le module choisi (donc, cinéma, conversation ou *chatbot*, respectivement sections 6.4.3, 6.4.4 ou 6.4.5) peut décider d'une ou plusieurs actions. Il construit alors un acte de communication représentant le sens intentionné, et cet acte de communication est ensuite confié à la partie "action" de la chaîne de traitement (c'est-à-dire les modules d'expression, répartition multimodale et énonciation).

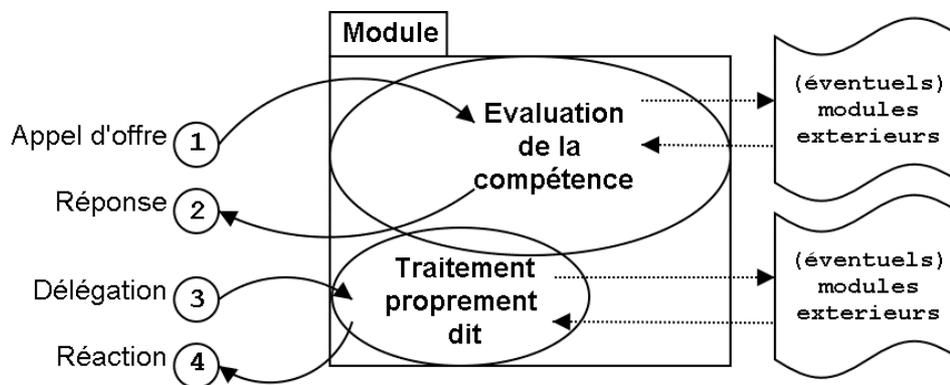


FIG. 6.6 – Fonctionnement général d'un module de raisonnement

Bien que cette fonctionnalité ne soit pas implémentée dans notre agent, il est prévu que celui-ci puisse aussi déclencher les modules de raisonnement par des messages liés à des temporisations, permettant ainsi l'apparition de comportements n'ayant pas de lien direct avec un événement de l'environnement. D'un point de vue extérieur, une telle fonctionnalité permettrait l'apparition de comportements proactifs.

6.4.2 Sélection

La stratégie sous-jacente à la sélection du module de raisonnement est basée sur un appel d'offre. Lors de l'arrivée d'un nouveau message, le module de sélection le soumet à chaque module de raisonnement et leur demande un estimation de leur

compétence ou de leur intérêt au traitement de ce message. Le traitement du message est ensuite confié au module qui a déclaré être le plus capable ou le plus intéressé.

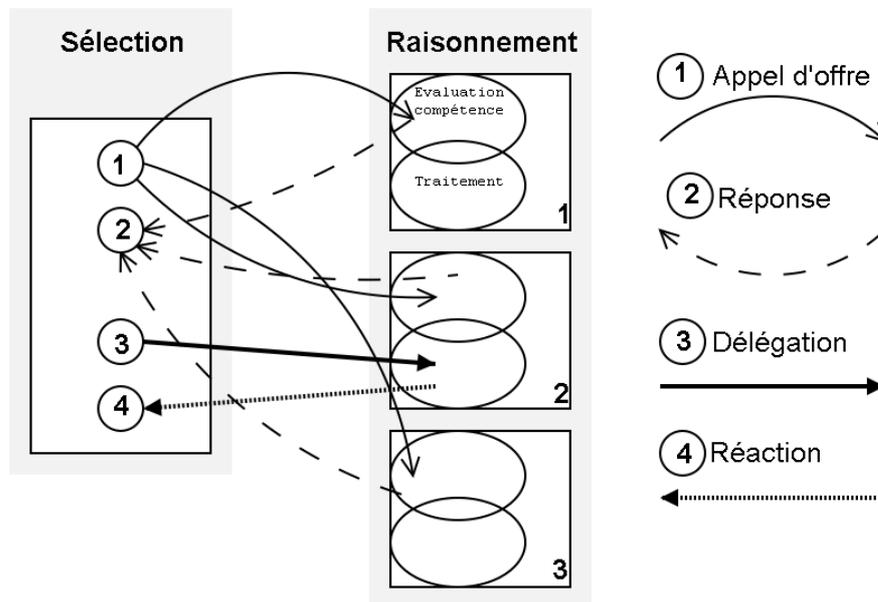


FIG. 6.7 – Séquence de raisonnement

Le séquençage détaillé d'un cycle de raisonnement est reporté sur le schéma 6.7. Lorsqu'un message arrive à l'étape de raisonnement, le module de sélection le transmet à tous les modules et demande lequel souhaite le plus être chargé du traitement (étape n° 1). Chaque module répond en renvoyant une valeur comprise entre 0 et 1 (étape n° 2). Dans l'exemple présenté, c'est le module numéro 2 qui remporte le traitement. Le module de sélection lui retransmet donc à nouveau le message, cette fois-ci pour le traiter (étape n° 3). Enfin, le module numéro 2 transmet les (éventuels) résultats au module de sélection (étape 4).

Chaque module est libre d'implémenter une solution de son choix pour mesurer son aptitude au traitement du message entrant. Nous avons posé un ensemble d'échelles de valeurs, pour obtenir un comportement cohérent au niveau de l'ensemble de l'agent. Lors de la définition des valeurs de la table 6.11, nous avons cherché à respecter trois objectifs :

1. donner une priorité supérieure aux modules ayant prévu l'arrivée d'un tel message (c'est-à-dire, lorsque les attentes du module sont respectées), ceci permettant de privilégier le suivi d'une conversation ;
2. donner une priorité directement proportionnelle à la spécificité du traitement, ceci permettant de privilégier les traitements spécifiques par rapport aux plus génériques (module Cinéma vs. module Chatbot par exemple) ;
3. donner une priorité plus importante aux énoncés plus (subjectivement) importants dans la conversation (gestion des insultes vs. négociation du film)

Steve (voir page 73) présente aussi une étape équivalente à notre "sélection", mais cette étape est plus complexe, car elle décide de l'orientation de la réponse (argumenter, répondre, ...), c'est donc un *comportement* qui est choisi et non pas un cadre dans lequel traiter le message (vente de films pour notre module cinéma ou gestion de la conversation pour notre module conversation). Nous avons pour notre

Valeurs	Type de message	Module	Remarques
0.95→1	Administration	Tous	Ce type de message peut être destiné à tout module de la chaîne de traitement.
0.90	Ruptures sociales	Conversation	Voir Goffman et le territoire (p. 107)
0.80→0.85	Enchaînement attendu	Cinéma	Voir les attentes (pp. 124 et 150)
0.75→0.80	Message spécifique “cinéma”	Cinéma	Priorité au rôle principal.
0.70	Échec de l’analyse syntaxique	Conversation	Tous les mots ont été reconnus.
0.60		Chatbot	La phrase telle qu’elle est entrée est reconnue par le <i>chatbot</i>
0.55→0.60		Chatbot	Une phrase à trous du <i>chatbot</i> est reconnue
0.35	Échec lexical	Conversation	Un seul mot inconnu.
0.25		Chatbot	Réponse du chatbot sur n’importe quelle phrase.
0.22	Fonction par défaut du rôle Cinéma	Cinéma	Tentative de reprise. Jamais déclenchée, à cause de la réponse systématique du <i>chatbot</i> .
0.20	Échec de l’analyse lexicale	Conversation	Plusieurs mots inconnus. Jamais déclenchée, à cause des deux réponses systématiques précédentes.

TAB. 6.11 – Grille de répartition des valeurs de sélection

part préféré une approche privilégiant le choix d'un traitement, plutôt que le choix d'un comportement.

Cette étape de sélection peut être cruciale dans la gestion de l'équilibre entre le rôle et la personnalité de l'agent. À ce niveau, il serait en effet possible de décider de comportements purement liés à la personnalité, car les aspects rôle et personnalité peuvent avoir des influences équivalentes sur le comportement. Nous avons, presque arbitrairement, choisi de n'associer le rôle qu'au traitement des messages entrants et à la décision d'action (modules *sélection* et *traitement*), laissant la personnalité s'exprimer au niveau de la forme que peut prendre une action (modules *interprétation* et *expression*). Le module "conversation" présente un cas limite. Par exemple, il doit réagir aux insultes, ou, de manière plus générale, aux dérapages sociaux. Ce genre de comportement dépend essentiellement de la personnalité, tant dans la détection de l'insulte (tel agent sera plus sensible que tel autre) que dans les réactions. Cette situation peut être traitée, en accord avec les modèles des chapitres 4 et 5, au niveau de l'étape d'interprétation (et de manière symétrique en expression), en influant sur le mode d'accomplissement, le degré de puissance ou les altérations. Ainsi, le module conversation reste relativement indépendant de la personnalité²². Des agents plus complexes, ou pour lesquels un comportement plus humanisé serait souhaité, devraient intégrer un module de raisonnement dédié à la personnalité. Alors, la gestion des conflits entre personnalité et rôle (ou tâche à accomplir) deviendra délicate.

Les approches mises en œuvre au sein du module de sélection sont relativement simples, mais adaptées aux fonctionnalités de l'agent telles qu'elles existent à l'heure actuelle. Un système plus souple, basé sur la compétition ou la négociation entre les différents modules de raisonnement serait nécessaire avec l'arrivée d'une plus grande variabilité dans les fonctions des modules, surtout si ceux-ci sont développés de manière *réellement* indépendante. En effet, comme le montre bien la table 6.11, nous avons calculé les réponses données par les modules en connaissance des différentes situations conversationnelles possibles mais aussi des différents modules en présence.

6.4.3 Cinéma

Le module dit "Cinéma" est, parmi les trois modules de raisonnement mis en place, le module de rôle. C'est lui qui prend en charge l'ensemble des messages ayant trait à la recommandation de films, et donc, celui qui le plus directement participe à la tâche de l'agent. Il est aussi, des trois modules, le plus complexe (en termes d'architecture et de moyens de traitement mis en œuvre) et le plus complet (en terme de couverture des concepts présentés dans les chapitres 4 et 5).

a) Architecture

L'architecture logicielle de ce module est en grande partie le fruit d'une réflexion conjointe avec Bruno Celle [Celle, 2001]. Elle se compose d'une partie générique chargée de la gestion d'un focus conversationnel quelconque et d'un ensemble d'attentes quelconques elles aussi, associée à une partie la spécialisant dans le domaine des films.

Les figures 6.8 et 6.9 présentent les deux éléments qui composent le module de traitement dédié au cinéma. La figure 6.8 décrit un module *générique* permettant

²²Pour des raisons de simplicité de programmation, nous avons tout de même déporté certains traitements au sein du module conversation.

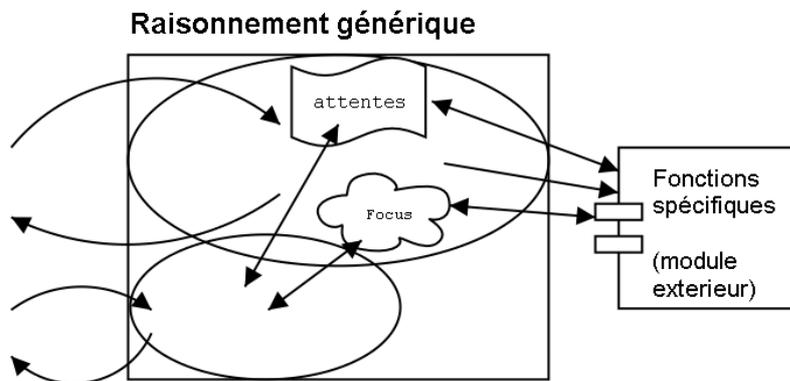


FIG. 6.8 – Les modules de traitement : partie générique (la figure 6.6 de la page 194 explicite les différents éléments du module)

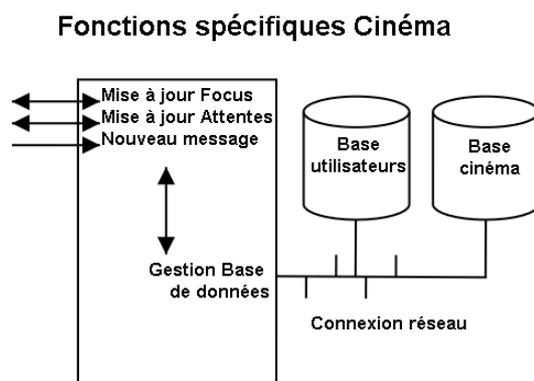


FIG. 6.9 – Les modules de traitement : partie spécifique cinéma

la gestion d'attentes et d'un focus, mais ne contient aucune donnée spécifique à un domaine. La figure 6.9 décrit quant à elle le module dans lequel le module générique ira chercher les fonctions lui permettant de se spécialiser. Ainsi, notre application propose une interface distinguant les processus abstraits de gestion de la dynamique de la conversation (attentes, focus) des connaissances spécifiques à un domaine (en l'occurrence, le cinéma).

Dans cette sous-section de présentation de l'architecture du module cinéma, nous distinguons trois parties : le focus conversationnel, la base de données et les attentes.

Focus conversationnel Le focus conversationnel est une liste de variables du type {type de critère, critère, relation}, comme nous l'avons dit dans la section 6.2.2. Les divers mécanismes de maintien de la cohérence du focus sont présentés dans la sous-section "Fonctionnement" (page 200). Le contenu de ce focus peut être directement convertit en une requête à destination de la base de données gérant les films. C'est cette requête qui fait le lien entre les contraintes (contenues dans les focus) et les préférences (le profil, contenu dans la base de données), par l'intermédiaire du critère d'ordonnement. Un exemple de traduction d'un focus vers une requête SQL est présenté dans la table 6.12 (page 202).

Base de données La base de données sur laquelle nous travaillons est segmentée en quatre grandes parties :

- Les films proprement dits. Cette partie contient l'ensemble des caractéristiques des films (acteurs, dates, origine, ...)
- Les similarités entre films. Cette partie contient, pour chaque paire de films, un valeur correspondant à leur similarité (fait partie de l'ensemble MMA).
- Les préférences des utilisateurs. Cette partie associe un vote à un film, et compose ce que nous appelons le profil des utilisateurs (fait partie de l'ensemble MMA).
- Les similarités entre utilisateurs. Cette partie contient, de même que pour les films, une valeur de similarité associée à chaque paire d'utilisateur (fait partie de l'ensemble MMA).

Cette base a été fournie telle quelle par Canal+ et nous n'avons apporté aucune modification sur la structure des tables existantes. La partie de la base sur laquelle nous nous appuyons comprend sept tables. Toutes les données sur les films (acteurs, dates, titres, etc.) sont stockées dans une même table. Trois tables sont consacrées au fonctionnement de MMA, à savoir une table pour les jugements portés par l'utilisateur, une table pour les similarité entre utilisateurs et une table pour les similarités entre films. Pour des raisons de facilité, nous avons construit une table associant aux films la moyenne de leurs notes. Enfin, deux tables nous permettent de noter les films possédés et refusés par l'utilisateur.

Attentes La partie s'occupant de la gestion des attentes se compose principalement d'une liste ordonnée des attentes en cours, d'un dictionnaire d'attentes et d'un moteur permettant de faire évoluer la liste d'attentes. Le dictionnaire contient l'ensemble des attentes possibles et est appelé lors de la construction des enchaînements d'attentes et lors de l'initialisation de la liste d'attentes (au début de l'interaction, puis au début de chaque tour de parole). Les attentes peuvent contenir des variables libres permettant de préciser l'événement déclenchant ou l'action associée, mais la création d'attentes à la volée n'est pas prévue.

b) Évaluation de la compétence (sélection)

Lors de la phase de sélection, le module rôle peut se déclarer apte suite, soit à la consultation de son focus, soit à la consultation de sa liste d'attentes, soit par défaut.

La consultation du focus peut elle-même prendre plusieurs formes. Nous avons dit dans le chapitre traitant du modèle de l'interaction (page 123) que deux types de messages étaient liés au focus : les *commentaires sur le focus* et les *tentatives de modification du focus*. Nous y ajoutons un troisième : la *proposition de solution*. De ces trois types de réactions liées au contenu du focus, une seule, au plus, sera proposée. Les *commentaires* mettant en évidence un conflit, sont prioritaires ; les *tentatives de modification* et les *solutions* ont une importance relative qui dépend du nombre de réponses possibles. Si peu de réponses (films respectant tous les critères de l'utilisateur et étant susceptibles de lui plaire) sont possibles, elles sont proposées ; si beaucoup sont disponibles, l'agent proposera plutôt l'ajout de critères restrictifs ; si aucune n'est disponible, l'agent demandera à son interlocuteur de relâcher certaines de ses contraintes.

La consultation des attentes, quant à elle, est relativement simple. Les attentes sont ordonnées, de façon à refléter leur priorité (voir le chapitre sur le modèle de l'interaction, page 125). Chaque attente contient une valeur d'*importance* et un *événement déclenchant*. La liste (ordonnée) des attentes est parcourue et la première attente déclenchable par l'événement entrant est sélectionnée et renvoie sa valeur d'*importance* au module de sélection.

Par ailleurs, le module rôle renvoie systématiquement une réponse non nulle, qui, si elle venait à remporter l'appel d'offre du module de sélection, entraînerait un comportement de tentative de reprise, avec une proposition faite d'après la consultation du focus.

En synthèse, trois mécanismes peuvent déclencher les réactions :

- le focus est systématiquement capable d'intervenir : soit il est cohérent et il peut proposer une solution ou un commentaire sur son contenu ; soit il contient un conflit, et il peut proposer une modification ;
- une attente peut éventuellement être déclenchée ;
- une réponse par défaut est systématiquement proposée.

Des trois réponses possibles (liée au focus, liée aux attentes, ou la réponse systématique), celle proposant la valeur la plus élevée est retenue et transmise au module de sélection comme réponse à l'appel d'offre.

c) Fonctionnement

L'étape de *fonctionnement* reprend l'essentiel des traitements faits en phase de sélection. Nous détaillons ici ces traitements, en séparant ceux ayant pour base le focus de ceux ayant pour base les attentes.

Focus Lorsqu'un message porteur d'une information à intégrer au focus arrive (messages appartenant à la classe sémantique²³ **InfoFocus**, ou, par extension, **InfoProfile**), cette information est extraite et confrontée au contenu du focus. Nous avons considéré plusieurs cas pouvant entraîner un conflit dans la mise à jour, selon les types de critères :

²³Voir page 172.

- **Acteurs.** Le focus n’a pas de limites quant au nombre de contraintes sur les acteurs. Le seul conflit envisagé concerne la situation pour laquelle l’interlocuteur réclame un acteur précis (ce qui correspond à un critère ressemblant à {NomActeur, Clint Eastwood, =}) qu’il avait auparavant explicitement refusé ({NomActeur, Clint Eastwood, <>}), ou *vice versa*. Dans cette situation, l’agent se contente de mettre à jour le focus, sans relever le conflit.
- **Réalisateurs & Genre.** Le focus ne peut contenir qu’une seule contrainte positive sur les réalisateurs. Si une deuxième contrainte arrive, l’agent détecte le conflit et l’exprime (“*Tu voulais auparavant un film de genre western. J’ai mis comédie à la place, comme tu viens de me le dire.*”).
- **Dates.** Nous n’avons mis aucune contraintes sur les dates. Si les diverses contraintes apportées par l’utilisateur rendent impossible la requête (un film d’avant 1970, ET datant des années 90, par exemple), ce conflit sera sanctionné par l’impossibilité de trouver un film satisfaisant toutes les requêtes de l’interlocuteur. Cette impossibilité entraînera la déclaration explicite du contenu du focus, laissant l’utilisateur libre de modifier des contraintes, particulièrement, évidemment, celles liées à la date.

