

Quatrième Workshop sur les
Agents Conversationnels Animés

25 et 26 novembre 2010 - Lille

Avant-propos

WACA 2010 constitue la quatrième édition du Workshop sur les Agents Conversationnels Animés (WACA). Après le succès des éditions de Grenoble (2005), Toulouse (2006) et Paris (2008), c'est cette année Lille qui accueille.

Les Agents Conversationnels Animés (ACA) -ou Embodied Conversational Agents en anglais- se situent à la rencontre des interfaces homme-machine, de l'intelligence artificielle, de l'informatique émotionnelle et de l'interaction homme-machine. Depuis une dizaine d'années, ils apparaissent de plus en plus comme des outils puissants pour l'assistance aux utilisateurs, les jeux vidéos, les environnements de conception participative, les environnements virtuels ou l'informatique ubiquitaire et l'enseignement à distance.

L'objectif de ce workshop est de recenser et de réunir les travaux déjà en cours sur les ACAs afin que les chercheurs des communautés scientifiques concernées puissent échanger leurs outils et leurs méthodes.

Cette année, le programme de WACA est composé de 16 présentations (12 longues, 3 courtes et 1 présentation invitée). Chacun des articles a été évalué par deux relecteurs. Les affiliations des auteurs reflètent l'aspect pluridisciplinaire du workshop : on y trouve des chercheurs en Informatique (en particulier en intelligence artificielle, en système multi-agents et en interaction homme-machine) en psychologie cognitive et en marketing. Les contributions sont organisées autour des axes de recherche suivants :

- Programmation de la personnalité et des émotions
- Détection et expression des comportements émotionnels
- Texte & ACA
- ACA vendeurs

Nous tenons à remercier chaleureusement les personnes qui ont contribué au succès de WACA 2010 : le comité de programme ainsi que les relecteurs supplémentaires et le comité d'organisation. Par ailleurs, l'inscription à cette édition est gratuite, c'est grâce au soutien financier d'Orange Labs, de la société Dialonics, du conseil régional Nord-Pas de calais et de l'Association Française d'Intelligence Artificielle. Nous les en remercions.

*Stéphanie Buisine
Nicolas Sabouret
Maxime Morge*

Sponsors



Comités

Co-présidents du comité de programme

Nicolas Sabouret (LIP6, Université Paris 6)
Stéphanie Buisine (LCPI, Télécom ParisTech)

Comité de programme

Carole Adam (Agents@RMIT, Melbourne)
Dia Aljrab (AsAnAngel, Paris)
Gérard Bailly (GIPSA-Lab, Grenoble)
Elisabetta Bevacqua (LTCI, Paris)
Stéphanie Buisine (LCPI, Paris)
Bjorn Gulsdorff (Artificial Solutions, Hambourg)
Marie-Jeanne Lesot (LIP6, Paris)
Muriel Lauvergne (Orange, Lannion)
Vincent Louis (Dialonics, Lannion)
Jean-Claude Martin (LIMSI-CNRS, Orsay)
Maxime Morge (LIFL, Lille)
Magalie Ochs (LTCI, Paris)
Alexandre Pauchet (LITIS, Rouen)
Sylvie Pesty (LIG, Grenoble)
Ken Prépin (LTCI, Paris)
Nicolas Sabouret (LIP6, Paris)
Jean-Paul Sansonnet (LIMSI-CNRS, Orsay)
Benoît Tabutiaux (AsAnAngel, Paris)

Comité d'organisation

Maxime Morge (LIFL, Université Lille 1)
Bruno Beaufls (LIFL, Université Lille 1)
Sameh Abdel-Naby (LIFL, Université Lille 1)
Yann Secq (LIFL, Université Lille 1)

Relecteur additionnel

Antoine Nongaillard (LIFL, Lille)

Liste des contributions

Programmation de la personnalité et des émotions	
Influence des Traits de Personnalité sur la Sélection des Rétroactions	
<i>Etienne de Sevin, Sylwia Julia Hyniewska et Catherine Pelachaud</i>	15
Programmation Orientée Émotion (EOP)	
<i>Kévin Darty et Nicolas Sabouret</i>	23
Un modèle de comportements affectifs fondé sur la Conservation de Ressources	
<i>Sabrina Campano, Nicolas Sabouret et Vincent Corruble</i>	31
Actes de conversation multimodaux et émotions	
<i>Jérémy Rivière et Sylvie Pesty</i>	39
Articles courts	
Expressing agents personality through ambient modalities	
<i>Jean-Paul Sansonnet et Yacine Bellik</i>	47
Expressive Gesture Model for Storytelling Humanoid Agent	
<i>Quoc-Anh Le et Catherine Pelachaud</i>	49
Un système multi-agents permettant une assistance homme-machine mutuelle dans un environnement ambiant	
<i>Hajer Sassi, José Rouillard, Jean-Claude Tarby et Grégory Renard</i>	51
Détection et expression des comportements émotionnels	
Combinaisons d’expressions vocales, faciales et posturales des émotions chez un agent animé : ce que perçoivent les utilisateurs	
<i>Céline Clavel, Laurence Devillers et Jean-Claude Martin</i>	53
Expressions Posturales de Tendances à l’Action	
<i>Ning Tan, Céline Clavel, Mathieu Courgeon et Jean-Claude Martin</i>	61
A la recherche d’indices de culture et/ou de langue dans les micro-événements audio-visuels de l’interaction face à face	
<i>Rosario Signorello, Véronique Aubergé, Anne Vanpé, Lionel Granjon et Nicolas Audibert</i>	69
NaviQuest : un outil pour naviguer dans une base de questions posées à un Agent Conversationnel	
<i>Philippe Suignard</i>	77

Texte & ACA

Classification of Wordnet Personality Adjectives in the NEO PI-R Taxonomy

François Bouchet et Jean-Paul Sansonnet 83

De la description de scénarios socio-émotionnels au narrateur virtuel expressif

Roman Miletitch, Nicolas Sabouret et Magalie Ochs 91

ACA Vendeurs

Agents virtuels vendeurs : que veulent les consommateurs ?

Mohamed Slim Ben Mimoun et Ingrid Poncin 99

Business-Driven Negotiation Strategies for Proactive Embodied Conversational Agents in E-Commerce

Sameh Abdel-Naby, Maxime Morge et Bruno Beaufls 107

Liste des contributeurs

Sameh ABDEL-NABY

`sameh.abdel-naby@lifl.fr`

Laboratoire d'Informatique Fondamentale de Lille (LIFL)

Université Lille 1/CNRS

Véronique AUBERGÉ

`veronique.auberge@gipsa-lab.grenoble-inp.fr`

Grenoble Images Parole Signal Automatique (GIPSA-Lab)

Laboratoire d'Informatique de Grenoble (LIG)

Nicolas AUDIBERT

`nicolas.audibert@gmail.com`

Laboratoire Informatique d'Avignon (LIA)

Université d'Avignon

Bruno BEAUFILS

`bruno.beaufils@lifl.fr`

Laboratoire d'Informatique Fondamentale de Lille (LIFL)

Université Lille 1/CNRS

Yacine BELLIK

`bellik@limsi.fr`

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI)

CNRS

Mohamed Slim BEN MIMOUN

`m.slim_benmimoun@skema.edu`

Lille School of Management Research Center (LSMRC)
Skema Business School

François BOUCHET

`bouchet@limsi.fr`

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI)

CNRS

Sabrina CAMPANO

`sabrina.campano@lip6.fr`

Laboratoire d'Informatique de Paris 6 (LIP6)

Université Pierre et Marie CURIE/CNRS

Céline CLAVEL

`celine.clavel@limsi.fr`

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI)

CNRS

Vincent CORRUBLE

`vincent.corruble@lip6.fr`

Laboratoire d'Informatique de Paris 6 (LIP6)

Université Pierre et Marie CURIE/CNRS

Mathieu COURGEON

`courgeon@limsi.fr`

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI)

CNRS

Kévin DARTY

`dartykevin@gmail.com`

Laboratoire d'Informatique de Paris 6 (LIP6)

Université Pierre et Marie CURIE/CNRS

Laurence DEVILLERS

`devil@limsi.fr`

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI)

CNRS

Lionel GRANJON

`lionel.granjon@gipsa-lab.grenoble-inp.fr`

Grenoble Images Parole Signal Automatique (GIPSA-Lab)

Sylwia Julia HYNIEWSKA

`sylwia.hyniewska@telecom-paristech.fr`
Telecom ParisTech

Quoc-Anh LE

`quoc-anh.le@telecom-paristech.fr`
Laboratoire Traitement et Communication de
l'Information (LTCI)
Telecom ParisTech/CNRS

Jean-Claude MARTIN

`martin@limsi.fr`
Laboratoire d'Informatique pour la Mécanique et les
Sciences de l'Ingénieur (LIMSI)
CNRS

Roman MILETITCH

`roman.miletitch@gmail.com`
Laboratoire d'Informatique de Paris 6 (LIP6)
Université Pierre et Marie CURIE/CNRS

Maxime MORGE

`maxime.morge@lifl.fr`
Laboratoire d'Informatique Fondamentale de Lille
(LIFL)
Université Lille 1/CNRS

Magalie OCHS

`magalie.ochs@gmail.com`
Laboratoire Traitement et Communication de
l'Information (LTCI)
Telecom ParisTech/CNRS

Catherine PELACHAUD

`catherine.pelachaud@telecom-paristech.fr`
Laboratoire Traitement et Communication de
l'Information (LTCI)
Telecom ParisTech/CNRS

Sylvie PESTY

`sylvie.pesty@imag.fr`
Laboratoire d'Informatique de Grenoble (LIG)

Ingrid PONCIN

`i.poncin@skema.edu`
Lille School of Management Research Center (LSMRC)
Skema Business School

Grégory RENARD

`gregory.renard@xbrainlab.com`
xBrainLab
Groupe Usilink

Jérémy RIVIÈRE

`jeremy.riviere@imag.fr`
Laboratoire d'Informatique de Grenoble (LIG)

José ROUILLARD

`jose.rouillard@univ-lille1.fr`
Laboratoire d'Informatique Fondamentale de Lille
(LIFL)
Université Lille 1/CNRS

Nicolas SABOURET

`nicolas.sabouret@lip6.fr`
Laboratoire d'Informatique de Paris 6 (LIP6)
Université Pierre et Marie CURIE/CNRS

Nicolas SABOURET

`nicolas.sabouret@lip6.fr`
Laboratoire d'Informatique de Paris 6 (LIP6)
Université Pierre et Marie CURIE/CNRS

Jean-Paul SANSONNET

`jps@limsi.fr`
Laboratoire d'Informatique pour la Mécanique et les
Sciences de l'Ingénieur (LIMSI)
CNRS

Hajer SASSI

`hajer.sassi@ed.univ-lille1.fr`
Laboratoire d'Informatique Fondamentale de Lille
(LIFL)
Université Lille 1/CNRS

Rosario SIGNORELLO

`rosario.signorello@gipsa-lab.grenoble-inp.fr`
Grenoble Images Parole Signal Automatique
(GIPSA-Lab)

Philippe SUIGNARD

`philippe.suignard@edf.fr`
EDF R&D

Ning TAN

`ntan@limsi.fr`

Laboratoire d'Informatique pour la Mécanique et les
Sciences de l'Ingénieur (LIMSI)
CNRS

Jean-Claude TARBY

`jean-claude.tarby@univ-lille1.fr`

Laboratoire d'Informatique Fondamentale de Lille
(LIFL)
Université Lille 1/CNRS

Anne VANPÉ

`anne.vanpe@gipsa-lab.grenoble-inp.fr`

Grenoble Images Parole Signal Automatique
(GIPSA-Lab)
Laboratoire d'Informatique de Grenoble (LIG)

Etienne DE SEVIN

`etienne.de-sevin@lip6.fr`

Laboratoire d'Informatique de Paris 6 (LIP6)
Université Pierre et Marie CURIE/CNRS

Influence des Traits de Personnalité sur la Sélection des Rétroactions

Etienne de Sevin¹, Sylwia Julia Hyniewska² and Catherine Pelachaud²
etienne.de-sevin@lip6.fr, {sylwia.hyniewska, catherine.pelachaud}@telecom-paristech.fr

¹LIP6 - UPMC
4 place Jussieu
75005 Paris

²CNRS - Telecom ParisTech
37/39, rue Dareau
75014 Paris, France

Résumé :

Notre but est de construire un agent conversationnel animé, capable de se comporter comme un interlocuteur lors d'une interaction avec un utilisateur en temps-réel. Pour être crédible, l'ACA doit avoir un comportement d'écoute différent (rétroactions) en fonction de sa personnalité. Dans ce papier, nous présentons un algorithme de sélection de rétroactions qui choisit en temps réel le type et la fréquence des rétroactions jouées par l'agent en fonction de sa personnalité. Une évaluation est effectuée sur l'adéquation entre les rétroactions générées par cet algorithme et les attentes des participants sur les comportements spécifiques de l'agent liés aux différentes personnalités.

Mots-clés : Rétroactions, Personnalités, ACAs.

Abstract :

Our aim is to build an Embodied Conversational Agent able to act as an interlocutor during an interaction with a user in real-time. To be credible, the ECA should have different listening behaviors (backchannels) depending on its personality. In this paper, we propose an action selection algorithm working to choose in real-time the type and the frequency of backchannels displayed by the ECA in accordance with its personality. An evaluation is performed to verify the adequacy between the generated backchannels by our algorithm and the participant expectations on the specific behaviours of the ECA linked to several personalities.

Keywords: Backchannels, Personality, ECAs.

1. Introduction

Lors de la création d'un agent conversationnel animé (ACA), la principale difficulté réside dans sa crédibilité non seulement dans son aspect visuel mais aussi dans son comportement [1]. En effet l'ACA doit

montrer des comportements appropriés en fonction de ses états internes comme la personnalité et de ses états externes [2].

Ce travail fait partie du projet européen SEMAINE. Il a pour but de construire un interlocuteur artificiel sensitif [3] et multimodal capable d'exhiber des comportements verbaux et non-verbaux appropriés de façon autonome lors d'une interaction avec un utilisateur. Basé sur une approche psychologique, quatre personnages ont été définis : pessimiste, extraverti, fiable et agressif. Chacun a ses propres stratégies individuelles de dialogue et possède ses propres rétroactions [4]. Dans cette perspective, il est important d'avoir des comportements d'écoute crédibles.

Les traits de personnalité influencent les comportements et sont une caractéristique fondamentale d'un individu. Nous basons notre travail sur une approche dimensionnelle de la personnalité. Pour limiter la complexité de notre modèle, nous nous focalisons sur deux dimensions: extraversion-introversion et stabilité-instabilité émotionnelle en utilisant la représentation d'Eysenck [5] (voir figure 1). Ces deux dimensions se retrouvent dans toutes les théories majeures des traits de personnalité comme le modèle OCEAN [6]. Nous pouvons ainsi formuler des prédictions sur le comportements de l'ACA en terme d'imitations ou de quantité de mouvements en fonction de sa personnalité.

L'empathie est la capacité de partager ou d'interpréter correctement les émotions et les sentiments des autres [7]. Des chercheurs ont montré qu'en général l'imitation permet d'avoir une interaction plus agréable car elle augmente le sentiment d'empathie [8]. Les individus empathiques imitent beaucoup plus les postures, les gestes et les expressions faciales que les individus non empathiques [9]. Selon Eysenck [10], l'instabilité émotionnelle est corrélée négativement avec l'empathie. On peut donc en déduire qu'un haut niveau de stabilité émotionnelle peut être corrélée avec une tendance élevée à imiter.

Des recherches ont aussi montré qu'un haut niveau d'extraversion est associé avec de nombreux gestes, des mouvements de tête plus fréquents et à une augmentation générale des mouvements [11].

Dans cet article, nous nous intéressons à l'influence de la personnalité sur les comportements d'écoute d'un agent interlocuteur [12]. Nous proposons un algorithme de sélection qui choisit en temps-réel le type et la fréquence des rétroactions, en fonction de la personnalité de l'ACA. Nous associons ces deux variables des rétroactions (type et fréquence) aux deux dimensions de la personnalité: extraversion et stabilité émotionnelle. Ainsi les personnages instables auront une tendance plus faible à imiter que les stables et les personnages extravertis montreront plus souvent des rétroactions que les introvertis.

La prochaine section est une brève description de l'état de l'art. La section 3 présente notre algorithme de sélection des rétroactions en fonction de la personnalité de l'ACA. Nous détaillons nos deux hypothèses : l'extraversion est associée avec la fréquence des rétroactions alors que la stabilité émotionnelle est associée avec le type des rétroactions. Dans la section 4, nous détaillons l'évaluation que nous avons effectuée pour vérifier nos hypothèses et dans la section 5, nous discutons des résultats.

2. Etat de l'art

Plusieurs modèles de rétroactions pour les agents conversationnels animés ont été étudiés. Thórisson [13] a développé une tête parlante, appelée Gandalf, capable de produire en temps réel des rétroactions pendant la conversation avec un utilisateur. L'agent interlocuteur [14], développé à ICT, produit des rétroactions basées sur une analyse en temps réel des comportements non verbaux du locuteur (comme les mouvements de la tête et les postures du corps) et des caractéristiques acoustiques extraites de la voix du locuteur. Kopp [15] propose un modèle pour la génération incrémentale des rétroactions. Le système est basé à la fois sur un modèle probabilistique, qui définit un ensemble de règles pour déterminer les occurrences des rétroactions, et sur un modèle simulé qui perçoit, comprend et évalue les entrées sensorielles.

3. Algorithme de sélection des rétroactions pour un ACA interlocuteur

Le travail proposé dans cet article est une partie d'un système préexistant pour la génération de rétroactions d'un ACA interlocuteur. Dans cette architecture, deux types de rétroactions ont été implémentés : les imitations et les rétroactions réactives [1].

Dans le projet SEMAINE, quatre agents virtuels interlocuteurs ont été définis avec chacun leurs propres traits de personnalité. Poppy est extravertie et joyeuse; Spike est agressif et pugnace; Prudence est fiable et pragmatique; et Obadiah est pessimiste et déprimé. Nous pouvons placer les traits de personnalités des agents par rapport aux deux dimensions d'Eysenck (voir figure 1): extraversion et stabilité émotionnelle [5].

Notre but est de tester les liens entre les personnalités et les variables des rétroactions (fréquence et type) que peut modifier l'algorithme de sélection. Pour se concentrer

seulement sur les comportements et éviter d'avoir à prendre en compte des variables supplémentaires, nous avons utilisé seulement un modèle de visage (Prudence) pour les quatre traits de personnalités. Nous formulons deux hypothèses :

H1 : la dimension extraversion peut être associée linéairement avec la fréquence des rétroactions. Poppy (extravertie) va montrer plus souvent des rétroactions qu'Obadiah (pessimiste) [11].

H2 : la dimension stabilité émotionnelle peut être associée linéairement avec le type de rétroactions. Plus la stabilité émotionnelle est grande plus l'ACA imitera l'utilisateur. Prudence (fiable) va montrer plus d'imitations que Spike (agressif) [8].

Si la tendance à imiter est négative, l'algorithme de sélection choisira de montrer plus de rétroactions réactives. Avec ces deux hypothèses, nous pouvons définir la fréquence et le type de rétroactions pour chaque personnalité des agents. En effet en localisant celle-ci sur la représentation d'Eysenck [5] et en translatant directement sa position sur notre graphe, on obtient les valeurs du nombre (fréquence) et du type (tendance à imiter) des rétroactions que l'on peut utiliser dans notre algorithme de sélection des rétroactions. Par exemple, Obadiah (pessimiste), fera peu de rétroactions (-0.75) et seulement quelques imitations (-0.6). On peut déduire de la même façon les valeurs pour les personnalités décrites dans la représentation d'Eysenck (voir tableau 1).

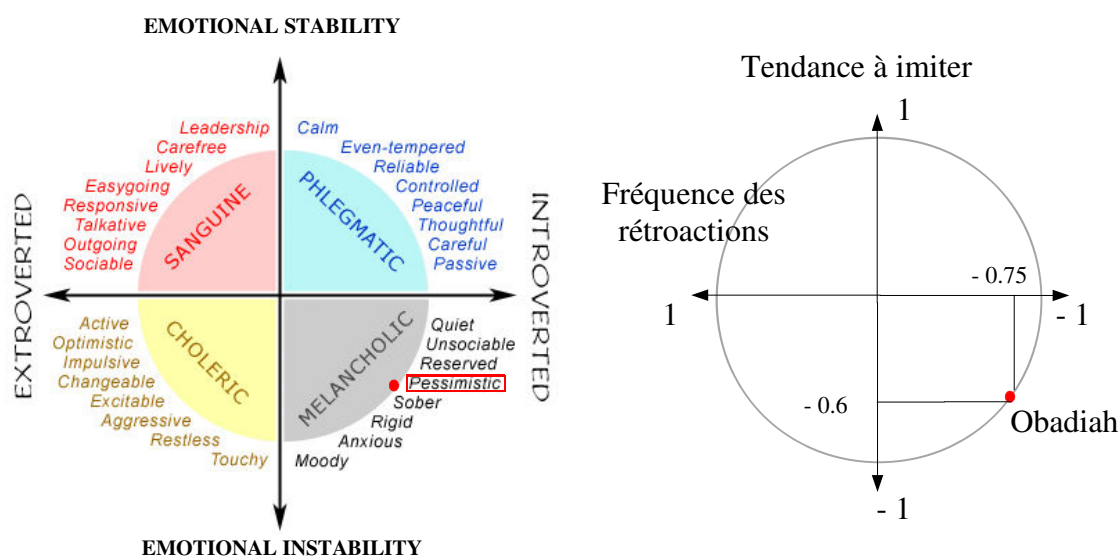


FIG. 1 - La représentation des personnalités d'Eysenck à deux dimensions [5] et notre graphe pour en déduire la fréquence et le type (tendance à imiter) des rétroactions pour chaque personnalité.

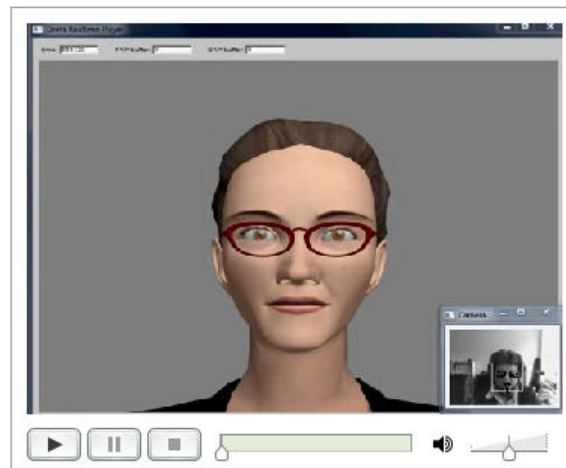
Exemple de déduction pour Obadiah.

TAB. 1 - Déduction des valeurs du type (tendance à imiter) et de la fréquence des rétroactions pour les quatre personnalités (comprises entre [-1 ; 1]).

	Obadiah (pessimiste)	Poppy (extravertie)	Prudence (fiable)	Spike (agressif)
Tendance à imiter	- 0.6	0.25	0.90	- 0.85
Fréquence des rétroactions	- 0.85	0.95	- 0.5	0.55

Personnalité de l'agent: pragmatique

*Une histoire est racontée à l'agent virtuel. Celui-ci doit se comporter de façon pragmatique. Veuillez regarder les deux vidéos et **répondre aux questions par rapport à la personnalité de l'agent virtuel.** Merci.*



L'agent réagit-il de façon pragmatique ?

L'agent réagit-il suffisamment ?



L'agent réagit-il de façon pragmatique ?

L'agent réagit-il suffisamment ?

Quelle vidéo correspond le mieux à la personnalité pragmatique de l'agent ?

[Suite](#)

FIG. 2 - Exemple d'une page internet de notre évaluation où deux vidéos de l'agent en interaction avec un utilisateur sont évaluées par les participants. Dans l'une des deux vidéos, l'agent virtuel a des rétroactions correspondant à la personnalité pragmatique (fiable). L'autre correspond soit à la personnalité extravertie soit pessimiste dans la même condition.

4. Evaluation

Nous avons fait une évaluation sur internet pour avoir un grand nombre de participants. Notre corpus est composé de vidéos d'une femme qui raconte une histoire à notre agent interlocuteur en se basant sur des extraits d'une bande dessinée. Celui-ci montre des rétroactions d'écoute en fonction de sa personnalité. Nous évaluons deux variables : la fréquence et le type des rétroactions pour les quatre personnalités de l'agent : pessimiste, extravertie, fiable et agressive. Nous montrons aux participants des vidéos du corpus et nous leur demandons d'évaluer pour chaque vidéo si la fréquence et le type des rétroactions de l'agent sont en adéquation avec les quatre personnalités de l'agent (voir figure 2).

Chaque page d'évaluation est composée de la description de la personnalité de l'agent virtuel, d'un résumé des instructions et de deux vidéos. Pour chaque vidéo, les participants répondent sur une échelle de Likert si l'agent interlocuteur réagit de façon appropriée (compris entre "pas du tout" à "complètement") et suffisamment (compris entre "pas assez" à "trop") en fonction de la personnalité décrite. Ensuite, les participants doivent déterminer quelle vidéo (entre les deux présentées) correspond le plus à la personnalité de l'agent décrite dans l'énoncé.

Pour chaque personnalité, trois conditions sont testées en fonction de l'utilisation ou non de notre algorithme de sélection des rétroactions :
 C1 : variation de la fréquence des rétroactions (type des rétroactions = 0 sur notre graphe)
 C2 : variation du type des rétroactions (fréquence des rétroactions = 0)
 C3 : variation de la fréquence et du type des rétroactions

Une seule condition est évaluée par page pour les quatre personnalités (12 pages en tout). Sur chaque page, il y a une vidéo dans laquelle les rétroactions de l'agent correspondent à la personnalité décrite sur la page. L'autre vidéo

est choisie aléatoirement dans celles montrant l'autre dimension de la personnalité dans la même condition. Par exemple, si la dimension de la personnalité de l'agent est l'extraversion et la condition est de faire varier la fréquence des rétroactions, une des vidéos correspondra à Poppy (extravertie). L'autre vidéo sera choisie aléatoirement parmi les vidéos de l'autre dimension (stabilité émotionnelle), c'est à dire Spike (agressif) ou Prudence (fiable) dans la même condition. La position des deux vidéos est aléatoire pour empêcher les participants de reconnaître l'ordre des vidéos.

5. Résultats

Quatre vingt treize personnes (56 femmes et 37 hommes) ont participé à notre étude sur internet. 54 % des participants ont choisi la vidéo appropriée à la personnalité de l'agent décrite dans l'énoncé. La personnalité la plus reconnue est agressive (62%), puis extravertie (57%), pessimiste (53%), et fiable (52%).

Hypothèse H1 : La majorité des participants ont répondu que la fréquence des rétroactions de l'agent est appropriée pour les quatre personnalités. Ce n'était pas le choix par défaut et les participants ont choisi activement cette réponse. Une analyse de la variance a été réalisée (ANOVA), ainsi que des tests t pour échantillons appariés. Nous vérifions le lien entre la fréquence des rétroactions et la dimension extraversion (H1, voir section 3). Notre prédiction est que la condition C3 est mieux évaluée par les participants que la condition C2 en particulier pour Poppy (extravertie) et Obadiah (pessimiste).

Les réponses des participants à la question si l'agent réagit suffisamment montre un effet sur la personnalité et sur la condition (ANOVA, $p < .05$) mais pas sur l'interaction entre la personnalité et la condition (ANOVA, $p > .05$) (voir tableau 2). La variation de la fréquence des rétroactions pour Poppy (extravertie) et Prudence (fiable) ont été significatives (t test, $p < .05$) mais pas pour Obadiah (pessimiste) et Spike (agressif) ($p > .05$) (voir tableau 3).

TAB. 2 - Résultats du test ANOVA pour la condition H1

	n	F	p	
Personnalité	3	9.737	.000	*
Condition	3	18.032	.000	*
Personnalité*Condition	6	1.369	.225	

TAB. 3 - Résultats des tests *t* pour échantillons appariés pour la condition H1

Fréquence des rétroactions	n	$\bar{X}$	Test t	p	
Obadiah C2/C3	93	2.78/2.79	-,101	.460	
Poppy C2/C3	93	2.84/3.13	-2.457	.008	*
Prudence C2/C3	93	2.84/3.05	-1.818	.036	*
Spike C2/C3	93	2.62/2.81	-1.592	.057	

TAB. 4 - Résultats du test ANOVA de Friedman pour la condition H2

n	Chi-Square	df	p	
94	141.948	11	.000	*

TAB. 5 - Résultats des tests de Wilcoxon pour échantillons pour la condition H2

Type de rétroactions	n	$\bar{X}$	p	
Obadiah C1/C3	94	2.01/2.67	.000	*
Poppy C1/C3	94	2.75/2.69	.433	
Prudence C1/C3	94	2.83/2.77	.253	
Spike C1/C3	94	1.91/2.01	.182	

Les participants considèrent que la condition C3 pour Poppy and Prudence est meilleure que la condition C2. La variation du type de rétroactions pour les quatre personnalités est significative ($p < .05$). Les participants considèrent que la condition C3 est meilleure que la condition C1 pour toutes les personnalités.

Hypothèse H2 : Concernant le type de rétroactions, 40,6% des participants ont mieux évalué la vidéo correspondant à la personnalité décrite dans l'énoncé et 25.9% ont évalué de façon identique les deux vidéos présentées. À part Spike, les rétroactions ont été évaluées par les participants comme appropriées, en particulier celles de Poppy. Nous avons effectué un test de Friedman et des tests de Wilcoxon pour échantillons appariés pour

vérifier notre hypothèse H2 à propos du lien entre le type de rétroactions et la dimension stabilité émotionnelle (voir section 3). Notre prédiction est que la condition C3 est mieux évaluée par les participants que la condition C1 en particulier pour Prudence (fiable) et Spike (agressif).

La réponse des participants à la question si l'agent réagit de façon appropriée par rapport à la personnalité décrite dans l'énoncé est significative (test Friedman, $p < .05$) (voir tableau 4). La variation du type de rétroactions pour Obadiah (pessimiste) est significative ($p < .05$) mais pas pour les autres personnalités ($p > .05$) (voir tableau 5). Les participants considèrent, seulement pour Obadiah, que la condition C3 est meilleure que la condition C1.