Une fois les données intégrées dans le focus, l’agent construit une requête à destination de la base de données des films à partir de toutes les contraintes du focus (voir tableau 6.12).

Une fois la requête effectuée, le module regarde le nombre de résultats disponibles.

- Si ce nombre est élevé (plus de quinze dans la situation actuelle), trois réactions sont possibles (le choix entre ces trois réactions dépend d’un tirage aléatoire) : (1) l’agent propose à son interlocuteur de restreindre le focus (par exemple “*Veux-tu des acteurs particuliers dans ton film ?*”); (2) l’agent propose à son interlocuteur de préciser ses préférences (par exemple “*Y a-t-il des acteurs que tu préfères ?*”); (3) l’agent déclare le problème (par exemple “*Beaucoup de films correspondant à tes critères sont disponibles*”), sans orienter particulièrement la suite de la discussion, laissant l’interlocuteur particulièrement libre du mouvement suivant.
- S’il y a peu de solutions (moins de quinze), l’agent propose les quelques (deux à cinq) premiers films qu’il a obtenus en réponse à sa requête.
- S’il n’y a aucune solution, l’agent envoie deux messages, l’un pour indiquer l’échec de la requête (par exemple l’énoncé A19 du dialogue 3 : “*Tes critères sont trop restrictifs.*”), l’autre pour récapituler le contenu du focus (par exemple l’énoncé A18 du dialogue 3 : “*Le film que tu veux est d’un genre différent de western, tourné aux alentours de 1975, [etc.]*”).

Attentes Dès l’initialisation de l’agent, un ensemble d’attentes sont mises en place. Elles correspondent aux situations auxquelles l’agent est susceptible de réagir dès son activation (voir le tableau 6.13). À ces attentes s’ajoutent, au fur et à mesure de l’avancée de la conversation, d’autres attentes plus contextuelles.

On remarquera dans la table 6.13 les liens qui existent entre attentes et focus. Par exemple, lorsqu’une action à entreprendre nécessite la proposition d’une liste de films, il est nécessaire de considérer le focus. Dans cette situation, le focus est susceptible de réagir, non pas en fournissant une liste de films, mais en levant une erreur (pas de films disponibles par exemple). Ce mode de fonctionnement justifie le nom donné à la deuxième colonne du tableau 6.13, “action *souhaitée*”.

Lors de l’arrivée d’un nouvel événement, toutes les attentes sont parcourues

```

critere(nomReal, clint eastwood, =) ;
critere(nomActeur, gene hackman, <>) ;
critere(movieCat, western, <>) ;
critere(date, 1995, ><) ;
ordering(DBField, DbJug.Note, DESC) ;
critere(titreFilm, minuit dans le jardin du bien et du mal, <>)

```

Contenu du focus

```

SELECT DBSimF_PourInit.*, DbJug.Note
FROM DBSimF_PourInit
INNER JOIN DbJug ON DBSimF_PourInit.Id_Film = DbJug.Id_Fiche
WHERE
  (((Titre_Find NOT LIKE '%MINUITDANSLEJARDINDUBIENETDUMAL%') AND
  (Annee_Prod >= 1990 AND Annee_Prod <= 2000)AND
  (Id_Genre <>18) AND
  (Acteur1_nom NOT LIKE '%GENE HACKMAN' AND
  Acteur2_nom NOT LIKE '%GENE HACKMAN' AND
  Acteur3_nom NOT LIKE '%GENE HACKMAN' AND
  Acteur4_nom NOT LIKE '%GENE HACKMAN') AND
  (Real1_Nom LIKE '%CLINT EASTWOOD'))AND
  ((DbJug.Id_User = 316796 ) AND
  (DbJug.Possede <> 1) AND
  (DbJug.Refuse <> 1)))
ORDER BY DbJug.Note DESC

```

Requête SQL résultante

TAB. 6.12 – Conversion focus → requête SQL

Événement	(Ré)Action souhaitée
Information sur le focus	Envoi d'une liste de films
Information sur le profil	Envoi d'une liste de films
Demande de solution	Envoi d'une liste de films
Demande d'informations sur un film	Envoi d'informations
Demande directe d'un film	Clôture de négociation et début de la vente (non traitée dans notre application)
Demande de clôture	Clôture de conversation
Pénétration dans le rayon des films	Initialisation de la conversation
(peu importe)	Réponse systématique, basée sur une consultation du focus

TAB. 6.13 – Attentes initiales ayant une durée de vie infinie

pour vérifier si cet événement est susceptible de les déclencher. Si une attente peut être déclenchée, et que le module rôle est choisi par le module de sélection, alors l'action associée est entreprise par l'agent. Le tableau 6.14 rassemble les différentes actions possibles. Ces actions correspondent aux actions constatées lors de l'analyse de corpus d'interactions.

<code>sendList</code>	Proposition d'une liste de films
<code>giveInfos</code>	Envoi d'informations sur un film donné
<code>askConfirm</code>	Demande de confirmation du choix d'un film
<code>acceptMovie</code>	Action entreprise lorsque l'interlocuteur accepte un film
<code>refuseMovie</code>	Action entreprise lorsque l'interlocuteur refuse un film
<code>focusAnswer</code>	Demande de génération d'une action en fonction de la consultation du focus
<code>initConv</code>	Début de conversation
<code>askConstraints</code>	Demande de contraintes supplémentaires
<code>misc</code>	Réactions spécifiques

TAB. 6.14 – Actions associées aux attentes

À la fin du tour de parole la liste d'attentes est remise à jour. Tout d'abord, l'éventuelle attente qui a été déclenchée est supprimée de la liste. Ensuite, les durées de vie des attentes sont décrémentées, et les attentes dont la durée de vie atteint zéro sont supprimées. Puis, il existe une série d'attentes qui doivent être toujours prêtes (voir tableau 6.13) et l'on s'assure qu'elles sont effectivement là (en effet, bien qu'ayant une durée de vie infinie, elles peuvent être supprimées de la liste en ayant été déclenchées). Enfin, l'attente déclenchée pouvant elle-même spécifier des événements susceptibles d'arriver dans les tours de parole suivants (à la manière des enchaînements des protocoles), elle peut directement préciser une série d'attentes à ajouter. Ces dernières attentes sont placées en tête de liste, car, étant plus spécifiques — plus contextuelles — elles doivent avoir priorité sur les autres.

En résumé, les actions entreprises par le module Cinéma sont les suivantes :

- si l'agent a détecté un conflit sur le focus, il déclare ce conflit.
- si l'agent dispose d'un nombre de solutions raisonnables, il le propose à l'utilisateur.
- si l'agent estime que trop de solutions sont possibles, il cherche à restreindre l'espace de recherche.
- si l'agent est incapable de trouver une solution, il déclare son incompetence et/ou propose de lever certaines contraintes.
- si une attente est déclenchée, l'action à entreprendre y est directement associée.
- enfin, s'il s'agit de la réponse systématique, l'action à entreprendre est une tentative de reprise de la négociation. Dans les faits, jamais l'agent n'en arrive là car les réponses systématiques des autres modules répondent de manière plus forte à l'appel d'offre.

6.4.4 Conversation

Le module de conversation est chargé de l'ensemble des traitements tenant plus de la gestion de la conversation que de son contenu. Principalement, il s'occupe des

cas d'incompréhension (échec de l'analyse lexicale ou syntaxique), et des infractions sociales (problèmes liés à la face ou au territoire).

a) Architecture et évaluation de la compétence (sélection)

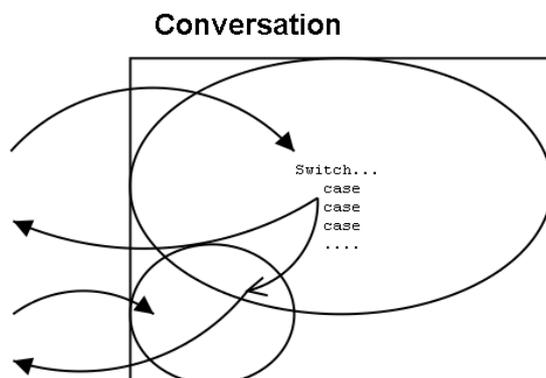


FIG. 6.10 – Les modules de traitement : la conversation (la figure 6.6 de la page 194 explicite les différents éléments du module)

L'architecture mise en place dans le module de conversation est la plus simple des trois modules de raisonnement. En effet, l'évaluation de la compétence du module se limite à une série de tests sur les différentes situations traitables par ce module (voir figure 6.10), tests effectués en ordre décroissant d'importance. Si la situation est traitable par ce module, la valeur associée au test est renvoyée, sinon, le module se déclare incompetent et renvoie zéro.

b) Fonctionnement

Le module de conversation tel qu'il existe remplit trois rôles : il s'occupe du territoire, des incompréhensions lexicales et syntaxiques et des infractions sociales²⁴.

Si le message à traiter est un déplacement, la nouvelle position de l'interlocuteur a été confrontée au territoire de l'agent en phase d'interprétation. Cette confrontation permet de déterminer la position de l'interlocuteur relativement au territoire : bien placé, trop près ou trop loin. Tant que l'interlocuteur est à la bonne distance, l'agent (du moins, le module de conversation) ne réagit pas aux déplacements. S'il se rapproche excessivement, l'agent se recule pour replacer l'interlocuteur dans une zone acceptable. Si l'interlocuteur s'éloigne de l'agent sans avoir explicitement rompu la conversation, l'agent le suivra pour le maintenir dans sa zone acceptable, entraînant l'apparition d'un comportement de suivi. Notons que le territoire de l'interlocuteur n'est pas pris en compte à cette étape, mais le sera lors de l'expression du déplacement.

Si le message à traiter est un échec de la part de l'étape de formalisation de la langue naturelle, deux cas sont pris en compte : l'échec lexical, lorsqu'un ou plusieurs mots ne sont pas reconnus, et l'échec syntaxique, lorsqu'il n'a pas été possible de

²⁴Nous nous limitons dans cette version de l'agent à la détection des insultes. Les infractions à la face de l'agent pourraient être gérées de la même façon que le territoire, car elles arrivent à cette étape sous la même forme (le traitement est effectué en phase d'interprétation) : en terme d'infraction à la modestie ou au ménagement.

trouver une structure à la phrase, bien que tous les mots aient été reconnus. Nous avons classé les différentes situations d'échec en fonction de leur gravité et de la facilité que pouvait avoir l'agent à rattraper la situation (cet ordonnancement apparaît implicitement dans le tableau 6.11). Les tentatives de reprise alors que plusieurs mots sont inconnus ont une priorité particulièrement basse de façon à pouvoir laisser la main, par exemple, à des modules qui ne partagent pas le même lexique, comme le module de *chat*. Elle n'est toutefois pas nulle, de façon à pouvoir tout de même signaler le problème à l'utilisateur si aucun autre module ne s'estimait compétent.

c) Conversation et rôle

Les situations à objectif apparemment purement social, comme l'extrait de dialogue avec REA présenté page 68, semblent indépendantes de toute tâche. Pourtant, si l'on s'intéresse aux raisons qui poussent REA (ou plutôt, ses concepteurs) à ce genre de comportement, il apparaît qu'il s'agit plus d'une fonctionnalité spécifique, ou d'une sous-fonctionnalité d'une tâche : REA a besoin de mettre l'utilisateur en confiance avant d'aborder des questions sensibles comme le salaire de l'acheteur, et elle le fait à travers le "Small talk" [Bickmore et Cassell, 2000], une discussion dont le cadre immédiat n'est pas la tâche mais les relations interpersonnelles, avec pour objectif de simplifier (voire tout simplement permettre) l'exécution de la tâche. La distinction avec une fonctionnalité sociale "générique" (respect de règles à la Goffman par exemple) n'est pas simple, et la répartition de ces fonctionnalités dans le module de conversation plutôt que dans un module de rôle est largement sujette à discussion.

En résumé, le module de conversation est techniquement le plus simple des trois modules de raisonnement. Il traite de l'ensemble des infractions sociales, et ses interventions les plus visibles ont trait au territoire. Il traite aussi des incompréhensions et est responsable des tentatives de récupération lors d'un échec de l'analyse de la langue naturelle (module de formalisation textuelle).

6.4.5 Chatbot

Le module chatbot est là pour traiter des discussions hors du champ du rôle de l'agent, traiter les situations ne nécessitant que des réactions du tac au tac, ou, plus simplement, permettre une réaction à tout message textuel (dans l'esprit même du mot *chat* : causer, bavarder, ...)

a) Architecture

L'essentiel du module chatbot est dû aux travaux de Yannick Chappellaz [Chappellaz, 2001] (voir figure 6.11). Son principe de fonctionnement est similaire à celui d'Eliza [Weizenbaum, 1966] : il dispose d'un dictionnaire, ensemble de paires de phrases, et lorsque la première phrase de cette paire est énoncée par son interlocuteur, il lui répond la seconde. Les phrases peuvent contenir un *joker* (représenté par une étoile), comme on peut le voir dans les exemples du tableau 6.15. Ce *joker* peut être retrouvé dans la phrase de réponse, après avoir subi des adaptations, principalement pronominales : transformation des "ma" en "ta" par exemple.

Pour apporter un peu de variabilité à son comportement, l'agent peut disposer de plusieurs phrases de sortie associées à la même phrase d'entrée. La sélection parmi les réponses possibles est alors aléatoire.

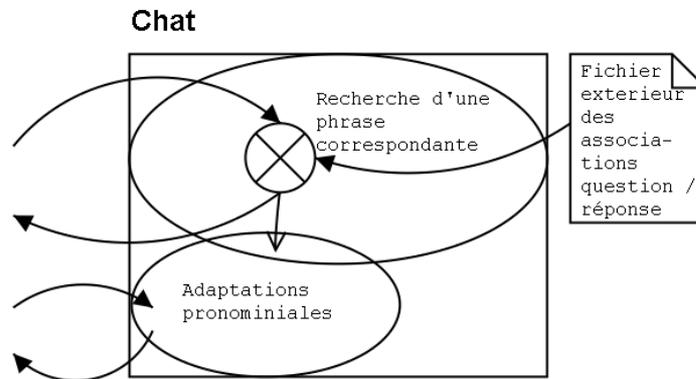


FIG. 6.11 – Les modules de traitement : le *chat* (la figure 6.6 de la page 194 explicite les différents éléments du module)

Entrée	Sortie
est ce que tu connais le *	# non, je ne sais pas ce qu'est le *.
est ce que tu connais la *	# non, je ne connais pas ce qu'est la *.
est ce que tu connais *	# non, je ne connais pas *.
mon nom est *	# * est un joli nom
je m'appelle *	# je t'appellerai donc *
dis *	# "*"
est tu *	# je ne sais pas vraiment si je suis * ou pas, mais peut etre que toi tu peux me le dire
*	# qu'est ce que "*" signifie?
*	# pardon ???
*	# je ne comprends pas ce que "*" veut dire!
bonjour	# bonjour!
salut	# salut, ca va?
merci	# De rien
merci	# Je t'en prie
merci	# Je suis là pour ca
merci	# C'est moi...
qu'est ce que tu fais?	# Je recommande des films
help	# Je recommande des films

TAB. 6.15 – Extrait des fichiers de configuration du module *chat*

La richesse du chatbot est directement liée à la richesse de son dictionnaire. De façon à enrichir simplement notre chatbot, Bruno Celle [Celle, 2001] a développé un traducteur permettant de convertir le dictionnaire du chatbot ALICE dans le format que manipule notre agent. Le choix d'ALICE est justifié par de nombreuses raisons : tout d'abord, son dictionnaire est au format AIML²⁵, ce qui en fait un format très riche et expressif, et disposant d'un grand nombre d'outils de manipulation (*parsers* par exemple); ensuite, ALICE est un projet important auquel participent de nombreux développeurs, et, commençant à se répandre, elle dispose de nombreux dictionnaires; enfin, ALICE a remporté en 2000 et 2001 le concours Loebner²⁶, montrant qu'un chatbot pouvait, malgré des mécanismes internes minimaux (mais à l'aide d'un dictionnaire énorme, puisqu'il contenait plus de 46000 associations question/réponse) participer à une conversation avec un être humain.

b) Évaluation de la compétence (sélection)

Lors de l'initialisation, l'agent charge en mémoire l'ensemble des correspondances question / réponse. Lorsqu'un nouveau message sur une modalité textuelle est perçu, le module de *chat* le confronte aux questions qu'il connaît. Nous distinguons quatre cas;

- la phrase en entrée correspond “mot pour mot” à une des questions, qui, de plus, ne contient pas de joker.
- la phrase en entrée correspond à une des questions, mais cette question contient un joker.
- la phrase en entrée correspond, si l'on peut dire, à une question, qui n'est composée *que* d'un joker. Ce cas correspond à la réponse par défaut du module chat : une question ne contenant que le joker sera déclenchée par *tout* message textuel.
- la phrase en entrée ne correspond à aucune question, pas même celles contenant des jokers (et, bien entendu, le dictionnaire ne comporte pas de “question” se limitant à un *joker*).

Selon la situation rencontrée, le module de chat s'estimera plus ou moins compétent au traitement du message entrant. La liste ci-dessus présente les situations dans l'ordre le plus favorable, allant du maximum (0,60 dans notre application) pour une correspondance absolue à zéro dans le cas de l'absence de correspondance. Cet ordre, une fois de plus, tient à notre choix de privilégier les traitements spécifiques par rapport aux traitements généraux.

c) Fonctionnement

Si le module de sélection décide de confier le traitement du message au module de *chat*, c'est que ce dernier a déjà été sollicité pour la sélection. Il a conservé de cette étape l'ensemble question/réponse associé au message, et n'a donc plus qu'à, le cas échéant, remplacer le *joker* dans la réponse. Ce remplacement est particulièrement simple, puisqu'il consiste en une simple recopie dans la réponse de la portion de texte correspondant au *joker* dans la question. Quelques adaptations pronominales sont possibles, par exemple le remplacement de “ma” en “ta”, et inversement.

²⁵AI pour Artificial Intelligence. AIML est basé sur le standard XML.

²⁶Concours basé sur une version adaptée du test de Turing, dans lequel un jury d'humains donnent des notes d'“humanité” aux chatbots avec lesquels ils conversent. <http://www.loebner.net/Prizef/loebner-prize.html>

Une fois la réponse complétée, le module de chat l'envoie. Ce message sera très directement déposé dans l'environnement, puisque, contenant un texte déjà sous une forme dépendante d'une modalité, les modules d'expression et de formalisation ne peuvent agir dessus, et que le module de répartition multimodale n'a qu'à aiguiller le message vers la modalité dans laquelle il est exprimé (ici, textuelle, par l'intermédiaire du canal de *chat*).

En résumé, le module de *chat* se limite à des considérations de surface du message. Il permet de traiter de manière simple un grand nombre de situations au prix d'un grand nombre d'associations question / réponse. Ce module est le plus facile à modifier pour prendre en charge de nouvelles situations conversationnelles, mais son manque de généralité patent et ses traitements limités à la surface du message ne le désignent que pour les tâches les plus basiques : réponse à une demande d'aide générale, entretien de la conversation, etc.

6.5 Analyses

Nous analysons dans cette section plusieurs interactions entre un humain et notre agent. Les retranscriptions présentées n'ont subi aucune retouche, particulièrement, les fautes d'orthographe²⁷ ou de syntaxe font partie de l'interaction. De même, les noms propres ou les titres apparaissent tantôt en majuscule, tantôt en minuscule, selon que l'information est issue des connaissances de l'agent (principalement sa base de données) ou de l'interlocuteur (réutilisation du rhème).

L'application choisie se prête plus à l'étude de la dynamique d'une conversation qu'à l'analyse des effets liés à l'interprétation et à l'expression. De plus, le module d'énonciation de langue naturelle tel qu'il a été implémenté n'est pas capable de suffisamment de nuances pour mettre en évidence les effets de l'expression et le module de formalisation de langue naturelle tel qu'il a été implémenté n'est pas capable de suffisamment de nuances pour permettre une interprétation. Ainsi, plusieurs informations sont présentes "en interne" mais n'apparaissent pas dans les comportements langagiers de l'agent. Par exemple, l'agent peut manipuler des messages spécifiant l'altération "politesse", mais n'a pas été doté de moyens de l'exprimer (vouvoiement, emploi de "excusez-moi", etc.).

Nous nous sommes donc limités à mettre en évidence les processus d'interprétation et d'expression à travers la gestion du territoire, en nous appuyant sur les travaux de Goffman. Ces effets sont présentés dans la dernière analyse.

Les interactions analysées se présentent dans les conditions suivantes :

Dialogue 1 un humain novice qui laisse l'agent mener la conversation (page 210).

Dialogue 2 un humain qui sait en grande partie ce qu'il cherche, mais qui se laisse guider par l'agent (page 214).

Dialogue 3 une négociation complexe où l'humain apporte un grand nombre de contraintes, revenant même sur des contraintes précédemment énoncées. Nous nous servons de cet exemple pour détailler le fonctionnement du focus (page 216).

²⁷On notera aussi que l'agent se permet à l'occasion quelques fautes. Il ne faut voir là ni une limitation des compétences en langue naturelle de l'agent ni une tentative de reproduction du style *chat*, mais tout simplement une manifestation de l'incompétence du programmeur à manier la langue française. . .

Dialogue 4 différentes situations d'échecs ou de ruptures (page 221).

Dialogue 5 une interaction quelconque, mais insistant sur les aspect non-langagiers de l'agent : déplacements, territoire et expressions corporelles (page 224).

6.5.1 Dialogue 1

Le dialogue que nous analysons dans cette section présente un humain novice qui laisse l'agent le guider.

H 01	a quoi tu sers ?
A 02	Je recommande des films
H 03	propose moi un film
A 04	Tu ne m'as pas encore donné de contraintes sur la film que tu voulais
A 05	Toutefois, je peux te proposer ceci :
A 06	Je vois LE ROI ET L'OISEAU, MON VOISIN TOTORO, HAIR, POPEYE et CHICKEN RUN qui pourraient aller...
H 07	je n'aime pas les dessins animés
A 08	Veux-tu des acteurs précis dans ton film ?
H 09	clint eastwood
A 10	J'ai PINK CADILLAC.
H 11	donne moi les plus récents
A 12	Je vois SPACE COWBOYS, JUGÉ COUPABLE, LES PLEINS POUVOIRS, SUR LA ROUTE DE MADISON et UN MONDE PARFAIT qui pourraient aller...
H 13	que sais tu sur space cowboys ?
A 14	, tourné en 2000 , Tommy Lee JONES joue dedans , Clint EASTWOOD joue dedans , réalisé par Clint EASTWOOD et le titre est SPACE COWBOYS .
H 15	JE vais le prendre
A 16	Tu veux prendre space cowboys, c'est ça ?
H 17	oui
A 18	Clôture de négociation

TAB. 6.16 – Dialogue n° 1

H 01	a quoi tu sers ?
A 02	Je recommande des films

L'humain initie l'interaction en demandant au bot quelles sont ses fonctionnalités. Ces phrases (du type “*help*”, “*aide*”, ou de manière plus complexe, “*à quoi sers tu ?*” “*qu'est-ce que tu fais*”) sont prises en charge par le module de *chat*.

En toute rigueur, l'aide devrait être distribuée entre les différents modules correspondant à des rôles de l'agent, et l'exploitation du contexte d'émission du message et des détails du message permettraient de décider du module auquel la demande d'aide s'adresse. Pour éviter de surcharger le module de formalisation de la langue naturelle, nous avons décidé d'intégrer dans le module de chat l'ensemble des phrases classiques de demande d'aide.

H 03	propose moi un film
A 04	Tu ne m'as pas encore donné de contraintes [...]
A 05	Toutefois, je peux te proposer ceci :
A 06	Je vois LE ROI ET L'OISEAU, [...]

Dans l'échange suivant, l'humain demande explicitement une liste de films (H 03). À cette requête, l'agent répond par une liste de films, puisque l'humain le lui a demandé explicitement, mais précise qu'il ne dispose d'aucune contraintes particulières sur le choix de ces films. En conséquence, les films proposés sont l'ensemble des films de la base de données. Dans cet exemple l'humain n'est pas connu de l'agent (i.e. l'agent ne dispose pas d'un profil pour cet interlocuteur), les films présentés sont ordonnés suivant les meilleurs notes de l'ensemble des films de la base²⁸. En effet, lorsqu'aucun profil n'est disponible, l'agent crée un profil générique en fonction des moyennes des notes apportées à tous les films de sa base.

H 07 je n'aime pas les dessins animés

En (H 07), l'humain apporte un jugement sur ses préférences, ce qui, au niveau de l'agent, entraîne (1) une mise à jour du profil (2) une mise à jour du focus. Lors de la mise à jour du profil (voir les résultats dans le tableau 6.17), l'agent consulte sa base de données pour en extraire une liste de films suivant deux critères. Tout d'abord, il rassemble les films correspondant aux critères sur lesquels porte le jugement (ici, les dessins animés); ensuite, il consulte la base qui rassemble les proximités entre films²⁹ de façon à rassembler les films proches. Les notes des films de ces deux listes sont ensuite ajustées en fonction de la nouvelle note (ici, "*Je n'aime pas*" est traduit par une note de 35%) et de la distance au(x) film(s) directement jugé(s) (la note des films les plus proches sera plus altérée que la note des films éloignés). Ainsi, une déclaration de préférence pour un film précis ("*j'ai adoré 1001 pattes*") permet de mettre à jour les préférences de l'utilisateur pour toute une série de films, sans quoi, l'information n'aurait qu'un intérêt limité. Dans le dialogue qui nous intéresse ici, cette mise à jour des films proches justifie la (légère) modification de note pour le film *Grease* (qui n'est pas un dessin animé), qui se retrouve derrière *Les Commitments* une fois la mise à jour effectuée (voir table 6.17).

Pour ce qui est de la mise à jour du focus, l'agent considère que toute évaluation *positive* (i.e. plus de 50%) correspond à un désir ("*j'aime tel type de film*" implique "*je veux tel type de film*") et toute évaluation *négative* (i.e. moins de 50%) correspond à un refus ("*je déteste tel type de film*" implique "*je ne veux pas tel type de film*"). Libre à l'interlocuteur de corriger explicitement ce choix, en enchaînant un "*j'aime John Wayne*" par un "*Je ne veux pas un film avec John Wayne*".

A 08 Veux-tu des acteurs précis dans ton film ?

Une fois mis à jour le profil et le focus, l'agent décide d'une action à entreprendre. Dans le cas que nous étudions, c'est le focus qui réagit, en effectuant une requête à la base de données. Les réponses étant en trop grand nombre (le seul critère étant qu'il ne doit pas s'agir d'un dessin animé), l'agent décide de proposer des contraintes supplémentaires (A08) (on notera que pour l'énoncé A06, l'agent avait donné la liste, bien que le nombre de réponses ait été plus élevé, mais il s'agissait alors de la réponse à la requête *explicite* d'une liste de films). Un tirage aléatoire pondéré sélectionne les *acteurs* comme proposition de contrainte (une stratégie de sélection plus efficace à cette étape serait de rechercher le critère le plus discriminant).

²⁸La base sur laquelle nous travaillons n'est qu'un extrait et la méthode choisie pour extraire cette partie a introduit quelques biais, entre autres, les dessins animés se retrouvent particulièrement mis en avant (cinq parmi les dix films les mieux notés).

²⁹Ces proximités sont précalculées et dépendent essentiellement de l'égalité de paramètres tels que metteur en scène, genre ou acteurs.

Profil par défaut		Mise à jour	
Note	Titre	Note	Titre
80	MON VOISIN TOTORO	78	HAIR
80	LE ROI ET L'OISEAU	75	POPEYE
78	HAIR	68	LES COMMITMENTS
75	POPEYE	Saut à l'enregistrement n° 1212	
72	CHICKEN RUN	47,1	LE ROI ET L'OISEAU
70	PRINCESSE MONONOKE	47,1	MON VOISIN TOTORO
68	GREASE	45	LES BLUES BROTHERS
68	LES COMMITMENTS	45	CHICKEN RUN
68	PINK FLOYD THE WALL	44,4	PRINCESSE MONONOKE

TAB. 6.17 – Evolution du profil de l'utilisateur (1)

A 08	Veux-tu des acteurs précis dans ton film ?
H 09	clint eastwood
A 10	J'ai PINK CADILLAC.

L'humain réagit à la proposition de contrainte (A08) en énonçant un simple nom (H09), ambigu car Clint Eastwood est à la fois un acteur et un réalisateur. Toutefois, lors de l'énonciation A08, l'agent a noté dans ses attentes qu'un nom d'acteur était une réaction probable. Ainsi, lors du traitement du tour H09 cette attente est déclenchée et l'agent traite "Clint Eastwood" comme un acteur. L'attente n'est là que pour lever cette ambiguïté puisqu'elle laisse ensuite le focus décider de l'action à entreprendre. Les deux informations contenues alors dans le focus (pas un dessin animé et Clint Eastwood comme acteur) contraignent suffisamment la requête pour que l'agent se permette une proposition. Il propose donc le film "Pink Cadillac" dans le tour A10, film pour lequel, effectivement, Clint Eastwood est acteur, et non réalisateur.

H 11	donne moi les plus récents
A 12	Je vois SPACE COWBOYS, JUGÉ COUPABLE, [...]

En H11, l'humain demande un réordonnement des solutions, qui sont implicitement classées suivant les notes contenues dans son profil. L'agent introduit dans sa requête la nouvelle contrainte d'ordonnement et propose une liste de films (A12).

H 13	que sais tu sur space cowboys ?
A 14	, tourné en 2000 , Tommy Lee JONES joue dedans , [...]

L'enchaînement H13-A14 est une digression dans la conversation, où l'humain demande des informations sur un film précis. Cet enchaînement question-réponse est géré au niveau du module Cinéma par une attente permanente. Nous avons en effet considéré que l'agent devait en permanence être capable de répondre à ce genre de demande de la part de l'utilisateur, car elles peuvent en grande partie conditionner son choix pour une réponse à une question immédiate, et doivent donc s'insérer dans la discussion, sans l'arrêter.

H 15	JE vais le prendre
------	--------------------

L'énoncé H15 fait immédiatement référence au dernier rhème, c'est-à-dire le film "Space Cowboys", il est donc transformé, dans l'étape d'interprétation de façon à contenir les mêmes informations que l'énoncé "*Je vais prendre Space Cowboys*".

H 15	JE vais le prendre
A 16	Tu veux prendre space cowboys, c'est ça ?
H 17	oui
A 18	Clôture de négociation

Les trois derniers énoncés (A16, H17 et A18) présentent une demande de confirmation, avant un aiguillage vers un système permettant la conclusion d'une vente éventuelle. Cette fonctionnalité n'est pas remplie par notre agent (elle le pourrait par l'ajout d'un module dédié par exemple), qui ne traite que la négociation du film.

6.5.2 Dialogue 2

Le dialogue que nous analysons dans cette section présente un humain qui sait en grande partie ce qu'il cherche, mais qui se laisse guider par l'agent.

H 01	J'aimerais bien un film drôle
A 02	Veux-tu des acteurs précis dans ton film ?
H 03	Woody allen
A 04	Je peux te proposer MEURTRE MYSTERIEUX A MANHATTAN et BROADWAY DANNY ROSE
H 05	je cherche un film des années 60
A 06	J'ai PRENDS L'OSEILLE ET TIRE TOI.
H 07	donne moi des infos sur PRENDS L'OSEILLE ET TIRE TOI
A 08	Voilà tes infos : , tourné en 1969 , Janet MARGOLIN joue dedans , Woody ALLEN joue dedans , réalisé par Woody ALLEN et le titre est PRENDS L'OSEILLE ET TIRE TOI .
H 09	non
A 10	Tu cherches un film qui ne soit pas PRENDS L'OSEILLE ET TIRE TOI , tourné aux alentours de 1965 , dans lequel joue woody allen et dont le genre est comedie , c'est ça ?
A 11	Je n'ai pas de film correspondant à tous tes critères...
H 12	tant pis
A 13	tant pis
A 14	Clôture de négociation
A 15	Au revoir !

TAB. 6.18 – Dialogue n° 2

H 01	J'aimerais bien un film drôle
------	-------------------------------

L'humain initie le dialogue en précisant directement une contrainte sur le film qu'il recherche (H01). Le lexique de la partie formalisation langue naturelle considère "drôle" comme un genre de film et y associe la glose "comédie", catégorie correspondante dans la description des films.

A 02	Veux-tu des acteurs précis dans ton film ?
------	--

L'agent intègre cette contrainte dans le focus conversationnel, et, après une requête à sa base de données estime que le nombre de réponses possibles pour les comédies (347 enregistrements) est trop grand pour être présenté. Il propose alors à l'humain de rajouter des contraintes et le guide vers une contrainte liée aux acteurs.

A 02	Veux-tu des acteurs précis dans ton film ?
------	--

H 03	Woody allen
------	-------------

La réponse de l'humain (H03) serait ambiguë hors contexte, car "Woody Allen" peut correspondre à la fois à un acteur et à un réalisateur. Une fois encore, les attentes permettent de lever l'ambiguïté et de prendre Woody Allen comme contrainte sur les acteurs (ce qui apparaît clairement lorsque, dans l'énoncé A10, l'agent déclare le contenu de son focus conversationnel) et non pas sur le réalisateur.

A 04	Je peux te proposer MEURTRE MYSTERIEUX A MANHATTAN et BROADWAY DANNY ROSE
------	---

Cette fois-ci, les informations contenues dans le focus permettent de cerner un ensemble plus restreint de cinq films, parmi lesquels l'agent propose le deux les mieux notés (A04)

A 04	Je peux te proposer MEURTRE MYSTERIEUX A MANHATTAN et BROADWAY DANNY ROSE
H 05	je cherche un film des années 60

Il est délicat de faire des suppositions sur les mécanismes internes à l'humain, on peut supposer par exemple qu'il n'avait pas déclaré à l'agent toutes les contraintes auxquelles il avait pensé, ou bien qu'à la vue des films proposés, il a pensé à un critère supplémentaire, toujours est-il que l'énoncé H05 ne fait pas directement référence à l'énoncé précédent, mais ajoute une contrainte supplémentaire, sur la période de réalisation du film.

A 06	J'ai PRENDS L'OSEILLE ET TIRE TOI.
H 07	donne moi des infos sur PRENDS L'OSEILLE ET TIRE TOI
A 08	Voilà tes infos : , tourné en 1969 , Janet MARGOLIN joue dedans , Woody ALLEN joue dedans , réalisé par Woody ALLEN et le titre est PRENDS L'OSEILLE ET TIRE TOI .
H 09	non

Les quatre échanges suivants montrent une légère digression, pendant laquelle l'humain va demander des informations complémentaires (H07-A08) avant d'apporter une réponse à l'agent (H09). Cette digression suivie d'une reprise de la discussion est possible grâce au temps de vie des attentes mises en place par l'agent lors de l'énoncé A06 (voir page 131 et suivantes).

A 06	J'ai PRENDS L'OSEILLE ET TIRE TOI.
	(un échange omis)
H 09	non
A 10	Tu cherches un film [...] et dont le genre est comedie , c'est ça ?
A 11	Je n'ai pas de film correspondant à tous tes critères...

La contrainte apportée par l'énoncé H09 (refus du film) entraîne que l'agent est incapable de trouver un film correspondant aux contraintes du focus conversationnel. Il déclare donc le contenu du focus (A10) ainsi que son incompétence (A11).

H 12	tant pis
A 13	tant pis
A 14	Clôture de négociation
A 15	Au revoir !

L'agent traite l'énoncé H12 comme une clôture de conversation.

6.5.3 Dialogue 3

L'extrait de dialogue que nous analysons dans cette section présente une négociation complexe où l'humain apporte un grand nombre de contraintes, revenant même sur des contraintes précédemment énoncées.

	(début omis)
H 05	je cherche un film d'eastwood
A 06	J'ai JUGÉ COUPABLE, MINUIT DANS LE JARDIN DU BIEN ET DU MAL et LES PLEINS POUVOIRS qu'en penses-tu ?
H 07	je préfère les westerns
A 08	Je peux te proposer IMPITOYABLE et PALE RIDER LE CAVALIER SOLITAIRE.
H 09	donne moi les plus anciens
A 10	J'ai L'HOMME DES HAUTES PLAINES, PALE RIDER LE CAVALIER SOLITAIRE et IMPITOYABLE.
H 11	que sais tu d'impitoyable ?
A 12	tourné en 1992 , Gene HACKMAN joue dedans , Clint EASTWOOD joue dedans , réalisé par Clint EASTWOOD et le titre est IMPITOYABLE .
H 13	je n'aime pas gene hackman
A 14	Je vois L'HOMME DES HAUTES PLAINES et PALE RIDER LE CAVALIER SOLITAIRE qui pourraient aller...
H 15	propose moi un film des années 70
A 16	Je peux te proposer L'HOMME DES HAUTES PLAINES.
H 17	je ne veux plus un western
A 18	Le film que tu veux est d'un genre différent de western , tourné aux alentours de 1975 , sans gene hackman et réalisé par clint eastwood.
A 19	Tes critères sont trop restrictifs.
H 20	donne moi un film des années 80
A 21	Je vois BIRD et CHASSEUR BLANC COEUR NOIR qui pourraient aller...
H 22	donne moi un film des années 90
A 23	Je peux te proposer UN MONDE PARFAIT et SUR LA ROUTE DE MADISON.
H 24	propose moi les meilleurs
A 25	J'ai JUGÉ COUPABLE et MINUIT DANS LE JARDIN DU BIEN ET DU MAL.
H 26	j'ai adoré MINUIT DANS LE JARDIN DU BIEN ET DU MAL
A 27	J'ai JUGÉ COUPABLE, UN MONDE PARFAIT, SUR LA ROUTE DE MADISON et SPACE COWBOYS qu'en penses-tu ?
H 28	donne moi un film de woody allen
A 29	Tu voulais auparavant un film dont le réalisateur était clint eastwood. J'ai mis woody allen à la place, comme tu viens de me le dire.
A 30	Je vois ESCROCS MAIS PAS TROP et COUPS DE FEU SUR BROADWAY qui pourraient aller...
H 31	non
A 32	J'ai MEURTRE MYSTERIEUX A MANHATTAN.
<i>suite page suivante...</i>	

H 33	d'accord
A 34	Tu veux prendre MEURTRE MYSTERIEUX A MANHATTAN, c'est ça ?
	(clôture identique aux exemples précédents)

TAB. 6.19: Dialogue n° 3

La prise de contact a été omise dans ce dialogue, de façon à se concentrer sur la constitution du focus. Au début de cet extrait, le focus conversationnel est vide. Dans la suite de l'analyse, les passages encadrés montrent le contenu du focus conversationnel au fur et à mesure de son élaboration. Ces extraits sont directement issus du fichier qui trace le fonctionnement de l'agent et comportent les éléments suivants :

- **ordering** spécifie le critère suivant lequel les réponses doivent être ordonnées. Ce critère peut être la note du film ou sa date de tournage. Les paramètres correspondent directement aux noms des champs de la base employés pour le tri, `DbJug.Note` pour les notes et `DBSimF_pourInit`. `Annee_Prod` pour les dates.
- **critere** précise les critères permettant de spécifier un film.