6. Discussion

Le but de cette évaluation est de vérifier si la variation du type et de la fréquence des rétroactions a un impact sur la perception de la personnalité des agents par les participants. La première hypothèse est partiellement vérifiée. Bien que les attributions soient plus hautes pour les vidéos avec la sélection de la fréquence des rétroactions que les alternatives, la différence n'a pas été significative pour certaines personnalités. La seconde hypothèse n'a été vérifiée qu'avec Obadiah (pessimiste).

Concernant l'hypothèse H1, la majorité des participants ont jugé que la fréquence des rétroactions est plus appropriée à la personnalité de l'agent, décrite dans l'énoncé de la page d'évaluation, lorsqu'elle est sélectionnée par notre algorithme. Cependant cela a été significatif que pour Poppy (extravertie) et Prudence (fiable). Notre prédiction pour Poppy est juste mais pas pour Obadiah (pessimiste). En effet quand un agent montre beaucoup de rétroactions, il est associé avec l'extraversion. Quand il fait moins de rétroactions que la normale, il n'est pas associé avec une personnalité pessimiste. Il aurait peut-être fallu vraiment diminuer le nombre des rétroactions par rapport à la normale pour pouvoir voir la différence. Pour Spike (agressive), les participants s'attendaient à une fréquence importante des rétroactions alors que dans notre prédiction, elle était moyenne. Nous avons aussi vu que le type de rétroactions a un effet sur la perception de la fréquence des rétroactions par les participants. Ceux-ci évaluent mieux la fréquence des rétroactions quand le type de rétroactions varie aussi.

Concernant l'hypothèse H2, nos prédictions pour Prudence (fiable) et Spike (agressif) ne se sont pas révélées justes. Nous pensons que les adjectifs décrivant les personnalités étaient trop ambigus. Plusieurs commentaires des participants rapportaient qu'ils n'ont pas compris l'adjectif "pragmatique" (fiable) et

qu'ils ne savaient pas vraiment comment les personnes pessimistes réagissent. Cette incompréhension des termes peut expliquer les mauvaises évaluations sur les variations du type de rétroactions. En effet s'ils n'avaient pas une idée claire sur comment l'agent devait se comporter, ils ne pouvaient pas évaluer correctement les vidéos. Ces termes ont besoin d'être plus explicites dans les prochaines évaluations. Le lien entre la stabilité émotionnelle et le type de rétroactions (tendance à imiter) nécessite d'autres évaluations pour être validé.

Nous avons utilisé un modèle neutre de visage de l'ACA (Prudence) pour éviter de donner aux participants des informations sur la personnalité de l'agent. Cependant ce visage peut être un biais pour exprimer certaines personnalités comme extravertie.

7. Conclusion

Dans cet article, nous évaluons notre algorithme de sélection des rétroactions fonctionnant en temps réel. Il choisit le type et la fréquence des rétroactions qui sont jouées par l'agent conversationnel animé en fonction de sa personnalité. Un comportement est évalué comme approprié à une personnalité quand 1) la fréquence des rétroactions est associée avec la dimension extraversion et 2) le type de rétroactions est associé avec la dimension stabilité émotionnelle. Le lien entre la fréquence et l'extraversion est vérifié pour les personnalités "extravertie" et "fiable". Le lien entre le type et la stabilité émotionnelle est vérifié seulement pour la personnalité "pessimiste". Pour ce dernier lien, les termes décrivant les personnalités doivent être clarifiés pour que les participants puissent l'évaluer correctement. Par exemple, le terme pragmatique (fiable) a été mal compris par les participants biaisant ainsi les résultats. Nous pouvons avancer que la sélection de la fréquence et du type des rétroactions par notre algorithme contribue à l'interprétation des comportements de l'agent en terme de traits de personnalité.

Remerciements

Cette recherche a été supportée par le programme européen Community's Seventh Framework (FP7/2007-2013) grant agreement No. 211486 (SEMAINE) et le projet Web 2.0 MyPresentingAvatar.

References

1. E. Bevacqua, M. Mancini, C. Pelachaud: A Listening Agent Exhibiting Variable Behaviour. In Prendinger, Helmut, Lester, James, Ishizuka, Mitsuru, eds. : *Proceedings of 8th International Conference on Intelligent Virtual Agents*, Tokyo, Japan, Vol. 5208, pp.262-269, 2008
2. E. de Sevin and C. Pelachaud, Real-time Backchannel Selection for ECAs according to User's Level of Interest. In *Proceedings of Intelligent Virtual Agents 2009, IVA'09*. Amsterdam, Holland. 2009.
3. E. Douglas-Cowie, R. Cowie, C. Cox, N. Amir, D. Heylen: The Sensitive Artificial Listener: an induction technique for generating emotionally coloured conversation. In : *LREC2008*. 2008.
4. M. McRorie, I. Sneddon, E. de Sevin, E. Bevacqua and C. Pelachaud. A Model of Personality and Emotional Traits. In *Proceedings of Intelligent Virtual Agents 2009, IVA'09*, Amsterdam, Holland. 2009.
5. H. J. Eysenck: Dimensions of personality: Criteria for a taxonomic paradigm. *Personality and Individual Differences*, Vol. 12, 773-79. 1991.
6. P. T. Costa and R. R. McCrae: Four ways five factors are basic. *Personality and Individual Differences*, Vol. 13, 653-665. (1992)
7. J. Decety and P.L. Jackson: The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, Nr 3, /s.71-100. 2004.
8. T.L. Chartrand, W. Maddux and J. Lakin: Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In *The new unconscious*, edited by R. Hassin, J. Uleman, & J.A. Bargh, pp. 334-361, New York: Oxford University Press. 2005
9. M. Sonnby-Borgstrom: Automatic mimicry reactions as related to differences in emotional empathy. *Scandinavian Journal of Psychology*, Vol. 43, 433-443 (2002)
10. S.B. Eysenck and H.J. Eysenck: Impulsiveness and Venturesomeness - Their Position in a Dimensional System of Personality Description. *Psychol Rep*, Vol. 43(3):1247-1255 (1978)
11. P. Borkenau and A. Liebler: Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, Vol. 62, 645-657. 1992.
12. E. Bevacqua, E. de Sevin, C. Pelachaud, M. McRorie and I. Sneddon. Building Credible Agents: Behaviour Influenced by Personality and Emotional Traits. In *Proceedings of International Conference on Kansei Engineering and Emotion Research 2010, KEER 2010*, Paris, France. 2010.
13. K. Thórisson: Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills. *PhD Thesis*, MIT Media Laboratory. 1996.
14. R. Maatman, J. Gratch and S. Marsella: Natural Behavior of a Listening Agent. In : *5th International Conference on Interactive Virtual Agents*. 2005.
15. S. Kopp, T. Stocksmeier: Incremental multimodal feedback for conversational agents. In : *IVA2007*, pp.139-146. 2007.

Programmation Orientée Émotion (EOP)

K. Darty ^a

dartykevin@gmail.com

N. Sabouret ^a

Nicolas.Sabouret@lip6.fr

^aUniversité Pierre et Marie CURIE
4 place Jussieu, 75005 Paris, France

Résumé :

Certains problèmes algorithmiques actuels se heurtent à une multitude d'états possibles, souvent dus à un environnement dynamique. L'être humain a comme solution l'utilisation d'émotions qui augmente sa réactivité et l'aide dans ses décisions. Nous proposons une méthode de résolution de problèmes algorithmiques par une modélisation du système émotionnel humain visant à améliorer la réactivité et les performances de la résolution.

Mots-clés : Programmation Orientée Émotion, émotion, résolution de problème

Abstract:

Some current algorithmic problems are facing a multitude of states, often due to a dynamic environment. The human being's approach to such problems consists of using emotions which increase his reactivity and help him in his decisions. We put forward an algorithmic problem solving method using a modelling of the human emotional system aiming at improving the resolution's reactivity and performance.

Keywords: Emotion Oriented Programming, emotion, problem solving

1 Introduction

Du point de vue des psychologues-cogniticiens, les êtres humains sont constamment en train de résoudre des problèmes à partir des informations fournies par leur environnement (via leurs perceptions) [11]. Les recherches ont montré que ces informations à analyser sont bien trop volumineuses pour que les problèmes soient résolus uniquement de façon rationnelle. Les émotions ressenties par l'être humain et d'autres animaux sont apparues au cours de l'évolution par la sélection naturelle et ont pour but d'améliorer la réactivité et ainsi les décisions [1]. Par la suite, Scherer [14] définit formellement l'émotion comme réponse d'un stimulus intervenant à des niveaux organiques notamment cognitif et physique. Sa théorie permet après évaluation d'un événement d'en déduire la nature et la force de l'émotion.

Les émotions permettent d'une part des réponses comportementales plus rapides et mieux adaptées et d'autre part une meilleure distribution de l'attention sur les informations les plus

récentes ou les plus importantes dans l'environnement. Ainsi, des décisions rapides, incertaines sont prises et c'est ce mécanisme qui permet des changements de comportement adaptés (dans une majorité de cas).

Lorsqu'on s'intéresse à la résolution de problème en informatique dans des environnements réels, comme c'est le cas en robotique ou dans le domaine des *Agents Conversationnels Animés* en interaction avec un être humain et dans des environnements virtuels de plus en plus riches, on se retrouve confronté aux mêmes difficultés. En effet, parce qu'il peut être très dynamique et contenir beaucoup d'informations, l'environnement conduit à une explosion combinatoire de l'espace d'états, ce qui rend difficile l'utilisation des algorithmes classiques issus de la recherche en Intelligence Artificielle (IA).

Dans ce contexte, l'utilisation de mécanismes similaires aux émotions humaines semble constituer une heuristique intéressante pour la résolution de problèmes complexes. Dans cet article, nous présentons un modèle informatique, inspiré des émotions, pour définir des programmes informatiques capables de résoudre des problèmes d'IA difficiles dans des environnements dynamiques et incertains.

Dans la section 2, nous présentons les travaux dans le domaine de l'informatique affective et des agents conversationnels animés qui ont servi de support à notre étude. Nous définissons ensuite dans la section 3 notre modèle de *Programmation Orientée Émotion (EOP)*. Nous illustrons son fonctionnement sur un exemple dans la section 4 et nous discutons des perspectives de recherche dans la section 5.

2 Étude de la problématique

Pour pouvoir aborder la question de la construction d'un modèle d'EOP, il faut étudier d'un côté les modèles d'émotions développés en psychologie et de l'autre les techniques classiques de résolution de problème en IA.

Notre objectif est de définir un outil de résolution de problèmes qui mette en œuvre de manière automatique ou semi-automatique des heuristiques inspirées du mécanisme “émotionnel” des être vivants. C’est pourquoi nous devons étudier à la fois les modèles issus de l’informatique affective et les modèles de résolution de problème en IA.

2.1 Informatique affective

Pour Lazarus[7], lorsqu’un événement survient, le sujet fait des évaluations cognitives qui entraînent la création d’une émotion. Ces émotions sont à l’origine d’efforts servant à garder le contrôle sur la situation. Finalement, le “*coping*” fait converger l’individu vers une solution qui se traduit par des comportements ou bien des décisions réponses au problème non trivial initial.

En informatique, les modèles d’évaluation cognitive (*appraisal*) des émotions, inspirés des travaux de Roseman [12] et Scherer [14], sont fréquemment utilisés. On distingue deux approches pour représenter et manipuler l’état émotionnel. L’approche dimensionnelle [9] considère l’émotion comme un vecteur dans un espace à n dimensions, l’ensemble des émotions constituant le tempérament. Par exemple, le *framework* de Mehrabian considère que les axes *Pleasure*, *Arousal* et *Dominance* (*PAD*) forment une base suffisante pour décrire une émotion. L’approche catégorielle considère au contraire qu’il est possible d’énumérer l’ensemble des émotions. Le modèle d’*appraisal* d’Ortony, Clore et Collins (*OCC*) [10] se situe dans cette approche et définit chaque catégorie d’émotion comme une conséquence de multiples cognitions et interprétations de l’environnement. Plus récemment, la thèse de Gebhard [4] regroupe ces deux approches dans un modèle computationnel d’évaluation cognitive : la partie évaluation se fait grâce à *OCC* pour classifier en type d’émotion et la partie évolution des humeurs qui se fait via des états émotionnels *PAD* simulés continuellement.

Nous proposons ici de nous inspirer de ce modèle pour définir notre solveur de problème : les événements de l’environnement induisent des évaluations cognitives qui débouchent sur des émotions. Ainsi, les émotions “synthétisent” l’état de l’environnement et permettent de réduire l’ensemble des règles comportementales à décrire pour la résolution du problème.

De plus, les émotions interviennent directement dans la perception de l’environnement en modifiant le mécanisme d’attention [16]. Dans notre modèle, nous utiliserons les émotions et l’humeur pour sélectionner les perceptions dans l’environnement et réduire ainsi l’espace de recherche. Ainsi, nous proposons de transformer le principe classique de perception-action des agents [13] en influant sur l’équilibre entre ces deux dernières.

2.2 Résolution de problème en IA

Du point de vue de l’IA, un problème peut être considéré comme un ensemble d’objectifs et de données [8]. La résolution d’un problème consiste alors à passer d’un état initial vers un état solution tout en minimisant le temps de calcul et en respectant les contraintes de l’environnement (qui définissent les transitions possibles entre les états). Dès 1969, certains théoriciens pensent approcher un solveur de problème général par décomposition en sous-problèmes [2] mais la définition du problème reste une étape complexe et cette méthode se voit limitée aux problèmes bien définis.

Par la suite, les systèmes de production puis à base de règles [17] font leur apparition avec pour formalisme *Condition* $\Rightarrow$ *Action* et comme base le *modus ponens*. Ils sont capables de modéliser de nombreux phénomènes expérimentaux comme *SOAR* avec son modèle computationnel de l’espace-problème [6]. Leur succès provient de la modélisation simple des connaissances disponibles et des procédures existantes. Cependant, ces solveurs déterministes s’adaptent mal à un environnement dynamique : c’est toujours le même cheminement vers la solution qui sera suivi.

Les approches à base de planification, comme le modèle *STRIPS* [3], s’appuient sur une analyse des buts et des actions de transition pour construire un parcours de l’espace d’état avec des heuristiques. Cette opération de planification est confrontée à plusieurs difficultés : pour qu’elle aboutisse, il est nécessaire de connaître parfaitement l’ensemble des règles de transition qui régissent l’environnement, ce qui est impraticable dès que la perception est limitée ou que l’environnement est trop riche ou trop dynamique.

Plus récemment, l’apprentissage par renforcement [15] a montré qu’il était possible pour les

agents d'explorer leur environnement pour dégager, de manière non-supervisée, un chemin vers la solution. Mais une fois encore, ces algorithmes ne peuvent pas faire face à un trop grand nombre d'actions ou de perceptions, ni à un environnement dynamique.

C'est pourquoi nous proposons, tout en nous appuyant sur les mécanismes de perception-action qui sont ceux de l'IA et des approches de résolution de problèmes à base de règle, d'utiliser les émotions comme une heuristique pour le parcours de l'espace d'états. Nous ne prétendons pas ici que les émotions sont nécessairement une bonne approche mais nous pensons qu'il existe une classe de problèmes, caractérisée par le grand nombre de perceptions et d'actions possibles, ainsi que par des changements trop fréquents dans l'environnement, pour lesquels les émotions peuvent permettre aux algorithmes classiques de converger.

3 Présentation du modèle EOP

Nous présentons dans cet article un modèle de programmation orientée émotion (EOP) qui a pour objectif de décrire et de résoudre des problèmes en utilisant un mécanisme inspiré des émotions. Notre modèle est donc composé de deux parties : la description du problème (langage d'EOP) et le moteur de résolution.

Le modèle EOP n'est donc pas un solveur général de problème, comme STRIPS [3]. Il demande un travail au programmeur : celui de choisir comment son agent interprète émotionnellement les variables de l'environnement pour générer ses émotions et de déterminer la réaction (*coping* [7]) à ces émotions pour sélectionner les actions.

3.1 Architecture générale

Le fonctionnement général du modèle est résumé dans la figure 1. L'environnement (1) envoie des perceptions au moteur EOP (2). Les perceptions sont traduites en émotions par les *Threads de perception* (2b). Ces émotions sont utilisées en entrée par les *Threads de décision* (2b) pour sélectionner les actions qui modifient l'environnement. L'équilibre entre perception et décision est géré par le contrôleur (2a) en fonction de l'humeur (qui elle-même dépend des émotions générées).

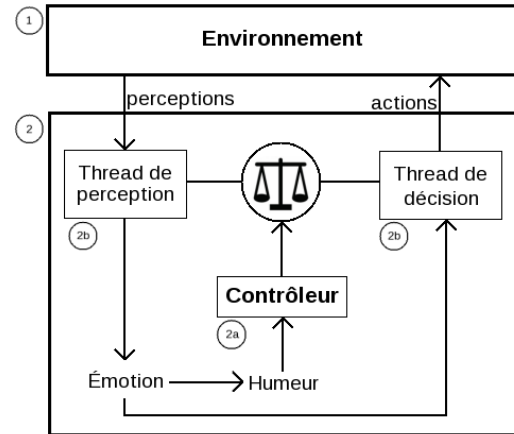


FIG. 1 – Blocs fonctionnels du modèle EOP

Modèle EOP. Le moteur EOP met en jeu des émotions et des humeurs. Elles sont représentées dans l'espace dimensionnel PAD de Mehrabian [9] dont la sémantique dans notre modèle EOP est la suivante :

- *Pleasure* représente le sentiment de proximité du but. Un fort plaisir équivaut à un état "proche" de l'état but (la mesure de proximité devant être définie par le programmeur).
- *Arousal* exprime le sentiment de nouveauté dans les perceptions. Un faible *arousal* correspond à un parcours dans lequel les états restent similaires.
- *Dominance* est une indication du sentiment de contrôle de l'agent sur la situation. Une forte dominance équivaut à une plus grande capacité d'action sur l'environnement.

Comme dans les travaux de Gebhard [4], les émotions peuvent être catégorisées (chaque catégorie d'émotion pouvant être calculée par une *PerceptionThread* différente). Le nombre et la nature de ces catégories n'est pas prédéfini (c'est au programmeur de les choisir).

L'humeur de l'agent-solveur est aussi définie en utilisant le modèle PAD et évolue en fonction des émotions déclenchées au cours de l'exécution (cf. section 3.4). De plus, le solveur sera paramétré par une humeur initiale $\overrightarrow{h_{init}}$ et des paramètres de sensibilité $\overrightarrow{h_{sens}}$ pour chacune des trois composantes.

Définition du problème. Le problème à résoudre est modélisé dans l'environnement : celui-ci contient l'ensemble des variables du problème et des actions par lesquelles les threads de décision peuvent agir dessus. De plus, l'environnement pouvant être dynamique, il dispose d'une

fonction de “mise à jour” qui définit son changement d’état en dehors de toute action externe.

Pour résoudre ce problème, le programmeur définit un ensemble de *threads* : les threads de perception (*PerceptionThreads*) et les threads de décision (*DecisionThreads*). Les premières servent de créateur d’émotions à partir des variables l’environnement. Les secondes utilisent les émotions pour déclencher une action de l’environnement.

Toutes les threads sont caractérisées par :

- Des conditions d’exécution qui définissent dans quels états la thread (perception ou action) peut être appliquée ;
- Une priorité calculée en fonction de l’humeur.

Fonctionnement par cycle. Le moteur de résolution fonctionne de manière cyclique. À chaque cycle, il choisit un thread parmi les thread disponibles (c’est-à-dire dont les conditions d’exécution sont réunies) en fonction de leur priorité, puis il l’exécute. Cela peut conduire à l’apparition d’émotions (*perceptionThread*) ou à des modifications de l’environnement (*decisionThread*). Le moteur exécute ensuite la fonction de mise à jour de l’environnement puis réévalue la disponibilité de toutes les threads en fonction du nouvel état. Il met à jour l’humeur en fonction de l’humeur actuelle et des émotions éventuellement produites et calcule la nouvelle valeur de priorité pour chaque thread.

Dans les sections suivantes, nous présentons formellement le modèle *EOP*. Nous présentons tout d’abord la description de l’environnement (section 3.2). Nous définissons ensuite (section 3.3) le modèle de programmation des threads de perceptions et de décisions. Nous décrivons enfin dans la section 3.4 le moteur de résolution qui contrôle l’exécution de notre solveur d’*EOP* : l’interprète *EOP*.

3.2 Description de l’environnement

L’environnement rassemble les données du problème ainsi que le problème en lui-même. Il contient les actions et les variables ainsi que des informations sur l’état du problème telles que sa solvabilité. L’environnement E est défini par un ensemble de variables V , un ensemble d’opérations O sur ces variables, un processus P_{env} et un ensemble F de fonctions d’évaluation du problème :

$$E = \langle V, O, P_{env}, F \rangle$$

Nous notons $V_e \subset V$ l’ensemble des variables externes, utilisables par le programmeur dans la description des threads. Pour chaque $v \in V$, nous notons D_v le domaine de valeurs de la variable v .

Chaque opération $o \in O$ est caractérisée par ses préconditions et ses effets :

$$o = \langle Prec_o, Eff_o \rangle$$

où $Prec_o$ est une fonction booléenne sur l’ensemble des variables : $Prec_o : V \rightarrow \mathbb{B}$. Eff_o est un ensemble ordonné de couples (var, val) , $var \in V$ et $val : V \rightarrow D_{var}$ est une fonction de l’ensemble des variables qui construit la nouvelle valeur de var .

Par exemple, si on souhaite faire un système *EOP* de conduite automobile assistée, on peut avoir une opération de bas niveau comme “changer de file”, “freiner”, *etc.* Les préconditions portent alors sur la présence d’autres véhicules, leurs vitesses, de panneaux de signalisation...

Les fonctions d’évaluation $f \in F$ du problème travaillent sur l’ensemble V et servent principalement au programmeur dans la définition des threads émotionnels, pour définir ses heuristiques.

Par exemple, dans le problème du labyrinthe, on peut souhaiter utiliser une fonction de distance qui donne la distance à vol d’oiseau au but (sans tenir compte des murs) :

$$d : (x_{agent}, y_{agent}, x_{but}, y_{but}) \rightarrow \mathbb{R}$$

Mais il existe quatre fonctions qui sont utilisées directement par le moteur de résolution pour évaluer l’état de la résolution du problème :

- la fonction *solved* : $V \rightarrow \mathbb{B}$ qui vaut *true* si et seulement si l’environnement est dans l’état but du problème.
- la fonction *solvable* : $V \rightarrow \mathbb{B}$ qui vaut *false* si et seulement si l’environnement est dans un état qui rend impossible la résolution du problème (par exemple, si l’explorateur a été mangé par un monstre).
- la fonction *distance* : $V_e^2 \rightarrow \mathbb{R}^+$ qui définit la dissimilarité entre deux états :

$$distance(v, v') = 0 \Leftrightarrow v = v'$$

distance(v, v') croît en fonction des différences de valeurs des variables.

- la fonction *goal* : $V_e \rightarrow \mathbb{R}^+$ qui définit la distance au but. Cette distance peut intégrer des

optimums locaux (par exemple, l'explorateur est à côté d'un trésor, d'un point de ravitaillement alors qu'il a soif, *etc*) et vaut 0 pour tout état but :

$$goal(v) = 0 \Leftrightarrow solved(v)$$

Ces fonctions doivent être définies lors de la description du problème dans l'environnement.

Processus. Un processus p est un ensemble d'opérations $O' \subset O$, munies de préconditions et d'effets, travaillant sur un sous-ensemble $V_p \in V$ et fonctionnant de manière cyclique.

Les préconditions des opérations doivent être mutuellement exclusives (au plus une seule action est éligible à chaque instant) ¹ :

$$\forall(o_i, o_j) \in p^2, Prec_{o_i} \wedge Prec_{o_j} = \perp$$

À chaque cycle, le processus exécute l'action éligible (s'il y en a une).²

L'environnement a un processus p_{env} mais on aura aussi des processus dans les threads de perception et de décision.

Exemple. Le problème classique du Wumpus [13] permet d'illustrer de façon concrète le fonctionnement de l'environnement : lorsqu'un monstre aperçoit le joueur dans un couloir, le processus permet de changer la position du monstre vers le joueur pour qu'il lui court après.

3.3 Programmation Orientée Émotion

Soit $\mathcal{T}$ l'ensemble des threads, décomposé en perceptions $\mathcal{P}$ et décisions $\mathcal{D}$. Pour chaque thread $t \in \mathcal{T}$, on note $p_t \in \mathbb{R}$ sa priorité et $c_t \in \mathbb{B}$ son éligibilité (déterminée par ses conditions d'exécution).

Les priorités des threads éligibles sont normalisées pour en faire des probabilités :

$$\sum_{\forall t \in \mathcal{T}. c_t = \top} norm(p_t) = 1$$

Puis un thread est sélectionné aléatoirement en respectant leur probabilité respective :

$$\forall t \in \mathcal{T}. c_t = \top, p(select = t) = norm(p_t)$$

¹Cela permet de ne pas avoir à choisir d'ordre dans l'exécution des actions, qui est un problème courant en systèmes multi-agents ou dans les algèbres de Gurevich [5].

²Nous verrons dans la suite de cet article comment les cycles sont contrôlés au niveau du moteur de résolution (section 3.4).

Comme dans le modèle de Gebhard [4], chaque émotion $e \in \mathcal{E}$ est caractérisée par un vecteur dans l'espace PAD :

$$\overrightarrow{PAD}^{(e)} = (P^{(e)}, A^{(e)}, D^{(e)}) \in [-1, 1]^3$$

Soit v l'état courant et v' l'état précédent :

– $P^{(e)} = f(goal(v) - goal(v'))$ est la valeur sur l'axe *pleasure*.

La fonction f est une exponentielle inverse définie de la manière suivante :

$$f(x) = \begin{cases} 1 - e^{-x} & \text{si } x \geq 0 \\ e^x - 1 & \text{sinon} \end{cases}$$

Ainsi, un plaisir proche de 1 correspond à un réel progrès en direction du but. Le plaisir diminue de manière exponentielle dès que cet écart décroît. Symétriquement, le déplaisir est d'autant plus grand que le but s'éloigne rapidement.

– $A^{(e)} = h(g(distance(v, v')))$ est la valeur sur l'axe *arousal*

La fonction $g : \mathbb{R}^+ \rightarrow [0, 1[$ est définie de la manière suivante : $g(x) = 1 - e^{-x}$. Cela permet d'accentuer aux bords l'impact de la distance sur l'*arousal*.

La fonction $h : [0, 1] \rightarrow [-1, 1]$ est, dans notre implémentation actuelle, définie par $h(x) = 2x - 1$ (elle transpose l'ensemble linéairement). Elle devrait en fait dépendre de l'écart-type entre deux états v et v' voisins dans l'espace d'état.³

– $D^{(e)} = h(\frac{card(t \in \mathcal{D}. c_t = \top)}{card(\mathcal{D})})$ est la valeur sur l'axe *dominance*. Elle utilise la même fonction h que dans la formule de $A^{(e)}$ pour transposer linéairement l'ensemble de valeurs. Elle devrait en fait dépendre du nombre d'actions valides (*i.e.* $c_t = \top$) en moyenne dans chaque état.

3.4 Émotions et Humeur

Les threads de perception permettent de générer des émotions. $\forall t \in \mathcal{P}$, nous notons $E(t) \subset \mathcal{E}$ l'ensemble des émotions produites par t .

Émotions. Chaque émotion $e \in \mathcal{E}$ est caractérisée non seulement par sa valeur dans le vecteur $\overrightarrow{PAD}^{(e)}$, comme énoncé précédemment, mais aussi par une date d'apparition $c_d^{(e)}$, une date de

³Nous pensons qu'en réalité, l'*arousal* devrait aussi dépendre du nombre de thread de perception actives par rapport au nombre de thread de perception possibles.

fin de vie $c_f^{(e)}$, définies par le programmeur, ainsi qu'un ensemble de variables sources $vs^{(e)} \subset V_e$ qui interviennent dans la construction de l'émotion (par exemple, une émotion de peur peut être provoquée par l'apparition d'un danger, la perte d'une ressource, etc...) :

$$e = \langle \overrightarrow{PAD}^{(e)}, c_d^{(e)}, c_f^{(e)}, vs^{(e)} \rangle$$

De plus, $\forall e \in \mathcal{E}$, nous notons $t_p(e) \in \mathcal{T}$ le thread créateur de l'émotion e .

Deux émotions e et e' sont considérées comme similaires quand leur valeurs dans l'espace PAD , leur ensemble de variables sources et le thread créateur sont identiques :

$$e \sim e' \Leftrightarrow \begin{cases} \overrightarrow{PAD}^{(e)} = \overrightarrow{PAD}^{(e')} \\ vs^{(e)} = vs^{(e')} \\ t_p(e) = t_p(e') \end{cases}$$

Au cycle τ , s'il existe une émotion e "vivante" ($\tau \in [c_d^{(e)}, c_f^{(e)}]$) et si un thread crée une émotion $e' \sim e$, alors e' est fusionnée à e pour créer une nouvelle émotion :

$$\langle \overrightarrow{PAD}^{(e)}, \min(c_d^{(e)}, c_d^{(e')}), \max(c_f^{(e)}, c_f^{(e')}), vs^{(e)} \rangle$$

Lorsque le cycle de fin de vie c_f est atteint, l'émotion meure : elle disparaît des données du moteur de résolution.⁴

Humeur. Le moteur de résolution maintient à jour une "humeur" de l'agent qui permet au programmeur de gérer les priorités des threads de perception et de décision.⁵ L'humeur $\vec{h}$ est caractérisée par un vecteur dans l'espace PAD initialisé à la valeur $\vec{h}_{init}$.

À chaque cycle τ , l'humeur est mise à jour en deux étapes. Premièrement, elle se rapproche des émotions vivantes :

$$\vec{h}_{new} = \vec{h} + \frac{1}{evol_h} \times (\vec{h}_{sens} \cdot \Sigma_E - \vec{h})$$

où $evol_h \in \mathbb{R}^+$ représente le pas de progression de l'humeur en direction des émotions vivantes

⁴En fait, l'émotion est seulement supprimée de la liste des émotions vivantes. Cependant les émotions les plus récentes sont conservées dans une file appelée "mémoire émotionnelle" afin que le programmeur puisse les utiliser dans la définition des threads de décision. Ainsi, la décision de l'agent peut dépendre d'émotions ressenties par le passé.