Contenu du focus : <code>ordering(DBField, DbJug.Note, DESC)</code> (initial)
--

H 05	je cherche un film d'eastwood
A 06	J'ai JUGÉ COUPABLE, MINUIT DANS LE JARDIN DU BIEN ET DU MAL et LES PLEINS POUVOIRS qu'en penses-tu ?

En H05, l'utilisateur pose comme critère que le réalisateur du film doit être Clint Eastwood. L'emploi d'un analyseur lexico-syntaxique en phase de formalisation, permet ici de distinguer les films "d'Eastwood" d'avec les films "avec Eastwood", les premiers caractérisant un réalisateur, les seconds un acteur. Le nombre de films d'Eastwood étant réduit, l'agent propose les trois premiers (A06).

Contenu du focus : <code>critere(nomReal, clint eastwood, =) ;</code> (après H05) <code>ordering(DBField, DbJug.Note, DESC)</code>

H 07	je préfère les westerns
A 08	Je peux te proposer IMPITOYABLE et PALE RIDER LE CAVALIER SOLITAIRE.

En H07, l'utilisateur déclare une préférence, ce qui va donc modifier son profil, tout en ajoutant la contrainte dans le focus. La modification de profil a déjà été détaillée dans l'analyse du dialogue n° 1.

Contenu du focus : <code>critere(nomReal, clint eastwood, =) ;</code> (après H07) <code>ordering(DBField, DbJug.Note, DESC) ;</code> <code>critere(movieCat, western, =)</code>

H 09	donne moi les plus anciens
A 10	J'ai L'HOMME DES HAUTES PLAINES, PALE RIDER LE CAVALIER SOLITAIRE et IMPITOYABLE.

En H09, l'utilisateur demande à l'agent de changer l'ordonnancement des réponses qui lui sont présentées, le comportement par défaut triant les films par ordre de notes décroissantes.

Contenu du focus : (après H09)	<code>critere(nomReal, clint eastwood, =) ;</code> <code>critere(movieCat, western, =) ;</code> <code>ordering(DBField, DBSimF_pourInit.Annee_Prod,</code> <code>ASC)</code>
-----------------------------------	---

H 11	que sais tu d'impitoyable ?
------	-----------------------------

A 12	tourné en 1992 , Gene HACKMAN joue dedans , Clint EASTWOOD joue dedans , réalisé par Clint EASTWOOD et le titre est IMPITOYABLE .
------	---

Les échanges H11 et A12 sont identiques à ceux vus dans les analyses précédentes.

H 13	je n'aime pas gene hackman
------	----------------------------

A 14	Je vois L'HOMME DES HAUTES PLAINES et PALE RIDER LE CAVALIER SOLITAIRE qui pourraient aller...
------	--

Après avoir vu les informations sur le film "Impitoyable" (A12), l'utilisateur déclare une nouvelle préférence (ou plutôt une animosité). Une fois de plus (voit énoncé H07), le profil de l'utilisateur est modifié : tous les films où joue Gene Hackman, ainsi que les films qui en sont proches, voient leur note baissée, et l'agent note dans son focus que l'utilisateur ne souhaite pas de film où joue Gene Hackman.

Contenu du focus : (après H13)	<code>critere(nomReal, clint eastwood, =) ;</code> <code>critere(movieCat, western, =) ;</code> <code>ordering(DBField, DBSimF_pourInit.Annee_Prod,</code> <code>ASC) ;</code> <code>critere(nomActeur, gene hackman, <>)</code>
-----------------------------------	--

H 15	propose moi un film des années 70
------	-----------------------------------

L'utilisateur continue à ajouter des contraintes; en H15, il précise une période de temps.

Contenu du focus : (après H15)	<code>critere(nomReal, clint eastwood, =) ;</code> <code>critere(movieCat, western, =) ;</code> <code>ordering(DBField, DBSimF_pourInit.Annee_Prod,</code> <code>ASC) ;</code> <code>critere(nomActeur, gene hackman, <>);</code> <code>critere(date, 1975, ><)</code>
-----------------------------------	---

H 17	je ne veux plus un western
------	----------------------------

En H17, l'utilisateur décide de lever une des contraintes qu'il avait précédemment donnée. Cet énoncé n'est pas traité par l'agent comme un conflit dans son focus, bien que l'information entrante, `critere(movieCat, western, <>)` soit incompatible avec l'information présente, `critere(movieCat, western, =)`. L'agent estime en effet qu'il s'agit de la mise à jour d'un critère, et non de son remplacement (comme cela sera le cas pour l'énoncé H28).

Contenu du focus : (après H15)	<pre>critere(nomReal, clint eastwood, =) ; ordering(DBField, DBSimF_pourInit.Annee_Prod, ASC) ; critere(nomActeur, gene hackman, <>) ; critere(date, 1975, ><) ; critere(movieCat, western, <>)</pre>
-----------------------------------	---

A 18	Le film que tu veux est d'un genre different de western ,[...]
------	--

A 19	Tes critères sont trop restrictifs.
------	-------------------------------------

H 20	donne moi un film des années 80
------	---------------------------------

(...)	
-------	--

H 22	donne moi un film des années 90
------	---------------------------------

Arrivé à ce point, l'agent ne trouve aucun film respectant l'ensemble des contraintes, il récapitule donc l'ensemble des contraintes qu'il a rassemblées (A18) et déclare son incompétence (A19). L'utilisateur demande donc successivement une modification de la date du film, des années 70 aux années 80 (H20) puis 90 (H22). De la même façon que nous l'avons dit précédemment, ceci est vu par l'agent comme une mise à jour d'un critère, et non comme un conflit dans son focus.

Contenu du focus : (après H22)	<pre>critere(nomReal, clint eastwood, =) ; ordering(DBField, DBSimF_pourInit.Annee_Prod, ASC) ; critere(nomActeur, gene hackman, <>) ; critere(movieCat, western, <>) ; critere(date, 1995, ><)</pre>
-----------------------------------	---

H 24	propose moi les meilleurs
------	---------------------------

En H24, l'utilisateur demande de changer à nouveau l'ordre de présentation de résultats.

Contenu du focus : (après H24)	<pre>critere(nomReal, clint eastwood, =) ; critere(nomActeur, gene hackman, <>) ; critere(movieCat, western, <>) ; critere(date, 1995, ><) ; ordering(DBField, DbJug.Note, DESC) ;</pre>
-----------------------------------	--

H 26	j'ai adoré MINUIT DANS LE JARDIN DU BIEN ET DU MAL
------	--

L'énoncé H26 entraîne, comme pour chaque déclaration de préférence, la mise à jour du profil de l'utilisateur et du focus. Toutefois, comme le critère est un film, le traitement est un peu particulier : l'agent exclut de sa requête tous les films pour lesquels le client a exprimé explicitement un jugement.

Contenu du focus : (après H26)	<pre>critere(nomReal, clint eastwood, =) ; critere(nomActeur, gene hackman, <>) ; critere(movieCat, western, <>) ; critere(date, 1995, ><) ; ordering(DBField, DbJug.Note, DESC) ; critere(titreFilm, minuit dans le jardin du bien et du mal, <>)</pre>
-----------------------------------	--

H 28	donne moi un film de woody allen
A 29	Tu voulais auparavant un film dont le réalisateur était clint eastwood. J'ai mis woody allen à la place, comme tu viens de me le dire.
A 30	Je vois ESCROCS MAIS PAS TROP et COUPS DE FEU [...]

Dans l'énoncé suivant (H28), l'utilisateur ajoute à nouveau un critère qui entre en conflit avec ce qu'il avait dit auparavant. L'agent relève le conflit et précise son attitude vis à vis de ce conflit (A29), puis continue l'interaction (A30). Cette stratégie a été choisie, plutôt qu'une demande de confirmation systématique pourtant plus en accord avec la notion de co-construction de la conversation, pour maintenir la fluidité de la conversation. L'utilisateur est tenu au courant des actions de l'agent sur son focus, libre à lui d'intervenir explicitement pour remettre Clint Eastwood comme contrainte sur le réalisateur, ou pour exprimer plus précisément qu'il souhaite un film dans lequel Woody Allen serait acteur, par exemple.

Contenu du focus : (après H26)	<pre>critere(nomActeur, gene hackman, <>) ; critere(movieCat, western, <>) ; critere(date, 1995, ><) ; ordering(DBField, DbJug.Note, DESC) ; critere(titreFilm, minuit dans le jardin du bien et du mal, <>) ; critere(nomReal, woody allen, =) ;</pre>
-----------------------------------	---

A 30	Je vois ESCROCS MAIS PAS TROP et COUPS DE FEU SUR BROADWAY [...]
H 31	non

Dans le dernier énoncé de l'utilisateur que nous analyserons ici (H31), l'utilisateur refuse les films que l'agent lui propose. Ce refus correspond à une attente qui entraîne l'ajout des films refusés dans le focus. Ainsi, tout se passe comme si l'utilisateur avait explicitement dit "*Je ne veux pas* « coups de feu sur Broadway »" et "*Je ne veux pas* « escrocs mais pas trop »".

Contenu du focus : (final)	<pre>critere(nomActeur, gene hackman, <>) ; critere(movieCat, western, <>) ; critere(date, 1995, ><) ; ordering(DBField, DbJug.Note, DESC) ; critere(titreFilm, minuit dans le jardin du bien et du mal, <>) ; critere(nomReal, woody allen, =) ; critere(titreFilm, COUPS DE FEU SUR BROADWAY, <>) ; critere(titreFilm, ESCROCS MAIS PAS TROP, <>)</pre>
-------------------------------	---

6.5.4 Dialogue 4

L'extrait de dialogue que nous analysons dans cette section présente différentes situations d'échecs ou de ruptures.

H 01	donne moi un film de Depardieu
A 02	Qu'est ce que "donne moi un film de Depardieu" signifie ?
H 03	'scuse, je voulais dire "avec depardieu"
A 04	Je ne comprends pas ce que "'scuse, je voulais dire "avec depardieu"' veut dire!
H 05	Donne moi un film avec depardieu
A 06	Je vois MANON DES SOURCES et JEAN DE FLORETTE qui pourraient aller...
H 07	j'en vauX d'autres
A 08	Je ne connais pas le mot "vauX". Est il correctement écrit ?
H 09	j'en veuX d'autres
A 10	Le film que tu veuX est different de JEAN DE FLORETTE , different de MANON DES SOURCES et avec elisabeth depardieu .
A 11	Tu cherches quelque chose de trop précis, je n'arrive pas à trouver de film
H 12	je veuX un film avec GERARD depardieu
A 13	Tu cherches un film dans lequel joue gérard depardieu , qui ne soit pas JEAN DE FLORETTE , qui ne soit pas MANON DES SOURCES et dans lequel joue elisabeth depardieu , c'est ça ?
A 14	Je n'ai pas de film correspondant à tous tes critères...
H 15	je ne veuX pas un film avec elisabeth depardieu
A 16	Es-tu sûr de l'orthographe de "elisabeth" ?
H 17	Tu serais pas un peu con ???
A 18	Un peu de tenue!
H 19	je ne veuX pas un film avec elisabeth depardieu
A 20	Tu veuX un film d'un réalisateur particulier ?
(La suite de l'interaction est omise)	

TAB. 6.20: Dialogue n° 4

H 01	donne moi un film de Depardieu
A 02	Qu'est ce que "donne moi un film de Depardieu" signifie ?

Le premier énoncé (H01) n'est pas analysable car le module de formalisation attend à la place du mot "Depardieu" un nom que les traits sémantiques peuvent classer comme réalisateur. L'analyse s'arrête donc sur un échec syntaxique : tous les éléments de la phrase ont été reconnus dans le lexique, mais aucune structure grammaticale correcte ne permet de les rassembler. La réaction de l'agent (A02) est prise en charge par le module Conversation.

H 03	'scuse, je voulais dire "avec depardieu"
A 04	Je ne comprends pas ce que "'scuse, [...]"

L'échange suivant (H03-A04), en apparence identique, est en fait traité par le module *chat*. Il y a deux échecs lexicaux car les mots "voulais" et "scuse" ne sont pas contenu dans le lexique. Lorsqu'un seul mot manque, le module de conversation tente une reprise, comme on peut le voir dans l'énoncé A08. Si plusieurs mots ne sont pas reconnus, le module s'estime être si peu compétent qu'il laisse la main au module *chat*, ce qui est le cas ici.

A 06	Je vois MANON DES SOURCES et [...]
H 07	j'en vauX d'autres
A 08	Je ne connais pas le mot "vauX". Est il correctement écrit ?
H 09	j'en veuX d'autres

L'énoncé H07 contient un mot inconnu, l'agent le signale à l'utilisateur, qui est alors libre de reformuler l'ensemble de son énoncé s'il pense que le mot est correct mais inconnu de l'agent, ou bien de corriger la faute, ce qui est le cas en H09. Cet énoncé (H09) est compris comme un refus et entraîne le retrait des films précédemment proposés. Notons une fois de plus que cette action est possible grâce aux attentes prévues lors de la préparation de l'énoncé A06.

A 10	Le film que tu veuX est [...] avec elisabeth depardieu .
A 11	Tu cherches quelque chose de trop précis[...]

Suite aux énoncés A10 et A11, l'utilisateur réalise un *quiproquo* sur le nom de Depardieu. L'utilisateur pensait à Gérard Depardieu, tandis que lors de la phase d'analyse, la stratégie de résolution des ambiguïtés se limite à la sélection de la première analyse correcte³⁰, et cette première analyse correcte s'est faite avec Élisabeth Depardieu et non Gérard. Élisabeth Depardieu jouant effectivement elle aussi dans Jean de Florette et Manon des sources, le *quiproquo* a duré plusieurs tour de parole avant d'être détecté.

H 12	je veuX un film avec GERARD depardieu
------	---------------------------------------

En H12, l'utilisateur cherche à résoudre le *quiproquo*, en précisant qu'il désire un film avec Gérard Depardieu, mais l'agent ne considère cet énoncé que comme une simple nouvelle contrainte sur le focus, et l'intègre donc, sans pour autant supprimer la contrainte sur Élisabeth Depardieu.

H 15	je ne veuX pas un film avec elizabeth depardieu
A 16	Es-tu sûr de l'orthographe de "elizabeth" ?

L'utilisateur est donc obligé de déclarer explicitement son désir de supprimer la contrainte sur Élisabeth Depardieu, ce qu'il tente en H15. Il fait malheureusement une faute d'orthographe sur Élisabeth, qu'il orthographie avec un "z". De la même façon que pour l'énoncé H07, l'agent signale que l'utilisateur a employé un mot qui n'est pas dans son lexique.

H 17	Tu serais pas un peu con ???
A 18	Un peu de tenue !

³⁰Ce choix extrêmement réducteur au regard de la richesse de la conversation humaine est dû à un choix technologique. Nous nous sommes concentrés, lors de la réalisation, sur les aspects de la dynamique de la conversation, la résolution des ambiguïtés étant secondaire.

Sur l'échange H17-A18, le module de conversation relève une infraction sociale, ici une insulte³¹, et la fait remarquer à l'utilisateur. Ce comportement, d'ordre social, est pris en charge par le module conversation qui maintient une liste d'insultes et de tournures de phrases agressives. Le mécanisme de reconnaissance est toutefois suffisamment précis pour réagir à des énoncés comme (H17) tout en laissant passer des énoncés comme "*j'ai bien aimé* « le dîner de cons »". La réaction aux infractions sociales se limite au premier mouvement des rituels réparateurs tels que Goffman les a identifiés : la sommation³² (L'offensé attire l'attention de l'offenseur sur le point désobligeant).

H 19	je ne veux pas un film avec elisabeth departieu
A 20	Tu veux un film d'un réalisateur particulier ?

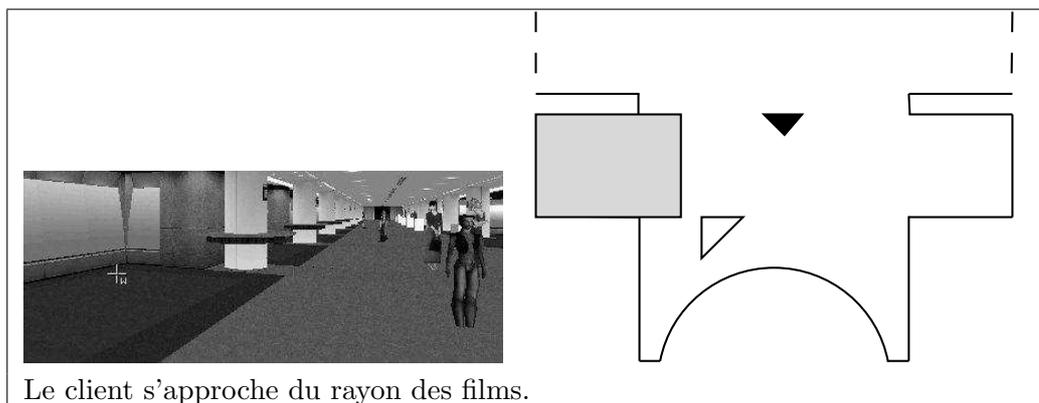
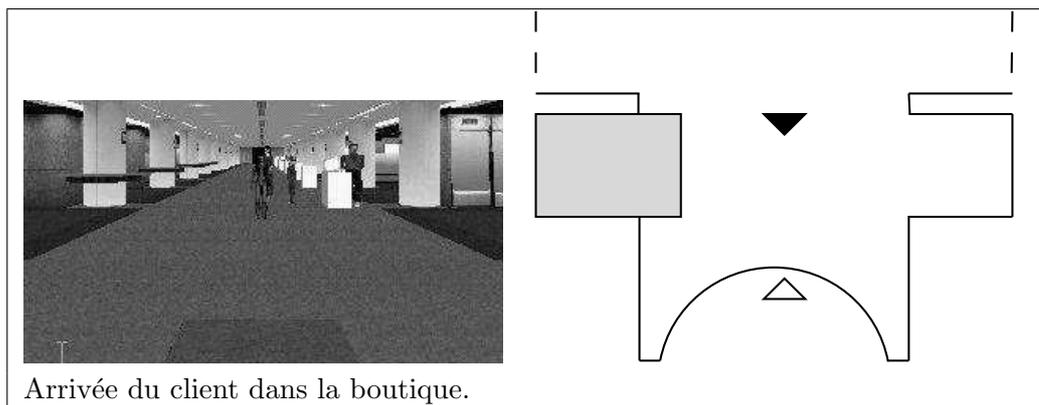
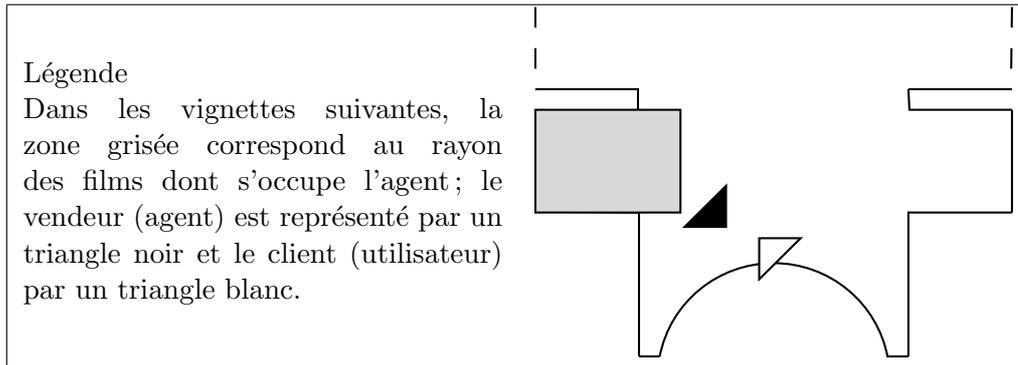
Enfin, à partir de l'énoncé H19, où l'utilisateur répète, sans la faute d'orthographe, l'énoncé H15, la discussion se poursuit d'une manière équivalente à celle qui a pu être présenté dans les dialogues précédents.

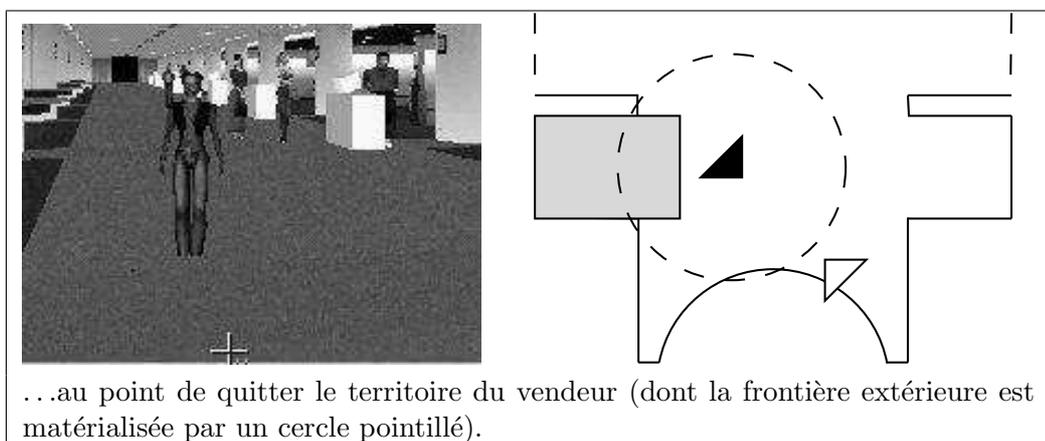
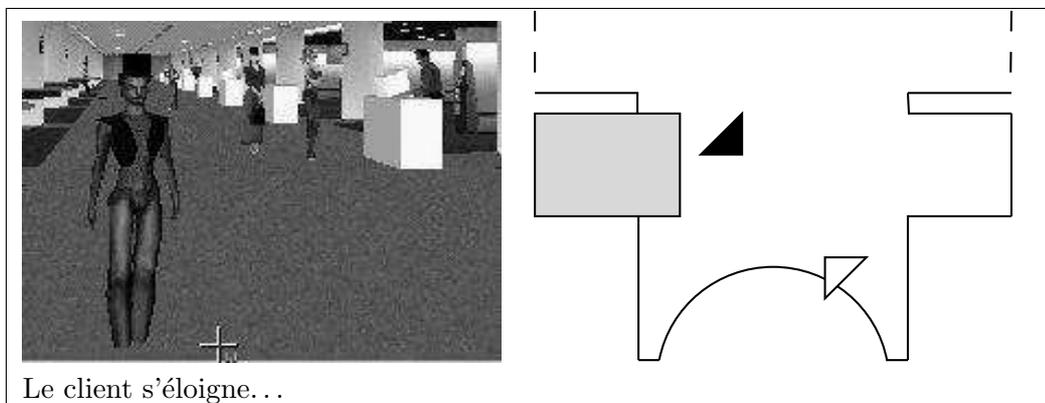
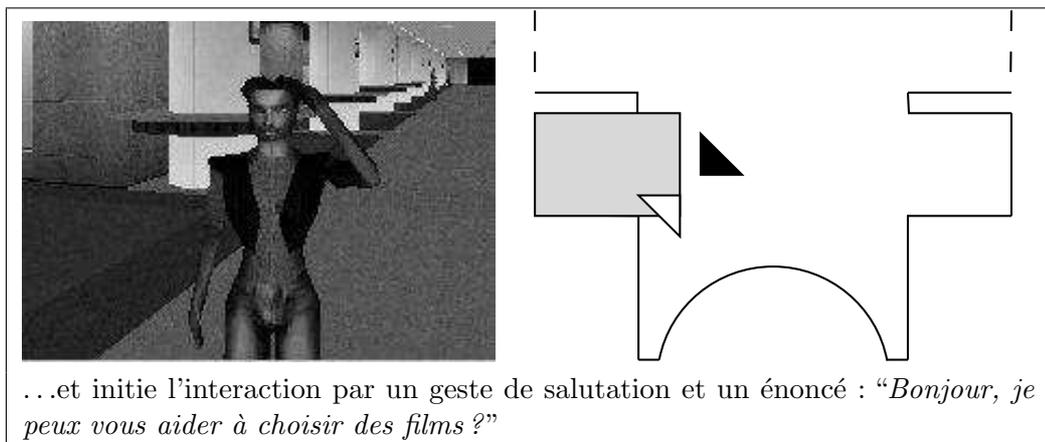
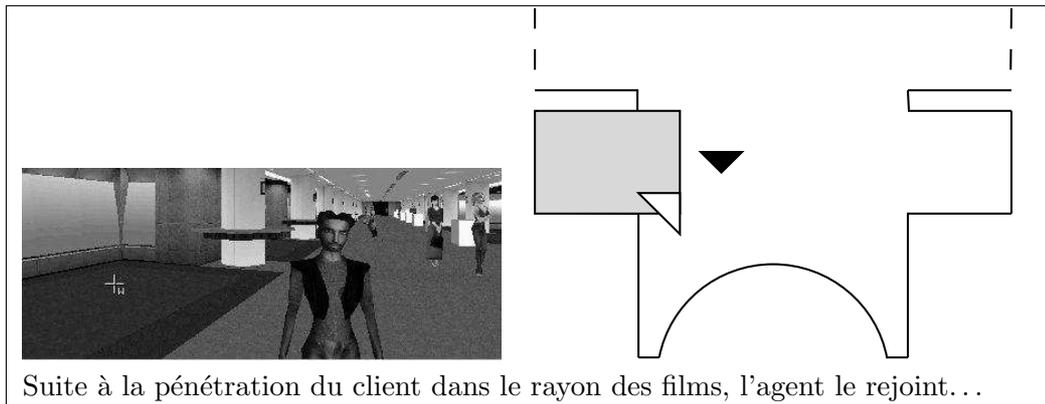
³¹Au vu des échanges précédents, et de l'erreur, apparemment insignifiante, bloquant le traitement de l'agent, l'irritation de l'utilisateur est bien compréhensible. . .

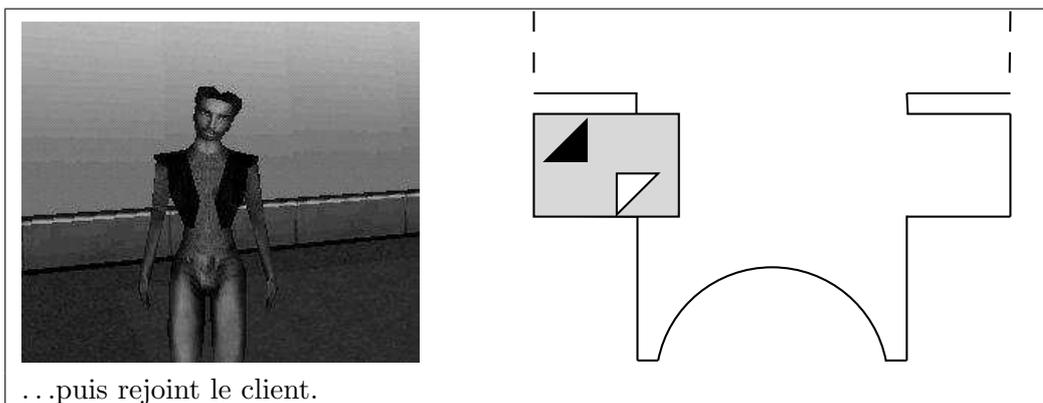
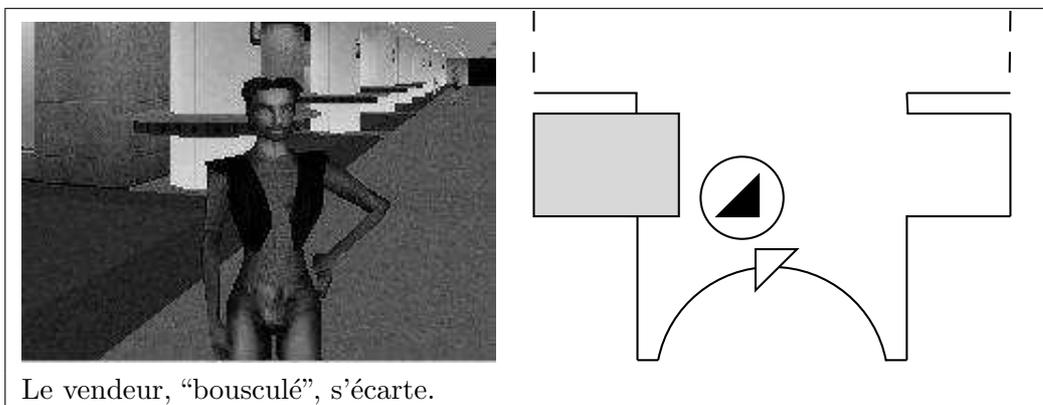
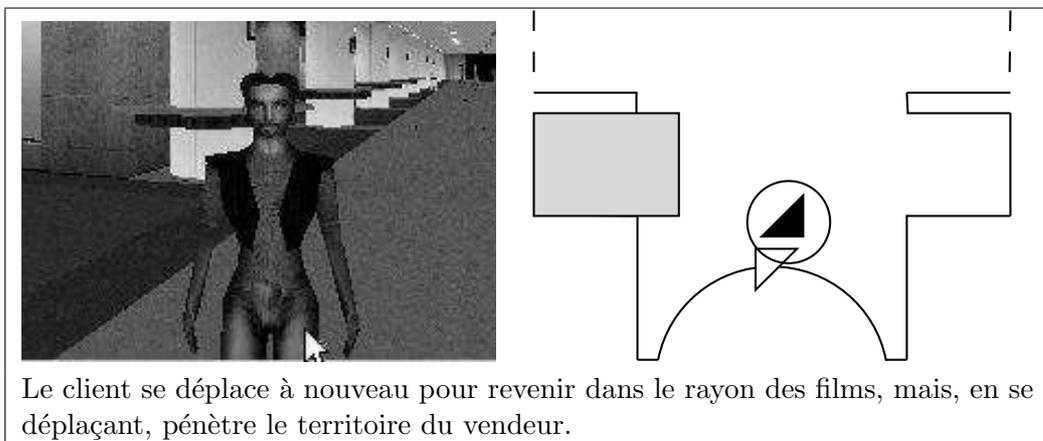
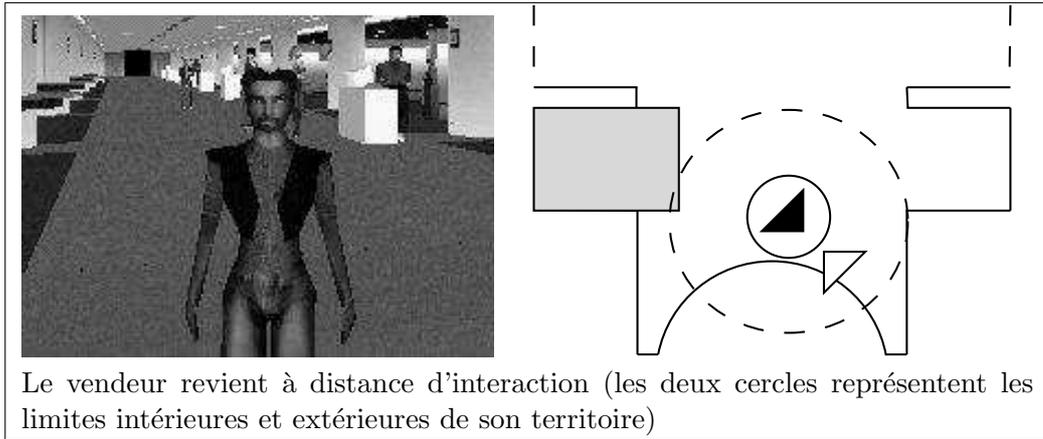
³²Les mouvements suivants étant *l'offre* (l'offenseur propose à l'offensé quelque chose pour réduire la faute. Il peut s'agir d'une explication, d'une pénitence, d'une prière. . .), *l'acceptation* (l'offensé se satisfait de l'offre proposée et accepte de passer l'éponge sur la faute) et enfin *la gratitude envers ceux qui ont pardonné* (un dernier tour pour remercier, et s'assurer que nous sommes reconnaissants envers notre interlocuteur pour nous avoir permis de remettre la situation en ordre).

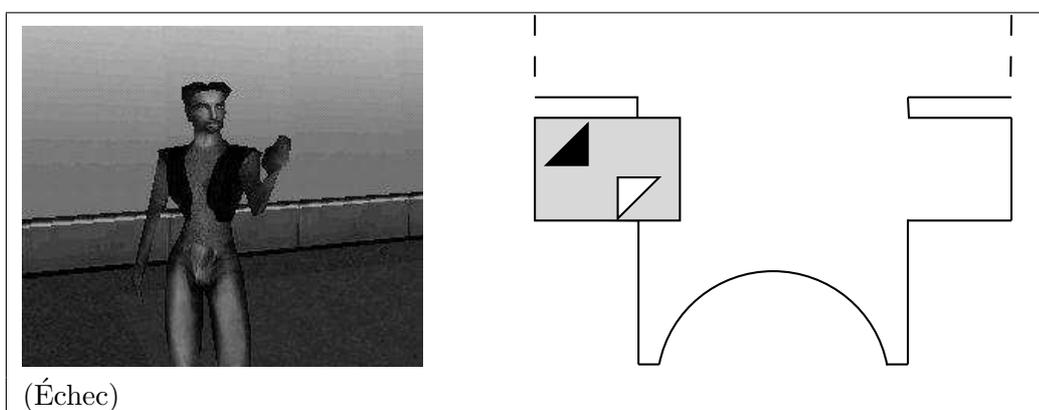
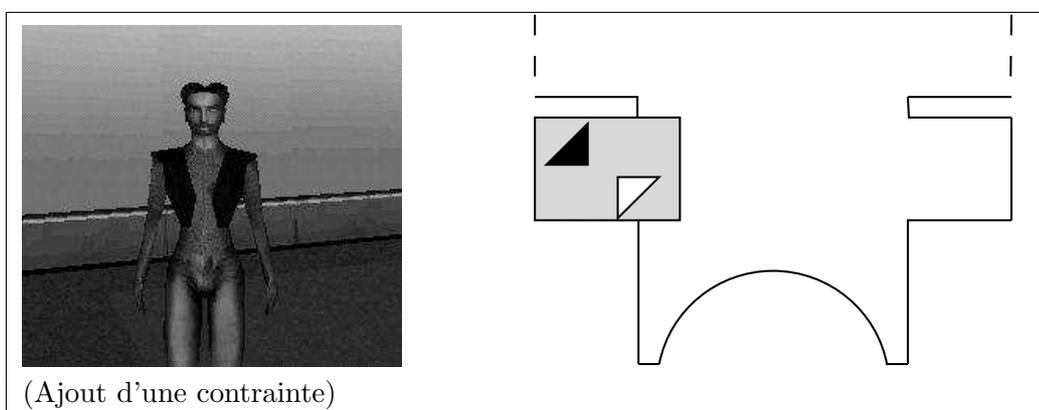
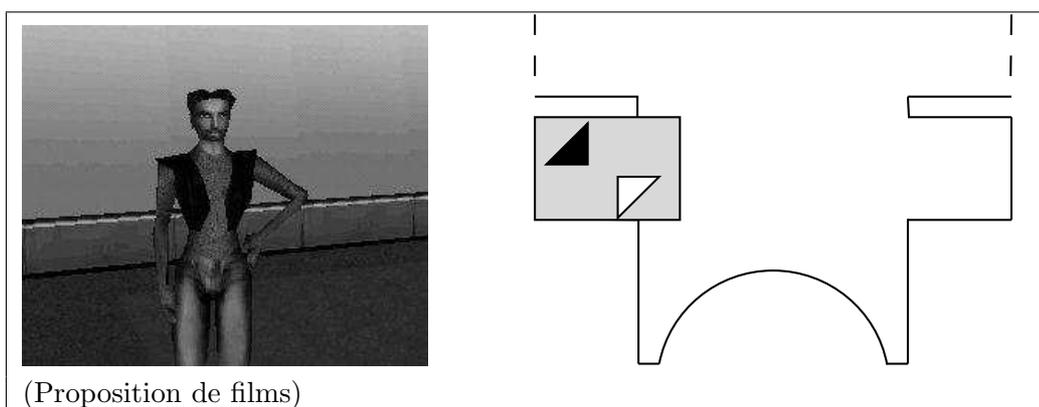
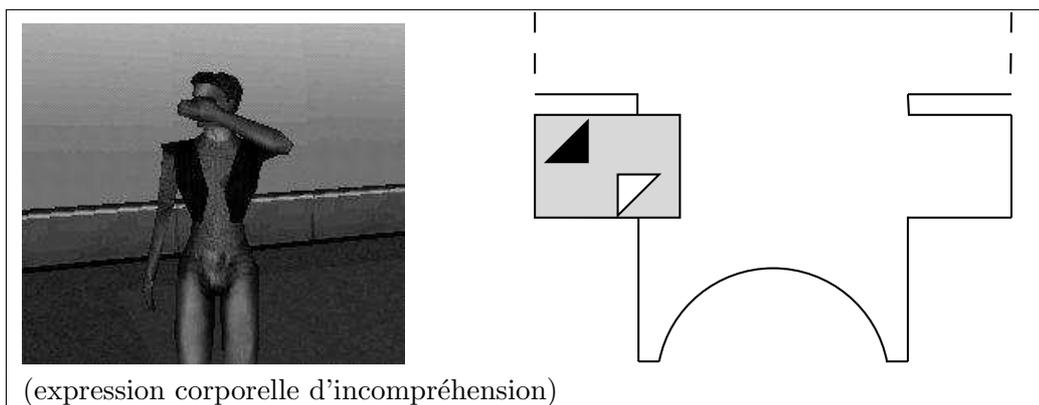
6.5.5 Dialogue 5