⁵À terme, notre objectif est que le programmeur ne définisse pas directement la priorité p_t des threads mais que celle-ci soit calculée automatiquement par le moteur à partir de fonctions caractéristiques.

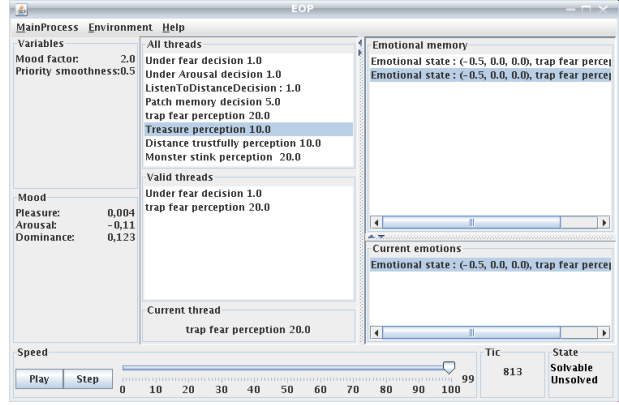


FIG. 2 – Moteur de résolution

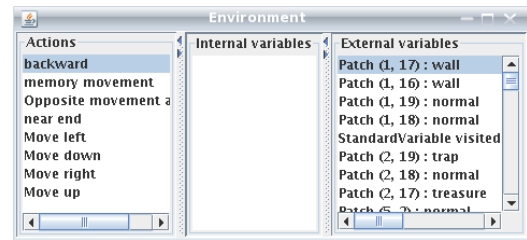


FIG. 3 – Environnement

et Σ_E est la somme des émotions vivantes :

$$\Sigma_E = \sum_{\forall e \in \mathcal{E}, \tau \in [c_d^{(e)}, c_f^{(e)}]} \overrightarrow{PAD}^{(e)}$$

Ensuite, elle se rapproche de $\vec{h}_{init}$ suivant un facteur de convergence $conv_h$:

$$\frac{(\vec{h}_{new} - \vec{h}_{init})}{conv_h} + \vec{h}_{init}$$

4 Implémentation et évaluation

La structure du modèle *EOP* a été implémentée en *Java 1.5*. Une interface simple, illustrée sur la figure 2, permet de visualiser l'humeur (à gauche), les threads disponibles, éligibles et leurs priorités (au centre) ainsi que les émotions déclenchées (à droite). Une barre de vitesse permet de contrôler l'exécution de moteur *EOP*. Enfin, il est possible de visualiser l'état de l'environnement dans une autre fenêtre (figure 3).

Exemple. Nous avons décidé de tester notre modèle *EOP* sur un problème de type *Wumpus* (labyrinthe avec des obstacles, des pièges et un ou plusieurs buts). La figure 4 montre l'*IHM* de visualisation de cet environnement.



FIG. 4 – Labyrinthe

Notre agent doit aller de la case de départ jusqu'à la case d'arrivée (drapeau) tout en récoltant le plus possible de trésors (en rouge sur l'IHM), en évitant des pièges statiques (en noir) et des monstres (Wumpus, représentés par un carré jaune sur l'IHM) qui le poursuivent dès qu'il s'approche.

Notre environnement contient donc trois ensembles de variables internes T (liste des positions des trésors), P (liste des positions des pièges) et W (liste des positions des Wumpus) ainsi que les variables externes M (liste des positions des murs), (x, y) (la position courante de l'agent), (x_{but}, y_{but}) (la position but), t le nombre de trésors ramassés, ainsi que des variables pour la perception : *piège visible* (pv), *monstre visible* (mv) et *trésor visible* (tv).

Le processus de l'environnement met à jour les variables de perception : lorsque l'agent se situe à moins d'une case d'un trésor ou d'un piège. Pour les monstres, la distance est de 3 cases. De plus, il déplace les monstres d'une case en direction de l'agent si celui-ci est situé sur la même ligne ou sur la même colonne, quelque soit la distance (sans obstacle).

Les opérations possibles sur l'environnement sont les déplacements de l'agent d'une case vers une case voisine accessible et l'action de ramasser un trésor. Chaque déplacement a pour précondition *solvable* et l'absence de mur dans la direction du déplacement. Ramasser un trésor a pour précondition la présence d'un trésor (c'est-à-dire que $(x, y) \in T$).

Les fonctions *solved*, *solvable*, *distance* et *goal* sont définies de la manière suivante :

- *solved* : $x = x_{but} \wedge y = y_{but}$ (l'agent est arrivé sur la case but) ;
- *solvable* : $(x, y) \notin P \cup W$ (l'agent n'est pas sur un piège ou sur un monstre) ;
- *distance* : $\delta \cdot \sqrt{(x - x')^2 + (y - y')^2} + \alpha \cdot |pv - pv'| + \beta \cdot |mv - mv'| + \gamma \cdot |tv - tv'|$ où α , β , γ et δ représentent le facteur d'im-

portance pour chaque critère (piège, monstre, trésor, distance à l'arrivée). Dans notre implémentation, nous avons pris $\alpha = \beta = \gamma = 2$ et $\delta = 1$.

- *goal*(v) = *distance*(v, v_{but}) avec $\alpha = \beta = 2$, $\gamma = 1$, $\delta = \max_{distance}^{-1}$ où $\max_{distance}$ est la longueur de la diagonale du labyrinthe.

Solution EOP. Pour tester la résolution de ce problème en utilisant une approche *EOP*, nous avons choisi des threads de perception et de décision simples : générer de la peur, fuir, s'approcher du but, etc.

Par exemple, nous avons une thread de perception t_{peur} qui génère une émotion lorsqu'on perçoit un piège ou un monstre. Cette perception produit une émotion de *pleasure* négatif, d'*arousal* positif et de *dominance* neutre. Sa priorité dépend de la composante *dominance* de l'humeur : moins l'agent se sent dominant, plus il va être attentif aux dangers potentiels. Les données mémorisées dans l'émotion sont les coordonnées de la menace, ce qui permet de les utiliser dans les threads de décision pour fuir.

Évaluation en cours. L'exemple décrit a été implanté dans une version intermédiaire de notre moteur *EOP*, qui n'intègre pas toutes les composantes du modèle décrit ici. Nous sommes actuellement en train de ré-implanter ce modèle pour pouvoir faire des comparaisons précises.

Nous avons implanté un algorithme de parcours de chemin en largeur (algorithme de Floyd) pour comparer cet algorithme classique avec notre modèle *EOP*. Nous avons mené une première expérience avec 100 labyrinthes de 10×10 cases générés aléatoirement avec deux Wumpus présents et les probabilités de piège, trésor et mur sur chaque case fixées à 5%, 5% et 40% respectivement. En utilisant le modèle *EOP*, l'agent atteint la case d'arrivée dans 27% des cas. Le nombre de trésors trouvés par parcours (même en cas d'échec) est en moyenne de 23% du nombre total de trésors.

De manière assez peu surprenante, notre modèle converge dans un nombre raisonnable de situations là où les algorithmes classiques sont voués à l'échec (Floyd ne peut pas gérer la dynamique des Wumpus dans l'environnement). Évidemment, il faudrait comparer à d'autres méthodes de recherches heuristiques (telles que A^*) et discuter l'expressivité de notre modèle pour pouvoir évaluer correctement l'approche

EOP. Nous n'en sommes qu'au début de nos recherches.

5 Conclusion et perspectives

Le modèle de Programmation Orientée Émotion présenté est encore en stade de développement. Les premières expérimentations montrent que cette méthode est une technique de résolution de problèmes qui donne des résultats intéressants : son fonctionnement lui permet de s'adapter en environnement dynamique et complexe. Une phase d'évaluation plus poussée (en cours) doit nous permettre de définir précisément la classe de problèmes qu'il peut traiter.

Pour généraliser notre modèle et ainsi le compléter, nous voudrions réduire la partie à réaliser par le programmeur au strict minimum. Nous pourrions, par exemple, changer la méthode de définition des priorités - pour l'instant laissée à la charge du programmeur - en demandant uniquement à quelles caractéristiques de l'émotion le thread doit réagir.

À terme, nous pensons que l'*EOP* peut définir une nouvelle approche pour la résolution de problèmes en environnement dynamique et en temps contraint, tirant parti à la fois des méthodes heuristiques en résolution de problèmes en *IA* et de la souplesse et de la réactivité qu'offrent les émotions chez les être vivants.

6 Remerciements

Merci à Cindy Mason du *MIT* pour nous avoir autorisé à reprendre l'idée originale de la programmation orientée émotions qu'elle avait initié. Merci à Guillaume Carré qui a travaillé sur la première version de notre moteur *EOP*.

Références

- [1] C. Darwin, P. Ekman, and P. Prodger. *The expression of the emotions in man and animals*. Oxford University Press, USA, 2002.
- [2] G. Ernst and A. Newell. *GPS : A case study in generality and problem solving*. Academic Pr, 1969.
- [3] R. Fikes and N. Nilsson. STRIPS : A new approach to the application of theorem proving to problem solving. *Artificial intelligence*, 2(3-4) :189–208, 1971.
- [4] P. Gebhard. Alma : A layered model of affect. *International Conference On Autonomous Agent*, pages 29–39, 1996.
- [5] Y. Gurevich. Sequential abstract-state machines capture sequential algorithms. *ACM Transactions on Computational Logic (TOCL)*, 1(1) :111, 2000.
- [6] J. Laird, A. Newell, P. Rosenbloom, C.-M. U. P. P. A. INTELLIGENCE, and P. PROJECT. Soar : An architecture for general intelligence. 1987.
- [7] R. Lazarus and S. Folkman. *Stress, appraisal, and coping*. Springer, 1996.
- [8] R. Mayer. Thinking, problem solving, cognition. 1983.
- [9] A. Mehrabian. Pleasure-arousal-dominance : A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4) :261–292, 1996.
- [10] A. Ortony, G. Clore, and A. Collins. *The cognitive structure of emotions*. Cambridge Univ Pr, 1990.
- [11] J.-F. Richard, S. Poitrenaud, and C. Tijus. Problem-Solving Restructuration : Elimination of Implicit Constraints. *Cognitive Science*, 17 :497–529, 1993.
- [12] I. Roseman. Appraisal determinants of emotions : Constructing a more accurate and comprehensive theory. *Cognition & Emotion*, 10(3) :241–278, 1996.
- [13] S. Russell and P. Norvig. *Artificial intelligence : a modern approach*. Prentice hall, 2009.
- [14] K. Scherer, H. Wallbott, and A. Summerfield. *Experiencing emotion : A cross-cultural study*. Cambridge University Press ; Paris : Maison des Sciences de l'Homme, 1986.
- [15] R. Sutton and A. Barto. *Reinforcement learning : An introduction*. The MIT press, 1998.
- [16] J. Taylor and N. Fragopanagos. Modelling the interaction of attention and emotion. In *2005 IEEE International Joint Conference on Neural Networks, 2005. IJCNN'05. Proceedings*, pages 1663–1668.
- [17] D. Waterman. Generalization learning techniques for automating the learning of heuristics. *Artificial Intelligence*, 1(1-2) :121–170, 1970.

Un modèle de comportements affectifs fondé sur la Conservation de Ressources

S. Campano N. Sabouret V. Corruble
sabrina.campano@lip6.fr nicolas.sabouret@lip6.fr vincent.corruble@lip6.fr

Laboratoire d'Informatique de Paris 6
4, place Jussieu
75005 Paris – FRANCE

Résumé :

Cet article définit une architecture visant à doter des agents virtuels autonomes de comportements affectifs proches de ceux des êtres humains, ou crédibles. Il est fondé sur un principe de conservation et d'acquisition de ressources, qui permet d'utiliser des principes génériques adaptables à un environnement virtuel peuplé par de nombreux agents.

Mots-clés : affect, émotion, agent virtuel, comportement

Abstract:

This paper aims at defining a model for autonomous virtual agents in order to provide them with affective behaviors close to those of human beings, i.e. *believable* behaviors. It is based on a principle of conservation and acquisition of resources, which allows the use of generic principles that are adaptable to virtual environments populated by numerous agents.

Keywords: affect, emotion, virtual agent, behavior

1 Introduction

Dans cet article nous nous intéressons à la simulation de comportements humains crédibles en milieu urbain. Notre travail s'inscrit dans le cadre du projet Terra Dynamica, qui vise à doter des agents de comportements autonomes et crédibles afin d'animer une ville virtuelle réaliste. Plus précisément notre objectif est de définir un modèle affectif du comportement humain, rendant compte de situations comme des disputes dans une file d'attente, ou des réactions de fuite face à un danger. Il doit pouvoir être appliqué à des situations urbaines, et être déployé sur un grand nombre d'agents. Afin de répondre à ce problème, nous adoptons une approche qui s'inspire des travaux du psychologue S.E. Hobföll sur la conservation des ressources [Hob89], et de ceux de la psychologue L. F. Barrett sur la nature des catégories émotionnelles [Bar06]. Nous proposons un modèle basé sur l'acquisition et la conservation des ressources qui ne modélise pas explicitement les émotions, mais repose sur des règles génériques permettant de produire des comportements affectifs. Le terme "ressource" doit être compris comme un concept général utilisé pour décrire des objets matériels, tout comme des objets

psychologiques, sociaux ou intellectuels. L'estime de soi, l'image auprès des autres, l'acquisition de connaissances sont des exemples de ressources. Nous espérons que le modèle développé pourra reproduire des comportements affectifs sans nécessiter la modélisation de catégories d'émotions, souvent associée à plusieurs problèmes difficiles comme le choix des émotions à prendre en compte, les conditions de déclenchement de chacune d'elles et leur mode d'action. Nous faisons donc l'hypothèse que les émotions sont des catégorisations de comportements ou de sensations, mais qu'elles ne sont pas nécessairement les composantes des processus à l'origine de ces comportements. Nous prenons donc le parti que l'utilisateur pourra interpréter lui même des émotions à partir de l'observation du comportement des agents produit par notre architecture, en leur accordant une crédibilité affective.

Nous présenterons dans la partie 2 une bibliographie sur les travaux théoriques ayant servi de support au modèle présenté, puis nous introduirons le principe du modèle dans la partie 3. La partie 4 expliquera comment est calculé le souhait pour une ressource et quel est son effet sur un agent, et la partie 5 portera sur les mécanismes de décision qui permettent la sélection du comportement.

2 Bibliographie

Lorsqu'on souhaite modéliser des comportements émotionnels chez un agent virtuel, une première approche possible consiste à identifier les catégories d'émotions que l'on veut prendre en compte, leurs conditions de déclenchement, ainsi que les comportements qu'elles induisent. Cependant une étude bibliographique des théories psychologiques sur les émotions aboutit rapidement à une difficulté majeure : il n'existe pas de consensus parmi les psychologues sur le nombre d'émotions existantes, ni sur leur fonction, ni sur leurs conséquences [OT90, Sch99]. James A. Russell identifie seulement deux types d'affects, la valence positive et né-

gative [Rus03], et attribue les différenciations entre d'autres types d'émotions à des différences contextuelles qui ne sont pas d'ordre émotionnel. Paul Ekman distingue six émotions en se fondant sur des mesures du système nerveux autonome, c'est à dire des changements physiologiques, qui permettraient de les distinguer [Ekm99]. Pourtant, dans une revue consacrée au sujet, Robert W. Levenson conclut clairement qu'il n'existe pas de preuves en faveur d'une distinction des émotions sur la base des mesures du système nerveux autonome [Lev92]. Silvan S. Tomkins quant à lui dresse une liste de neuf affects et trois valences en adoptant une approche fonctionnelle des émotions [Tom62, Tom63].

Quelles que soient les méthodes d'observation utilisées, mesures du système nerveux autonome, imagerie cérébrale, observation du comportement, il n'existe aucun ensemble de patrons qui puissent correspondre exactement à des catégories d'émotions [Bar06]. Seules deux affirmations semblent communes à chaque théorie : (i) les émotions répondent à un impératif d'adaptation et sont orientées vers un but (ii) il existe un état affectif de valence positive, et un état affectif de valence négative. À côté de ces deux grands principes, pourquoi existe-t-il autant de divergences dans la catégorisation des émotions ? Une réponse intéressante a été apportée par la psychologue L. F. Barrett [Bar06] : si l'absence de patrons émotionnels est plutôt la règle que l'exception, c'est peut-être parce qu'il n'y en pas. Selon elle les émotions sont des conceptualisations et non des entités de notre système affectif. Les humains font l'expérience des émotions de la même façon qu'ils font l'expérience d'une couleur ou du comportement de quelqu'un : ils utilisent leurs connaissances pour décomposer et conceptualiser l'information qui leur parvient, en adoptant une approche ascendante.

Si les émotions sont effectivement des catégorisations, elles ne sont donc pas suffisantes, ni peut-être même nécessaires, pour produire les comportements que nous qualifions comme émotionnels. Il faut donc se demander quels processus peuvent être à l'origine de ces comportements. La théorie du psychologue S.E. Hobföll sur la conservation des ressources constitue une piste de travail qui nous semble intéressante en ce sens [Hob89]. Pour lui, la protection et l'acquisition de ressources sont les seuls éléments nécessaires pour expliquer l'état de stress et de bien être d'un

individu. Le principe de ce modèle est que les personnes luttent pour retenir, protéger, construire leurs ressources, et que ce qui les menace est la perte effective ou potentielle de ces précieuses ressources. Lorsque les individus ne sont pas confrontés à des éléments stressants, ils cherchent à acquérir des surplus de ressources afin de compenser des pertes futures. Des exemples de ressources citées par Hobföll sont le contrôle, l'estime de soi, le statut socio-économique, l'emploi, les êtres aimés, les croyances. Elles sont donc de type diversifié et peuvent s'appliquer à l'ensemble des éléments de l'environnement dans lequel évolue un agent. Ce lien fonctionnel entre l'affect et les ressources se retrouve également dans la théorie de la psychologie sociale des groupes de J. W. Thibaut et H. H. Kelley [TK86]. Selon eux les dyades se forment lorsque deux individus s'apportent un ratio récompense/coût satisfaisant.

Suivant l'argumentation de J.A. Russell et L.F. Barrett, la diversité d'expression des émotions tient plus à la complexité de l'environnement qu'à leur nature elle-même. Les modèles que nous avons présentés permettent d'expliquer les phénomènes affectifs sous un angle fonctionnel et nous pensons qu'ils peuvent permettre d'effectuer des généralisations couvrant de nombreuses situations et rendant compte d'une grande partie des comportements affectifs. C'est l'hypothèse que nous souhaitons valider.

3 Principe du modèle proposé

Plutôt que de se fonder sur des catégories émotionnelles, nous proposons un modèle pour le comportement affectif d'agents autonomes basé sur la théorie de la conservation des ressources de Hobföll. Cette théorie psychologique, n'étant pas modélisée et n'ayant pas vocation à être implémentée, a nécessité un travail important d'adaptation, de complétion, et de modélisation.

3.1 Principe général

Le modèle présenté dans ce document repose sur le principe suivant : les agents souhaitent des ressources qu'ils vont tenter d'acquérir, et une fois ces ressources acquises ils cherchent à les protéger. Une ressource souhaitée active des actions d'acquisition, et une ressource menacée active des actions de protection. Les actions envisageables pour acquérir ou protéger

une ressource dépend du type de la ressource. Par exemple toute ressource r de type *énergie comestible* s'obtient en réalisant l'action de manger r . Un agent peut donc manger un gâteau pour gagner la ressource *énergie comestible*, mais aussi parler avec un autre agent pour gagner la ressource *relation sociale*, ou doubler dans une file d'attente pour gagner la ressource *meilleure position* dans la file. Chaque ressource souhaitée est associée à un degré de manque qui exprime la force avec laquelle l'agent souhaite la ressource. Ce degré peut évoluer au cours du temps de façon définie, ou bien selon les effets des actions de l'agent comme “*manger un gâteau*”, d'autres agents comme “*voler le gâteau*”, ou d'événements naturels comme “*un éclair foudroie le gâteau*”.

Nous avons porté notre choix sur l'utilisation de variables discrètes à valeur symbolique afin d'éviter les problèmes associés aux variables continues dans les modèles motivationnels et affectifs. L'utilisation de ces dernières se révèle souvent délicat lorsqu'il s'agit de paramétrer un modèle, de déterminer des seuils, d'effectuer des calculs ayant un sens sur le plan psychologique, et d'interpréter les significations des valeurs en sortie de calcul. Une approche discrète offre une gradation moins riche, mais une définition plus claire des paramètres et une meilleure prédiction et compréhension de leurs effets.

3.2 Architecture

Soit A l'ensemble des agents et R l'ensemble des ressources du monde Ω . $\forall i \in A$, l'architecture affective de i est composée de quatre piles de ressources actualisées au temps t :

- $RS_i(t)$, les ressources souhaitées par i ;
- $RA_i(t)$, les ressources acquises de i ;
- $RM_i(t)$, les ressources menacées de i ;
- $RP_i(t)$, les ressources que i a perdues.

Soit $DV \in \mathbb{N}$ un ensemble fini de valeurs. $\forall r \in RS_i(t)$, nous notons $\mu_i^r(t) \in DV$ le manque, ou degré de souhait, éprouvé par i pour une ressource r au temps t .

Les piles de ressources sont initialisées avant le début de la simulation. Elles peuvent être vides, remplies selon une configuration par défaut, ou selon un scénario spécifique. Une configuration par défaut pourrait fixer $\forall i \in A$, $RS_i(t) = \{Nourriture, Repos, Boisson\}$, avec $\mu_i^{Nourriture}(0) = \mu_i^{Repos}(0) = \mu_i^{Boisson}(0) = 2$,

où 2 correspond à un souhait modéré. Une configuration pour un scénario pourrait fixer pour $i \in A$, $RS_i(t) = RS_i(t) \cup \{TicketDeTrain\}$, avec $\mu_i^{TicketDeTrain}(0) = f(HeureDeDepart)$, où plus l'heure de départ du train est proche plus le degré de souhait pour le ticket de train est élevé.

Soit O l'ensemble des objets de l'environnement avec $A \subset O$. $\forall o \in O$, $R(o)$ est l'ensemble des ressources offertes par l'objet o , incluant o lui même car un objet peut-être considéré comme une ressource à lui seul, lorsque par exemple un agent s'y attache¹. $\forall r \in R(o)$ il existe un degré de compensation $C_i^r(o, t) \in [-DV; DV]$ qui indique le degré avec lequel la ressource r de l'objet o peut diminuer ou augmenter un manque de l'agent i pour r au temps t . Par exemple on peut fixer $\forall i \in A$, $C_i^{Nourriture}(hamburger, t) < C_i^{Nourriture}(carotte, t) < 0$, signifiant que la ressource *Nourriture* de l'objet *hamburger* diminue plus le manque de nourriture chez i que la ressource *Nourriture* de l'objet *carotte*.

En cours de simulation l'ajout, la suppression et la hiérarchisation des ressources dans les piles d'un agent i sont prises en charge par le Gestionnaire de Pile (voir figure 1). Ce module réorganise les piles à chaque cycle d'exécution en prenant en compte le degré de manque de i pour des ressources, les ressources disponibles dans l'environnement et leur degré de compensation pour i , les menaces sur les ressources de i , et l'acquisition ou la perte de ressources de i .

3.3 Réalisation d'une action

Soit $A_i(t)$ les actions pouvant être réalisées par l'agent i au temps t . Lorsqu'une action $a \in A_i(t)$ est réalisée elle a des effets sur les ressources d'un ou plusieurs agents qui sont les patients de l'action notés $patients(a) \subset A$. $\forall j \in patients(a)$, $E(a, j, t')$ est l'ensemble des ressources gagnées, perdues ou protégées au temps t' pour chaque agent j suite à la réalisation de l'action a par i . Une action peut avoir des effets immédiats ou des effets plusieurs pas de temps après sa réalisation. Une action ayant un effet immédiat recherché par l'agent peut être considérée comme une *action pulsionnelle*, et une action ayant un effet reporté peut-être considérée comme une *action contrôlée*.

¹Les mécanismes d'attachement ne sont pour le moment pas pris en compte dans le modèle.

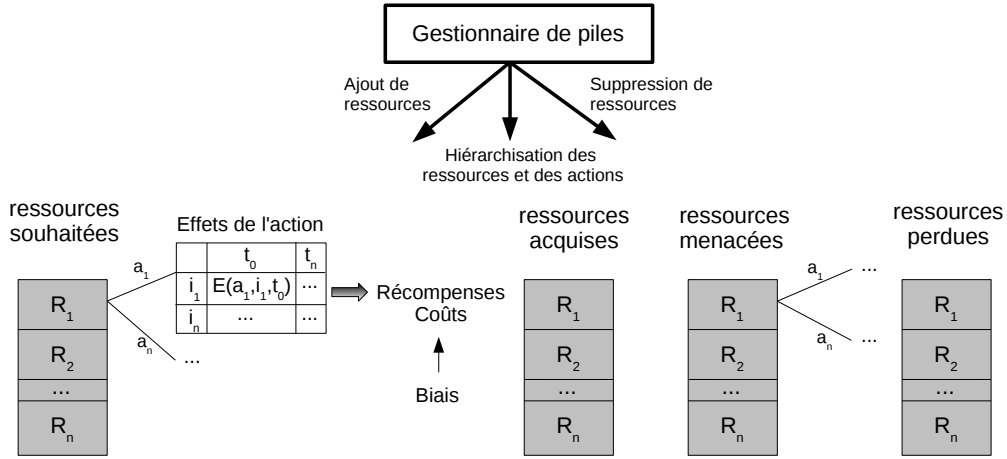


FIG. 1 – Architecture Générale.

On définit A_r^+ et A_r^- les ensembles d'actions ayant respectivement pour effet l'acquisition et la protection de r par l'agent qui réalise l'action. Si $r \in RS_i(t)$ alors $A_i(t) \leftarrow A_i(t) \cup A_r^+$, et si $r \in RM_i(t)$ alors $A_i(t) \leftarrow A_i(t) \cup A_r^-$. L'agent peut donc réaliser des actions d'acquisition vis à vis des ressources qu'il souhaite et des actions de protection vis à vis de ses ressources menacées.

Le gain et la perte d'une ressource $r \in R(o)$ par l'agent i au temps t sont associés à un revenu $r_i^r(t)$. Cette valeur est calculée par rapport au degré de compensation $C_i^r(o, t)$ et au degré de manque $\mu_i^r(t)$. Plus le degré de manque chez i pour r est élevé et plus la compensation de la ressource r pour i est importante, plus l'acquisition ou la perte de r provoquera un gain ou un coût élevé.

Un agent a des connaissances sur ses degrés de manque, et sur les degrés de compensation des ressources des objets du monde. Il peut donc anticiper les gains et les coûts. Ces connaissances peuvent être modifiées par l'optimisme, l'excitation, le bien-être, et l'égoïsme ou l'altruisme de l'agent. Nous développons ces paramètres dans la section 5.3.

4 Degrés de manque et de compensation

Soit un agent i , $\forall r \in RS_i(t)$, nous notons $\mu_i^r(t) \in DV$ le manque éprouvé par i pour une ressource r au temps t . Soit O l'ensemble des objets du monde Ω , $\forall o \in O, \forall r \in R(o)$ nous notons $C_i^r(o, t) \in [-DV; DV]$ le degré de compensation de la ressource r de l'objet o sur le

manque de i au temps t . $C_i^r(o, t)$ influence $\mu_i^r(t)$ lorsque o est acquis par i .

4.1 Calcul de la valeur du manque

La valeur d'un manque pour une ressource peut suivre des variations intrinsèques ou extrinsèques. Une variation intrinsèque est uniquement fonction d'intervalles de temps, c'est à dire que la valeur du manque change après un certain temps écoulé sans modification. Il s'agit typiquement de manque pour des ressources comme la nourriture, la boisson, le repos, dont la valeur croît progressivement jusqu'à une nouvelle compensation. Une variation extrinsèque a lieu lorsqu'une ressource correspondant à la variable est acquise ou perdue, et sa valeur est déterminée en fonction de la valeur de compensation de la ressource.

Variation intrinsèque. Soit RI l'ensemble des ressources dont la valeur de manque $\mu_i^r(t)$ est associée à une variation intrinsèque. On se donne une fonction $f_r(x, I(x))$ qui renvoie la nouvelle valeur de manque en fonction de la valeur courante x du manque et de l'intervalle de temps $I(x)$ en heures passé à cette valeur x . $\forall r \in RI, \forall i \in A, \mu_i^r(t+1) \leftarrow f_r(\mu_i^r(t), I(\mu_i^r(t)))$.

exemple : soit *Nourriture* $\in RI$, $f_N(x, I(x))$ la fonction de variation intrinsèque associée à cette ressource, et $\max(DV)$ la valeur maximum de l'ensemble $DV \in \mathbb{N}$.

$$f_N(x, I(x)) = \begin{cases} \max(DV), & \text{si } x = \max(DV) \\ x + 1, & \text{si } I > 4 \\ x, & \text{sinon} \end{cases}$$

Si $\max(DV) = 5$, $\mu_i^r(t) = 3$, et qu'il s'est écoulé 5 heures depuis la dernière modification du manque de l'agent i , alors $\mu_i^r(t+1) = 4$ et le compteur $I(\mu_i^r(t+1))$ est initialisé à 0. Le besoin de l'agent i pour une ressource *Nourriture* a donc augmenté en passant de 3 à 4 au bout de cinq heures. La fonction de variation proposée dans cet exemple est simple, mais il est possible de la rendre plus réaliste en prenant par exemple pour base un modèle biologique.

Variation extrinsèque. $\forall o \in O$, les objets du monde, chaque ressource $r \in R(o)$ a un attribut de compensation sur un agent i , noté $C_i^r(o, t) \in \{-DV; DV\}$. Une valeur négative diminue un manque, tandis qu'une valeur positive augmente un manque. Ce degré varie au cours du temps. Par exemple dans le cas d'un aliment qui se périme, on peut fixer un degré de compensation qui tend vers 0 pour la ressource *Nourriture*, et un degré de compensation qui tend vers $\max(DV)$ pour la ressource *Santé* de l'agent.