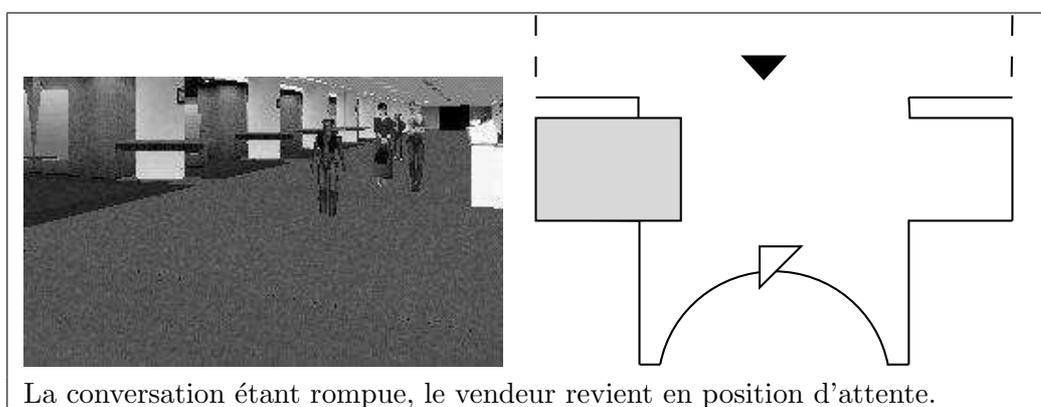
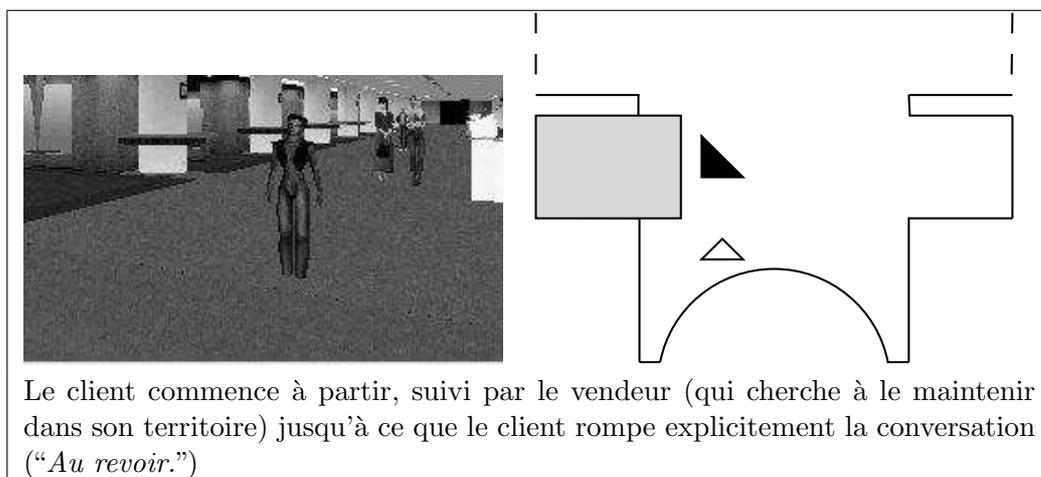
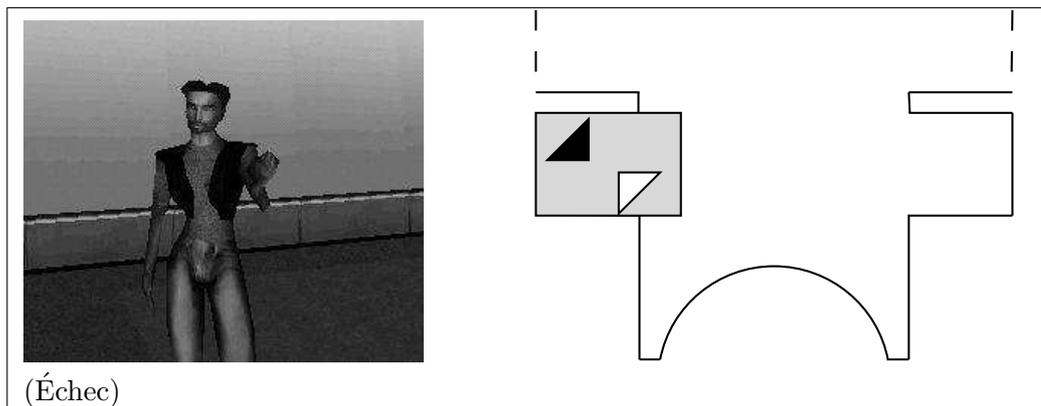
Dans cette analyse, nous ne détaillons pas les échanges verbaux entre l'humain et l'agent — l'interaction est quelconque — pour nous concentrer sur les aspects spatiaux (déplacements et expressions corporelles). L'interaction s'est déroulée de manière classique (connexion de l'agent et de l'utilisateur sur le serveur), alors qu'un logiciel d'acquisition vidéo fonctionnait sur l'ordinateur d'où l'humain se connectait. Les images sont extraites de cette séquence vidéo, et sont donc une vue subjective du client.











6.6 Considérations techniques

D'un point de vue technique, les réalisations sont faites en utilisant deux langages de programmation : Visual Basic (VB) et Visual C++ (VC++). VB est utilisé pour la création des interfaces graphiques et l'encapsulation de modules développés en VC++. Les modules développés sont de deux types : des DLL³³ et des ActiveX³⁴.

³³Dynamic Library Link. L'équivalent de bibliothèques qui ne sont pas liées à la compilation mais au cours de l'exécution.

³⁴Les ActiveX sont des DLL "actives", des composants, qui permettent aussi l'envoi de message, c'est-à-dire qu'elles ne se contentent pas de renvoyer des résultats à des appels de fonctions.

La base de données sur le cinéma est un extrait de la base complète de Canal+, qui comporte 1256 films distincts, plus de 600 metteurs en scène et plus de 2500 acteurs. À Grenoble, cette base est stockée au format MS Access et le programme y accède à travers une liaison ODBC réseau. À Paris, le format d'origine de la base est inconnu (et sans importance) et le programme y accède par l'intermédiaire, à nouveau, d'une liaison ODBC, mais cette fois-ci servie par MS SQL-server. La configuration de l'agent pour passer d'un site à un autre ne dépend que de quelques paramètres stockés dans des fichiers de configuration au format texte facilement éditables.

Les agents se connectent à un serveur de mondes virtuels Blaxxun par l'intermédiaire d'un ActiveX. Ils reçoivent de cet ActiveX des événements et lui envoient des actions. Parmi les événements reçus, les messages émanant des autres connectés sont traités de manière particulière. Les événements envoyés par le serveur Blaxxun peuvent être :

- arrivée / départ d'un avatar ;
- déplacement ;
- ouverture / fermeture d'une discussion privée ;
- réception d'un message (les expressions corporelles transitent par le même canal que le *chat*).

Les actions possibles sont :

- déplacement ;
- changement de scène ;
- déclenchement d'une expression corporelle ;
- ouverture / fermeture d'une discussion privée ;
- émission d'un message.

L'architecture a été pensée de façon à pouvoir recevoir un nombre arbitraire de modules de raisonnement, chacun libre d'employer les ressources et traitements que souhaite le développeur. À cet effet chaque module de raisonnement doit se présenter sous la forme d'une DLL. L'ajout ou le retrait d'un des modules s'effectue par l'édition d'un fichier d'initialisation, et la mise à jour d'un module par le simple remplacement de l'ancienne DLL par la nouvelle.

Il est clair, en comparant les fonctionnements des modules chat, conversation et cinéma que l'application ne contraint en rien les mécanismes de traitement. Seule l'interface de la DLL doit être respectée. Ainsi, le code source de PC-PATR a été recompilé pour former une DLL (travail en grande partie réalisé par Yannick Chapellaz [Chapellaz, 2001]) voir section 6.4.5, et le module de cinéma est basé sur une DLL proposant des fonctionnalités génériques de gestion de focus et d'attentes (travail en grande partie réalisé par Bruno Celle [Celle, 2001]). Une autre DLL, spécifique au domaine du cinéma s'y connecte pour former un module de traitement fonctionnel, mais la DLL générique pourrait tout à fait être spécialisée dans un tout autre domaine. Une telle architecture nous permet de distinguer d'un côté la dynamique créée autour des attentes et du focus et de l'autre le domaine de discours auquel les attentes et le focus s'appliquent (voir section 6.4.3).

Conclusions

*Le savant doit ordonner ; on fait la science
avec des faits comme une maison avec des
pierres. Mais une accumulation de faits
n'est pas plus une science qu'un tas de
pierres est une maison.*

HENRI POINCARÉ, la science et
l'hypothèse.

Nous avons présenté dans ce mémoire la progression qui nous a amenés à la réalisation d'un agent conversationnel incarné, c'est-à-dire un agent doté de capacités de dialogue et s'intégrant dans un monde virtuel. Ce résultat participe au cadre plus vaste de la création de communautés mixtes où agents artificiels et humains mettraient en commun leurs compétences. Tout au long du cheminement qui nous a mené de l'étude des communautés au développement d'un agent, nous avons mis en évidence des limitations, parfois choisies, parfois imposées, parfois d'ordre technique, parfois d'ordre théorique. La première section de cette conclusion présente l'essentiel de ces limitations. Dans la deuxième section, nous présentons les extensions possibles, théoriques ou pratiques, à nos travaux. Ces extensions rassemblent à la fois une série d'éléments pratiques déjà présentés dans les chapitres théoriques mais non intégrés dans l'agent et des questions plus théoriques portant parfois sur la structure même des modèles que nous avons présentés. Enfin, dans une troisième et dernière section, nous concluons sur les approches pluridisciplinaires.

1 Limitations

1.1 Limitations d'ordre technologique

a) Limitations dans les outils

Les limitations les plus importantes liées aux technologies employées sont en rapport avec le serveur de communauté virtuelle et l'interface client développés par la société Blaxxun et sur lesquels repose le Deuxième Monde.

La première limitation a trait aux capacités d'action dans l'environnement. Les agents ne peuvent que s'y déplacer et les objets actifs, intégrés par Blaxxun après les débuts de nos travaux, ne sont pas manipulables directement par les agents. De plus, le contrôle de l'agent sur sa représentation, l'avatar, est limité, puisqu'il ne peut que déclencher des expressions corporelles préenregistrées.

La seconde limitation a trait aux capacités perceptives. L'agent a une perception limitée de son environnement, à savoir, sa position sous forme de coordonnées par rapport à un repère presque arbitraire. À moins de mettre en place de coûteux

mécanismes d'analyse de scènes VRML, il ne sait donc pas où sont les obstacles par exemple et ne traite pas les collisions. Les seules autres perceptions de l'agent sont des événements correspondant à des actions (déplacement, expression corporelle ou émission d'un message) des autres membres. Dans le contexte particulier de l'infrastructure Blaxxun, toute perception est le résultat de l'action d'un autre agent.

Outre les limitations issues de l'emploi de la technologie Blaxxun, le choix de technologies Microsoft comme les ActiveX, les classes MFC (bibliothèques C++) et les connexions aux bases de données par ODBC nous imposent, respectivement, l'emploi d'Internet Explorer comme interface Web, l'emploi de Visual Studio comme environnement de développement et de compilation, et un fonctionnement sous Windows.

b) Limitations dans les objectifs atteints

Les limitations dont nous faisons état ici correspondent à la différence entre les spécifications et le résultat obtenu. Globalement, la structure interne de l'agent rassemble les différents éléments dont nous avons montré l'importance théorique, et, globalement toujours, le comportement de l'agent est conforme à nos attentes.

Dans les détails, plusieurs étapes de traitement sont succinctes, mais elles ont toutefois été mises en place au sein de l'agent (1) en prévision des extensions qui seront apportées ultérieurement et (2) pour correspondre au plus près au modèle théorique présenté. Le programme se veut une plate-forme d'expérimentation et il était important de présenter une architecture complète, quitte à ce que certains modules soient présents sans remplir entièrement leur fonction. Par exemple, les fonctionnalités principales des modules d'interprétation et d'expression sont liés au territoire et aux anaphores alors que la théorie leur accorde un champ d'application bien plus vaste. De même, l'analyse et la génération de langue naturelle sont basées sur des mécanismes élémentaires (analyse lexico-syntaxique et phrases à base de patrons) mais leurs capacités sont suffisantes pour mettre en évidence les mécanismes liés à la dynamique de la conversation.

Au niveau de son comportement, l'agent dispose de l'essentiel de ce qui était souhaité. Si les interactions sont parfois délicates, c'est principalement pour des raisons quantitatives : une incompréhension de mots usuels ou la non-reconnaissance de situations d'échec par exemple. Ces limitations peuvent se résoudre par la complétion de lexique ou l'ajout d'attentes au module de conversation, pour reprendre les exemples précédemment cités, et ne nécessitent pas la mise en place de nouveaux processus.

Plus subjectivement, notre plus grand regret est de n'avoir pu tester notre agent au sein de la version publique du Deuxième Monde pour le confronter aux utilisateurs, et ce, pour des raisons liées à la politique de l'entreprise qui ont conduit à la suppression de l'activité autour du Deuxième Monde fin 2001.

1.2 Limitations d'ordre théorique

Le principal écueil théorique auquel nous avons été confronté est certainement celui du sens. Nos hypothèses les plus discutables sont les suivantes :

- **Les premières étapes de perception d'un message n'y trouvent pas un sens, mais un potentiel de sens.** Cette hypothèse a été posée par Alain

Trognon et Christian Brassac lorsqu'ils ont voulu apporter une dimension dialogique à la théorie des actes de langages. Nos travaux, implicitement basés sur les actes de langages puisque liés à l'ACL-FIPA, portent sur l'aspect dialogique de l'échange de messages, il est donc naturel de continuer dans la lignée d'Alain Trognon et de Christian Brassac, et donc d'adopter cette hypothèse.

- **Un sens est représentable sous une forme symbolique et manipulable sous cette forme.** Le paradoxe de la chambre chinoise de Searle met en évidence les limitations de la manipulation de symboles indépendants d'un contenu sémantique. Nous suivons pour notre part les hypothèses du mouvement *cognitivist* pour lequel les processus mentaux (donc, par exemple, ceux liés à des contenus sémantiques) peuvent être basés sur des manipulations de symboles. Nous avons déjà abordé cette question dans la section 4.3.
- **La représentation d'un sens peut être faite en employant le formalisme des actes de langages.** Daniel Vanderveken prétend que "*Toute pensée conçue par un sujet humain est en principe exprimable par les moyens de son langage lors de l'accomplissement d'actes de discours.*" ([Vanderveken, 1988], préface). Nous nous sommes appuyés sur cette affirmation pour justifier l'emploi du formalisme des actes de langages pour représenter les sens que nous manipulons.
- **Le formalisme des actes de langages peut être adapté à la représentation de toute action.** Puisque notre modèle d'interaction passe uniquement par l'environnement¹, les messages transitent impérativement par des actions et des perceptions. Une représentation unifiée entre action et communication est donc non seulement souhaitable, mais théoriquement justifiée. Des travaux récents de Daniel Vanderveken portent sur la logique de l'action et nous espérons donc bientôt disposer d'un cadre unifié *et philosophiquement valide* pour les actions, qu'elles soient langagières ou non².

De manière connexe aux problèmes que nous venons de soulever se pose, dans les mécanismes d'interprétation et d'expression, le problème du point de vue : le message émis est sensé porter des informations sur le locuteur, informations exploitées en phase d'interprétation par l'allocutaire. Mais, si le locuteur dispose lui aussi d'une phase d'expression, le message est aussi porteur d'informations sur son destinataire. Comment distinguer ces deux types d'informations ? Jusqu'où pousser les mécanismes d'enrichissement contextuel ? Nous revenons ici sur les problèmes de références circulaires déjà présentés dans ce manuscrit.

Enfin, les approches que nous avons suivies sont, malgré tout, fortement ancrées dans le domaine de l'intelligence artificielle, et plus particulièrement, des systèmes multi-agents. Quelques remarques sur nos travaux ont révélé qu'il serait intéressant d'approfondir nos liens avec d'autres domaines. Principalement, nos travaux auraient certainement bénéficié d'une considération plus poussée des domaines de l'IHM (nous nous sommes limités à l'étude des agents conversationnels) et du traitement de la parole (nous nous sommes limités à une modalité textuelle).

¹Contrairement aux modèles qui distinguent un canal privilégié pour la communication directe entre agents.

²Une étape suivante serait la possibilité de représenter tout événement, quelle qu'en soit l'origine (i.e., qu'un agent soit identifié ou non comme dans le cas d'une porte qui claque) de façon à unifier perception et action.

2 Extensions

De nombreuses extensions sont possibles aux travaux que nous venons de présenter. Nous nous contenterons ici des principales.

Tout d'abord, sans intervenir au niveau de l'architecture de l'agent, il est souhaitable de tester notre agent avec une base de données différente de celle sur le cinéma, mais restant dans un domaine fortement connexe, comme le catalogue étoffé d'un grand magasin. Cette étape permettrait de bien cerner l'impact d'un tel changement dans les différentes parties de l'agent (mise à jour du lexique dans le module de formalisation de la langue naturelle, nouvelles attentes dans le module de rôle ou rôle supplémentaire, etc.). Une fois cette première étape accomplie, il serait nécessaire d'intégrer une base de données plus fondamentalement différente, mais en restant dans le cadre d'une consultation d'expert (recommandation de voyages, diagnostic médical, support technique, etc.). Enfin, la réalisation d'un agent multi-plates-formes passera par la re-programmation de cet agent en Java. Sylvie Brincourt, dans le cadre de son stage d'ingénieur CNAM travaille sur ces points [Brincourt, 2003].

Plusieurs modules de l'agent que nous avons réalisés ne remplissent leur fonction que de manière minimale, bien que suffisante pour mettre en évidence les points que nous voulions présenter. Les modules nécessitant le plus une "mise à niveau" sont les modules de formalisation et d'énonciation en langue naturelle³ et le module de sélection. Les améliorations de ce dernier (qui se limite pour le moment, rappelons-le, à un simple appel d'offre) seront indispensables si le nombre de modules de rôle augmente ou que l'on cherche à réutiliser les rôles dans différents agents. Un système basé sur la négociation entre les modules nous semblerait une bonne approche.

Une amélioration intéressante serait que l'agent puisse changer de sujet de discussion, donc de focus, et que ce changement s'effectue à travers une discussion. C'est-à-dire, que le focus lui-même, en tant que cadre général du dialogue, devienne un sujet de négociation. Ce genre de fonctionnement récursif (utiliser un focus pour négocier les paramètres d'un focus) apportera de nouvelles et importantes questions, tant théoriques que pratiques : peut-on tout simplement négocier les paramètres d'un focus comme les paramètres d'un film ? un agent possède-t-il plusieurs focus ? si oui, comment les gérer, les remplir, définir celui qui est pertinent à tel instant précis de l'interaction ? si non, comment gérer les changements de sujets et distinguer les simples digressions des changements de sujet de discussion ? Que deviennent les attentes ? sont-elles liées à un focus donné (i.e. à un thème de conversation) ou plus globalement à la conversation ? etc.

Nous avons à plusieurs reprises parlé de la personnalité de l'agent, que nous avons décomposé en trois grandes tendances, le profil psychologique, la dynamique émotionnelle et la conscience sociale. Nous nous sommes essentiellement limités à la conscience sociale. Compléter la personnalité par les deux autres paramètres et analyser leur influence sur l'interprétation et l'expression reste à faire. Des travaux en cours, menés dans l'équipe par Patricia Jaques, portant sur les agents pédagogiques émotionnels [Jaques, prévu 2003] pourraient déboucher sur un modèle de dynamique émotionnelle intégrable dans notre agent.

Un point particulièrement important à régler sera le passage au multi-utilisateurs, qui sera loin d'être trivial. Par exemple, le mode de communication n'étant pas adressé, les modules d'interprétation (et d'expression) se retrouveront chargés de l'identification des destinataires, en intégrant le nom du destinataire au message, en

³Émilie Benoit a récemment travaillé sur l'étape d'énonciation (génération), l'améliorant grandement [Benoit, 2002].

dirigeant le regard, en venant se placer face à lui, etc. La gestion du focus deviendra aussi plus délicate, puisqu'il est associé à un sujet de discussion, et non à un interlocuteur. De manière très liée les interactions entre plusieurs agents artificiels devront aussi être mises en place.

Enfin, la dernière extension sera la tentative d'utilisation des nos proposition de modèle dans des situations conversationnelles relevant d'autres types de dialogues argumentatifs : débats, enseignement, négociation ou critique, pour reprendre certains de types rassemblés par Douglas Walton [Walton, 1990].

Finalement, notre agent n'est qu'une étape sur le chemin de la création d'agents pour des communautés mixtes. Nous aimerions qu'un jour il quitte notre laboratoire pour, comme il est prévu, s'intégrer à une communauté mixte où des humains et d'autres agents pourront profiter de ses services...