Lorsqu'un agent i acquiert un objet o , $\forall r \in R(o)$, $\theta = \mu_i^r(t) + C_i^r(o, t)$ et

$$\mu_i^r(t+1) = \begin{cases} \mu_i^r(t), & \text{si } |\theta| > \max(DV) \\ \theta, & \text{sinon} \end{cases}$$

exemple : soit un agent i , avec $\mu_i^{\text{Plaisir}}(t) = 2$ et $\mu_i^{\text{Nourriture}}(t) = 4$. On définit l'objet h représentant un hamburger avec $R(h) = \{\text{Plaisir}, \text{Nourriture}\}$, $C_i^{\text{Plaisir}}(h, t) = -2$ et $C_i^{\text{Nourriture}}(h, t) = -3$. L'acquisition de l'objet h par i produira les affectations $\mu_i^{\text{Plaisir}}(t) \leftarrow 0$ et $\mu_i^{\text{Nourriture}}(t) \leftarrow 1$. L'agent i n'éprouve plus de manque pour la ressource *Plaisir*, et son manque pour la ressource *Nourriture* a diminué.

Lorsqu'un agent i perd un objet o_1 , $\forall r \in R(o_1)$ s'il n'existe pas d'objet o_2 dans la pile de ressources acquises $RA_i(t)$ tel que $r \in R(o_2)$, alors $\theta = \mu_i^r(t) - C_i^r(o_1, t)$. C'est à dire que si l'agent n'a plus de ressource r le manque de l'agent augmente si la compensation de r par o_1 était < 0 , et qu'il diminue elle était > 0 . En revanche $\forall r \in R(o_1)$ s'il existe $o_2 \in RA_i(t)$ tel que $r \in R(o_2)$, alors $\theta = \min_{o_2 \in RA_i(t)} \mu_i^r(t) + C_i^r(o_2, t)$. Cela signifie que si l'agent a encore une ressource r c'est la ressource r de l'objet $o \in RA_i(t)$ offrant le degré de manque minimum à l'agent qui est prise en compte.

4.2 Calcul du coût et de la récompense apportée

Lorsque le manque d'un agent i pour une ressource augmente cela engendre un coût pour i , et lorsqu'un manque diminue cela engendre une récompense. Le calcul du revenu d'une ressource pour un agent i au temps t est donné par :

$$r_i^r(t) = (\mu_i^r(t) - \mu_i^r(t+1)) * \mu_i^r(t)$$

Si $r_i^r(t) < 0$ cela signifie que le manque a augmenté, et si $r_i^r(t) > 0$ cela signifie que le manque a diminué. Plus le manque est élevé au temps t plus le revenu est majoré. Nous faisons en effet l'hypothèse qu'un état de manque élevé grossit l'attention de l'agent sur les coûts et les récompenses qu'il subit.

Le revenu d'une ressource entre en compte dans la sélection de l'action. L'agent peut anticiper le revenu engendré par l'acquisition ou la perte d'une ressource r s'il connaît le degré de compensation de r . Notre modèle n'intégrant pas de processus d'apprentissage, tous les agents connaissent les valeurs de compensation des ressources des objets de l'environnement.

5 Gestion des piles et sélection de l'action

Le gestionnaire de pile contrôle l'entrée et la sortie des ressources dans chacune des piles, ainsi que leur hiérarchisation à l'intérieur des piles. Chaque ressource est associée à un ensemble d'actions en fonction de la pile à laquelle elle appartient. Les ressources en sommet de pile sont traitées en priorité par l'agent.

5.1 Ajout et suppression de ressources dans les piles

Il existe pour chaque pile de ressources d'un agent i des conditions qui déterminent l'ajout ou la suppression d'une ressource r dans la pile. Ces conditions sont évaluées à chaque pas de temps. Soit $patients(a)$ les patients d'une action a . $\forall i \in patients(a)$ on note $R^+(a, i, t)$ et $R^-(a, i, t)$ l'ensemble des ressources respectivement gagnées et perdues au temps t pour i après la réalisation de l'action a , $AC(t)$ l'ensemble des actions en cours de réalisation au temps t , et $AR(t)$ l'ensemble des actions dont

la réalisation s'est terminée au temps t . Une action interrompue n'est pas considérée comme une action réalisée.

Pile des ressources souhaitées $RS_i(t)$. Une ressource est souhaitée par un agent lorsque le degré de manque de l'agent pour la ressource est > 0 . Elle n'est plus souhaitée lorsqu'il est égal à 0.

- Ajout : $\forall r \in R$, si $(\mu_i^r(t) > 0 \text{ et } r \notin RS_i(t))$, alors $RS_i(t) \leftarrow RS_i(t) \cup r$
- Suppression : $\forall r \in RS_i(t)$, si $\mu_i^r(t) = 0$ alors $RS_i(t) \leftarrow RS_i(t) - r$

Pile des ressources acquises $RA_i(t)$. Une ressource est acquise par un agent i lorsqu'une action ayant pour effet de faire gagner la ressource à l'agent a été réalisée. Elle n'est plus acquise lorsqu'une action ayant pour effet de faire perdre la ressource à l'agent a été réalisée.

- Ajout et suppression : $\forall a \in AR(t)$, $\forall i \in patients(a)$, $RA_i(t) \leftarrow RA_i(t) \cup R^+(a, i, t) - R^-(a, i, t)$

Pile des ressources menacées $R_{menacees_i}$. Une ressource d'un agent i est menacée s'il y a une action en cours ayant pour effet de faire perdre la ressource à i . Elle n'est plus menacée lorsqu'elle est effectivement perdue, ou lorsqu'il y n'y a plus d'action en cours menaçant cette ressource.

- Ajout : $\forall a \in AC(t)$, $\forall i \in patients(a)$, $RM_i(t) \leftarrow RM_i(t) \cup R^-(a, i, t)$
- Suppression : (i) $\forall a \in AR(t)$, $\forall r \in R^-(a, i, t)$, si $r \in RM_i(t)$ alors $RM_i(t) \leftarrow RM_i(t) - r$ (ii) $\forall r \in RM_i(t)$, si $r \notin \bigcup_{a \in AC(t)} R^-(a, i, t)$ alors $RM_i(t) \leftarrow RM_i(t) - r$.

Pile des ressources perdues $R_{perdues_i}$. Si une action réalisée a pour effet de faire perdre une ressource r à un agent i alors r est ajoutée à $RP_i(t)$. Si une action réalisée a pour effet de faire gagner une ressource r à i et que $r \in RP_i(t)$, alors r est supprimée de $RP_i(t)$.

- Ajout : $\forall a \in AR(t)$, $\forall i \in patients(a)$, $RP_i(t) \leftarrow RP_i(t) \cup R^-(a, i, t)$
- Suppression : $\forall a \in AR(t)$, $\forall i \in patients(a)$, $RP_i(t) \leftarrow RP_i(t) - R^+(a, i, t)$

$\forall r \in RP_i(t)$, $p(r)$ est la durée de persistance pendant laquelle l'agent i garde en mémoire la

perte de cette ressource. Si la durée de la persistance est dépassée, r est supprimée de la pile. Si la perte d'une ressource r a provoqué un coût maximum pour i elle n'est jamais retirée de la pile des ressources perdues, c'est à dire que $p(r) = +\infty$, à moins qu'elle ne soit acquise à nouveau. Les ressources contenues par $RP_i(t)$ ont une influence sur le bien-être de l'agent (voir 5.3).

5.2 Activation des actions

Chaque ressource r est associée à un ensemble d'actions réalisables pour l'obtenir, et un ensemble d'actions réalisables pour la protéger. Soit $A_i(t)$ les actions activées de i , c'est à dire les actions pouvant être réalisées par l'agent i au temps t . On définit A_r^+ et A_r^- les ensembles d'actions ayant respectivement pour effet l'acquisition et la protection de r par l'agent qui réalise l'action.

$$A_i(t) = \bigcup_{r \in RS_i(t)} A_r^+ \cup \bigcup_{r \in RM_i(t)} A_r^-.$$

L'agent peut donc réaliser des actions d'acquisition vis à vis des ressources qu'il souhaite et des actions de protection vis à vis de ses ressources menacées. Les actions contenues dans $A_i(t)$ sont concurrentes. On peut désactiver certaines actions chez un agent pour le priver de la possibilité de réaliser l'action, par exemple s'enfuir du restaurant sans payer. Une action désactivée ne fait jamais partie de $A_i(t)$.

5.3 Biais appliqués aux revenus d'une ressource

Par défaut, tout agent i connaît les effets d'une action $a \in A_i(t)$ pour tout $j \in patients(a)$, c'est à dire les ressources gagnées $R^+(a, j, t)$, les ressources perdues $R^-(a, j, t)$ et les ressources protégées $R^\sim(a, j, t)$ de l'agent j (avec $i \subset j$). Il connaît donc aussi $\forall j \in patients(a)$, les revenus de j consécutifs à chaque action $a \in A_i(t)$. Soit $Revenus(a, j)$ l'ensemble des revenus anticipés par i qui seraient obtenus par j suite à la réalisation de l'action $a \in A_i(t)$. Certains paramètres affectifs de l'agent i peuvent supprimer ou amplifier des revenus de coût ou de récompense dans $Revenus(a, j)$. Si un revenu est supprimé i ne pourra pas en tenir compte dans la sélection de son action.

Optimiste/Pessimiste. Soit un agent i , on note : $optimisme_i(t) =$

$\{\text{optimiste}, \text{neutre}, \text{pessimiste}\}$ son état d'optimisme, fixe ou variable aléatoirement au cours du temps. Si $\text{optimisme}_i(t) = \text{optimiste}$: $\forall r \in \text{Revenus}(a, j)$, si $r < 0$ alors $\text{Revenus}(a, j) \leftarrow \text{Revenus}(a, j) - r$, c'est à dire que tout revenu de coût est supprimé. Si $\text{optimisme}_i(t) = \text{pessimiste}$: $\forall r \in \text{Revenus}(a, j)$, si $r > 0$ alors $\text{Revenus}(a, j) \leftarrow \text{Revenus}(a, j) - r$, c'est à dire que tout revenu de récompense est supprimé.

Bien-être. L'état de bien-être d'un agent i noté $BE_i(t) \in \{\text{mauvais}, \text{neutre}, \text{bon}\}$, avec $\text{mauvais} < \text{neutre} < \text{bon}$, influence sa sensibilité à la récompense. Si l'état de bien-être d'un agent est mauvais, les récompenses qu'il peut obtenir sont amplifiées. Un être humain dans un état de dépression peut en effet chercher à améliorer son humeur en acquérant des ressources qui vont lui procurer un signal positif [FC96]. La variable de bien-être d'un agent est déterminée en fonction de trois autres valeurs de bien être dépendantes des piles de ressources de l'agent :

- $BE_{RS_i(t)}$: $\forall r \in RS_i(t)$ si $\exists \mu_i^r(t) = \max(DV)$, c'est à dire si un état de manque de l'agent est à sa valeur maximum, alors $BE_{RS_i(t)} \leftarrow \text{mauvais}$. Soit un seuil $s \in DV \setminus \max(DV)$, si $\forall r \in RS_i(t)$, $\mu_i^r(t) \leq s$, aucun degré de manque ne dépasse le seuil s , alors $BE_{RS_i(t)} \leftarrow \text{bon}$. Dans les autres cas, $BE_{RS_i(t)} \leftarrow \text{neutre}$.
- $BE_{RM_i(t)}$: si $RM_i(t) \neq \{\emptyset\}$, alors $BE_{RM_i(t)} \leftarrow \text{mauvais}$, sinon $BE_{RM_i(t)} \leftarrow \text{bon}$.
- $BE_{RP_i(t)}$: si une ressource $r \in RP_i(t)$ a engendré un coût maximum pour i lors de sa perte alors $BE_{RP_i(t)} \leftarrow \text{mauvais}$. Si $RP_i(t) = \{\emptyset\}$, alors $BE_{RP_i(t)} \leftarrow \text{bon}$. Dans les autres cas, $BE_{RP_i(t)} \leftarrow \text{neutre}$

Si $\text{optimisme}_i = \text{optimiste}$ alors $BE_i(t) \leftarrow \max\{BE_{RS_i(t)}, BE_{RM_i(t)}, BE_{RP_i(t)}\}$.
 sinon $BE_i(t) \leftarrow \min\{BE_{RS_i(t)}, BE_{RM_i(t)}, BE_{RP_i(t)}\}$.
 Si $BE_i(t) = \text{mauvais}$, toutes les récompenses que i peut obtenir sont majorées par un coefficient $\alpha_{BE} > 1$. $\forall r \in \text{Revenus}(a, j)$, si $r > 0$ alors $r \leftarrow r * \alpha_{BE}$.

Excitation. L'excitation d'un agent i au temps t est notée $\text{excitation}_i(t) \in \{\text{active}, \text{inactive}\}$. Soit h un horizon de temps, et $d(r)$ la date à la-

quelle un revenu r est obtenu. Si $d(r) < h$, r est considéré comme un revenu un court terme, et si $d(r) \geq h$, r est considéré comme un revenu à long terme. Une excitation active supprime les revenus à long terme. Si $\text{excitation}_i(t) = \text{active}$, alors $\forall r \in \text{Revenus}(a, j)$, si $d(r) \geq h$ alors $\text{Revenus}(a, j) \leftarrow \text{Revenus}(a, j) - r$.

$\text{excitation}_i(t)$ devient *active* lorsque l'agent anticipe ou reçoit un revenu maximum. Après un certain temps déterminé par une constante c , $\text{excitation}_i(t)$ devient à nouveau inactive. Cela signifie qu'après avoir été activée, l'excitation peut encore avoir un effet sur le choix des actions de l'agent pendant un certain temps même si ces actions n'ont pas pour conséquence d'activer l'excitation.

Egoïsme/Altruisme. Les revenus de l'agent i sont associés à un coefficient α_{ego^i} , et les revenus des autres agents à un coefficient α_{alt^i} . Ce coefficient représente l'importance que i accorde à ses propres revenus et aux revenus des autres agents patients de son action. Ces coefficients sont utilisés dans la sélection de l'action. Par exemple si $\alpha_{alt^i} = 0$, i n'accorde aucune importance aux revenus des autres agents.

5.4 Sélection de l'action

Soit un agent i , $A_i(t)$ son ensemble d'actions activées, et $p(a)$ les patients d'une action a . $\forall a \in A_i(t)$, $\forall j \in p(a)$, $E(a, j, t')$ contient les effets de l'action, c'est à dire les ressources acquises, perdues et protégées de j au temps t' suite à la réalisation de l'action a par i . $\forall e \in E(a, j, t')$, on note $r(e, j, t')$ le revenu consécutif à l'effet. Nous calculons $\forall a \in A_i(t)$ l'utilité de l'action a dont la réalisation peut commencer au temps $t + 1$ que nous notons $u(a, i, t + 1)$:

$$u(a, i, t + 1) = \left(\alpha_{alt^i} * \sum_{j \in p(a) \setminus i} \sum_{e \in E(a, j, t)} r(e, j, t) \right) + \left(\alpha_{ego^i} * \sum_{e \in E(a, i, t)} r(e, i, t) \right)$$

L'utilité d'une ressource $r \in \{RS_i(t) \cup RM_i(t)\}$ est égale à l'utilité maximum de son ensemble d'actions A_r . $A_r = A_r^+$ si $r \in RS_i(t)$, et $A_r = A_r^-$ si $r \in RM_i(t)$. Nous définissons l'utilité de la ressource r :

$$u(r) = \max_{a \in A_r} \{u(a, i, t + 1)\}.$$

Les ressources sont classées dans les piles par utilité décroissante, avec l'utilité maximum en tête de pile. Lorsqu'un agent i parcourt une pile, l'action réalisée par i au temps $t + 1$ est la première action a qui vérifie $u(a, i, t + 1) > 0$ et $u(a, i, t + 1) = \text{Max}_{a' \in A_r} \{u(a', i, t + 1)\}$. Si pour une ressource r il existe plusieurs actions $\in A_r$ d'utilité maximum, alors une action d'utilité maximum est choisie aléatoirement parmi A_r . Une utilité < 0 étant un coût, s'il n'existe pas d'utilité positive l'agent ne réalise pas d'action. Le parcours de la pile $RM_i(t)$ est prioritaire par rapport à la pile $RS_i(t)$, un agent cherchant toujours à protéger ses ressources avant d'en acquérir d'autres.

6 Conclusion et perspectives

Nous avons présenté dans cet article une architecture dont l'objectif est de permettre la génération de comportements affectifs chez des agents virtuels dans un environnement réaliste. Nous espérons que ce modèle permettra à un observateur humain de catégoriser les comportements des agents en émotions, alors même que ces catégories ne sont pas des composants du modèle. Si notre hypothèse se vérifie, nous disposerons d'un modèle généralisable et facilement adaptable à un grand nombre d'environnements peuplé par des agents virtuels. Les utilités des actions ont une valeur fonctionnelle qui dépend des propriétés des ressources et des souhaits de l'agent, elles ne sont pas fixées arbitrairement à la main. Secondairement, une validation de ce modèle permettrait d'apporter un soutien en faveur des théories psychologiques de L.F. Barrett et S.E. Hobföll.

Cet article est issu d'un travail préliminaire en cours d'implémentation dont l'architecture de base sera complétée ou modifiée après la première procédure d'évaluation. Nous comptons introduire ultérieurement la notion de mémoire d'un agent et utiliser une méthode d'apprentissage par renforcement qui déterminera son attraction ou sa répulsion pour certaines ressources. La persévérance d'un agent face à l'échec dans l'acquisition d'une ressource, et son niveau de frustration face à un blocage seront également intégrés. Certaines procédures pour le moment simplifiées nécessiteront sûrement des améliorations comme la visibilité des paramètres de décision, et le calcul des coûts et des récompenses qui devraient prendre un compte la notion d'effort. Nous introduirons également des dimensions sociales comme l'at-

tachement, et nous étendrons notre échelle au niveau du groupe et de la foule. Enfin une typologie rigoureuse des ressources et des types de comportements qui leur sont associés sera spécifiée.

Références

- [Bar06] L.F. Barrett. Solving the emotion paradox : Categorization and the experience of emotion. *Personality and social psychology review*, 10(1) :20, 2006.
- [Ek99] P. Ekman. Basic emotions. *Handbook of cognition and emotion*, pages 45–60, 1999.
- [FC96] R.J. Faber and G.A. Christenson. In the mood to buy : Differences in the mood states experienced by compulsive buyers and other consumers. *Psychology and Marketing*, 13(8) :803–819, 1996.
- [Hob89] S.E. Hobföll. Conservation of resources. *American Psychologist*, 44(3) :513–524, 1989.
- [Lev92] R.W. Levenson. Autonomic nervous system differences among emotions. *Psychological Science*, pages 23–27, 1992.
- [OT90] A. Ortony and T.J. Turner. What's basic about basic emotions. *Psychological review*, 97(3) :315–331, 1990.
- [Rus03] J.A. Russell. Core affect and the psychological construction of emotion. *Psychological review*, 110(1) :145–172, 2003.
- [Sch99] K.R. Scherer. Appraisal theory. *Handbook of cognition and emotion*, pages 637–663, 1999.
- [TK86] J.W. Thibaut and H.H. Kelley. *The social psychology of groups*. Transaction Publishers, 1986.
- [Tom62] S. Tomkins. *Affect Imagery Consciousness-Volume I the Positive Affects*. Springer Pub Co, 1962.
- [Tom63] S. Tomkins. *Affect imagery consciousness : The negative affects*. Springer Pub Co, 1963.

Actes de conversation multimodaux et émotions

J. Rivière

S. Pesty

Jeremy.Riviere@imag.fr

Sylvie.Pesty@imag.fr

Laboratoire d'Informatique de Grenoble (LIG)

Maison Jean Kuntzmann - 110 av. de la Chimie - Domaine Universitaire de Saint-Martin-d'Hères
38041 Grenoble cedex 9 – FRANCE

Résumé :

Cet article se place dans le cadre de dialogues entre un humain et un Agent Conversationnel Animé (ACA). Nous discutons ici de la nécessité de mettre en place un langage spécifique aux interactions entre l'homme et l'ACA, et du rôle central que jouent les émotions au sein de ces interactions. Dans ce but, nous proposons une bibliothèque d'*actes de conversations multimodaux* basée sur les actes de discours et l'expression des émotions liées aux actes. Nous nous intéressons plus particulièrement aux émotions dites *complexes*, résultant d'un raisonnement de l'agent. Ce langage de conversation multimodal sera le moyen pour l'ACA d'exprimer son intention communicative de façon multimodale : verbale et non verbale.

Mots-clés : Interaction, Dialogue, Actes de discours, Agents Conversationnels Animés, Emotions

Abstract:

This article falls within the realm of dialog between human and Embodied Conversational Agent (ECA). We discuss here about the need to have a language which is peculiar to human-ECA interaction and based on the main role of emotion in human communication. In order to define this language, a multimodal conversation acts library is proposed, based notably on speech acts and the expression of emotions linked to these acts. This study focuses in particular on complex emotions, arising from agents' reasoning. This multimodal conversation language will be used in future to express the communicative intention of an Embodied Conversational Agent able to communicate through all modalities, verbal and non-verbal.

Keywords: Interaction, Dialog, Speech act, Embodied Conversational Agent, Emotion

1 Introduction

De nos jours, l'augmentation des interactions entre l'humain et la machine et au sein même de la machine entre agents autonomes, fait du dialogue un verrou scientifique majeur. De ce fait, certaines recherches dans le domaine des systèmes de dialogue humain-agent (et plus généralement de dialogue homme-machine) se sont orientées vers l'annotation de dialogue ([8], [19]) pour la compréhension du discours, la reconnaissance de structures et la génération de dialogue, tandis que d'autres recherches s'orientent vers la communication en langage naturel. Avec l'arrivée des agents virtuels et dans le contexte du dialogue humain-agent virtuel, où les aspects expressifs et multimodaux

sont premiers, il s'agit maintenant d'augmenter les capacités conversationnelles des agents en cherchant à étroitement relier les aspects verbaux aux aspects non-verbaux (expressions faciales, gestes, regard, ...). Notre problématique se concentre sur le rôle central de l'émotion dans la communication et plus généralement dans les dialogues, ce rôle ayant été montré dans de nombreux travaux tant en psychologie qu'en informatique ([3], [20], [21]). Pour améliorer ces interactions tant en termes de crédibilité que d'efficacité, c'est-à-dire de succès de la communication, et rendre les agents virtuels plus crédibles et convaincants pour l'humain, de récents travaux ont doté les agents de capacités d'exprimer les émotions, pour en faire des Agents Conversationnels Animés (ACA ou ECA- Embodied Conversational Agent en anglais) ([22], [9]).

L'objectif de nos travaux est la mise en place d'un langage qui dépasse le simple échange d'information (Inform - Request) dans les dialogues entre agents virtuels et humains, en incluant la dimension émotionnelle. La proposition ici présentée consiste à concevoir pour les agents virtuels une bibliothèque d'*actes de conversation multimodaux*, basés d'une part sur les actes de conversation d'Alexandra Berger ([5]), dans la lignée de la théorie des actes de discours ([26]), et basés d'autre part sur les émotions dans l'interaction et la multimodalité de leur expression. Cette bibliothèque doit fournir à l'agent virtuel les émotions liées à un acte, et plus particulièrement les émotions que nous appellerons *complexes*, définie notamment sur le raisonnement contrefactuel détaillé en 4.1. Ainsi, lorsque l'agent choisit un acte de conversation, il formulera non seulement un énoncé mais exprimera également, de manière multimodale, l'ensemble des émotions intimement liées à cet acte. Ce *langage de conversation multimodal* s'intégrera par la suite dans l'ACA Greta, développé au LTCI¹ à Télécom-ParisTech par Catherine Pelachaud ([22]), et dont nous détaillons l'architecture en 5.1.

¹Laboratoire Traitement et Communication de l'Information

La première partie de cet article présente les principaux moyens d'expression des émotions utilisés par les ACA. Nous verrons que, dans la majorité des travaux, la modalité langagière ne fait pas le lien entre émotions et dialogue : les émotions de l'ACA sont principalement exprimées par la prosodie, mais sont indépendantes de l'énoncé. Dans le but de faire le lien entre dialogue et émotions, nous nous intéressons dans la deuxième partie à deux des principales méthodes d'annotation de dialogue et d'émotions entre humain et agent virtuel. En effet, ces travaux nous renseignent sur les émotions d'une part et sur les actes d'autre part qui interviennent dans les interactions ACA-humains. Dans la troisième partie, nous rappelons succinctement les actes de conversation entre agents développés par A. Berger sur lesquels nous basons notre proposition, puis nous détaillons nos *actes de conversation multimodaux* en montrant le lien entre les actes de langage et les *émotions complexes* dont nous donnons la définition. Enfin, nous montrons une première implémentation de plusieurs actes de conversations multimodaux avec l'ACA Greta et les principales perspectives de travail.

2 Expression d'émotions chez les ACA

La plupart des travaux sur l'expression des émotions par un ACA se focalisent sur les modalités faciale et gestuelle ([9], [4]). L'ACA Greta ([22]) que nous utilisons en est un bon exemple. Greta se base sur le système des FACs² de Ekman et Friesen ([10]) pour définir les expressions faciales, et permettre de contrôler l'animation indépendamment de la géométrie du visage. Les émotions ainsi exprimées sont les émotions "de base" de Ekman ([11]) ou des émotions dérivées, facilement reconnaissables et universelles. Le regard et les mouvements de la tête sont également pris en compte. Les gestes de Greta, et notamment ceux des bras et des mains, sont synchronisés à la parole et définis en quatre phases : la préparation, l'apogée (le *stroke*), la suspension puis la rétraction du mouvement. Le *stroke* doit spécifiquement correspondre à l'intensité maximale de l'intonation. Une des difficultés à ce niveau consiste à lier différents mouvements entre eux d'une manière fluide et naturelle.

On peut distinguer deux niveaux dans la modalité langagière qui nous intéresse : l'énoncé en

lui-même et la prosodie, qui est l'intonation de la voix lors de l'énoncé suivant une ou plusieurs émotions. La majorité des travaux sur les ACA se concentrent sur la prosodie, en ajustant l'intensité de la voix suivant l'énoncé d'une part et les émotions d'autre part. Les émotions et l'énoncé sont ainsi indépendants l'un de l'autre, ce qui ne garantit pas la cohérence d'un énoncé par rapport à une émotion : par exemple, on pourrait avoir un agent qui affirme être très heureux tout en exprimant une émotion de grande tristesse.

3 Annotation d'émotion et annotation de dialogue

Cette partie présente deux travaux majeurs dans les domaines d'annotation d'émotions et d'annotation de dialogue à partir d'actes de discours. Ces deux approches nous éclairent respectivement sur le rôle des émotions et celui des actes dans les interactions, ce qui met en évidence l'absence de lien entre actes et émotions.

Le groupe de travail "Emotion Markup Language" (EmotionML [2]) du W3C³ propose un langage adapté à l'annotation d'émotions, la reconnaissance automatique d'émotions au travers du comportement de l'utilisateur, et la génération d'émotions. Ce langage est basé sur des balises xml permettant de classer les émotions suivant les théories de l'évaluation ([23]), les théories dimensionnelles ([17]) ou encore la théorie de Frijda, qui définit une émotion suivant une tendance à l'action ([12]). Ainsi, l'émotion de peur peut se traduire par exemple :

- En théorie de l'évaluation :
`"<appraisal name="novelty" value="0.9"/>
 <appraisal name="intrinsic-pleasantness"
 value="0.1"/>"`
- Dans la théorie de Frijda :
`"<action-tendency name="avoidance" value="0.8"/>"`

De plus, à ces émotions peuvent être liées des modalités d'expression ("`<...modality="face voice"> <category name="satisfaction"/>`"), et une intensité ("`<intensity value="0.1" confidence="0.8"/>`"), à laquelle est attribuée ici une certaine confiance. Ces différentes balises offrent ainsi un langage générique, permettant d'annoter les émotions suivant les principales

²Facial Action Coding System

³World Wide Web Consortium - <http://www.w3.org/>

théories et de spécifier les modalités d'expression de ces émotions.

Cependant, dans le cas de la génération d'émotions, l'énoncé et l'émotion sont indépendants : l'extrait de code figure 1 illustre cet absence de lien avec un exemple de génération d'émotions. "Do you need help ?" est le texte à énoncer avec une émotion de doute dont la catégorie appartient à l'ensemble défini par "everydayEmotions". Cette émotion est à exprimer avec une intensité de 0.4, par le ton de la voix. Dans cet exemple, on peut imaginer que le texte pourrait être remplacé par une phrase qui serait incohérent de prononcer avec une émotion de doute : par exemple, "there is no doubt in my mind that I'm right".

```
<speak>
<s>
<emo:emotion>
<emo:category set="everydayEmotions"
name="doubt"/>

<emo:intensity value="0.4"/>
</emo:emotion>

Do you need help?
</s>
</speak>
```

FIG. 1 – Exemple de génération d'émotion en EmotionML

L'identification des actes de discours reste une des méthodes les plus utilisées dans l'annotation et l'analyse de dialogue. Plusieurs travaux ([7], AMI Project⁴) s'inspirent de la théorie des actes de discours de Searle et de la taxonomie qu'il propose pour l'analyse du dialogue. Cette taxonomie est formée de cinq classes d'actes : les actes assertifs, directifs, engageants, déclaratoires et expressifs. Elle varie selon les travaux, mais une correspondance peut généralement s'établir entre les taxonomies. Dans les travaux basés sur cette taxonomie, l'identification et la classification des actes dans un énoncé s'effectuent de manière unidimensionnelle : à un énoncé correspond une seule annotation, un seul acte. Ce type d'annotation a été remis en cause notamment par Allen et Core ([8]), qui ont proposé une taxonomie multidimensionnelle, DAMSL. Un énoncé peut ainsi être lié à plusieurs actes, qui appartiennent à des fonctions communicatives "forward" ("Assert", "Offer", "Commit",...) ou "backward" ("Accept", "Signal-

Understanding", "Agreement",...). Harry Bunt montre quant à lui ([19]) que le fait d'avoir plusieurs dimensions facilite l'annotation et permet une analyse plus fine du discours. Ces différentes annotations indiquent le rôle des émotions dans la structure du dialogue et la gestion du sujet de la conversation, mais ne permettent pas d'identifier les émotions associées aux actes de discours.

Dans l'optique de mettre en place une base pour un langage émotionnel, et en lien avec la théorie des actes de discours de Searle ([24]), il est nécessaire de faire un lien entre les actes de discours, qui sont les unités premières de signification du langage, et les émotions.

4 Proposition : Actes de Conversation Multimodaux

Notre proposition s'appuie sur les actes de discours, et se situe à la croisée de l'expression d'émotion et de la capacité langagière de l'agent. En effet, dans les travaux ci-dessus il n'y a pas de lien entre ce que dit l'agent et les émotions qui sont exprimées : le lien entre énoncé et émotion n'est pas clairement établi ou tout du moins il n'est pas formalisé. Nous cherchons à éclairer ce lien et ainsi donner la possibilité à l'agent, lors de l'usage d'un acte de conversation d'exprimer également les émotions intrinsèques à cet acte. Nous entendons par "émotions intrinsèques" les émotions inhérentes à l'acte, indépendamment de tout facteur extérieur. Nous verrons par exemple que l'acte *S'excuser* "porte" l'émotion de regret ou encore que l'acte *Se plaindre* "porte" l'émotion de déception. Ainsi, nous tentons de répondre aux questions suivantes : comment va s'exprimer multimodalement un agent qui souhaite promettre quelque chose ? Comment peut-il être convaincant ? Quelles sont les conditions requises sur son état émotionnel pour qu'il soit sincère ?

La première partie de cette section introduit les actes de conversation définis par A. Berger suivant la théorie des actes de discours de Vanderveken. Notre proposition, présentée en deuxième partie, s'appuie sur ces actes pour identifier les composantes qui portent les émotions. Des premiers *actes de conversation multimodaux* sont alors définis à partir de la formalisation des *émotions complexes* que nous avons développée. Nous nous intéressons aux *émotions complexes* car, contrairement aux émotions primaires, qui sont réactives et qui se

⁴<http://corpus.amiproject.org/>

traduisent le plus souvent par des expressions faciales prototypiques, les *émotions complexes* sont le résultat d'un raisonnement réfléchi et s'expriment le plus souvent selon plusieurs modalités et principalement par la modalité langagière qui nous intéresse ici.

4.1 Définition et formalisation des *émotions complexes*

En ce qui nous concerne, nous nous intéressons à une catégorie bien particulière des émotions : celles basées d'une part sur le raisonnement contrefactuel de l'agent, et d'autre part sur son raisonnement à propos de sa responsabilité, de ses capacités et celles des autres agents, et des normes sociales. Le raisonnement contrefactuel consiste à comparer la réalité et une vue imaginaire de ce qui aurait pu se produire ([14]), et induit des émotions telles que le regret ou la culpabilité. Ce choix s'explique par l'importance de ces émotions dans le raisonnement de l'agent, comme le montrent les théories de l'évaluation ([23], [25]).

La formalisation proposée a été réalisée en coopération avec l'équipe LiLaC de l'IRIT⁵, dans le cadre du projet ANR CECIL⁶. Elle se base sur trois opérateurs : l'opérateur $Goal_i\varphi$ (l'agent i a pour but que φ soit vrai), l'opérateur $Ideal_i\varphi$ (idéalement pour i , φ devrait être vrai) et l'opérateur $Resp_i\varphi$ (l'agent i est responsable de φ) ([6], [15]). Un quatrième opérateur, $Bel_i\varphi$, représente la croyance qu'a l'agent i en φ . Nous définissons explicitement huit *émotions complexes* : le regret, la déception, la culpabilité, le reproche, la satisfaction morale, l'admiration, la réjouissance et la reconnaissance ou gratitude. Par exemple, l'émotion de culpabilité peut correspondre à un regard porté par l'agent sur ce qu'il a fait et ce qu'il aurait pu faire à un instant donné. De même, le reproche peut être dû à la comparaison des capacités d'un agent et de ce qu'il a fait par un autre agent. Le tableau 1 montre la formalisation de ces *émotions complexes* pour un agent i , concernant une situation φ .

Comme nous le verrons par la suite, nous définissons les *actes de conversation multimodaux* à partir de cette formalisation des *émotions complexes*. Cette formalisation apporte la preuve

théorique de notre proposition et permettra par la suite de servir de base au raisonnement de l'agent : à partir de ses connaissances, son environnement et des événements passés, un agent BDI⁷ ou EBDI⁸ ([18], [1]) peut déduire l'état émotionnel dans lequel il se trouve et communiquer de façon émotionnelle.

4.2 Actes de Conversation

Les travaux d'A. Berger ([5]) ont été menés dans le but de dépasser le simple échange d'information entre agents. Ils ont permis de mettre en place un *Langage de Conversation*, basé sur la théorie des actes de discours de Searle ([24]) et composé de trente-deux actes de conversation définis formellement. Ces actes ont été sélectionnés à partir des définitions des verbes performatifs données par Vanderveken ([26]) pour leur pertinence dans l'utilisation qu'un agent peut en faire, dans des situations comme l'assistance à l'utilisateur, la négociation, la coopération etc... Ces actes appartiennent aux cinq classes d'actes de discours définies par la théorie, qui correspondent aux cinq différentes façons fondamentales d'utiliser le langage.

Chacun de ces actes est de la forme $F(P)$, où P est la proposition énoncée et F la *force illocutoire* qui détermine dans quel but l'énoncé est formulé ([26]). La force illocutoire F est constituée de six composantes qui expriment chacune un aspect différent de l'acte. Ces composantes sont les conditions à respecter pour que l'acte soit performé avec succès et satisfaction, suivant *le but illocutoire* de l'acte qui représente l'usage que l'on fait de l'acte, l'état mental du locuteur s'il est sincère (*conditions de sincérité*), le contexte d'énonciation (*conditions préparatoires, conditions sur le contenu*), la manière dont le but doit être atteint (*mode d'atteinte*) et le *degré de puissance* de l'acte. De cette façon, les actes de conversation basiques tels que *Inform* et *Demand* ont été définis, mais aussi *Conseiller*, *Approuver* ou encore l'acte *S'excuser* dont le but illocutoire est expressif.

4.3 Lien avec les émotions et définition des *Actes de Conversation Multimodaux*

Notre travail a permis d'identifier le lien qui existe entre les actes de conversation détaillés ci-dessus et les émotions. Ce lien est obtenu

⁵Institut de Recherche en Informatique de Toulouse - <http://www.irit.fr/~Equipe-LiLaC>

⁶projet CECIL - Complex Emotions in Communication, Interaction, and Language - ANR-08-CORD-005

⁷Belief, Desire, Intention

⁸Emotional BDI

$\wedge$	$Goal_i\varphi$	$Goal_i\neg\varphi$	$Ideal_i\varphi$	$Ideal_i\neg\varphi$
$Bel_i Resp_i\varphi$	$Rejouissance_i\varphi$	$Regret_i\varphi$	$SatisfactionMorale_i\varphi$	$Culpabilite_i\varphi$
$Bel_i Resp_j\varphi$	$Reconnaissance_{i,j}\varphi$	$Deception_{i,j}\varphi$	$Adhesion_{i,j}\varphi$	$Reproche_{i,j}\varphi$

TAB. 1 – Formalisation des émotions *complexes* pour un agent i vis-à-vis de sa responsabilité (ou la responsabilité d'un agent j) par rapport à une situation φ .

à partir de l'étude de la théorie des actes de discours et consiste en trois composantes de la force illocutoire F de ces actes. Ainsi, si l'on considère par exemple l'acte de conversation *S'excuser*, l'émotion du regret est exprimée par **la condition de sincérité** de l'acte, qui est la condition minimale pour que le locuteur soit sincère. D'après Vanderveken ([26]) : *S'excuser, c'est exprimer ses regrets à un allocataire (condition de sincérité) à propos de l'état de chose représenté par p en présupposant (condition préparatoire) que l'on est responsable de cet état de chose et qu'il est mauvais pour l'allocataire*. Cet acte est un cas particulier de l'expression de l'émotion complexe du regret : vis-à-vis d'autrui, il s'agit d'exprimer du regret *de ne pas avoir satisfait le but d'autrui p* . On peut également exprimer du regret vis-à-vis de soi-même par rapport à la situation présente p .

Nous nous basons sur la formalisation logique des *émotions complexes* présentée précédemment pour exprimer l'émotion portée par la condition de sincérité. Ainsi, la condition de sincérité de l'acte *S'excuser* se traduit formellement par :

$$Regret_i(Goal_j\varphi \wedge \neg\varphi)$$

i se croit responsable du fait que le but de j ne soit pas atteint.

De manière plus générale, les conditions de sincérité portent les émotions à exprimer lors de l'acte. Ces conditions forment un lien permanent entre émotions et actes dans le sens où l'expression de l'émotion lors de l'acte est nécessaire et suffisante pour que le locuteur soit sincère. Nous définissons les *actes de conversation multimodaux* que nous appelons "neutres", car exprimant une seule *émotion complexe* avec un degré neutre, contrairement à des actes qui pourraient exprimer une composition de ces émotions ou des actes avec un *degré de puissance* plus élevé, comme détaillé ci-dessous. L'équipe LiLaC a défini l'opérateur $Assert_{i,j}\varphi$, représentant l'usage assertif de la théorie des actes de discours, car exprimer un état mental peut se

traduire comme informer quelqu'un de son état mental. Cet opérateur permet donc de traduire formellement : " i informe j que φ " ou, dans la logique des annonces ([13], symbolisée par !), "le but de i est que j croit φ ".

$$Assert_{i,j}\varphi \stackrel{d\acute{e}f}{=} Goal_i Bel_j\varphi !$$

L'acte de conversation multimodal *S'excuser* s'écrit donc formellement :

$$S'excuser_{i,j}\varphi \stackrel{d\acute{e}f}{=} Assert_{i,j}(Regret_i(Goal_j\varphi \wedge \neg\varphi))$$

Nous proposons une taxonomie des *actes de conversation multimodaux* "neutres" suivant les différentes *émotions complexes* qu'ils expriment. Chaque acte est défini formellement suivant l'opérateur $Assert_{i,j}$ et l'émotion complexe que l'agent i exprime à un autre agent j : comme ces actes sont des actes de communication, il y a toujours un locuteur et un allocataire. Le tableau 2 présente les huit actes ainsi définis, dont la seconde forme de l'expression de l'émotion du regret (regretter vis-à-vis de soi-même).

L'émotion à exprimer lors de l'acte est donc incluse dans la condition de sincérité, mais celle-ci n'est pas la seule composante à porter les émotions dans les actes de conversation. En effet, une seconde composante permet de déterminer la force avec laquelle s'exprime une émotion, par rapport à un acte plus modéré : c'est la composante de **degré de puissance**. En général, le degré de puissance d'un acte s'exprime dans un énoncé en utilisant des adverbes tels que sincèrement, franchement... Il forme des liens entre différents actes qui expriment une même émotion : par exemple, récriminer est plus fort que protester qui lui-même est plus fort que désapprouver ([26]). Le degré de puissance peut donc nous permettre d'enrichir notre bibliothèque d'*actes de conversation multimodaux* : on peut imaginer qu'une émotion de reconnaissance avec une forte intensité amènera

$Assert_{i,j}(\dots)$	Acte de conversation multimodal
$Goal_i\varphi \wedge Bel_iResp_i\varphi$	$\stackrel{déf}{=} Assert_{i,j}(Rejouissance_i\varphi) \stackrel{déf}{=} \mathbf{SeFeliciter}_{i,j}\varphi$
$Goal_i\varphi \wedge Bel_iResp_j\varphi$	$\stackrel{déf}{=} Assert_{i,j}(Reconnaissance_{i,j}\varphi) \stackrel{déf}{=} \mathbf{Remercier}_{i,j}\varphi$
$Goal_i\neg\varphi \wedge Bel_iResp_i\varphi$	$\stackrel{déf}{=} Assert_{i,j}(Regret_i\varphi) \stackrel{déf}{=} \mathbf{Regretter}_{i,j}\varphi$
$Goal_i\neg\varphi \wedge Bel_iResp_j\varphi$	$\stackrel{déf}{=} Assert_{i,j}(Deception_{i,j}\varphi) \stackrel{déf}{=} \mathbf{SePlaindre}_{i,j}\varphi$
$Ideal_i\varphi \wedge Bel_iResp_i\varphi$	$\stackrel{déf}{=} Assert_{i,j}(SatisfactionMorale_i\varphi) \stackrel{déf}{=} \mathbf{Reconnaître}_{i,j}\varphi$
$Ideal_i\varphi \wedge Bel_iResp_j\varphi$	$\stackrel{déf}{=} Assert_{i,j}(Adhesion_{i,j}\varphi) \stackrel{déf}{=} \mathbf{Adherer}_{i,j}\varphi$
$Ideal_i\neg\varphi \wedge Bel_iResp_i\varphi$	$\stackrel{déf}{=} Assert_{i,j}(Culpabilite_i\varphi) \stackrel{déf}{=} \mathbf{Culpabiliser}_{i,j}\varphi$
$Ideal_i\neg\varphi \wedge Bel_iResp_j\varphi$	$\stackrel{déf}{=} Assert_{i,j}(Reproche_{i,j}\varphi) \stackrel{déf}{=} \mathbf{Reprocher}_{i,j}\varphi$

TAB. 2 – Actes de conversation multimodaux définis avec l’opérateur Assert et exprimant une émotion complexe

l’agent à utiliser un acte tel que complimenter, vanter ou louer plutôt que remercier.

Une dernière composante, le **mode d’accomplissement**, détermine comment le but illocutoire doit être accompli et les modalités d’expression des émotions contenues dans l’acte. En effet, si l’on considère l’exemple du verbe performatif Protester définis par Vanderveken ([26]) : *Protester, c’est exprimer d’une façon formelle (mode d’accomplissement) sa désapprobation.... D’une façon formelle* peut se traduire par le langage (tournures élaborées, rigueur,...) mais également par la gestuelle, la posture ou l’expression du visage. Ainsi, les modes d’accomplissement permettent de déterminer les modalités d’expression des différentes émotions, que ce soit la modalité langagière ou d’autres modalités. On peut imaginer que cette composante pourra être par la suite influencée par exemple par la personnalité de l’ACA, le type de relation qu’il entretient avec l’utilisateur ou simplement par les modalités qu’il peut exprimer. Un agent peut donc, à la sortie de son raisonnement, choisir l’acte qui correspond à sa volonté d’action : à cet acte sont liées les émotions et les modalités d’expression de celles-ci, que l’agent va alors exprimer de façon multimodale.

Nous proposons ici neuf premiers *actes de conversation multimodaux* "neutres" exprimant les *émotions complexes*, dont la forme particulière de l’expression de l’émotion du regret qui induit deux actes, *S’excuser* et *Regretter*. Ces actes constituent les premiers actes de notre bibliothèque. De prochains actes, construits à partir des degrés de puissance et des compositions d’émotions, restent à définir. L’étape sui-

vante sera la formalisation des autres classes d’actes de conversation d’A. Berger, c’est-à-dire les actes engageants tels que promettre et accepter, les actes directifs tels que exiger et conseiller, etc... La différence entre ces actes et ceux que nous avons définis est que ces derniers ont pour but d’exprimer l’*émotion complexe* liée, alors que pour les autres classes l’émotion pourra être liée mais ne sera pas le but de l’acte : on peut imaginer par exemple que des actes comme exiger ou promettre ne sont pas détachés de toute émotion pour être convaincants. Les quatre *actes de conversation multimodaux* *S’excuser*, *Reprocher*, *Se plaindre* et *Culpabiliser* ont été expérimentés avec l’agent Greta, dans le but d’identifier les limites et les prochaines perspectives de travail concernant la mise en place de notre bibliothèque dans l’architecture de Greta, laquelle est détaillée dans la section suivante.

5 Première implémentation avec Greta et perspectives de travail

Cette partie présente l’architecture de Greta et l’implémentation de l’*acte de conversation multimodal S’excuser*. Cette première implémentation ne prend pas en compte la formalisation proposée qui pourra être utilisée par la suite comme base de raisonnement de l’ACA, mais a pour but de déterminer les modifications à effectuer sur l’architecture de Greta de manière à y ajouter notre bibliothèque d’*actes de conversation multimodaux*.

5.1 Architecture de Greta et implémentation

L'ACA Greta est conforme au standard SAIBA ([27]) qui est une initiative de recherche internationale dont le but est de définir une architecture logicielle commune pour la génération du comportement d'agents virtuels. L'architecture SAIBA est composée de trois modules. Le premier module *Intent Planning* décide de l'intention communicative de l'agent à partir de ses buts, de ses croyances et de son état émotionnel. Cette intention communicative est encodée en langage FML⁹ ([16]) qui est un langage de balisage pour être interprété par le module *Behavior Planning*. Ce module gère un certain nombre de modalités de communication dans le but d'exprimer l'intention communicative de l'agent : la parole, l'expression du visage et les gestes sont ainsi calculés. Ces différentes modalités traduisent le comportement verbal et non-verbal de l'agent, qui, une fois ces modalités encodées en BML¹⁰ ([27]), est réalisé par le dernier module *Behavior Realization*.

Notre travail consiste à intervenir au niveau des deux premiers modules. Lors de la planification de l'intention communicative, l'agent choisit, dans la bibliothèque d'*actes de conversation multimodaux*, l'acte qui correspond le mieux à son intention de communiquer. A cet acte est associé un ensemble d'émotions qui seront exprimées de manière concomitante. Nous avons implémenté les quatre *actes de conversation multimodaux* que nous avons définis. L'extrait de code en langage FML figure 2 illustre ici comment a été implémenté l'acte *S'excuser* et l'*émotion complexe* du regret associée.

```
<tm id="tm1"/>
  I apologise
<tm id="tm2"/>
  for my absence
<tm id="tm3"/>
<pitchaccent id="xpa1" type="Lstar" start="s1:tm1"
end="s1:tm2" level="high" importance="1" />
<performative id="p1" type="announce"
start="s1:tm1" end="s1:tm3"/>
<emotion id="e1" type="remorse"
start="s1:tm1" end="s1:tm3"
regulation="felt" importance="1" intensity="1" />
```

FIG. 2 – Texte à énoncer divisé par des marqueurs de temps

Ces informations sont ensuite traduites par le module *Behavior Planning* en BML, pour la

réalisation du comportement effectuée par le dernier module. Ainsi, l'émotion *remorse* correspond à une certaine expression du visage, à une certaine posture du corps, et est exprimée durant l'énonciation de l'acte.

5.2 Conclusions et perspectives de travail

Cette implémentation sur l'ACA soulève un certain nombre de problèmes qui illustrent une partie des modifications à réaliser dans l'architecture de Greta. Dans l'exemple de l'acte *S'excuser*, la séparation de l'énoncé en plusieurs parties par les marqueurs de temps est à faire de façon manuelle, ainsi que la spécification des parties à accentuer. L'*émotion complexe* du regret n'est pas intrinsèque à l'acte et doit également être explicitée manuellement. Par ailleurs, un certain nombre d'approximations sont faites pour les besoins de l'implémentation (émotion de *remords*, performatif *announce*). En l'état actuel donc, le langage FML ne permet pas d'exprimer toutes les informations définies dans les *actes de conversation multimodaux*, comme celles portées par le mode d'accomplissement et le degré de puissance. La création de nouvelles balises dans le langage FML est donc nécessaire pour transmettre ces nouvelles informations au module *Behavior Planning*, et pouvoir choisir directement pour l'intention communicative un *acte de conversation multimodal* de la bibliothèque, ce qui permettra d'obtenir implicitement les émotions liées à cet acte. De ce fait, des modifications sur le module *Behavior Planning* sont à réaliser, dans le but de le rendre capable d'interpréter les nouvelles balises qu'il reçoit en FML, et traduire les émotions dans plusieurs modalités.

Le travail décrit dans cet article n'est que le premier pas vers l'implémentation d'une bibliothèque la plus complète possible d'*actes de conversation multimodaux*, générique et ouverte. Nous avons actuellement défini neuf *actes de conversation multimodaux* présentés en 4.3 et implémenté quatre d'entre eux avec Greta, dans le but d'identifier les modifications à faire dans son architecture. Cette bibliothèque devra rendre possible par la suite l'expression de l'ensemble des trente-deux actes du *langage de conversation*. Elle sera ensuite implémentée dans l'architecture SAIBA, ce qui permettra l'évaluation de cette bibliothèque auprès d'utilisateurs grâce à des scénarii de tests. Cette évaluation devra vérifier d'un côté la crédibilité de l'agent en s'appuyant sur des critères de re-

⁹Function Markup Language

¹⁰Behavior Markup Language

connaissance d'émotions, et de l'autre l'amélioration de l'interaction entre agents humains et virtuels, c'est-à-dire la performance de l'ACA, et pourra permettre de valider notre hypothèse théorique.

Remerciements

Ce travail a bénéficié d'une aide de l'Agence Nationale de la Recherche portant la référence ANR-08-CORD-005.

Références

- [1] Carole Adam. *The emotions : from psychological theories to logical formalization and implementation in a BDI agent*. Thèse de doctorat, Institut National Polytechnique de Toulouse, Toulouse, France, juillet 2007.
- [2] Paolo Baggia and al. Emotion markup language (emotionml) 1.0. 2009.
- [3] Joseph Bates. The role of emotion in believable agents. *Commun. ACM*, 37(7) :122–125, 1994.
- [4] Christian Becker and Ipke Wachsmuth. Simulating the emotion dynamics of a multimodal conversational agent. In *In Proceedings Tutorial and Research Workshop on Affective Dialogue Systems (ADS-04), LNAI 3068*, pages 154–165. Springer, 2004.
- [5] Alexandra Berger. *La communication entre agents de communautés mixtes : Un Langage de Conversation Expressif pour agents artificiels*. Thèse de doctorat, Institut National de Polytechnique de Grenoble et Université du Québec à Trois-Rivières, décembre 2006.
- [6] Jean-François Bonnefon, Dominique Longin, and Manh Hung Nguyen. A Logical Framework for Trust-Related Emotions. *Electronic Communications of the EASST, Formal Methods for Interactive Systems 2009*, 22, 2009.
- [7] Jean Caelen. Stratégies de dialogue. In Chaib-Draa Brahim Philippe Mathieu Herzig, Andreas, editor, *Actes des Secondes Journées Francophones des Modèles Formels de l'Interaction*, page 29/39. Cépaduès, 2003.
- [8] Mark G. Core and James F. Allen. Coding dialogs with the DAMSL annotation scheme. In *Proc. of the Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*, Cambridge, MA, 1997.
- [9] Matthieu Courgeon, Jean-Claude Martin, and Christian Jacquemin. Marc : a multimodal affective and reactive character. In *Proceedings of The 1st workshop on Affective Interaction in Natural Environments (AFFINE)*, 2008.
- [10] P. Ekman and W.V. Friesen. The facial action coding system : A technique for the measurement of facial movement. In *Consulting Psychologists*, 1978.
- [11] Paul Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3) :169–200, 1992.
- [12] Nico H. Frijda. *The Emotions*. Cambridge University Press, 1986.
- [13] Nadine Guiraud, Andreas Herzig, and Emiliano Lorini. Speech acts as announcements (regular paper). In *Dagstuhl Seminar on Information processing, rational belief change and social interaction*, Dagstuhl, Germany, number 1862-4405. Science Publications, novembre 2009.
- [14] Daniel Kahneman and Dale T. Miller. Norm theory : Comparing reality to its alternatives. *Psychological Review*, 93(2) :136–152, 1986.
- [15] Emiliano Lorini and François Schwarzentruher. A logic for reasoning about counterfactual emotions. In C. Boutilier, editor, *International Joint Conference on Artificial Intelligence (IJCAI)*, Pasadena (USA), pages 867–872. AAAI Press, 2009.
- [16] M. Mancini and C. Pelachaud. The FML - APML language. In *The First FML workshop, AAMAS'08, Estoril, Portugal*, 2008.
- [17] Albert Mehrabian. Pleasure-arousal-dominance : A general framework for describing and measuring individual differences in temperament. *Current Psychology*, 14(4) :261–292, December 1996.
- [18] David Pereira, Eugénio Oliveira, and Nelma Moreira. Formal modelling of emotions in bdi agents. pages 62–81, 2008.
- [19] Volha Petukhova and Harry Bunt. A multidimensional approach to multimodal dialogue act annotation. In *Proceedings Seventh International Workshop on Computational Semantics (IWCS-7)*, pages 142–153, 2007.
- [20] J. Piaget. *Les émotions*, pages 75–95. Delachaux et Niestle, Neuchâtel-Paris, 1989.
- [21] Rosalind W. Picard. *Affective Computing*. The MIT Press, Cambridge, Massachusetts, 1997.
- [22] I. Poggi and al. Greta : A believable embodied conversational agent, 2005.
- [23] Klaus R. Scherer. *Emotion*, pages 151–195. Wiley-Blackwell, 2001.
- [24] John R. Searle. *Speech acts : an essay in the philosophy of language / John R. Searle*. Cambridge University Press, London :, 1969.
- [25] Aaron Sloman. Beyond shallow models of emotion. In *Cognitive Processing : International Quarterly of Cognitive Science*, pages 177–198, 2001.
- [26] Daniel Vanderveken. *Les actes de discours*. Editions Pierre Mardaga, Liège, Bruxelles, 1988.
- [27] Hannes Vilhjalmsson and al. The behaviour markup language : recent developments and challenges. *Lecture notes in artificial intelligence*, 4722 :99–11, 2007.

Expressing agents personality through ambient modalities

J. P. Sansonnet* Y. Bellik*
jps@limsi.fr bellik@limsi.fr

*LIMSI-CNRS BP 133 F-91403 Orsay cedex

Abstract:

In this study, we explore the possibility to express the emotions and personality traits of an ambient assistant agent through the specific output modalities of the ambient environment as an alternative to the modalities provided by an animated graphic character.

Keywords: Ambient modalities, Personality expression.

1 Agents in an ambient room

In previous work, we have implemented Conversational Assistant Agents (CAA) in an ambient environment, namely the IRoom project at LIMSI-CNRS [1]. Presently, two main modalities support the agent/user interactions:

— *Natural language* for control/command and assistance is based on Speech Recognition (SR) and Text to Speech (TTS) output.

— *Personification* is based on the display of animated characters on various screens, as illustrated by the layout, shown in Figure 1.

For the user, the agent is viewed as consubstantial of the ‘ambient as a person’ (say the *agent-room*), entailing that beyond rational assistance about control/command, the user expects the agent to reveal a psychological behavior. Although, natural language and graphic animation provide a straightforward basis for expressing the psychology of the agent, one can look for **alternative modalities** as a support for it support. Such alternative modalities can be provided by ambient devices listed in Table 1-middle that provide ‘atmosphere components’ such as described in Table 1-bottom.

2 Agent’s personality expression

Research on human psychology has exhibited two main criteria for classifying personality descriptors: first intrinsic features *vs* interpersonal relations and second static traits *vs* dynamic emotional states. Here, we will restrict to **intrinsic descriptors** and consider both the expression of emotional states and trait handling, in the context of ambient output modalities.

Emotions Typically, emotional states refer to Paul Ekman’s six basic emotions (see Table 2-

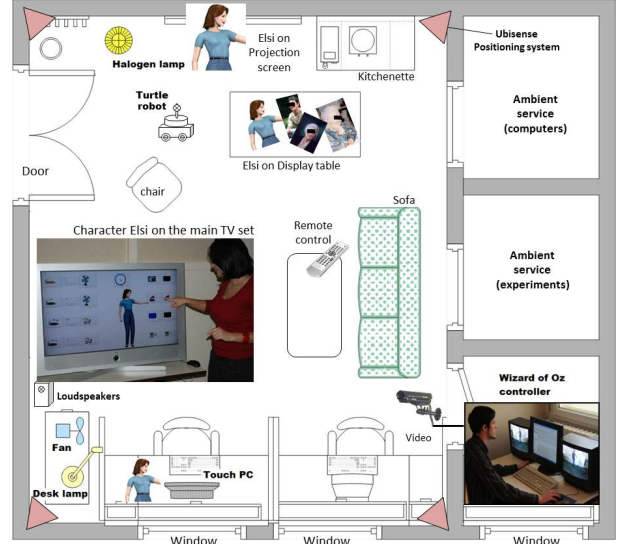


Figure 1: Layout of LIMSI IRoom project.

Table 1: Output modalities of the IRoom.

Character display Text to Speech Screens	<i>also used for information display</i> Agents’ oral expression TV, mural screen, touch PC <i>etc.</i>
Devices Air control Light control Sound control Static appliances* Robots	<i>producing an output effect</i> fan, heater, cooler*, scent dispenser* lamps, electric curtains* music loudspeakers, alarms coffee-machine, cooker, fridge <i>etc.</i> autonomous moving machines
Atmosphere Luminance Music (<i>backgd</i>) Alarm Temperature* Scent* Devices	<i>main components</i> level, color (hot, cold, red, green...) and dynamics (waves, flash) level and mood (chill, cheer, sad...) level, type and dynamics (bip, honk...) level level and theme (spring, gas, sweat...) force and specific action

*Not yet implemented.

left) even if other authors, *e.g.* Frijda, have proposed distinct models.

Traits For more than 20 years, research on personality traits [3] focuses on the classification of psychological notions. Presently the prevalent paradigm is the Five Factor Model (FFM) [2], resulting in five large classes of traits called *factors* (also noted in short, O.C.E.A.N.).

Table 2: Expression of emotional states.

Mental states	Luminance	Music	Alarm	Temp.	Scent	Device
None	= neutral	= chill	0	=	0	= unspecific
Joy	+ hot	+ cheer	0	=	+ spring	+ Robot.move
Sadness	- cold	- sad	0	-	0	- Robot.move
Fear	+ red <i>blink</i>	0	+ danger <i>rep</i>	+	+ gaz	+ Robot.hide
Surprise	+ neutral <i>flash</i>	0	+ oops <i>once</i>	=	0	0 Robot.stop; Fan.stop
Anger	+ red	+ harsh	+ rap <i>rep</i>	+/-	+ sweat	+ Fan.run
Disgust	- Gloomygreen	0	0	-	0	= unspecific

0 is none = is neutral (as set by user) + is higher than neutral or none (- is lower) rep is repetitive.

Expression of emotional states Considering the output modalities of the IRoom, we fill the features (level, theme, type, *etc.*) of the atmosphere components (defined in Table 1-bottom) in order to express Ekman's emotions as illustrated in Table 2. A given emotion can be expressed either in redundant mode (all features of the atmosphere components are activated) or in partial mode (at least one feature is activated). Table 2 reveals two main results: a) any atmosphere component is used, at least three times b) any emotion can rely on several modalities (at least three). This shows that ambient output modalities can support a form of expression for basic emotions. *Note:* that it does not imply that people would actually perceive them.