3 En conclusion : de l'intérêt de la pluridisciplinarité

Notre objectif était de travailler à un modèle d'interaction adapté à la fois aux humains et aux agents artificiels. L'approche que nous avons suivie a consisté à étudier les modèles d'interaction déjà existant pour ces deux types d'entités (à travers principalement les domaines de la psychologie de l'interaction et de la linguistique d'une part et des systèmes multi-agents d'autre part), puis à en synthétiser un modèle adapté aux deux. Deux écueils existent dans une telle démarche, le premier apparaît au moment de l'étude des différents domaines, le second au moment de leur mise en relation.

Le premier écueil est particulièrement simple : certaines hypothèses implicites à un domaine d'étude peuvent échapper à un néophyte, qui emploiera de manière erronée les notions liées à ce domaine. Nous pensons avoir échappé à cet écueil dans nos travaux, n'employant que les notions que nous estimions avoir étudiées et comprises suffisamment en détail.

Le second écueil est plus délicat. En effet, rien ne permet d'affirmer que la mise en relation de deux concepts développés dans des domaine distincts puisse donner un troisième concept adapté aux deux domaines simultanément. Si l'on se reporte à la figure 1, cela signifie qu'il n'existe pas forcément d'intersection entre les deux domaines. En tout état de cause, il semble raisonnable de penser que ce nouveau concept sera moins bien adapté dans chacun des domaines que ne l'étaient les concepts spécifiques (c'est en grande partie la justification des sections 3.3 et 3.4, "Nécessités côté humain" et "Nécessités côté agent"). Sur la figure, cela est matérialisé par le fait qu'aucun des ensembles de modèles ne contient totalement l'autre. De plus, il peut arriver que la confrontation de modèles issus de domaines différents entraîne l'émergence de propriétés qui n'appartenaient à aucun des modèles, ce qui serait le cas d'un modèle correspondant au cercle pointillé de la figure. Nous avons étudié les modèles de l'interaction humaine et ceux de l'interaction entre agents, et nous en avons tiré, *a priori*, un modèle qui appartiendrait à l'intersection des deux domaines.

Ce problème peut être considéré d'une manière plus vaste :

1. À partir de quand, lors de la résolution d'un problème, faut-il arrêter d'adapter les modèles existants pour se mettre à chercher une solution novatrice⁴ ? Des

⁴Comme le dit Abraham Maslow, psychologue Américain, "Pour la personne qui n'a qu'un marteau, tout problème prend l'apparence d'un clou."

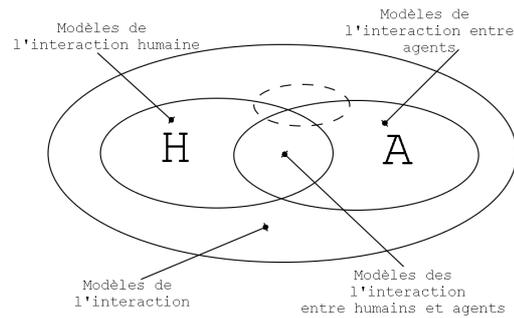


FIG. 1 – Modèles génériques ou d'intersection ?

exemples classiques disent que l'ampoule électrique n'est pas née de l'évolution de la bougie ou que l'avion a commencé à fonctionner efficacement lorsqu'on a arrêté d'essayer de lui faire battre des ailes.

2. Dans quelle mesure un modèle valide dans un domaine est-il transposable à un autre domaine? Cette question est particulièrement pertinente dans le cadre des approches pluridisciplinaire où parfois, les modèles sont tout simplement *recopiés* d'un domaine à l'autre. Les exemples classiques des limites d'une telle approche sont la roue et l'hélice, probablement parmi les moyens les plus efficaces de locomotion terrestre et aquatique, mais qui ne sont pas nés de la copie de solutions de locomotion naturelles.

Je tiens à conclure ce mémoire de doctorat en répondant moi-même à la question qui servait de clôture à mon mémoire de DEA, il y a déjà quatre années de cela. Non, le modèle humain n'est pas le meilleur possible. Mais la richesse des systèmes naturels font de ces derniers une source d'inspiration inépuisable.

Entre :
Ce que je pense,
ce que je veux dire,
ce que je crois dire,
ce que je dis,
ce que vous avez envie d'entendre,
ce que vous croyez entendre,
ce que vous entendez,
ce que vous avez envie de comprendre,
ce que vous comprenez,
Il y a dix possibilités qu'on ait des difficultés à communiquer.
Mais essayons quand même. . .

B. WERBER, "*Le père de nos pères*"

Bibliographie

- R. ALAMI, S. FLEURY, M. HERRB, F. INGRAND et F. ROBERT. *Multi Robot Cooperation in the Martha Project*. *IEEE Robotics and Automation Magazine* (Special Issue on "Robotics & Automation in the European Union"), tome 5, n° 1 (1998).
- Ludovic ALIDRA. *Les Langages D'interaction Dans Les Systèmes Multi-Agents*. Mémoire de DEA / Master, Institut polytechnique de Sévenans (1997).
- Françoise ARMENGAUD. *La Pragmatique*. Que Sais-Je? PUF (1985).
- Aaron ARMSTRONG et Edmund DURFEE. *Dynamic Prioritization of Complex Agents in Distributed Constraint Satisfaction Problems*. Dans *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI97)* (1997).
- J-L AUSTIN. *Quand Dire, C'est Faire*. Points. Seuil (1962).
- Christof BAEIJS. *Fonctionnalité Émergente Dans Une Société D'agents Autonomes. Étude Des Aspects Organisationnels Dans Les Systèmes Multi-Agents Réactifs*. Thèse de doctorat, INP Grenoble (1998).
- Gene BALL et Jack BREESE. *Relating Personality and Behavior : Posture and Gestures*. Dans *International Workshop on Affect in Interactions (IWAI'99)* (1999). Version restreinte de [Ball et Breese, 2000].
- Gene BALL et Jack BREESE. *Emotion and Personality in a Conversational Agent*. Dans [Cassell *et al.*, 2000d], chapitre 7, pp. 189–219.
- Gene BALL, Dan LING, David KURLANDER, John MILLER, David PUGH, Tim SKELLY, Andy STANKOSKY, David THIEL, Maarten VAN DANTZICH et Trace WAX. *Lifelike Computer Characters : The Persona Project at Microsoft Research*. Dans [Bradshaw, 1997], chapitre 10, pp. 191–222.
- Joseph BATES. *The Nature of Character in Interactive Worlds and The Oz Project*. Rapport technique CMU-CS-92-200, School of Computer Science, Carnegie Mellon University (CMU) (1992).
- Joseph BATES, A. Bryan LOYALL et W. Scott REILLY. *Broad Agents*. Dans *Proceedings of the AAAI Spring Symposium on Integrated Intelligent Architectures*, (1991).
- Joseph BATES, A. Bryan LOYALL et W. Scott REILLY. *An Architecture for Action, Emotion, and Social Behavior*. Dans *MAAMAW'92* (1992). (publié auparavant en 1992 comme rapport technique CMU-CS-92-144).
- Émilie BENOIT. *Génération Automatique D'énoncés En Langue Naturelle Pour un Agent Conversationnel*. Mémoire de maîtrise llce d'allemand, mention industrie de la langue, INPG, laboratoire Leibniz, Grenoble (2002).

- Alexandra BERGER. *De la Théorie Des Actes de Langages À la Conversation Entre Agents Logiciels*. Mémoire de maîtrise sciences du langage, mention industrie de la langue, INPG, laboratoire Leibniz, Grenoble (2001).
- Timothy BICKMORE et Justine CASSELL. “How About This Weather?” *Social Dialogue with Embodied Conversational Agents*. Dans *Proceedings of AAAI Fall Symposium on Socially Intelligent Agents* (2000).
- Olivier BOISSIER. *Modèles et Architectures D’agents*. Dans [Briot et Demazeau, 2001], pp. 71–108.
- Abdenour BOUZOUANE, Carl DIONNE, Ilie STIHARU-ALEXE et Denis GAGNÉ. *Jeu de Rôle Virtuel à Base D’agents Intelligents*. Dans *JFIADSMA ’98. Systèmes Multi-Agents, de L’interaction À la Socialité*, pp. 147–161. Hermes, Paris (1998).
- J. BRADSHAW, S. DUTFIELD, P. BENOIT et J. WOOLLEY. *KAoS : Toward an Industrial-Strength Open Agent Architecture*. Dans [Bradshaw, 1997], pp. 375–418.
- Jeffrey BRADSHAW, éd. *Software Agents*. AAAI / MIT Press (1997).
- Christian BRASSAC. *L’interaction Inter-Agents : Non Littéralité et Processualité*. Dans *Deuxièmes Journées Francophones IAD-SMA* (1994).
- Christian BRASSAC. *L’imprévisibilité de la Conversation, Une Constructibilité Par Défaut* (Mai 1995).
- Christian BRASSAC et Sylvie PESTY. *La “Pelouse Fournilière”. De la Coaction À la Coopération..* Dans *Quatrièmes Journées Francophones IAD-SMA* (1996).
- Christian BRASSAC et Alain TROGNON. *Analyse de Conversations et Théorie Des Actes de Langage*. *Cahiers de la linguistique Française*, tome 13 : pp. 62–76 (1992).
- Jack BREESE et Gene BALL. *Modelling Emotional State and Personality for Conversational Agents*. Rapport technique MSR-TR-98-41, Microsoft research (1998).
- Sylvie BRINCOURT. *Analyse et Réalisation D’un Agent Conversationnel Générique (Titre Provisoire)*. Rapport technique, Conservatoire National des Arts et Métiers / Leibniz-IMAG (2003). Mémoire d’ingénieur CNAM (En cours, soutenance prévue en Juin 2003).
- Jean-Pierre BRIOT et Yves DEMAZEAU, éd. *Principes et Architecture Des Systèmes Multi-Agents*. Hermes, Paris (2001).
- J. CASSELL, M. ANANNY, A. BASU, T. BICKMORE, P. CHONG, D. MELLIS, K. RYOKAI, J. SMITH, H. VILHJÁLMSSON et H. YAN. *Shared Reality : Physical Collaboration with a Virtual Peer*. Dans *ACM CHI 2000 Conference Proceedings* (2000a).
- J. CASSELL et T. BICKMORE. *External Manifestations of Trustworthiness in the Interface*. *Communications of the ACM*, tome 43, n° 12 (2000).
- J. CASSELL, T. BICKMORE, M. BILLINGHURST, L. CAMPBELL, K. CHANG, H. VILHJÁLMSSON et H. YAN. *Embodiment in Conversational Interfaces : Rea*. Dans *Proceedings of the ACM CHI* (1999a).
- Justine CASSELL. *Nudge Nudge Wink Wink : Elements of Face-to-Face Conversation for Embodied Conversational Agents*. Dans [Cassell et al., 2000d], chapitre 1, pp. 1–27.

- Justine CASSELL, Tim BICKMORE, Lee CAMPBELL, Hannes VILHJÁLMSSON et Hao YAN. *Human Conversation as a System Framework : Designing Embodied Conversational Agents*. Dans Justine CASSELL, Joseph SULLIVAN, Scott PREVOST et Elizabeth CHURCHILL, édés., *Embodied Conversational Agents*, chapitre 2, pp. 29–63. MIT Press (2000b).
- Justine CASSELL, Timothy BICKMORE, Hannes VILHJÁLMSSON et H. YAN. *More Than Just a Pretty Face : Affordances of Embodiment*. Dans *Proceedings of International Conference on Intelligent User Interfaces* (2000c).
- Justine CASSELL, J. SULLIVAN, Scott PREVOST et Elizabeth CHURCHILL, édés. *Embodied Conversational Agents*. MIT Press (2000d).
- Justine CASSELL et K.R. THÓRISSON. *The Power of a Nod and a Glance : Envelope Vs. Emotional Feedback in Animated Conversational Agents*. *Applied Artificial Intelligence*, tome 13 : pp. 519–538 (1999).
- Justine CASSELL, Obed E. TORRES et Scott PREVOST. *Turn Taking Vs. Discourse Structure : How Best to Model Multimodal Conversation*. Dans WILKS, éd., *Machine Conversations*. The Hague : Kluwer (1999b).
- Bruno CELLE. *Analyse et Programmation D'un Comportement D'agent Virtuel*. Rapport de stage IUT d'informatique, INPG, Laboratoire Leibniz, Grenoble (2001).
- Brahim CHAIB-DRAA, Imed JARRAS et Bernard MOULIN. *Systèmes Multi-Agents : Principes Généraux et Applications*. Dans [Briot et Demazeau, 2001], pp. 27–70.
- Patrick CHAMPAGNE. *La Sociologie*. Les Essentiels. Milan (1997).
- Yannick CHAPPELLAZ. *Conception et Implémentation D'un Vendeur Virtuel Sous Forme D'agent*. Rapport de stage IUT informatique, INPG, Laboratoire Leibniz, Grenoble (2001).
- Guillaume CHICOISNE. *Conversation et Relations Sociales Pour Des Agents Moins Artificiels*. Mémoire de DEA / Master, INPG (1998).
- Guillaume CHICOISNE. *Interaction Conversationnelle Entre Internautees et Agents Artificiels Partageant un Monde Virtuel*. Dans *JFIADSMA 2000* (2000).
- Guillaume CHICOISNE et Sylvie PESTY. *Modèle de Conversation et Agents Rationnels Socialement Corrects*. Dans *Atelier Thématique TALN 1999 "La Langue Dans L'interaction Homme-Machine"*, pp. 91–104 (1999).
- Guillaume CHICOISNE et Sylvie PESTY. *The Puppeteer Behind the Avatar*. Dans *Sketches and Applications, Proceedings of the ACM SIGGRAPH 2000*, p. 246. ACM SIGGRAPH (2000).
- Anne COLLINOT, Laurent PLOIX et Alexis DROGOU. *Application de la Méthode Cassiopée À L'organisation D'une Équipe de Robots*. Dans Jean-Pierre MÜLLER et Joël QUINQUETON, édés., *Journées Francophones Intelligence Artificielle Distribuée et Systèmes Multi-Agents*. Hermès, Paris (1996).
- Yves DEMAZEAU. *Steps Toward Multi-Agent Programming*. Dans *IWMAS-97* (1997).
- Yves DEMAZEAU. *Next Agents' World*. Dans *ASAI 2000*, pp. 11–13 (1999).
- Alexis DROGOU. *Systèmes Multi-Agents Situées* (2000). Dossier d'habilitation à diriger des recherches.

- Alexis DROGOUL et Sébastien PICAULT. *MICRobES : Vers Des Collectivités de Robots Socialement Situés*. Dans *Actes Des 7èmes Journées Francophones Intelligence Artificielle Distribuée et Systèmes Multi-Agents (JFIADSMA '99)*. Hermès (1999).
- Emmanuel DUBOIS, Laurence NIGAY et Jocelyne TROCCAZ. *Combinons Le Monde Virtuel et Le Monde Réel – Classification et Principes De Conception*. Dans *Actes Des Rencontres Jeunes Chercheurs En IHM*, pp. 31–34 (2000).
- Amal EL FALLAH. *Modèles de Coordination D'agents Cognitifs*. Dans [Briot et Demazeau, 2001], pp. 139–176.
- Renee ELIO, Afsaneh HADDADI et Ajit SINGH. *Task Models, Intentions, and Agent Conversation Policies*. Dans *Pacific Rim International Conference on Artificial Intelligence*, pp. 394–403. Springer Verlag, Berlin (2000).
- Jacques FERBER. *Les Systèmes Multiagents : Vers une Intelligence Collective*. InterEditions, Paris (1995).
- Kelly FERNANDES. *Systèmes Multi-Agents Hybrides : Une Approche Pour la Conception de Systèmes Complexes*. Thèse de doctorat, Université Joseph Fourier, Grenoble (2001).
- T. FININ, J. WEBER, G. WIEDERHOLD, M. GENESERETH, R. FRITZON, J. MCGUIRE, S. SHAPIRO et C. BECK. *Specification of the KQML Agent-Communication Language*. Rapport technique, DARPA Knowledge Sharing Initiative, External Interface Working Group (1994).
- Humbert FIORINO. *Élaboration de Conjectures Par Des Agents Coopérants*. Thèse de doctorat, École nationale supérieure de l'aéronautique et de l'espace (1998).
- FIPA. *Agent Communication Language*. Rapport technique OC00003, FIPA (Foundation for Intelligent Physical Agents) (1997). www.fipa.org/specs/fipa00003/OC00003.pdf.
- FIPA. *FIPA-ACL Message Structure Specification*. Rapport technique XC00061E, FIPA (Foundation for Intelligent Physical Agents) (2001a). <http://www.fipa.org/specs/fipa2000.tar.gz>.
- FIPA. *FIPA Communicative Act Library Specification*. Rapport technique XC00037H, FIPA (Foundation for Intelligent Physical Agents) (2001b). <http://www.fipa.org/specs/fipa2000.tar.gz>.
- FIPA. *FIPA Interaction Protocol Library Specification*. Rapport technique XC00025E, FIPA (Foundation for Intelligent Physical Agents) (2001c). <http://www.fipa.org/specs/fipa2000.tar.gz>.
- FIPA. *FIPA SL Content Language Specification*. Rapport technique XC00008G, FIPA (Foundation for Intelligent Physical Agents) (2001d). <http://www.fipa.org/specs/fipa2000.tar.gz>.
- Jerry A. FODOR. *La Modularité de L'esprit. Essai sur la Psychologie Des Facultés*. Editions de Minuit (1983).
- Leonard FONER. *What's an Agent Anyway? A Sociological Case Study*. Rapport technique Agents Memo 93-01, MIT Media Lab, Boston (1993). <http://www.media.mit.edu/people/foner/Julia/>.
- Erving GOFFMAN. *La Mise En Scène de la Vie Quotidienne*, tome 1 & 2. Minuit, Paris (1973).
- Erving GOFFMAN. *Les Rites D'interaction*. Minuit, Paris (1974).