Expression of personality traits In the R&B architecture [4] we have proposed a framework to express traits in terms of their various influences/alterations over the rational process of an agent achieving a plan. We will just give an illustration of this approach through the example of a common personality adjective like 'lazy' that can be classified in the FFM model as part of the factor **C** (standing for Conscientiousness). The main Wordnet gloss of *lazy* = "*shy of work or exertion*" indicates that a lazy agent is subject to the four following influences on its rational behavior:

1) *Plan modification* to avoid performing a required action, either by providing the user with a rational **Rebuke** or with a less hard to perform **Alternative** action.

2) *Action modality* to perform a required action in a **Partial** manner or in a **Slack** manner.

Table 3 gives the values of the four influences associated with *lazy*. The four actions can be altered by one or more influences; respectively, each influence is activated, at least in two cases. This example shows how one can proceed from personality adjectives to rational process alterations.

3 Further work

There has been a lot of works about expressing emotions through embodied (especially human) characters. The issue raised here was: can emotions be expressed in non-humanlike modalities? If proved it could bring an interesting light for ECA research itself. Moreover, this preliminary proposition is an incentive for building experiments in the IRoom project in order to evaluate the actual perception of the subjects, especially the level of modal redundancy required.

References

- [1] Y. Bellik, I. Rebaï, E. Machrouh, Y. Barzaj, C. Jacquet, G. Pruvost, and J. P. Sansonnet. Multimodal interaction within ambient environments: an exploratory study. *In Proc. of INTERACT 2009*, 5727(1):89–92, 2009.
- [2] L. R. Goldberg. An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59:1216–1229, 1990.
- [3] Oliver P. John, Richard W. Robins, and Lawrence A. Pervin, editors. *Handbook of personality*. The Guilford Press, third edition edition, 2008.
- [4] Jean Paul Sansonnet and François Bouchet. Joint handling of rational and behavioral reactions in assistant conversational agents. In *European Conference on Artificial Intelligence ECAI'10*, Lisbon, 2010.

Table 3: Influences of *lazy* on four actions.

Actions	Rebuke	Altern.	Partial	Slack
Open curtains	too shiny!	lamp on	yes	yes
Play music	.	.	yes	.
Set timeout	.	post it	.	.
Clean floor	bag full battery low	.	yes	yes

Expressive Gesture Model for Storytelling Humanoid Agent

Le Quoc Anh and Catherine Pelachaud

{quoc-anh.le, catherine.pelachaud}@telecom-paristech.fr
CNRS, LTCI Telecom ParisTech, France

Abstract : We aim at developing an expressive gesture model for the humanoid robot Nao and the virtual agent Greta during a storytelling application.

Keywords: Humanoid robot, expressive gesture, Nao, Greta, ECA, BML, FML, SAIBA.

1 Introduction

The work is a part of the GVLEX project whose objective is to equip agents with the capacity of reading story agreeably through expressive verbal and non-verbal behaviors. While other partners of the project deal with expressive voice and linguistic aspects, our task focus on expressive behaviors, especially on expressive gesture, that will be displayed by the humanoid robot Nao developed by Aldebaran [6] and by the virtual agent Greta [5].

2 System Overview

The system takes as input a text (i.e. a story) to be said by the agent. The text has been enriched with information on the manner the text ought to be said (i.e. with which communicative acts it should be said). The behavioural engine bases on information extracted from the enriched text, such as the structure, the various semantic and pragmatic elements as well as the emotion content of the story to select the multimodal behaviors to display and to synchronize the verbal and nonverbal behaviors of the agents.

The first attempt is to develop our existing agent system, Greta - a general purpose ECA (Embodied Conversational Agent) using and modular architecture which follows SAIBA framework [3], to control multimodal signals (i.e. expressive gestures, speech, etc) in a precise context: telling a story expressively for children. In this architecture, the agents are driven by two representation languages, one at the intention level FML (Function Markup Language) [2] and one at the behavior level

BML (Behavior Markup Language) [3,4].

Both FML and BML are XML languages and they do not refer to any particular parameters of an agent. We also use BML to describe repertoire of behaviors, called lexicon. The robot and the agent do not have the same behavior capacities (e.g. the robot can move its legs and torso but does not have facial expression and very limited hand movements). That is why a proper lexicon should be elaborated for each agent based on the storytelling video corpus [7] in order to fit the storytelling application. Moreover, the respective elements in both lexicons should have the same meaning. To ensure that, behavior invariant should be elaborated. By Calbris's taxonomy, description dimensions of a gesture might be divided into two parts: invariant and variant, in which the invariant dimensions should be consistent meanwhile the variant dimensions can be adjusted without changing the meaning of the gesture [1]. For instance, in the "refuse" gesture, orientation of the palm towards ahead is one invariant dimension while the distance between the hand and the body is one variant dimension. BML will be extended to include information related to behavior invariant. An overview of the system is illustrated in Figure 1.

3 Behavior Engine

The Greta platform, which is used to control behaviors of the Nao robot and the virtual agent, bases on the interfaces of the multimodal generation SAIBA framework (Situation, Agent, Intention, Behavior, Animation). Its structure consists of three separated modules[3]:

Intent Planning: Planning a communicative intent that agent will convey.

Behavior Planning: Planning multimodal behaviors that agent will realize.

Behavior Realization: Realize the planned behaviors.

Two processing stages, Intent Planning and the Behavior Planning, are interconnected by an interface described in FML, a representation language encodes the communicative and expressive intent that the agent aims at transmitting. That may be emotional states, beliefs or goals, etc. [2].

Meanwhile, two processing stages, Behavior Planning and Behavior Realization, are linked by another interface described in BML, another representation language specifies multimodal behaviors with constraints to be realized by agent. The description of BML may be occurrence of behaviors, relative timings of behaviors, form of behavior or reference to extended animation scripts, conditions, events, feedbacks, etc [3.4].

Last, the behavior is keyframed. Each keyframe is described by the value of all the articulators of the robot (the virtual agent correspondingly).

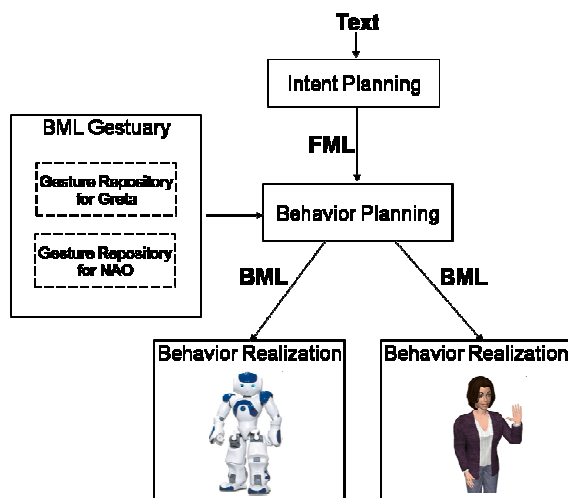


Fig. 1: The Greta platform controlling the behaviors of the Nao robot (left) and virtual agent (right).

4 Conclusion

In this article, we present an overview of a common system that controls expressive gesture for the humanoid robot Nao and the virtual agent Greta. In order to solve different behavior capacities of the agents, we use two lexicons, one for the robot and one for the virtual agent where their respective entries convey the

similar meaning.

5 Acknowledgments

We thank Andre-Marie Pez, Elisabetta Bevacqua and Radoslaw Niewiadomski for their help. This work has been funded by GVLEX project (www.gvlex.com).

References

- [1] Calbris, G. 1983. *Contribution à une analyse sémiologique de la mimique faciale et gestuelle française dans ses rapports avec la communication verbale*. PhD Thesis, Volume II.
- [2] Heylen, D., Kopp, S., Marsella, S., Pelachaud, C., Vilhjalmsen, H., 2008. *The Next Step Towards a Functional Markup Language*, Intelligent Virtual Agents, IVA'08, Tokyo.
- [3] Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thorisson, K., Vilhjalmsen, H., 2006. *Towards a Common Framework for Multimodal Generation in ECAs: The Behavior Markup Language*. 6th Int. Conf. on Intelligent Virtual Agents.
- [4] Vilhjalmsen, H., Cantelmo, N., Cassell, J., Chafai, N.E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A.N., Pelachaud, C., Ruttkay, Z., Thorisson, H. van Welbergen, R. van der Werf, 2007. *The Behavior Markup Language: Recent Developments and Challenges*, Intelligent Virtual Agents, IVA'07, Paris.
- [5] Niewiadomski, R., Bevacqua, E. Mancini, M., Pelachaud, C. 2009, *Greta: A Interactive Expressive ECA System*. The 8th Int. Conf. on Autonomous Agents and Multiagent Systems.
- [6] Gouaillier, D., Hugel, V. et al, 2009. *Mechatronic design of NAO humanoid*. IEEE Int. Conf. on Robotics and Automation, Kobe, Japan.
- [7] J.-C. Martin. *The contact video corpus*, 2009.

Un système multi-agents permettant une assistance homme-machine mutuelle dans un environnement ambiant

H.Sassi^{1,2}
Hajer.sassi
@ed.univ-lille1.fr

J. Rouillard¹
Jose.rouillard
@univ-lille1.fr

J-C. Tarby¹
Jean-claude.tarby
@univ-lille1.fr

G. Renard²
gregory.renard
@xbrainlab.com

¹Laboratoire d'Informatique Fondamentale de Lille
Université des Sciences et Technologies de Lille
59655 Villeneuve d'Ascq cédex – France

²xBrainLab-Groupe Usilink – Euratechnologies
165 Avenue de Bretagne
59000 Lille - France

Résumé :

L'objectif de nos travaux de recherche est de pouvoir aider les utilisateurs dans leurs tâches quotidiennes tout en mettant les nouvelles technologies à leurs services et en assurant un usage simple et pédagogique. Nous avons choisi de travailler sur la notion de système multi-agents (SMA) et plus précisément sur les agents introspectifs qui ont la capacité d'analyser et de raisonner sur leurs propres compétences. Cet environnement se compose des agents réels (les humains), des agents virtuels et de tout objet pouvant être référencé par l'environnement. Les agents virtuels apprennent les habitudes et les connaissances des agents réels afin qu'ils ouvrent des nouvelles perspectives d'apprentissage, et ils réunissent leurs savoir-faire afin d'agir plus efficacement (collaboration). Ceci en combinant le monde réel et le monde virtuel à travers une assistance homme-machine mutuelle.

Mots-clés : assistance mutuelle, agent introspectif, informatique ubiquitaire, interaction homme/machine, interaction machine/machine, système multi-agents.

Abstract :

The objective of our research is to better help users with their daily tasks and simultaneously make technologies more accessible while ensuring a simple and educational use. We decided to work on the concept of multi-agent systems (SMA) and more particularly on introspective agents which have the ability to analyze and reason by their own. This situation is made up of real agents (humans), virtual agents as well as all devices which can be labeled as environmental. Virtual agents pick up the habits and knowledge of live agents in order to broaden new learning perspectives, and they combine their expertise to act more effectively (collaboration). This can be possible by linking the real world to the virtual world by way of mutual human-machine assistance.

Keywords: mutual assistance, introspective agent, ubiquitous computing, human/machine interaction, machine-machine interaction, multi-agent systems.

1 Introduction

Depuis quelques années nous nous trouvons face à un monde numérique très riche en

information. Chaque individu a la possibilité de faire partie de réseaux sociaux informatisés, plus ou moins complexes à gérer. La possibilité de se faire représenter par des agents logiciels, capables de se coordonner dans de tels environnements est séduisante mais difficile à mettre en œuvre pour plusieurs raisons : confidentialité, sécurité, gestion des connaissances, prise de décisions de manière autonome, collaboration etc. Par ailleurs, le domaine de l'intelligence ambiante (AmI) propose une vision plus distribuée des entités (capteurs et effecteurs) pouvant participer à l'amélioration des services proposés aux utilisateurs (cas de la domotique par exemple).

2 État de l'art

La littérature fait état de plusieurs problèmes et verrous scientifiques liés à cette problématique, tels que l'adaptation individuelle/collective, la gestion de l'interaction bidirectionnelle simultanée agent/environnement, la nature des interactions qui peut être un choix ou une contrainte et le partage efficace des informations entre agents. L'AmI est décrite comme un environnement ayant l'aptitude de percevoir, de raisonner, d'agir et d'interagir avec les objets du quotidien, qui intègrent les technologies de l'information. Ces environnements sont basés sur l'éclatement des systèmes informatiques dans une multitude d'objets hétérogènes dans leur nature et leur mode de communication [6].

L'évolution actuelle du web (cf. web 2.0 et 3.0 orientés « social » et « sémantique ») tend à améliorer la qualité des services interrogés par les internautes. Ceci à travers le développement

d'un certain nombre d'outils et de services parmi lesquels on peut citer les réseaux sociaux, définis comme un ensemble d'individus ou organisations sociale reliées entre elles par des interactions sociales [1]. Un ensemble d'individus peut être aussi représenté par un ensemble d'agents introspectifs (SMA), qui ont la capacité d'interagir avec leur environnement ambiant. Ce passage entre ces deux mondes est un point que nous étudions plus spécifiquement. Les approches actuellement proposées dans l'AmI sont pour la plupart centralisées autour d'un contrôleur central [3, 4] ou d'un moteur d'exécution de plans [5], ce qui rend le système faillible si une panne survient au niveau du contrôleur. Et même les approches décentralisées existantes ne traitent que les besoins saisis par l'utilisateur [2].

3 Problématique

3.1 Scénario

Imaginons que, dans une maison intelligente, un utilisateur interroge son agent afin qu'il l'aide à gérer au mieux son chauffage. Ce dernier peut rencontrer un ensemble de problèmes, parmi lesquels, le fait que l'agent ne puisse pas savoir répondre, s'il n'est pas un expert du domaine ; que la demande de l'utilisateur soit trop imprécise, si elle a un sens large, etc. Un agent doit chercher la réponse la plus adéquate possible et qui s'adapte le mieux aux besoins de l'utilisateur, mais celle-ci peut également contenir des informations confidentielles. Les interactions basées sur des actions communicatives mutuellement échangées sont à la base de cette problématique.

3.2 Solutions envisagées

Un agent qui apprend ses connaissances d'un être humain ne pourra jamais être un expert dans tous les domaines, mais pour répondre au mieux, l'agent peut interroger ses accointants [2]. Ceci sera réalisé en implémentant une approche dynamique d'interaction inter-agents et en se basant sur des outils développés par l'équipe interne USILINK (la plate-forme 2.5 du IAS et le langage Athena). Pour que l'agent soit capable de satisfaire les besoins des utilisateurs, exprimés au sens large, il doit intégrer un module d'apprentissage capable d'étudier leurs comportements et de générer des nouvelles connaissances. Cela permet à l'agent d'affiner ses recherches et à l'utilisateur de s'exprimer librement. Dans le cas où l'agent

trouve une réponse acceptable, mais ne peut pas l'utiliser, pour des raisons de confidentialité, par exemple, nous proposons de traiter le problème selon des règles de degré de proximité. Exemple : un utilisateur est en train de naviguer sur internet, il a trouvé une offre intéressante de voyage, il veut réserver immédiatement mais il ne peut pas joindre son épouse par téléphone, pour lui demander son avis. L'utilisateur contacte son représentant virtuel pour lui trouver l'information. Ce dernier analyse la requête et commence à échanger avec l'agent destinataire qui lui annonce la confidentialité des données trouvées. Par conséquent, les deux agents déclencheront un test d'identification de degré de proximité. Le résultat de l'identification prouve qu'ils sont conjoints, l'agent destinataire peut alors libérer l'information.

4 Travaux en cours et perspectives

Nos travaux se focalisent actuellement sur la conception d'une approche dynamique capable d'inférer les besoins de l'utilisateur et de coordonner les compétences de chaque entité composant l'environnement. Nous étudierons dans un second temps l'intérêt de la multi-canalité, de la multi-modalité et de l'auto-duplication (des agents) dans un environnement ambiant.

Références

- [1] J. A. Barnes, « Class and Committees in a Norwegian Island Parish », *Human Relations*, n° 7, 1954.
- [2] Y. Charif, Chorégraphie dynamique de services basée sur la coordination d'agents introspectifs, *une thèse de doctorat en informatique*, Université Pierre et Marie Curie, 2007.
- [3] A. Gárate, N. Herrasti, and A. López. GENIO : an Ambient Intelligence Application in Home Automation and Entertainment environment. In *Proc. Joint Conference on Smart Objects and Ambient Intelligence*, pages 241–256, 2005.
- [4] T.A. Lashina. Intelligent Bathroom. In *European Symposium on Ambient Intelligence (EUSAI 2004)*, 2004.
- [5] M. Vallée, F. Ramparany, and L. Vercouter. Composition flexible de services d'objets communicants. In *Ubimob'05*, volume 120, pp 85–192, 2005.
- [6] M. Weiser. The Computer for the 21st Century. *Scientific American*, 165(3): 94–104, 1991.

Combinaisons d'expressions vocales, faciales et posturales des émotions chez un agent animé : ce que perçoivent les utilisateurs

Céline Clavel
celine.clavel@limsi.fr

Laurence Devillers
devil@limsi.fr

Jean-Claude Martin
martin@limsi.fr

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
B.P. 133
91403 ORSAY CEDEX – FRANCE

Résumé :

L'objectif de ce papier est de présenter l'évaluation d'un système de communication à distance médiatisée par un avatar capable d'exprimer les émotions détectées dans la voix de l'utilisateur. Un module de reconnaissance audio des émotions exprimées dans la voix et une librairie d'expressions non verbales émotionnelles ont ainsi été développés. Le focus de ce papier porte plus particulièrement sur l'impact de l'intégration de ces deux composants sur la perception émotionnelle et sur la qualité perçue de la synchronisation audio/vidéo.

Mots-clés : Agent conversationnel, émotion, détection, voix, comportement non verbal

Abstract:

The aim of this paper is to present the evaluation of a mediated communication system by a virtual agent who can express emotions which are detected in the voice of the user. An audio recognition module of expressed emotions in the voice and a library of nonverbal emotional expressions have been developed. The interest of this paper focuses on the impact of integration of these two components on the emotional perception and the perceived quality of the audio / video.

Keywords: Virtual agent, emotion, detection, voice, non-verbal behavior

1 Introduction

Communiquer revient à échanger des signaux verbaux et non verbaux comme les expressions faciales ou posturales, la gestualité, l'intonation de la voix... Une partie de ces signaux se rapporte à la dimension émotionnelle. Comprendre les émotions véhiculées par autrui est un enjeu majeur de la communication. En effet, l'émotion est considérée comme une construction sociale dans la mesure où elle est constituée d'un ensemble de réponses prescrites qui se réfèrent à des normes ou à des attentes partagées [3 ; 15]. Ainsi, les expressions émotionnelles permettent d'établir, maintenir,

cesser une relation sociale ou de donner du sens sur la nature de la relation établie [12]. Elles favorisent ou dissuadent également l'expression de certains comportements chez autrui.

Aujourd'hui, une nouvelle forme d'interaction se développe. L'homme est amené de plus en plus fréquemment à interagir avec des agents conversationnels animés. Ce sont des personnages animés de telle sorte qu'ils sont capables d'aider un utilisateur du web dans la résolution de diverses tâches : environnements virtuels de formation, commerce électronique, réservation de billets...

Alors que plusieurs études ont déjà évalué la perception par des utilisateurs d'expressions d'émotions dans des modalités isolées (parole, expressions faciales ou postures), nous connaissons encore peu de choses sur la manière dont ces expressions sont perçues lorsqu'elles sont combinées. Le système que nous avons utilisé pour effectuer cette étude est un prototype développé dans le cadre du projet Affective Avatars¹. Il permet d'animer en temps réel des agents animés capables d'exprimer les émotions détectées dans la voix de l'utilisateur. La voix de l'utilisateur sert donc d'interface de contrôle de l'expressivité de l'avatar.

Pour se faire, nous décrivons dans la section 2 les modules du prototype Affective Avatar qui ont été utilisées. La section 3 présente l'évaluation du système développé et plus particulièrement l'examen de la cohérence perçue entre l'expressivité vocale de l'utilisateur et l'expressivité corporelle de

¹ Cette étude a été soutenue par le projet "ANR-07-TLOG-Affective Avatars – Cap Digital"

l'avatar qui le représente. Enfin nous terminons par une conclusion et les perspectives qu'ouvre cette étude.

2. Description du système

Le système proposé permet d'animer un agent virtuel à partir des émotions qui sont détectées dans la voix de l'utilisateur. Deux modules interviennent dans le traitement des émotions.

2.1. Module de détection des émotions

En parole, on distingue en général deux types d'information : les informations linguistiques et paralinguistiques. Concernant ces dernières, les différents indices peuvent être regroupés en trois catégories : les indices suprasegmentaux (fréquence fondamentale, énergie, durée, contour intonatif), les indices segmentaux (articulation, durée) et les indices intra-segmentaux (qualité vocale). Parmi les propriétés considérées comme pertinentes pour caractériser les manifestations émotionnelles dans la voix, les indices prosodiques sont les plus communément utilisés. De plus, les modèles les plus souvent employés pour la détection des émotions sont les SVM (Support Vector Machine), les GMMs (Mélange de Gaussiennes), les HMMs (Modèles de Markov cachés) les kNN, (K plus proches voisins), et les arbres de décision. Les résultats obtenus par ces différents classificateurs sont souvent comparables [4].

Toutefois, à l'heure actuelle, les études sur la détection des émotions dans la voix présentent un certain nombre de limites. Tout d'abord, la plupart d'entre elles se sont focalisées sur un ensemble minimal d'émotions : positives et négatives [14], émotion vs. état neutre [5], frustration/colère vs. neutre/amusé [2]. Il existe très peu de systèmes de détection des émotions qui vont au-delà de la classification de 4 à 5 émotions de base (très différentes). Les taux de détection deviennent très bas sur des émotions nuancées. Peu d'études sont menées avec des données issues de corpus enregistrés dans des contextes réels. Or l'étude de voix spontanées montre que les émotions sont souvent mélangées ou masquées dans les interactions

naturelles et que les émotions spontanées sont produites à travers des indices de différents niveaux : acoustiques, prosodiques, des « affect bursts » (rire, toux, souffle), disfluences, indices lexicaux et dialogiques. Ainsi le challenge des études sur les comportements émotionnels à partir de données naturelles réside dans la multiplicité d'indices à la fois linguistiques et paralinguistiques témoignant des émotions. Il s'agit donc de construire des modules de détection des émotions en temps réel robustes aux variabilités d'expressivité prosodiques et linguistiques des locuteurs.

Pour cela, il faut utiliser un grand nombre de paramètres pour caractériser les émotions et développer des corpus de données naturelles où l'apprentissage se réalise à partir d'un grand nombre de voix (une centaine dans le cadre de cette étude). Ainsi, deux corpus ont été collectés et annotés pour servir de base d'apprentissage (3 heures de parole) et un corpus de test de 15 locuteurs a été collecté. Le schéma d'annotation utilisé est composé de catégories verbales et de dimensions. Ce schéma permet d'annoter des émotions complexes. Les modèles (arbres de décisions, SVM) ont été construits soit à partir des dimensions (comme l'activation), soit à partir de catégories émotionnelles (3, 4 ou 5 émotions).

Les résultats montrent peu de différences entre la librairie OpenEar (plateforme de référence pour la détection des émotions). Un protocole 10-folds cross-validation «Speaker – Indépendant» a été mis en place. L'évaluation finale a été menée avec un corpus de test collecté auprès de 15 personnes [6 ; 16 ; 17].

2.2. Module de l'expressivité non verbale de l'agent

De nombreuses études ont considéré les expressions faciales d'émotion [10] mais aussi posturales [9 ; 18]. Plusieurs travaux se sont appuyés sur ces données afin de nuancer les comportements expressifs de l'avatar en fonction des émotions exprimées [1 ; 11 ; 7]

Cet état de l'art nous a servi à définir des spécifications des expressions non-verbales (expressions faciales, posture, mouvement) de

quelques émotions de base (colère, peur, tristesse, joie, surprise). Ces spécifications informelles ont permis la conception des expressions faciales et animations posturales pour un personnage virtuel. Ces animations combinant expressions faciales et posturales ont ensuite été validées par des tests perceptifs [8].

2.3. Interfaçage des deux modules

Les postures et expressions du visage qui ont été créées sont rassemblées dans une bibliothèque de gestes expressifs. Les transitions entre animations sont gérées à travers un graphe d'états découpé en "zones émotionnelles". On est ainsi en mesure de piloter automatiquement un personnage à partir de la voix en synchronisant à la fois des gestes associés aux périodes de parole et de silence et les gestes expressifs associés aux émotions.

3. Evaluation

Les différents composants du système ayant été évalués indépendamment [16 ; 6 ; 17 ; 8], nous connaissons leurs points forts et leurs limites. Ainsi concernant l'expressivité non verbale, les expressions de colère, joie, et tristesse avec deux niveaux d'intensité sont bien reconnues aussi bien quand les modalités sont présentées de manière indépendante, expression du visage seule ou expression posturale seule que lorsque les expressions faciale et posturale sont combinées. La limite principale concernant l'expressivité non verbale est que la dimension temporelle n'a pas été prise en compte. L'objectif de l'étude que nous présentons dans cette section vise à étudier comment l'agent est perçu quand les deux principaux composants sont présents (expressivité audio et expressivité visuelle). Nous avons donc réalisé un test perceptif à partir d'enregistrements audio-vidéo du système pour mesurer les émotions perçues au cours de ces clips vidéo et la qualité perçue de l'interaction audio/vidéo.

3.1. Méthode

3.1.1. Participants

20 participants dont 8 femmes et 12 hommes ont passé l'expérimentation. La moyenne d'âge

de l'échantillon est de 27 ans. Les participants ont un niveau d'étude compris entre bac +4 et doctorat. Ils sont tous issus d'un milieu informatique.

3.1.2. Matériel

Pour évaluer l'intégration des deux composants, il nous a semblé intéressant de manipuler le *contenu sémantique du message verbal* de l'agent et le *niveau de concordance des différentes modalités expressives*. Ainsi nous nous intéressons à l'impact de la présence d'un indice sémantique dans le message verbal de l'agent sur la perception du sujet. La perception est-elle favorisée lorsque l'agent dit « je suis un agent expressif en colère » (présence d'un indice sémantique) par rapport à la condition où l'agent dit « je suis un agent virtuel expressif » (absence d'un indice sémantique). La seconde variable manipulée correspond à la concordance entre les modalités expressives et nous nous interrogeons sur l'impact de la non concordance sur la perception. Lorsque la voix, le message et l'expression non verbale ne concordent pas, quel impact cela a-t-il sur la perception ? Le tableau 1 présente l'ensemble des conditions expérimentales proposées au cours de cette étude.

Type de message linguistique	Concordance entre l'audio et la vidéo	Audio	Vidéo
Présence d'un indice sémantique (triste, en colère, joyeux)	concordants	colère	colère
		joie	joie
		triste	triste
	non concordants	colère	joie
		triste	neutre
		colère	neutre
		joyeux	triste
		joyeux	colère 1
		joyeux	colère 2
Absence d'indice sémantique (je suis un agent virtuel expressif)	concordants	colère	colère
		joie	joie 1
		joie	joie 2
		triste	triste
	non concordants	neutre	neutre
		joie	colère
		triste	colère

TAB. 1 Ensemble des conditions expérimentales

Ainsi le matériel expérimental se compose de 16 enregistrements vidéo dont 7 enregistrements où l'agent exprime un message linguistique sémantiquement neutre (*absence d'indice sémantique sur l'émotion exprimée*) et 9 enregistrements où l'agent véhicule un

message linguistique sémantiquement émotionnel (*présence d'un indice sémantique sur l'émotion exprimée*).

Parmi les clips vidéo où il n'y pas d'indice sémantique dans le message linguistique de l'agent, 5 animations sont *concordantes* : Le locuteur avait l'intention d'exprimer de la joie dans sa voix, le système a bien détecté cette émotion, l'agent exprime donc cette émotion par son visage et sa posture. Deux autres animations sont *non concordantes*. Le locuteur avait l'intention d'exprimer une émotion mais c'est une autre émotion qui a été détectée par le système et c'est donc cette autre émotion que l'agent exprime par ses expressions faciales et ses postures.

Parmi les clips vidéo où il y a un indice sémantique dans le message linguistique de l'agent, 3 animations sont *concordantes*. 6 autres animations sont *non concordantes*.

3.2. Hypothèses et mesures

3.2.1. Les émotions perçues

La première série de mesures concernait les émotions perçues par les sujets au cours de chaque séquence vidéo. Les participants devaient indiquer dans quelle mesure ils percevaient de la colère, de la peur, de la surprise, de la tristesse, de la joie, du dégoût et de la neutralité pour chaque clip vidéo via des échelles de Likert de 0 « pas du tout » à 4 « énormément ».

Notre but était ici de tester si la perception émotionnelle des utilisateurs pour les animations où la concordance audio/vidéo est vérifiée, est moins confuse que pour les animations où la concordance audio/vidéo n'est pas respectée. On s'attendait à ce que ce résultat soit d'autant plus vrai que les sujets étaient confrontés à un message émotionnel.

Pour les animations non concordantes, il s'agissait de voir l'impact des modalités expressives sur la perception. Le message linguistique sémantique est-il prévalent sur les messages facial et corporel ? Que se passe-t-il en l'absence d'indice sémantique dans le message linguistique ? Les sujets perçoivent-ils

le décalage audio vidéo ou s'appuient-ils exclusivement sur les messages facial et corporel pour élaborer leur jugement ?

3.2.2. La qualité de l'interaction audio/vidéo

Le second groupe de mesure visait à recueillir des informations sur le ressenti subjectif des sujets quant à la qualité de la synchronisation audio/vidéo. 8 items ont été élaborés à cette fin. Il s'agissait de savoir si les sujets percevaient et étaient gênés par le léger décalage entre l'audio et la vidéo induit par la détection automatique des émotions.

Nous nous attendions à ce que les sujets lorsqu'ils étaient confrontés aux animations *non concordantes*, jugent l'interaction audio vidéo de moins bonne qualité que lorsqu'ils étaient confrontés à des animations où la *concordance* audio vidéo était respectée. Nous faisons l'hypothèse que ce résultat serait d'autant plus vrai que les sujets étaient dans la condition « présence d'un indice sémantique ».

3.3. Protocole

Les participants ont passé le test et répondu individuellement au questionnaire. La moitié des participants a été confrontée à la version «*présence d'un indice sémantique*» concordante ou non, le personnage indique dans son message l'émotion qu'il est censé exprimer. L'autre moitié a été confrontée à la condition ne donnant *pas d'indice dans le message verbal* de l'agent sur son état émotionnel. Le premier groupe de sujets a donc visionné 9 vidéos et le second 7.

Les passations se sont réalisées de manière individuelle sur un écran d'ordinateur. A l'issue de chaque vidéo, les participants devaient répondre à un questionnaire.

3.4. Résultats

3.4.1. Résultats concernant la perception émotionnelle

Perception émotionnelle des animations dans la condition : présence d'un indice sémantique

Le tableau 2 présente une vue synthétique de la perception émotionnelle moyenne de chaque

animation *cohérente* (la même information émotionnelle est censée être véhiculée par les différentes modalités expressives : voix, contenu linguistique et expression non verbale). Les résultats ont ensuite été analysés via des T de Student.

Animations			Emotions Perçues						
émotion détectée dans la voix	émotion indiquée dans le message linguistique	émotion exprimée par l'agent via les expressions faciales et posturales	joie	peur	colère	tristesse	dégout	surprise	neutralité
colère	colère	colère	0,7	0	3	0	0,9	0	0,5
joie	joie	joie	2,8	0	0,1	0	0	0,1	0,9
tristesse	tristesse	tristesse	0,3	0	0	3,1	0,9	0	0

TAB. 2 Perception émotionnelle moyenne des animations cohérentes en «présence d'un indice sémantique»

Les scores de perception pouvaient varier de 0 à 4. On observe que la perception émotionnelle de ces trois animations est conforme à ce qui était attendu. Les animations qui ont une certaine cohérence multimodale sont donc relativement bien perçues par les sujets.

Concernant *les animations non cohérentes* (des informations différentes sont véhiculées dans la voix, le message linguistique et les expressions non verbales de l'agent), le tableau 3 présente les résultats concernant la perception émotionnelle moyenne.

Animations				Emotions Perçues						
Número Animation	émotion détectée dans la voix	émotion indiquée dans le message linguistique	émotion exprimée par l'agent via les expressions faciales et posturales	joie	peur	colère	tristesse	dégout	surprise	neutralité
1	joie	colère	joie	2,70	0,30	0,20	0,00	0,10	0,60	0,50
2	neutre	colère	neutre	0,60	0,20	0,40	2,00	0,90	0,00	0,90
3	colère 1	joie	colère 1	1,70	0,00	1,70	0,00	0,50	0,10	0,30
4	colère 2	joie	colère 2	1,00	0,00	2,30	0,00	0,70	0,20	0,40
5	triste	joie	triste	0,30	0,00	0,00	3,10	1,60	0,00	0,30
6	neutre	triste	neutre	0,70	0,20	0,50	2,10	1,10	0,10	0,80

TAB. 3 : Perception émotionnelle moyenne des animations non cohérentes en «présence d'un indice sémantique»

Lorsque la correspondance entre les différentes modalités expressives n'est pas respectée, c'est-à-dire que l'émotion indiquée dans le message linguistique de l'agent n'est pas détectée dans la voix et donc n'est pas exprimée par le comportement non verbal de l'agent, cela ne favorise pas la perception émotionnelle.

Ces résultats suggèrent que les sujets ne s'appuient pas principalement sur le message linguistique comme on aurait pu s'y attendre

mais élaborent plutôt leur perception à partir de l'expression non verbale de l'agent. Ces résultats semblent conforter nos spécifications concernant l'expression non verbale des agents qui est bien reconnue.

Enfin concernant les animations 3 et 4, où l'agent se déclare « joyeux » et exprime deux intensités différentes de colère (colère 1 et colère 2), les sujets perçoivent à la fois de la colère et de la joie. Cependant les sujets attribuent plus de colère et moins de joie à l'animation 4 qui exprime d'un point de vue non verbal une colère d'intensité plus forte. Ce résultat confirme nos spécifications concernant l'intensité expressive de la colère.

Perception émotionnelle des animations dans la condition : absence d'un indice sémantique

Les sujets sont confrontés ici à un agent qui n'indique pas dans son message linguistique l'émotion qu'il exprime, mais son expression est concordante à l'émotion que le locuteur avait l'intention d'exprimer. Le tableau 4 présente ces résultats. L'émotion la plus perçue au cours de chaque clip vidéo est l'émotion cible, c'est-à-dire l'émotion traitée dans la voix et exprimée par le comportement non verbal de l'agent. Les sujets, même en l'absence d'un indice sémantique, sont capables de percevoir les émotions traitées par le système et véhiculées par l'agent via son expressivité non verbale.

Cependant, les résultats soulignent également le fait que pour toutes les animations, les sujets perçoivent un peu de joie. Les deux animations censées véhiculer des émotions négatives (les animations 1 et 4 du tableau 4) sont perçues comme véhiculant un peu de joie. Ce résultat peut être compris par le fait que l'agent au cours de chaque clip vidéo exprime une émotion via une séquence d'animations composées de l'animation neutre puis de l'animation expressive pour terminer par une expression neutre. L'apex de l'animation globale correspond à l'animation expressive. Toutefois, l'animation neutre n'est pas tout à fait neutre, en effet l'agent utilisé est censé être un agent assistant donc plutôt convivial et son expression

neutre est donc empreinte d'un léger sourire. Aussi lorsqu'il n'y pas d'indice sémantique dans le message linguistique de l'agent, les sujets semblent porter beaucoup d'attention à l'ensemble de son expressivité non verbale, traiter plus d'informations et de ce fait percevoir plus d'émotions.

Animations			Emotions Perçues						
Numéro Animation	émotion détectée dans la voix	émotion exprimée par l'agent via les expressions faciales et posturales	joie	peur	colère	tristesse	dégoût	surprise	neutralité
1	Colère	Colère	1,00	0,10	2,50	0,00	0,10	0,40	0,20
2	Joie	Joie 1	3,50	0,00	0,00	0,00	0,00	0,40	0,60
3	Joie	Joie 2	3,20	0,00	0,20	0,10	0,10	0,90	0,10
4	Triste	Triste	1,30	0,20	1,20	2,40	0,70	0,10	0,60
5	Neutre	Neutre	1,20	0,00	0,00	0,10	0,00	0,00	2,60

TAB. 4 : Perception émotionnelle moyenne des animations cohérentes pour la condition «absence d'un indice sémantique»

Lorsque le système n'a pas détecté l'émotion que le locuteur avait l'intention d'exprimer, la perception émotionnelle en est affectée (tableau 4). La mauvaise détection du système induit de la confusion et le message émotionnel véhiculé par l'agent n'est pas clairement interprété par les utilisateurs qui perçoivent un mélange d'émotions.

Animations				Emotions Perçues						
Numéro Animation	émotion intentionnée dans la voix	émotion détectée dans la voix	émotion exprimée par l'agent via les expressions faciales et posturales	joie	peur	colère	tristesse	dégoût	surprise	neutralité
1	joie	colère	colère	1,80	0,00	2,50	0,20	0,30	0,00	0,40
2	triste	colère	colère	0,00	0,20	2,30	1,40	0,90	0,00	0,80

TAB. 5 Perception émotionnelle moyenne des animations **incohérentes** en «l'absence d'un indice sémantique»

Comparaison de la perception émotionnelle moyenne des animations en condition présence d'un indice sémantique vs en condition absence d'un indice sémantique

L'analyse des résultats via des T de Student concernant la comparaison des deux conditions révèle que la perception émotionnelle ne varie pas en fonction des conditions expérimentales pour les animations cohérentes. Quelle que soit la nature du message linguistique (absence vs présence d'un indice sémantique), les sujets rapportent une perception émotionnelle équivalente.

Concernant les animations non cohérentes, la comparaison ne peut porter que sur les animations combinant de la joie dans la voix et de la colère dans l'expression faciale et posturale de l'agent. C'est la seule combinaison qui se retrouve dans les deux conditions expérimentales. A nouveau quelle que soit la condition expérimentale, la perception de ces animations incongruentes ne varie pas. La présence de l'indice sémantique ne favorise pas la perception du message véhiculé dans la voix.

3.4.2. Résultats concernant la perception de la qualité de la synchronisation audio/vidéo

La qualité de la synchronisation est calculée à partir des scores obtenus aux 8 items évaluant la qualité de l'interaction audio/vidéo. Cet indice peut prendre des valeurs comprises entre 0 et 32. Un score proche de 32 rapporte une évaluation très satisfaisante de la synchronisation audio/vidéo et un score proche de 0 au contraire montre que les sujets sont très critiques à l'égard de la synchronisation audio/vidéo.

Perception de la qualité de la synchronisation audio-vidéo pour les animations cohérentes en présence d'un indice sémantique

Le tableau 6 présente ces résultats pour les trois animations qui véhiculent la même émotion via les différentes modalités expressives. Seule l'animation de la joie à un niveau de satisfaction relativement satisfaisant. La qualité moyenne est de 18,80 ce qui est supérieure à la moyenne (16). Pour les deux autres animations, les sujets jugent la qualité de la synchronisation comme peu satisfaisante.

Animations			Qualité Globale
Emotion détectée dans la voix	Emotion indiquée dans le message linguistique	Emotion exprimée par l'agent via les expressions faciales et posturales	
colère	colère	colère	11,00
joie	joie	joie	18,80
tristesse	tristesse	tristesse	13,90

TAB.6 Qualité moyenne de la synchronisation audio/vidéo pour les animations cohérentes en «présence d'un indice sémantique»

Perception de la synchronisation audio-vidéo

pour les animations cohérentes en l'absence d'un indice sémantique

Le tableau 7 présente ces résultats pour les cinq animations qui véhiculent la même émotion via les différentes modalités expressives. Trois animations sur cinq obtiennent un niveau de satisfaction relativement satisfaisant (supérieur à la moyenne).

Animations			Qualité Globale
Numéro Animation	Emotion détectée dans la voix	Emotion exprimée par l'agent via les expressions faciales et posturales	
1	Colère	Colère	17,50
2	Joie	Joie 1	18,30
3	Joie	Joie 2	14,60
4	Triste	Triste	10,80
5	Neutre	Neutre	18,90

TAB.7 Qualité moyenne de la synchronisation audio/vidéo pour des animations cohérentes en «l'absence d'un indice sémantique»

Comparaison de la qualité moyenne perçue de la synchronisation audio-vidéo des animations en condition présence d'un indice sémantique vs en condition absence d'un indice sémantique

Les deux sections précédentes semblent suggérer que les sujets ont jugé le système comme plus performant (c'est-à-dire comme proposant des animations correspondant mieux aux expressions verbales) quand l'agent ne proposait pas d'indice sémantique dans son message linguistique que l'inverse. Les différentes animations cohérentes en condition «absence d'un indice sémantique» obtiennent de meilleurs scores en qualité de synchronisation que les animations cohérentes en condition «présence d'un indice sémantique». Ces résultats ont été analysés via des T de Student afin de vérifier que ces différences sont significatives. Concernant les animations relatives à la joie et à la tristesse, aucune différence ne s'observe. La présence ou l'absence d'un indice émotionnel dans le message verbal de l'agent n'influe pas sur la qualité perçue de la synchronisation audio-vidéo pour ces animations. Ce résultat ne se vérifie pas pour les animations relatives à la colère. Les sujets jugent la qualité meilleure

dans la condition « absence d'indice » que dans la condition « présence d'indice ».

Ces résultats nous questionnent quant à la temporalité du comportement non verbal associé à l'expression multimodale de l'émotion. Il est possible que selon la nature de l'émotion, la temporalité de l'expression varie. Le système de détection/animation développé, nécessite une certaine latence entre la détection des émotions dans la voix de l'utilisateur et l'expression non verbale associée. Cette latence devient moins frappante quand le discours du locuteur ne se limite pas à une seule phrase ce qui n'est pas le cas dans le protocole expérimental réalisé. Ces résultats suggèrent qu'on attendrait une expression rapide pour un événement relatif à de la colère alors qu'à des événements reliés à la tristesse ou la joie, l'expression pourrait arriver plus tardivement. Ceci expliquerait pourquoi les sujets jugent la qualité de l'interaction audio/video plus mauvaise lorsqu'ils sont confrontés à un agent affirmant exprimer de colère et mettant relativement du temps à l'exprimer que lorsqu'ils sont face à un agent n'affirmant pas quelle émotion il exprime. Le message linguistique empreint d'un indice sémantique concernant l'expression de l'agent induirait plus d'attente chez les utilisateurs concernant l'expressivité de l'agent et cela aurait un impact différent sur la qualité perçue selon les émotions traitées.

4. Discussion et perspectives

Cette évaluation perceptive concernant l'intégration du système de détection émotionnelle et d'expression émotionnelle non verbale suggère que les sujets perçoivent mieux les animations émotionnelles quand un indice sémantique est présent dans le message verbal de l'agent. Cependant, cet indice dégrade la perception de la synchronisation audio vidéo en créant probablement chez l'utilisateur des attentes quant à l'expressivité de l'agent.

Cette étude a été réalisée auprès d'un petit échantillon et devrait être reproduite afin de confirmer ces premiers résultats. Enfin, elle questionne au sujet de la temporalité, de la

dynamique de l'expression non verbale et de la qualité de mouvement (Wallbott 1998) et les relations entre la forme des gestes et les émotions [13].

Références

- [1] André, E., Klesen, M., Gebhard, P., Allen, S., Rist, T. (2000). *Exploiting Models of Personality and Emotions to Control the Behavior of Animated Interface Agents* In J. Rickel (Ed.), Workshop on "Achieving Human-Like Behavior in Interactive Animated Agents". 3-7.
- [2] Ang, J., Dhillon, R., Krupski, A., Shriberg, E. & Stolcke, A. (2002.). *Prosody-Based Automatic Detection of Annoyance and Frustration in Human-Computer Dialog*. Proceedings of International Conference on Spoken Language Processing, 2037-2040.
- [3] Averill, J. R. (1980). *A constructivist view of emotion*. In R. Plutchick & H. Kellerman (Eds), *Emotion, theory, research, and experience: Theories of emotions*. New York: Academic Press.
- [4] Batliner, A. Hacker, Ch. Steidl, S., Nöth, E., S. D'Arcy, Russell, M & Wong, M. (2004). *"You stupid ting box"- children interacting with the AIBO robot: A cross-linguistic emotional speech corpus*. 4th international Conference on Language Resources and Evaluation, 171-174.
- [5] Batliner, A., Fisher, K., Huber, R., Spilker, J. & Noth, E. (2003). *How to Find Trouble in Communication*. Speech Communication, 40, 117-143.
- [6] Brendel, M., Zaccarelli, R., Devillers, L. *"Building a system for emotions detection from speech to control an affective avatar"*, LREC 2010
- [7] Buisine, S., Abrilian, S., Niewiadomski, R., Martin, J.-C., Devillers, L. and Pelachaud, C. (2006). *Perception of Blended Emotions: from Video Corpus to Expressive Agent*. IVA'2006, 93-10
- [8] Clavel, C., Plessier, J., Martin, J.C., Ach, L., Morel, B. (2009). *Combining Facial and Postural Expressions of Emotions in a Virtual Character*. IVA'09, 287 - 300.
- [9] Coulson, M. (2004). *Attributing emotion to static body postures Recognition accuracy Confusions and viewpoint dependence*. Journal of nonverbal behavior, 28 (2), 117-139
- [10] Ekman, P. and Friesen, W. (1975). *Unmasking the face. A guide to recognizing emotions from facial clues*, Prentice-Hall Inc., Englewood Cliffs, N.J.
- [11] Gillies, M., Crabtree, I. B. and Ballin, D. (2006). *Individuality and Contextual Variation of Character Behaviour for Interactive Narrative*. AISB Workshop
- [12] Kaiser, S. Wehrle, T. & Schenkel, K. (2009) *Expression faciale des émotions*. In D. Sander, & K. Scherer, (Eds). *Traité de psychologie des émotions*. Paris : Dunod, 77-108
- [13] Kipp, M., Martin, J.-C. (2009). *Gesture and Emotion: Can basic gestural form features discriminate emotions?* ACII-09
- [14] Lee, C.M., Narayanan, S. & Pieraccini, R. (2002). *Combining acoustic and language information for emotion recognition*. Proc. International Conference on Spoken Language Processing, 873-876.
- [15] Parkinson, B. (1996). *Emotions are social*. British Journal of Psychology
- [16] Rollet, N., Delaborde, A., Devillers, L., (2009). *"Protocol CINEMO: The use of fiction for collecting emotional data in naturalistic controlled oriented context"*, ACII 2009
- [17] Schuller, B., Rollet, N., Zaccarelli, R., Devillers, L. (2010). *"CINEMO: A French Spoken Language Resource for Complex Emotions: Facts and Baselines"* LREC
- [18] Wallbott, H. G. (1998). *Bodily expression of emotion*. European Journal of Social Psychology (28), 879-896.

Expressions Posturales de Tendances à l'Action

Ning. Tan

Céline. Clavel

Mathieu. Courgeon

Jean-Claude. Martin

ntan@limsi.fr

celine.clavel@lim
si.fr

courgeon@limsi.fr

martin@limsi.fr

LIMSI-CNRS,

BP 133, 91403 Orsay Cedex, France

Résumé :

L'un des enjeux de la recherche sur les agents virtuels est de spécifier l'expression multimodale des émotions complexes. Les recherches sur l'expression d'émotion se sont plus particulièrement focalisées sur certaines modalités : expressions faciales, regard ou gestes. Au contraire, les expressions posturales de l'émotion ont reçu moins d'attention, de même que les autres composantes de l'émotion, telle que les tendances à l'action. Dans cet article, nous proposons une approche pour concevoir et évaluer les expressions posturales des tendances à l'action. Nous avons conçu les expressions posturales des tendances à l'action en utilisant la plateforme d'animation de personnages virtuels MARC. Ces postures ont été spécifiées à partir des données issues de la littérature et des annotations manuelles d'un corpus vidéo de tendances à l'action. 5 couples d'images d'expressions posturales ont été évalués en termes de tendances à l'action et de catégories d'émotion discrète [20]. Les sujets ont reconnu les expressions posturales des tendances à l'action suivantes: *attend*, *disappear from view* et *exuberant*. La perception des catégories d'émotion a été également conforme aux prédictions. Ces résultats suggèrent que les expressions posturales peuvent être utiles pour exprimer les tendances à l'action.

Mots-clés : Tendance à l'Action, Emotion, Corpus Multimodal, Posture, Agent Virtuel

Abstract :

One of the challenges of virtual character research is the specification of reliable and discriminative features of complex emotions in multiple modalities. Whereas expressions of different emotion categories in facial expressions, gaze, and gestures were explored in several studies, postural expressions of emotion and other components of emotions received less attention. In this paper, we propose an approach for designing and evaluating postural expressions of action tendencies. We designed postural expressions of several action tendencies using MARC, our virtual character animation platform. These postures were informed by data from the literature completed with the manual annotations of an exploratory video corpus. A decoding study tested 5 pairs of static pictures of postural expressions in terms of action tendencies as well as discrete emotion categories [20]. Subjects reliably

recognized the postural expressions of the following action tendencies: attending, disappear from view and exuberant. Perception of emotion categories was also consistent with psychological predictions about action tendencies. These results suggest that postural expressions can be useful to express action tendencies in virtual characters.

Keywords: Action Tendency, Emotion, Multimodal Corpus, Posture, Virtual Agent

1 Introduction

An *emotion* can be seen as “an episode of interrelated, synchronized changes in five components in response to an event of major significance to the organism” [29]. These five components are: the cognitive processing (function of evaluation of the objects and of the events), the physiological changes (function of system regulation), the action tendencies (function of preparation and direction of action), the motor expression (function of communication of reaction and behavioral intention) and the subjective feeling (function of monitoring of internal state and organism-environment interaction). *Action tendencies* play a role in the preparation and the direction of action at the motivation level. On the one hand, they consist of a readiness to execute action: they involve the activation of a class of possible responses selected out of a human's response repertoire. On the other hand, they consist of “readiness to achieve or maintain a given kind of relationship with the environment” [29]. Thus, they involve orientation toward a present or forthcoming state.

Although action tendency is an important component of emotions, few studies have explored how they can be simulated and expressed in virtual agents. Yet, this would allow an agent to reveal affect and enable the

user to possibly predict its future actions. Indeed, posture was observed to be a relevant modality to express self reported aversion, openness, irritation, happiness, and self-confidence [34]. Ekman suggests that body actions might provide information about the intensity of the felt emotion [17]. Wallbott observed some discriminative features of emotion categories in both posture and movement quality [34]. Video corpora and motion capture have been collected to get detailed information on the postural expression of emotion but these data were not analyzed in terms of action tendencies [4] [5].

In the study that we present in this paper, we explore whether postural expressions might convey discriminative and reliable features of action tendencies. Our long-term motivation is to consider the perception of body postures and movements in social scenes featuring multiple interacting characters. A secondary more pragmatic goal is to extend the posture library of our MARC platform with postural expressions of emotions. These research goals raise several questions: How to code action tendencies? How to collect postural expressions of action tendencies? Which methodology should be defined to investigate how postural expressions of action tendencies are perceived?

2 Related Work

2.1 Action Tendencies

According to Frijda “action tendencies are states of readiness to execute a given kind of action, and thus an action tendency is defined by its end result aimed at or achieved” [19]. In other words, “action tendencies are states of readiness to achieve or maintain a given kind of relationship with the environment”. They aim at achieving changes to the actual situation. Action tendencies are thus “readiness to engage in or disengage from interaction with some goal object in some particular fashion”. During social interactions, action tendencies are perceivable structure that is intended to indicate an implicit quality about the people or their environment.

Different action tendencies correspond to different emotions. Some emotion categories might be better represented by appraisals (jealousy, surprise, hope), and others by action tendencies (disgust, despair, anxiety, anger) [20].

2.2 Postural Expressions in Humans and Virtual Characters

Darwin regarded emotions as predispositions to act adaptively and suggested that specific body movements are associated with each emotional state [12]. Ekman suggested that judgments based on the face lead to higher recognition accuracy and higher agreement among viewers when considering the emotion category [16], whereas the body would play a role in conveying the intensity of the emotion [17]. The role of the body in conveying emotions has been reinforced by recent neurosciences studies for example the role of body postures over facial expressions in cases of incongruent affective displays [13].

Ekman and Friesen suggested that static postures are more likely to convey gross affect (e.g. liking), whereas movements of the body are more likely to convey specific emotions [18]. Wallbott observed discriminative features of emotions both in static body postures and in the movement quality [34]. Harrigan suggested that the body posture and the body actions help decoding more subtle facial and vocal affects [22]. Bull reported several studies about postural expressions [7]. Yet, he did not investigate the postural expressions of action tendencies. Finally, the Component Process Model of emotion considers the causal sequences of bodily changes [30]. Scherer predicts a series of bodily posture shifts following a process of multi-level sequential checking. Grammer et al. [21] propose an approach combining traditional behavior observation and modern anthropometric analysis. A set of photographs was taken with 100 actors exposing body postures. The body postures were then transferred to a 3D virtual environment. 50 subjects judged the affective states of people

who expressed postures. Five categories were mainly represented (aversion, openness, irritation, happiness and self-confidence). The authors generalized the body postures per affect category, and reconstructed postures using a virtual environment. The reconstructed postures were validated in a perception test with 50 observers.

Digital corpora have also been recorded to collect detailed information about the expression of emotions in different modalities. The Geneva Multimodal Emotion Portrayal database contains video recordings of portrayals of 18 emotions acted by 10 professional actors [3]. The AffectMe project collected a library of postural expressions of emotions using a VICON motion-capture system [6]. 13 subjects acted anger, fear, happiness and sadness. Their movements were recorded at 32 points of the body. 111 affective postures were collected and presented to 5 subjects who had to judge these postures according to emotion categories and affective dimensions (valence, arousal, potency and avoidance). The authors observed that some postural features provide information about some affective dimensions. For example, openness of the body seemed to be important to arousal dimension of affective [6].

Postural expression was also investigated in virtual characters. Ballin et al. [2] designed the Demeanour system for animating the postural and spatial behaviors of virtual agents for expressing affiliation and dominance. Luo et al. found that adding lower body motion did improve the believability of virtual characters. They added lower body movements including movements co-occurring with gesture, and idle movements that do not occur with gesture [27]. Egges designed an animation model based on a combination of motion synthesis from motion capture and a statistical analysis of prerecorded motion clips [15]. His model includes idle motions that are generated by sequencing prerecorded motion segments organized in a graph. To control the path followed through the graph, he used an

activation-evaluation emotion space, in which different emotional states are represented using a 2-dimensional vector.

Previous work has considered to which extent postural expressions convey emotions, but not its relation to action tendency. In the following sections, we propose an approach to represent action tendency in postural expressions.

3 Design

3.1 Corpus of Postural Expressions of Action Tendencies

TAB. 1 – The action tendencies selected for the study [20] [33].

Action tendencies	Description
Exuberant	I wanted to move, be exuberant, sing, jump, and undertake things.
Attending	I wanted to observe well, to understand, or I paid attention.
Antagonistic	I wanted to oppose, to assault; hurt or insult.
In command	I stood above the situation; I felt I was in command; I held the ropes.
Disappear from view	I wanted to sink into the ground, to disappear from the Earth, not to be noticed by anyone.

As explained in the previous section, the literature is scarce about postural expressions of action tendencies. Thus, we decided to collect additional knowledge by exploiting our PERMUTATION video corpus [9]. This corpus contains 100 video samples of American television comedy-drama series for a total duration of 42 minutes. It contains samples of rich emotional interactions in a variety of social situations. The corpus was presented to 200 participants, who attributed action tendencies based on the nonverbal behaviors of the characters. We selected five action tendencies (Tab. 1) which received a high agreement among subjects [9] (average ratings between 3.4 and 4 on a scale between 0 and 4).

We segmented and annotated the postures

observed in these clips using the Anvil tool [23]. We defined a multi-level scheme for annotating posture based on the Posture Scoring System [8] and on the annotation scheme for conversational gestures [24]. Our scheme includes the annotation of head, shoulder trunk and arm [32]. The annotations resulted in a set of specifications at different levels of body posture. For example, these annotations indicated that the actors tend to lean their shoulders forward while they were judged as “wanted to move, be exuberant, sing, jump, and undertake things”. Postures were annotated in 14 videos clips by two coders.

3.2 The MARC Virtual Agent Platform

MARC (Multimodal Affective and Reactive Characters) is the platform that we develop for conducting experimental studies using interactive virtual characters [11]. MARC includes a Body Posture editor. Edited postures are added to a posture sequence in a time line. The postures that we informally specified out of the corpus annotations described in the previous section were saved in an XML library. Animations of the virtual character are then triggered by BML messages at runtime [25]. MARC also features the possibility to include a 3D environment and multiple characters in the same scene.

4 Exploratory Study

Our previous experiment using the PERMUTATION corpus showed that people reliably assign some action tendencies to the video clips [9]. In the present study, we explore if subjects are able to perceive the action tendencies conveyed by postural expressions annotated in the corpus and replicated on the virtual agent Mary in a social interaction setting.

4.1 Stimuli

We designed 12 postures for our study. For each of the five selected action tendencies (Tab. 1), two postural expressions were specified out of the corpus annotations so that we could keep the best recognized posture.

More than two postural expressions would require a larger video corpus of expressions of action tendencies. One pair of neutral postures was also designed to control the experiment.

<i>Action tendencies</i>	<i>Postures</i>	
Attending		
Disappear from view		
In command		
Exuberant		
Antagonistic		
Neutral		

FIG.1 – The target postural expressions designed for the selected action tendencies.

We designed a 3D scene simulating a social interaction between two female characters standing face to face. Only the character facing the camera, displays emotional postures. The other one is visible from the back and is displayed only for helping the interpretation of the social interaction. It stands in a neutral postural expression. Fig.1 presents the set of target postural expressions designed for the selected set of action tendencies. The face was blurred to inhibit the influence of a neutral facial expression on users' perception. Although dynamics of postural expression is quite relevant for emotions [34], it requires additional explorations as we suggest in the future directions section. That is why we used static pictures in this study.

4.2 Participants

20 subjects (7 female, 13 male, aged 21-60, 79% European, 16% African, and 5% Asian) completed a paper questionnaire and supplied information about age, gender and culture.

4.3 Procedure

A questionnaire was set up to assess the extent to which subjects recognize the action tendencies in our pictures of postures. The questionnaire contains two parts. In the first part of the questionnaire we asked subjects to select the pictures of postures that match the written description of each action tendency. The written descriptions are those proposed by Frijda. The order of presentation of the action tendencies was randomized. For each action tendency, we showed 6 images: 2 target images (i.e. supposed to characterize the action tendency specified in the written description), and 4 distracting images (randomly selected out of the other 10 postural expressions). Subjects were instructed to choose one or two images representing the sample description (which means the second response is optional).

In the second part of the questionnaire, subjects were asked to assign one emotional label to each of the 12 postures. For each posture, they had to select one label out of the following list: sadness, joy, anger, anxiety, surprise, fear, irritation, shame, contempt,

guilt, disgust, pleasure, despair, pride. This list of emotion labels was used by Frijda in his study about the relation between action tendencies and emotion categories. This forced-choice method was used in several studies ([14], [17], [18], [31], [34]), in which they found that subjects agree above chance levels. Since different action tendencies correspond to different emotions [20], we expected that our results would be consistent with predictions drawn from psychology studies about action tendencies.

4.4 Results

Posture Recognition. Three action tendencies (*attending*, *disappear from view*, *exuberant*) had success rates above 50% for one of their two target images. The postural expressions designed for the action tendency *in command* was not correctly recognized (69% of the answers referred to other postures that the two target images). Tab. 2 provides for each action tendency the attribution rate for the different postures.

Emotion Attribution. Except for *in command*, we observed that the emotional attribution is quite consistent across the two target postures for each action tendency. The same emotions are attributed to both target posture for each action tendency. For *attending*, subjects mainly ascribed negative emotions like anxiety, anger, or irritation. These results showed that subjects assign related emotions to postures that are supposed to express *attending*. Regarding *antagonist*, we found the same pattern of results. Subjects mainly ascribed irritation, contempt, anger, and anxiety. Regarding *disappear from a view*, subjects mainly attributed shame and guilt. These two emotions are very close to each other in the circumplex model of Russell [28]. For *exuberant*, the emotions *joy*, *pride* and *pleasure* were selected to describe the two target postures. These emotions are all positive. Finally only for *in command*, we observed contradictory emotional results for the two target animations. The first target animation is perceived as conveying *pleasure*

and *pride*. The second target image is perceived as conveying anger.