- Erving GOFFMAN. *Façons de Parler*. Editions de Minuit, Paris (1981).
- Mark GREAVES et Jeffrey M. BRADSHAW, édés. *Workshop on Specifying and Implementing Conversation Policies* (1999).
- Mark GREAVES, Heather HOLBACK et Jeffrey BRADSHAW. *What Is a Conversation Policy ?*. Dans [Greaves et Bradshaw, 1999].
- Herbert Paul GRICE. *Logic and Conversation*. Dans P. COLE et J. L. MORGAN, édés., *Syntax and Semantics : Vol. 3 : Speech Acts*, pp. 41–58. Academic Press, New York (1975). (référence issue de notes de cours).
- S. HAMBRIDGE. *Netiquette Guidelines*. Rapport technique RFC 1855, Responsible Use of the Network (RUN) / Internet Engineering Task Force (IETF) (1995). <http://www.rfc-editor.org/>.
- Marc-Philippe HUGET. *Une Ingénierie Des Protocoles D'interaction Pour Les Systèmes Multi-Agents*. Thèse de doctorat, Université Paris IX - Dauphine (Juin 2001).
- Ricardo IMBERT et Angélica DE ANTONIO. *The Bunny Dilemma : Stepping Between Agents and Avatars*. Dans *Proceedings of the 17th Twente Workshop on Language Technology (TWLT 17)* (2000).
- Ricardo IMBERT, Angélica DE ANTONIO, María Isabel SÁNCHEZ et Javier SEGOVIA. *How Can Virtual Agents Improve Communication in Virtual Environments ?*. Dans *Proceedings of the Second Workshop on Intelligent Virtual Agents (VA99)* (1999a).
- Ricardo IMBERT, Angélica DE ANTONIO, Javier SEGOVIA et María Isabel SÁNCHEZ. *A Fuzzy Internal Model for Intelligent Avatars*. Dans *I3 Spring Days'99. Workshop on Behavior Planning for Life-Like Characters and Avatars* (1999b).
- Ricardo IMBERT, María Isabel SÁNCHEZ, Angélica DE ANTONIO et Javier SEGOVIA. *The Amusement Internal Modelling for Believable Behaviour of Avatars in an Intelligent Virtual Environment*. Dans *ECAI'98. Workshop in Intelligent Virtual Environments* (1998).
- Toru ISHIDA, éd. *Community Computing and Support Systems. Social Interaction in Networked Communities*. Springer (1998a).
- Toru ISHIDA, éd. *Community Computing. Collaboration over Global Information Network*. John Wiley & Sons Ltd (1998b).
- Patricia Augustin JAQUES. *Conception D'un Agent Pédagogique Qui Prenne En Consideration la Composante Affective Des Comportements de L'apprenant (Concepção de Um Agente Pedagógico Que Considera Os Aspectos Afetivos Da Interação Com O Aluno)*. Thèse de doctorat, UFRGS - PPGC (prévu 2003).
- Quentin JONES. *Virtual-Communities, Virtual-Settlements & Cyber-Archaeology : A Theoretical Outline*. *Journal of Computer Mediated Communication*, tome 3, n° 3 (1997).
- Mark KANTROWITZ. *Glinda : Natural Language Text Generation in the Oz Interactive Fiction Project*. Rapport technique CMU-CS-90-158, School of Computer Science, Carnegie Mellon University (1990).
- Fredéric KAPLAN et Luc STEELS. *Comment Les Robots Construisent Leur Monde : Expériences sur la Convergence Des Catégories Sensorielles*. Dans J-L. DES-SALLES, éd., *Journée ARC Evolution et Cognition*, pp. 13–18. ENST (2000).

- Margaret Thomas KELSO, Peter WEYHRAUCH et Joseph BATES. *Dramatic Presence*. Rapport technique CMU-CS-92-195, School of Computer Science, Carnegie Mellon University (CMU) (1992).
- Catherine KERBRAT-ORECCHIONI. *Théorie Des Faces et Analyse Conversationnelle*. Dans Joseph ISAAC, éd., *Le Parler Frais d'Erving Goffman*. Minuit (1989).
- Jean-Luc KONING et Sylvie PESTY. *Modèles de Communication*. Dans *Principes et Architecture Des Systèmes Multi-Agents*, pp. 109–137. Hermes, Paris (2001).
- Jaroslav KOZLAK, Yves DEMAZEAU et François BOUSQUET. *Multi-Agents Systems to Model the Fishbanks Game Process..* Dans *International Workshop of Central and Eastern Europe on Multi-Agent Systems (CEEMAS'99)*, pp. 154–162 (1999).
- Jaroslav KOZLAK. *Maintien de L'intégrité Fonctionnelle Dans Les Systèmes Multi-Agents Ouverts À Ressources Renouvelables*. Thèse de doctorat, INP Grenoble, cotutelle avec AGH Cracovie (2000).
- Henri LABORIT. *La Légende Des Comportements*. Flammarion, Paris (1994).
- Jérôme LEHUEN. *Un Modèle de Dialogue Dynamique et Générique Intégrant L'acquisition de Sa Compétence Linguistique. Le Système COALA*. Thèse de doctorat, Université de Caen (1997).
- Alan LEIGH. *Spiral Pyramids, a New Way of Looking at Communication*. Dans *Facilities Management '95*. Strathclyde Graduate School of business (1995).
- J.C.R. LICKLIDER. *Man-Computer Symbiosis. IRE transactions on humans factors in electronics*, tome HFE-1 : pp. 4–11 (March 1960). (reprint by Digital Equipment Corporation 1990).
- J.C.R. LICKLIDER. *The Computer as a Communication Device. Science and technology* (avril 1968). (reprint by Digital Equipment Corporation 1990).
- Fuhua LIN, Douglas NORRIE, Weiming SHEN et Rob KREMER. *A Schema-Based Approach to Specifying Conversation Policies*. Dans [Greaves et Bradshaw, 1999].
- A. Bryan LOYALL et Joseph BATES. *Personality-Rich Believable Agents That Use Language*. Dans *Proceedings of the First International Conference on Autonomous Agents* (1997).
- Michel MARCOCCIA. *La Netiquette : Analyse Sociopragmatique Des Règles de Savoir-Vivre sur Internet. Informations InCognito*, tome 13 : pp. 5–14 (1999).
- R. MCCRAE et P. COSTA. *The Structure of Interpersonal Traits : Wiggins Circumplex and the Five Factor Model. Journal of personality and social psychology*, tome 56 : pp. 586–595 (1989). (Référence issue de notes de cours en psychologie sociale).
- Paul MILGRAM et Fumio KISHIN. *A Taxonomy of Mixed Reality Visual Displays. IEICE Transactions on Information Systems*, tome E77-D, n° 12 (1994).
- Chip MORNINGSTAR et F Randal FARMER. *The Lessons of Lucasfilm's Habitat*. Dans Mickael BENEDIKT, éd., *Cyberspace, First Steps*. MIT Press (1990).
- Bernard MOULIN, Sylvain DELISLE et Brahim CHAIB-DRAA, éd. *Analyse et Simulation de Conversation : De la Théorie Des Actes de Discours Aux Systèmes Multiagents*. L'interdisciplinaire, Limonest (1999).
- Alex MUCCHIELLI. *Psychologie de la Communication*. PUF (1995).
- Roger MUCCHIELLI. *L'examen Psychotechnique*. ESF, Paris (1971).

- Roger MUCCHIELLI. *La Dynamique Des Groupes*. Formation Permanente En Sciences Humaines. ESF, Paris (1983).
- Jacqueline NADEL et Anne-Marie MELOT. *Théorie de L'esprit*. Dans Olivier HOUDÉ, Daniel KAYSER, Olivier KOENIG, Joëlle PROUST et François RASTIER, édés., *Vocabulaire de Sciences Cognitives*, pp. 395–397. PUF (1998).
- Katashi NAGAO. *Agent Augmented Reality : Agents Integrate the Real World with Cyberspace*. Dans [Ishida, 1998b], chapitre 7, pp. 207–244.
- Hideyuki NAKANISHI, Chikara YOSHIDA, Toshikazu NISHIMURA et Toru ISHIDA. *FreeWalk : A Three-Dimensional Meeting-Place for Communities*. Dans [Ishida, 1998b], chapitre 3, pp. 55–89.
- Clifford NASS, Katherine ISBISTER et Eun-ju LI. *Truth is Beauty : Researching Embodied Conversational Agents*. Dans [Cassell et al., 2000d], chapitre 13, pp. 374–402.
- Clifford NASS, Jonathan STEUER et Ellen R. TAUBER. *Computers are Social Actors*. Dans *Proceeding of the CHI Conference* (1994).
- Michio OKADA. *Muu : Artificial Creatures as an Embodied Interface*. Dans *SIGGRAPH 2000 Conference Abstracts and Applications*, p. 91. ACM Siggraph (2000).
- A. ORTONY, G. CLORE et A. COLLINS. *The Cognitive Structure of Emotions*. Cambridge university press (1988).
- Malcom R. PARKS et Kory FLOYD. *Making Friends in Cyberspace*. *Journal of Computer-Mediated Communication*, tome 1, n° 4 (1996).
- Sylvie PESTY, Christian BRASSAC et Pierre FERRENT. *Ancrer Les Agents Cognitifs Dans L'environnement*. Dans QUINQUETON, THOMAS et TROUSSE, édés., *Actes Des 5ème JFIADSMA* (1997).
- Laurence PHILLIPS et Hamilton LINK. *The Role of Conversation Policy in Carrying Out Agent Conversations*. Dans [Greaves et Bradshaw, 1999].
- C.S. PIERCE. *Ecrits sur Le Signe*. Seuil (1978). (tr. de G. Deledalle, référence et citations issues de notes de cours).
- Violaine PRINCE. *Vers Une Informatique Cognitive Dans Les Organisations*. Masson (1996).
- Philippe QUÉAU. *Le Virtuel. Vertus et Vertiges*. INA / Champ Vallon (1993).
- S. RAFAELI. *Interactivity : From New Media to Communication*. Dans *Sage Annual Review of Communication Research : Advancing Communication Science*, pp. 110–134 (1988). (Cité par Quentin Jones [Jones, 1997]).
- Elizabeth REID. *Cultural Formations in Text-Based Virtual Realities*. Mémoire de DEA / Master, University of Melbourne (1994).
- W. Scott Neil REILLY. *Belivable Social and Emotional Agents*. Thèse de doctorat, Carnegie Mellon University (1996).
- W. Scott Neil REILLY et Joseph BATES. *Building Emotional Agents*. Rapport technique CMU-CS-92-143, Carnegie Mellon University (1992).
- Alexandre Moretto RIBEIRO. *Un Modèle D'interaction Dynamique Pour Les Systèmes Multi-Agents*. Thèse de doctorat, Université Joseph Fourier - Grenoble I (2000).

- J. RICKEL, J. GRATCH, R. HILL, S. MARSELLA et W. SWARTOUT. *Steve Goes to Bosnia : Towards a New Generation of Virtual Humans for Interactive Experiences*. Dans *AAAI Spring Symposium on Artificial Intelligence and Interactive Entertainment* (2001).
- Jeff RICKEL et W. Lewis JOHNSON. *Task-Oriented Collaboration with Embodied Agents in Virtual Worlds*. Dans [Cassell et al., 2000d], chapitre 4, pp. 95–122.
- Anne-Marie ROCHEBLAVE. *Rôles et Statuts*. Dans *Encyclopædia Universalis*, pp. 114–116 (1994).
- Gérard SABAH. *L'intelligence Artificielle et Le Langage : Processus de Compréhension*, tome 2. Hermes, Paris (1989).
- Gérard SABAH. *Dialogue et Sciences Cognitives*. Dans D. LUZZATI, J.-C. BEACCO, R. MIR-SAMII, M. MURAT et M. VIVET, édés., *Le Dialogique*, pp. 323–346. Sciences pour la communication, Berne (1997a).
- Gérard SABAH. *Le Sens Dans Les Traitements Automatiques Des Langues*. *TA-information*, tome 38, n° 2 : pp. 91–133 (1997b). (Actes de la journée ATALA du 14 Décembre 1996 : "un demi-siècle de traitement automatique des langues : état de l'art").
- Jean-Paul SANSONNET et Erika VALENCIA. *Un Système de Représentation Des Connaissances Fondé sur L'accès Par Le Contenu Pour L'interaction Dialogique*. Rapport technique, LIMSI (2000).
- John SEARLE et Daniel VANDERVEKEN. *Foundations of Illocutionary Logic*. Cambridge University Press (1985).
- John R. SEARLE. *Speech Acts*. Cambridge University Press (1969).
- Claude SHANNON. *A Mathematical Theory of Communication*. *The Bell system technical journal*, tome 27 : pp. 379–423, 623–656 (1948). (reprint with corrections).
- V. SHEA. *Netiquette*. Albion books, San Francisco (1994). (Cité par Marcoccia [Marcoccia, 1999]).
- John SOWA. *Conceptual Structures : Information Processing in Mind and Machine*. Addison-Wesley (1984). (Référence issue de notes de cours).
- Luc STEELS et Frédéric KAPLAN. *Situated Grounded Word Semantics*. Dans T. DEAN, éd., *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI'99)*, pp. 862–867. Morgan Kaufmann (1999).
- Bruce STERLING, éd. *Mirrorshades : A Cyberpunk Anthology* (1987). Edition française : "Mozart en verres miroir", Denoël, 1987.
- Obed E. TORRES. *Producing Semantically Appropriate Gestures in Embodied Language Generation*. Mémoire de DEA / Master, MIT (1997).
- Gulden UCHYIGIT et Keith CLARK. *Agents That Model and Learn User Interests for Dynamic Collaborative Filtering*. Dans M. KLUSH, S. OSSOWSKY et O. SHEHORY, édés., *Proceedings of CIA 2002*, pp. 152–163. Springer-Verlag (2002).
- Francis VAN AEKEN. *Les Systèmes Multi-Agents Minimaux, un Modèle Adapté À L'étude de la Dynamique Organisationnelle Dans Les Systèmes Multi-Agents Ouverts*. Thèse de doctorat, INP Grenoble (1999).
- Daniel VANDERVEKEN. *Les Actes de Discours*. Pierre Mardaga, Liège (1988).
- Daniel VANDERVEKEN. *Meaning and Speech Acts*, tome 1 (Principles of Language Use). Combridge (1990a).

- Daniel VANDERVEKEN. *Meaning and Speech Acts*, tome 2 (Formal semantics of success and satisfaction). Combridge (1990b).
- Daniel VANDERVEKEN. *La Structure Logique Des Dialogues Intelligents*. Dans [Moulin *et al.*, 1999], chapitre 2.
- Walter VAN DE VELDE. *Co-Habited Mixed-Realities*. Dans Fumio HATTORI, éd., *Proceedings of the IJCAI'97 Workshop on Social Interaction and Communityware* (1997).
- Denis VERNANT. *Les Niveaux D'analyse Des Phénomènes Communicationnels : Sémantique, Pragmatique et Praxéologique*. Dans [Moulin *et al.*, 1999], pp. 101–131.
- H. VILHJALMSSON et J. CASSELL. *BodyChat : Autonomous Communicative Behaviors in Avatars*. Dans *Proceedings of ACM Second International Conference on Autonomous Agents* (1998).
- Hannes VILHJÁLMSOON. *Autonomous Communicative Behaviours in Avatars*. Mémoire de DEA / Master, MIT (June 1997).
- Diane VINCENT, Marty LAFOREST et Guylaine MARTEL. *Le Corpus de Montréal 1995. Adaptation de la Méthodologie Sociolinguistique Pour L'analyse Conversationnelle*. *Dialangue*, tome 6 : pp. 29–45 (1995).
- Douglas WALTON. *What is Reasoning ? What is an Argument ?*. *Journal of philosophy*, tome 87 : pp. 399–419 (1990).
- Gerhard WEISS, éd. *Multiagent Systems. A Modern Approach to Distributed Artificial Intelligence*. MIT Press (1999).
- Joseph WEIZENBAUM. *ELIZA - a Computer Program for the Study of Natural Language Communication Between Man and Machine*. *Communications of the ACM*, tome 9, n° 1 : pp. 36–45 (1966). <http://doi.acm.org/10.1145/365153.365168>.
- J.S WIGGINS, P. TRAPNELL et N. PHILLIPS. *Psychometric and Geometric Characteristics of the Revised Interpersonal Adjective Scales (IAS-R)*. *Multivariate Behavioral Research*, tome 23 : pp. 517–530 (1988). (Référence issue de notes de cours).
- Toshihiko YAMAKAMI et Gen-Ichi NISHIO. *Social Pattern Development Analysis : A Case Study in a Regional Community Network*. Dans [Ishida, 1998a], pp. 170–182.
- Xiaoquin ZHANG, Victor LESSER et Tom WAGNER. *Integrative Negotiation in Complex Organizational Agent Systems*. Dans Cristiano CASTELFRANCHI et W. Lewis JOHNSON, édés., *AAMAS 2002*, pp. 503–504. ACM Press (2002).

Annexe A

Documents électroniques

- [ActiveWorlds] Logiciels de création et gestion de communautés virtuelles basées sur un environnement 3D (client, serveur et backoffice).
<http://www.activeworlds.com/>
- [Blaxxun] Logiciels de création et gestion de communautés virtuelles basées sur un environnement 3D (client, serveur et backoffice). Technologie employée pour le Deuxième Monde.
<http://www.blaxxun.com/>
- [Colonicity] Communauté virtuelle basée sur la technologie Blaxxun.
<http://www.cybertown.com>
- [Kyoto] Digital City Kyoto. Communauté virtuelle basée sur la ville de Kyōtō (arrêt de la maintenance des pages depuis Octobre 2001).
<http://www.digitalcity.gr.jp/index-e.html>
- [2M] Deuxième Monde. Communauté virtuelle développée par CANAL+ et basée sur la technologie Blaxxun (fermée depuis le printemps 2002).
<http://www.2monde.com>
- [Fabric|ch] Artistes contemporains travaillant sur les environnements virtuels (technologie Blaxxun).
http://www.fabric.ch/La_Fabrique00
- [GNL] *Gesture and Narrative Language Group*. Groupe de recherche sur les agents conversationnels incarnés.
<http://gn.www.media.mit.edu/groups/gn/index.html>
- [Mauz] Site construit autour des communautés virtuelles basées sur la technologie Active Worlds.
<http://mauz.info/>
- [Ultima] Ultima On Line. Communauté virtuelle basée sur le jeu d'aventure Ultima.
<http://www.uo.com/>

DIALOGUE ENTRE AGENTS NATURELS ET AGENTS ARTIFICIELS.
UNE APPLICATION AUX COMMUNAUTÉS VIRTUELLES

Cette thèse s'inscrit dans le cadre de la constitution de communautés mixtes rassemblant des agents naturels (humains) et des agents artificiels. Dans cet objectif, la thèse présentée traite des interactions pouvant exister entre ces types d'agents lorsqu'ils sont mis en présence au sein d'une communauté virtuelle.

Les modèles d'agent et d'interaction proposés reposent sur l'hypothèse que chaque message est porteur d'un *potentiel* de sens —négociable— qu'il est nécessaire d'interpréter et non pas porteur d'*un* sens —unique et précis— qu'il suffit de décoder. Ceci implique les deux points centraux suivants : l'enrichissement contextuel du message et le focus conversationnel. Le premier point, l'enrichissement contextuel, associe un message avec son contexte de façon à permettre l'interprétation de ce message plutôt qu'un simple décodage comme cela est le cas dans les applications informatiques classiques. De manière duale, un processus d'expression remplace l'encodage. Le deuxième point, le focus conversationnel, représente l'objet de la discussion, lui aussi négociable et construit de manière collective par l'ensemble des interactants, leurs différentes interventions entraînant l'émergence du dialogue. Un agent conversationnel dédié à la recommandation de films, dans le cadre d'une boutique virtuelle 3D, constitue le cadre technologique et applicatif de la thèse défendue qui se conclut sur l'analyse de plusieurs dialogues entre cet agent et un client humain.

DIALOGUE BETWEEN NATURAL AGENTS AND ARTIFICIAL AGENTS
WITHIN A VIRTUAL COMMUNITY

In order to set up mixed communities in which natural agents (humans) and artificial agents are brought together, it is necessary to ensure smooth interactions.

This thesis presents models of interaction and agent, based on the hypothesis that a message carries a *potential* meaning —subject to discussion— which has to be interpreted, and not *one* meaning —set and precisely defined— which only needs to be decoded. This hypothesis implies the two cornerstones of this thesis : contextual enhancement and conversational focus. The first point, contextual enhancement, associates a message to its context in a process of interpretation, instead of a simple decoding process, as is usual in typical information systems. Symmetrically, an expression process replaces the encoding. The second point, conversational focus, is a representation of the topic of the interaction. It is collectively built by the people involved in the interaction and subject to discussion, which naturally leads to conversational follow-up.

A conversational agent for movies recommendation in a 3D virtual shop has been designed following the presented models, and several interactions between this agent and a human customer are analysed.

Systèmes Multi-Agents	Agents conversationnels	Communautés mixtes
Communautés virtuelles	Interaction personne-système	Dialogue
Pragmatique	Communication langagière	