TAB. 2 – Attribution rates for target and distracting postures for each action tendency.

	Target posture #1	Target posture #2	Distracters (other postures than targets)	TO-TAL
Attending	52%	10%	38%	100%
Disappear from view	67%	33%	0%	100%
Exuberant	54%	21%	25%	100%
In Command	25%	6%	69%	100%
Antagonistic	39%	27%	33%	100%

4.5 Discussion

We observe three main results: 1) Postures of three action tendencies (*attending*, *disappear from view*, *exuberant*) received a high recognition rate, 2) The emotional attribution is quite consistent across the two target postures for each emotion excepted *in command*, 3) The action tendency for which the error rate is the highest (*in command*) corresponds to postures for which subjects attributed few and different emotions categories. The results also suggest that the two parts of the questionnaire provide complementary information and are relevant from a methodological point of view.

Our results are consistent with the results reported by Frijda [20]. Only the results for action tendency “*in command*” are not consistent with those of Frijda [20]. The target postures designed for this action tendency are not well recognized. Subjects selected other postures as representing the action tendency *in command*. Given that the action tendency attribution is not reliable, it is not surprising that the emotional attributions appear inconsistent compared to those of Frijda.

Moreover, there are several possible

explanations to lower recognition rates for the second target image. First, in the recognition task, as we explained above, we required the subjects to choose one or two images representing the sample description, which means the second response is optional. This explains why one target image has been more associated to the sample description than the other one.

5 Conclusion and Future Directions

In this paper, we explored the relations between postural expressions and action tendencies. We defined an experimental methodology, and applied it to an exploratory perception study. Results suggest that postures expressed in a social interaction scene might provide information about action tendencies.

We will also work on the synchronization and the dynamics of different body members (probably using subtle slight movements) and consider discriminative features of movement quality as observed by Wallbott [34]. This will enable us to design and test animations of our virtual characters instead of the static pictures of the study presented in this paper. We also intend to use video and motion capture corpora to collect knowledge on such a dynamics. Frijda examined the complementary contributions of appraisal and action tendencies. These contributions might be conveyed by facial expressions of appraisal and postural expressions of action tendencies. We intend to evaluate the perception of congruent and incongruent nonverbal expressions in terms of emotions and cognitive appraisals (combining dynamic facial expressions and body postures). Few studies addressed the emotion perception based on an overview of both facial and bodily dynamic expressions[10]. Finally, posture might also be relevant for expressing other related social behaviors such as attitudes: a mental state of readiness for action that is organized through experience, impacting a directive and dynamic influence on behavior [1].

Acknowledgments

The work described in this paper was supported by the French Agence Nationale de la Recherche (ANR) under the OTIM project (ANR BLAN08-2_349062).

References

- [1] Allport, G. 1935. *Attitudes*. In *A Handbook of Social Psychology*, Clark University Press, Worchester, MA 798-844.
- [2] Ballin, D., Gillies, M., Crabtree, I.B. 2004. A Framework For Interpersonal Attitude And Non-Verbal Communication In *Proceeding of improvisational Visual Media Production, First European Conference on Visual Media Production* IEE. London, UK.
- [3] Bänziger, T., Pirker, H., Scherer, K.R. 2006. GEMEP - GENEVA Multimodal Emotion Portrayals: A corpus for the study of multimodal emotional expressions. In *Proceeding of Workshop "Corpora for research on emotion and affect". 5th International Conference on Language Resources and Evaluation LREC'06*. 15-19.
- [4] Bänziger, T., Scherer, K.R. 2007. Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus. In *Proceeding of Affective Computing & Intelligent Interaction ACII'07*. 476-487.
- [5] Bianchi-Berthouze, N., Kim, W.W. 2007. Darshak Patel: Does Body Movement Engage You More in Digital Game Play? and Why? . In *Proceeding of Affective Computing & Intelligent Interaction ACII'07*. 102-113.
- [6] Bianchi-Berthouze, N., Kleinsmith, A. 2003. *A categorical approach to affective gesture recognition*. Connection Science, 15(4), 259-269.
- [7] Bull, P. 1978. *The interpretation of posture through an alternative methodology to role play*. British Journal of Social & Clinical Psychology, 17(1), 1-6.
- [8] Bull, P.E. 1987. *Posture and gesture*. Pergamon Press, Elmsford, NY US.
- [9] Clavel, C., Martin, J.-C. 2009. PERMUTATION: A Corpus-Based Approach for Modeling Personality and Multimodal Expression of Affects in Virtual Characters, In *Proceeding of 13th International Conference on Human-Computer Interaction HCI*. 11, 211-220.
- [10] Clavel, C., Plessier, J., Martin, J.-C., Ach, L., Morel, B. 2009. Combining Facial and Postural Expressions of Emotions in a Virtual Character, In *Proceeding of 9th International Conference on Intelligent Virtual Agents IVA'09*, 287-300.
- [11] Courgeon, M., Martin, J.-C., Jacquemin, C. 2008. MARC: a Multimodal Affective and Reactive Charater. In *Proceeding of Workshop on Affective Interaction on Natural Environment*.
- [12] Darwin, C. 1955. *The expression of the emotions in man and animals*. Philosophical Library, Oxford England.
- [13] de Gelder, B., Snyder, J., Greve, D., Gerard, G., Hadjikhania, N. 2004. Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. In *Proceedings of the National Academy of Sciences*, 101(47), 16701-16706.
- [14] De Silva, P.R., Bianchi-Berthouze, N. 2004. *Modeling human affective postures: an information theoretic characterization of posture features*. Computer Animation and Virtual Worlds, 15, 269-276.
- [15] Egges, A. 2006. *Real-time Animation of Interactive Virtual Humans*. PhD Thesis, University of Geneva, Geneva.
- [16] Ekman, P. 1964. *Body position, facial expression, and verbal behavior during interviews*. The Journal of Abnormal and Social Psychology, 68(3), 295-301.
- [17] Ekman, P. 1965. *Differential communication of affect by head and body cues*. Journal of Personality and Social Psychology, 2(5), 726-735.

- [18] Ekman, P., Friesen, W.V. 1967. *Head and body cues in the judgment of emotion: a reformulation*. Perceptual And Motor Skills, 24(3), 711-724.
- [19] Frijda, N.H. 1987. *Emotion, cognitive structure, and action tendency*. Cognition and Emotion, 1(2), 115-143.
- [20] Frijda, N.H., Kuipers, P., ter Schure, E. 1989. *Relations among emotion, appraisal, and emotional action readiness*. Journal of Personality and Social Psychology, 57(2), 212-228.
- [21] Grammer, K., Fink, B., Oberzaucher, E., Atzmüller, M., Blantar, I., Mitteroecker, P. 2004. *The representation of self reported affect in body posture and body posture simulation*. Collegium antropologicum, 28(2), 159-174.
- [22] Harrigan, J.A., Rosenthal, R., Scherer, K.R. 2005. *The new handbook of methods in nonverbal behavior research*. Oxford University Press, New York, NY US.
- [23] Kipp, M. 2001. Anvil - A Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology* Eurospeech. 1367-1370.
- [24] Kipp, M., Neff, M., Albrecht, I. 2007. *An annotation scheme for conversational gestures: how to economically capture timing and form*. Language Resources & Evaluation, 41, 325-339.
- [25] Kopp, S., Krenn, B., Marsella, S., Marshall, A.N., Pelachaud, C., Pirker, H., Thórisson, K.R., Vilhjálmsón, H.H. 2006. *Towards a Common Framework for Multimodal Generation: The Behavior Markup Language*. In *Proceeding of 6th International Conference on Intelligent Virtual Agents IVA'06*.
- [26] Lazarus, R.S. 1991. *Progress on a cognitive-motivational-relational theory of emotion*. American Psychologist, 46(8), 819-834.
- [27] Luo, P., Kipp, M., Neff, M. 2009. *Augmenting Gesture Animation with Motion Capture Data to Provide Full-Body Engagement*, In *Proceeding of 9th International Conference on Intelligent Virtual Agents IVA'09*, 405-417.
- [28] Russell, J.A. 1980. *A circumplex model of affect*. Journal of Personality and Social Psychology, 39(6), 1161-1178.
- [29] Scherer, K.R., Perrig, W.J., Grob, A. 2000. *Emotional expression: A royal road for the study of behavior control*. In *Control of human behavior, mental processes, and consciousness: Essays in honor of the 60th birthday of August Flammer.*, Lawrence Erlbaum Associates Publishers, Mahwah, NJ US 227-244.
- [30] Scherer, K.R., Schorr, A., Johnstone, T. 2001. *Appraisal considered as a process of multilevel sequential checking*. In *Appraisal processes in emotion: Theory, methods, research.*, Oxford University Press, New York, NY US 92-120.
- [31] Schouwstra, S.J., Hoogstraten, J. 1995. *Head position and spinal position as determinants of perceived emotional state*. Perceptual & Motor Skills, 81(2), 673.
- [32] Tan, N., Ferre, G., Tellier, M., Cela, E., Morel, M.A., Martin, J.C., Blache, P. 2010. *Multi-level Annotations of Nonverbal Behaviors in French Spontaneous Conversations*, in *In Proceeding of Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, LREC Malta.
- [33] Tcherkassof, A. 1996. *La perception des expressions émotionnelles faciales*. PhD Thesis, Université de Paris X, Nanterre.
- [34] Wallbott, H.G. 1998. *Bodily expression of emotion*. European Journal of Social Psychology, 28, 879-896.

A la recherche d'indices de culture et/ou de langue dans les micro-événements audio-visuels de l'interaction face à face

Rosario Signorello¹ Véronique Aubergé^{1,2} Anne Vanpé^{1,2} Lionel Granjon¹ Nicolas Audibert³
{rosario.signorello, veronique.auberge, anne.vanpe, lionel.granjon}
@gipsa-lab.grenoble-inp.fr ; nicolas.audibert@gmail.com

¹ GIPSA-Lab, Département Parole et Cognition, UMR CNRS 5216, Grenoble, France

² Laboratoire d'Informatique de Grenoble, équipe GETALP, UMR CNRS 5217, Grenoble, France

³ Laboratoire Informatique d'Avignon, Université d'Avignon, Avignon, France

Résumé :

Cet article présente une étude sur l'identification perceptive, par des juges français et italiens, de la langue/culture, le degré de confiance accordé à ce choix, et, en parallèle, du degré d'extraversion. Les jugements sont faits sur des micro-événements acoustiques et visuels non langagiers (ex. bruits de bouche, *grunts*, *bursts*, *fillers* et interjections) de 6 sujets français, placés en interaction personne-machine. Ces événements ont été choisis et ordonnés en listes de 24 stimuli selon un contrôle croissant de la prosodie (à partir d'un non contrôle) et de leur distance des sons phonologiques. Pour chaque modalité de passation (Audio seul ou Visuel seul ou Audio-Visuel), des groupes différents de 15 juges, de chaque nationalité, ont été interrogés. Les résultats montrent une tendance, des juges des deux nationalités, à identifier les sujets comme français. Toutefois, alors qu'en Visuel seul le jugement est surtout fondé sur l'aspect des sujets (choix des juges invariable dès le début de l'expérimentation), en Audio seul et en Audio-Visuel l'apparition du contrôle prosodique correspond à un « point de stabilité » dans les choix langue/culture.

Mots-clés : culture, langue, extraversion, contrôle prosodique, micro-gestualité, interjections, *bursts*, *grunts*.

Keywords: culture, language, extraversion, prosody, micro-gestures, interjections, bursts, grunts.

1 Introduction

Beaucoup de travaux [ex. 2] ont été consacrés aux relations étroites entre langue et culture, aussi bien dans leurs aspects théoriques que dans des champs applicatifs comme l'enseignement des langues destiné à un usage pertinent en situation socio-culturelle. Les sciences du langage, mais aussi tous les travaux d'ingénierie qui s'adressent à la mise en situation réelle d'agents virtuels ou de robots humanoïdes, pointent sur un niveau de

granularité révélant des structures et des modèles caractéristiques du « rôle social » tenu dans chaque contexte situationnel : un banquier, un enseignant, un assistant, une mère s'adressant à son bébé, etc. Bien entendu la description de ces rôles (plus ou moins emprunts d'invariants) se construit par le filtre de la langue/de la culture dans lesquelles ils se définissent.

L'interaction est un processus dynamique contextualisé par quand, où, pourquoi, à qui et qui est en cours de communication langagière [1]. Lorsque l'on observe une interaction en train de se dérouler les caractéristiques comportementales de la personne (ses tendances génériques et ses spécificités de personnalité, de langue et de culture) deviennent certainement des paramètres fondamentaux pour décrire, comprendre, simuler et prédire ses actes de communication dans cette entité globale de l'interaction. L'interactant est en communication permanente, pendant ses tours de parole (bien sur) mais aussi en dehors ceux-ci, par des gestes et micro-gestes audibles et/ou visibles qui se réalisent dans ce que nous désignons comme *Feeling of Thinking* (FoT) : états mentaux (ex. la compréhension, la concentration), attitudes, affects sociaux (ex. le doute, la politesse), émotions (ex. la tristesse) et humeurs (ex. le stress) [3, 8, 9].

2 Objectifs de l'étude

Les comportements communicatifs d'un humain, en interaction face à face, mettent en

jeu autant ses composantes personnelles biophysiques et mentales (ex. sa personnalité) que le rôle sociétal qui situe son interaction (cela avec le filtre prégnant de sa langue et de sa culture). Il est supposé ici que l'observation des événements langagiers d'une personne, au cours d'une interaction, ne pourra être expliquée globalement (et donc à terme prédite ou reconnue automatiquement par des agents virtuels, par exemple) que si sont pris en compte tous ces degrés de granularités emboîtés : sa personnalité, son rôle sociétal au moment de l'interaction, sa langue/sa culture.

L'expérience perceptive qui est présentée ici est une étude très préliminaire qui se donne comme objectif de poser deux questions fondamentales :

- A. Peut-on discriminer deux langues/cultures, mêmes proches, à travers des micro-événements, audibles et visibles, mais non langagiers (bruits de bouche, *grunts*, *bursts*, *fillers*, interjections) du FoT ? ;
- B. Existe-t-il une interaction perceptive entre des faits de langue/culture et des indices de personnalité ?

Le problème posé ici est lié à la nature de ces micro-événements non langagiers : sont-ils de valeur informative variée par rapport aux critères d'identification de la langue/culture ? L'hypothèse forte est que la nature langagière des événements sonores est construite par le contrôle de la prosodie (i.e. contrôle « volontaire » de la durée, de la qualité et/ou de la rythmicité), et que les premiers faits de langage apparaissent ainsi bien avant que ces sons ne soient doublement articulés (controverse introduite par [8], inspirée par [7]). Nous voulons donc tenter de mettre en évidence cette possible identité langagière (ou au minima culturelle) de certains des micro-événements, et de surcroît, mesurer perceptivement lesquels commencent à contenir de telles informations *vs.* ceux qui n'ont pas de statut langagier. Nous avons retenu, comme paire à contraster, le français et l'italien, de typologie linguistique très proche et culturellement très familières. Ces deux

langues/cultures font partie de la même aire euro-méditerranéenne. En outre, il existe une opposition actuelle et claire des langues nationales (français *vs.* italien). Enfin, des stéréotypes de représentations collectives (ex. le trait d'une supposée extraversion dans l'interaction langagière des habitants proches de la méditerranée) sont bien installés et véhiculés par la croyance populaire.

3 Protocole expérimental

3.1 Participants

90 juges¹, 45 français et 45 italiens, ont participé au test de perception. Chaque groupe comptait 3 sous-groupes de 15 pour chaque modalité de passation du test : Audio seul (noté A), Visuel seul (noté V) et Audio-Visuel (noté AV). Chaque juge a passé le test dans une seule modalité. La faible connaissance de l'autre culture était une condition nécessaire (ex. ne pas être bilingue français-italien et/ou avoir vécu dans les deux pays).

3.2 Choix et organisation des stimuli

Les micro-événements étudiés ici sont extraits du corpus d'expressions émotionnelles authentiques Sound Teacher recueilli sur des sujets français grâce à la plate-forme magicien d'Oz E-Wiz [1], induisant des états affectifs « prédits », dans une tâche prétexte d'apprentissage phonétique des sons des langues du monde. Après l'expérimentation, les sujets impliqués auto-annotaient leurs propres enregistrements, désignant ainsi « naïvement » un large spectre d'états mentaux, cognitifs, attitudeaux et plus largement affectifs, qui a été regroupé en FoT [3]. Un étiquetage des formes acoustiques et gestuelles a été mené ensuite avec une méthodologie empruntée à l'éthologie [9]. Cela a été fait afin de ne pas inscrire, *a priori*, les données observées dans une théorie de la gestualité ou de la face, et de vérifier précisément, *a posteriori*, la capacité de ces

¹ Moyenne d'âge 26,33 ans. Juges français : 35 femmes, 10 hommes ; Juges italiens : 15 femmes, 35 hommes.

modèles à prendre en compte aussi bien les valeurs annotées par le sujet (qui dépassent largement le cadre des émotions) et les formes acoustiques et gestuelles ([3] puis [8]). Ces dernières sont désignées comme « subtiles » par certains auteurs [4, 5] sans pour autant couvrir l'étendue de celle présentes et validées dans Sound Teacher (depuis [1] à [9]).

Les sujets étaient tous français : 3 femmes (F-M, F-S, F-T) et 3 hommes (M-J, M-N, M-R), choisis pour une variabilité supposée (mais non vérifiée objectivement) de leur extraversion. 24 stimuli par sujet ont été sélectionnés, par expertise auditive de leur articulation et par analyse du signal acoustique [8]. Ils ont été ensuite ordonnés selon leur qualité prosodique croissante, cette prosodie étant supposée informative. Dès sons non phonétiques (supposés sans contrôle prosodique volontaire, ex. bruits de bouches), puis, avec un contrôle de plus en plus précis (faisant apparaître des sons non phonétiques avec contrôle prosodique, ex. expiration, qualité de voix) jusqu'aux phonèmes de la langue qui n'entrent pas dans le processus du morpho-lexique (ex. *fillers*, interjections). Ci-dessous en détail, du 1^{er} au 24^{ème} stimulus :

- 1^{er} stimulus $\rightarrow$ 10^{ème}/14^{ème} stimulus² : pas de contrôle prosodique volontaire des articulateurs (relâchement articulateur, respiration, avalement de la salive, interaction langue-lèvres, friction, aspiration salive) ;
- 10^{ème}/14^{ème} stimulus $\rightarrow$ 13^{ème}/20^{ème} stimulus : inspiration ou expiration contrôlée dans le temps, raclement de gorge contrôlé au niveau de leur durée, clic ou occlusions non phonétiques soit répétés soit suivis d'une expiration ou d'une inspiration, contrôle de qualité de voix dans le temps, etc. ;
- 13^{ème}/20^{ème} stimulus $\rightarrow$ 22^{ème}/24^{ème} stimulus : *fillers* (ex. « *mmmh* », « *euh* », « *fff* »),

- sons phonétiques (voyelles, syllabes) qui peuvent être ou non allophones du français ;
- 22^{ème}/24^{ème} stimulus $\rightarrow$ 24^{ème} stimulus : interjections (ex. « *ben* », « *ouh la la* », etc.), sons phonologiques du français.

3.3 Choix des modalités de présentation

Un point clé de cette étude a été de vérifier si les informations traitées visuellement (les traits statiques d'aspect vestimentaire, capillaire, d'accessoires ainsi que les traits dynamiques de gestualité, liés soit à la personnalité soit à des facteurs culturels et langagiers de communication) sont en rapport avec celles traitées acoustiquement (en particulier la dynamique de la prosodie testée ici). Il était donc fondamental que les facteurs d'apprentissage dans une modalité n'influencent pas une autre modalité. C'est pourquoi chaque modalité (A, V ou AV) a été jugée par des groupes différents (de 15 juges chacun), et ceci pour chacune des deux nationalités d'appartenance.

Pour vérifier, en particulier, que les signaux acoustiques (modalité A) contiennent, à partir d'un certain seuil, des indices informationnels propres aux formes acoustiques, il a été nécessaire de présenter les mêmes stimuli, dans le même ordre, en modalité V. La modalité AV, quant à elle, a permis de mesurer comment sont intégrées ces informations.

3.4 Procédure

L'interface du test affichait, dans sa première page, les instructions détaillées avec les consignes pour le bon déroulement du test. Les 24 stimuli de chaque sujet étaient présentés aux juges dans l'une des modalités possibles de présentation des stimuli (choisie au hasard par l'expérimentateur). L'ordre d'affichage des sujets était aléatoire et différent pour chaque juge. Après la présentation de chaque stimulus, un écran de réponse s'affichait :

- une première échelle de réponse permettait aux juges de choisir la langue/culture (soit français, soit italien) et donner en même

² Si deux valeurs sont utilisés (ex. 10^{ème}/14^{ème} stimulus) c'est parce qu'il n'a pas été possible de trouver exactement des stimuli aux caractéristiques articulatoires et acoustiques similaires pour chaque sujet. Toutefois, l'ordre supposé croissant a été respecté pour chaque sujet.

temps la valeur (entre 1 et 10) de la confiance accordée à leur choix ;

- Les juges déplaçaient ensuite un autre curseur sur un axe limité par « introversion » et « extraversion », sans autre valeur quantitative que le milieu noté « 0 ». L'objectif était ici de mesurer si les stimuli donnaient, ou non, aux juges « matière » à évaluer les sujets sur les traits de personnalité introverti vs. extraverti, mais aussi de vérifier si ce jugement était influencé par le jugement de langue/culture et s'il évoluait avec la progression des stimuli.

Dans chaque modalité de passation, afin de mesurer indirectement un potentiel facteur d'apprentissage, un jugement « a posteriori » global pour le sujet locuteur a été demandé aux juges, à la fin de la série de stimuli et sans support de stimulus. Cette mesure leurs a été demandée pour le choix catégoriel italien vs. français (avec le degré de confiance associé à cette réponse) et pour le degré d'extraversion perçu.

Ce test a été passé devant un ordinateur (résolution écran 1440x900 pixels) à l'aide d'un casque Sennheiser HD 25-13. Résolution des stimuli : Vidéo 720x540 pixels ; Audio 44.1 KHz, 16 bits.

4 Résultats

4.1 Discrimination

L'un des buts de notre étude était de savoir si les juges arrivaient à percevoir une différence, en terme d'appartenance culturelle, entre les sujets : ont-ils été reconnus comme des français ? Y-a-t'il eu un effet de la nationalité des juges sur ce choix ? La modalité de passation du test a-t-elle influencé la perception du sujet ? Et si oui, de quel sujet en particulier ? Les sujets ont-ils été eux-mêmes source d'influence sur la perception des juges ?

Globalement, pour les deux nationalités de juges (toutes modalités confondues), les sujets sont perçus à 58,4% comme français et à 41,6% comme des italiens ($\chi^2(1)=15.79$,

$p<.001$). Les juges français classifient mieux : ils identifient les sujets comme français à 60,2% ($\chi^2(1)=267.26$, $p<.001$) tandis que les juges italiens à 56,7% ($\chi^2(1)=116.80$, $p<.001$). Les réponses « a posteriori » confirment ces tendances : juges français 78,9% ($\chi^2(1)=90.13$, $p<.001$), juges italiens 72,6% ($\chi^2(1)=55.12$, $p<.001$).

4.2 Croisement des facteurs « modalité », « sujet » et « choix de langue/culture »

4.2.1 Modalité A (Audio seul)

Comme le montre le TAB. 1a les juges français, en moyenne, reconnaissent tous les sujets comme des français, significativement mieux que le hasard ($p<.001$, sauf M-R ($p<.05$)). Ce résultat est confirmé par le jugement « a posteriori » ($p<.001$, sauf M-J ($p>.05$)). Au contraire, les juges italiens (TAB. 1b) reconnaissent 4 sujets comme italiens (M-J, M-N, M-R ($p<.001$) et F-S ($p<.05$); résultats pour F-M et F-T pas significatifs, $p>.05$). Ces tendances changent radicalement en considérant les moyennes des réponses « a posteriori » : presque tous les sujets sont reconnus comme des français ($p<.001$), mis à part M-J qui continue à être perçu comme italien avec un pourcentage assez élevé (73,3%, $p<.001$).

Grace a ces résultats nous pourrions déjà avancer l'hypothèse qu'il y a de l'information dans l'audio mais qu'elle n'est pas suffisante pour discriminer la langue/culture (au moins pour les juges italiens qui semblent subir un biais cognitif, ex. l'attraction vers leur propre langue/culture d'appartenance).

4.2.2 Modalité V (Visuel seul)

Dans cette modalité, deux sujets (F-M et F-S) sont perçus avec des tendances similaires par les juges français et italiens. Les juges français perçoivent quatre sujets (F-T, M-J, M-N, M-R) comme des français ($p<.001$) et deux sujets (F-M et F-S) comme des italiens ($p<.001$) (cf. TAB. 1a). Les juges italiens reconnaissent deux sujets (M-N et M-R) comme des français

($p < .001$) et deux sujets (F-M ($p < .001$) et F-S ($p < .05$)), comme des italiens. Les résultats pour les sujets F-T et M-J ne sont pas significatifs ($p > .05$) (cf. TAB. 1b). Dans les réponses « a posteriori » les juges français aussi bien que les juges italiens perçoivent quatre sujets (F-T, M-J, M-N et M-R) comme des français ($p < .001$) et le sujet F-M comme italien ($p < .001$). Le résultat pour le sujet F-S n'est pas significatif ($p > .05$).

D'après ces résultats similaires nous pourrions avancer une hypothèse sur l'aspect général des locuteurs et/ou des *a priori* stéréotypiques d'aspect (traits somatiques, indices esthétiques, etc.) partagés par les deux cultures testées. Toutefois, il est nécessaire d'analyser l'évolution des réponses durant la progression des listes de stimuli (cf. 4.3), pour savoir si ces résultats doivent être reliés à l'aspect global du sujet ou aux stimuli dynamiques présentés en liste (soit par apprentissage quantitatif, soit par qualité informative – l'hypothèse « prosodique » étant donnée sur l'acoustique mais elle pourrait aussi être visible).

4.2.3 Modalité AV (Audio-Visuel)

Les juges français reconnaissent, en moyenne, trois sujets comme français (F-M, M-N, F-S, $p < .001$) et un sujet comme italien (M-R, $p < .05$) (résultats pour F-T et M-J non significatifs, $p > .05$) (cf. TAB. 1a). Par contre, dans les réponses « a posteriori » les sujets sont tous reconnus comme français ($p < .001$). Les juges français ne sembleraient pas suivre les performances de la modalité A. Nous interprétons cela par le fait que l'audio serait porteur d'informations et que les juges français pensent y puiser des indices. Par contre, avec l'ajout de la vidéo, ces indices ne seraient pas suffisamment discriminants pour diriger les juges français vers le bon choix. Les indices visuels auraient donc affecté leur perception.

Au contraire, les juges italiens ne reconnaissent que des français parmi les sujets (cf. TAB. 1b) (les réponses « a posteriori » confirment ces tendances). Ces résultats légitiment donc l'interprétation formulée pour

la modalité A : l'information contenue dans l'acoustique (insuffisante si présentée seule), ajoutée à celle contenue dans la visuelle, fait basculer les résultats vers la bonne réponse.

Comme pour la modalité V, ces interprétations doivent être confrontées à l'évolution des performances des juges dans la progression des stimuli (cf. 4.3).

TAB. 1 : Matrices de scores de choix langue/culture en pourcentage par modalité de passation (A, V, AV), choix langue/culture (*fra* vs. *ita*) et sujet. Analyse statistique : t-tests échantillons uniques, valeurs de test 50%.
Niveaux de significativité : * = $p < .001$;
^ = $p < .05$; ° = choix non significatif.

1a - Juges Français

	A		V		AV	
	fra	ita	fra	ita	fra	ita
F-M	77,2*	22,8	27,8	72,2*	73,3*	26,7
F-S	75*	25	42,8	57,2*	63,1*	36,9
F-T	66,1*	33,9	66,7*	33,3	51,4°	48,6°
M-J	66,4*	33,6	62,2*	37,8	49,7°	50,3°
M-N	68,1*	31,9	73,1*	26,9	70,8*	29,2
M-R	48,6°	51,4°	55,8*	44,2	44,7	55,3^
Total	66,9*	33,1	54,7°	45,3°	58,8*	41,2

1b - Juges Italiens

	A		V		AV	
	fra	ita	fra	ita	fra	ita
F-M	53,3°	46,7°	39,2	60,8*	73,3*	26,7
F-S	44,2	55,8^	44,4	55,6^	68,1*	31,9
F-T	48,9°	51,1	55^	45	66,4*	33,6
M-J	26,9	73,1*	55^	45	59,2*	40,8
M-N	40,3	59,7*	60,6*	39,4	66,1*	33,9
M-R	40,8	59,2*	91,9*	8,1	87,2*	12,8
Total	42,4	57,6*	57,7*	42,3	70,0*	30,0

4.3 Point de stabilité de la réponse : degré de confiance et choix langue/culture

La FIG. 1 montre l'évolution du degré de confiance qu'ont les juges sur leurs réponses en fonction de leur exposition aux stimuli. Le degré de confiance est modérément mais

significativement corrélé à l'ordre de présentation des stimuli dans les 3 modalités de présentation et pour chaque nationalité des juges. Cette corrélation étant toutefois faible en condition V : juges français (A ($r=.516$), V($r=.095$), AV ($r=.086$); corrélations significatives à $p<.0001$); juges italiens (A ($r=.289$); V ($r=.103$), AV ($r=.352$), corrélations significatives à $p<.0001$).

La comparaison, au moyen de tests χ^2 , des choix langue/culture effectués par les juges pour différentes séquences de stimuli (définies en découpant l'ensemble des 24 stimuli en 2, 3 ou 4 parties égales) confirme que cette augmentation du degré de confiance s'accompagne d'une évolution significative de la distribution des réponses.

Nous avons voulu ensuite vérifier si une stabilisation des réponses se manifestait à l'apparition des stimuli contrôlés prosodiquement. Nous avons donc extrait pour chaque triplet (juge, sujet, modalité) le « point de stabilité » (cf. TAB. 3), défini comme l'indice du stimulus à partir duquel le juge ne modifie plus son choix de langue/culture.

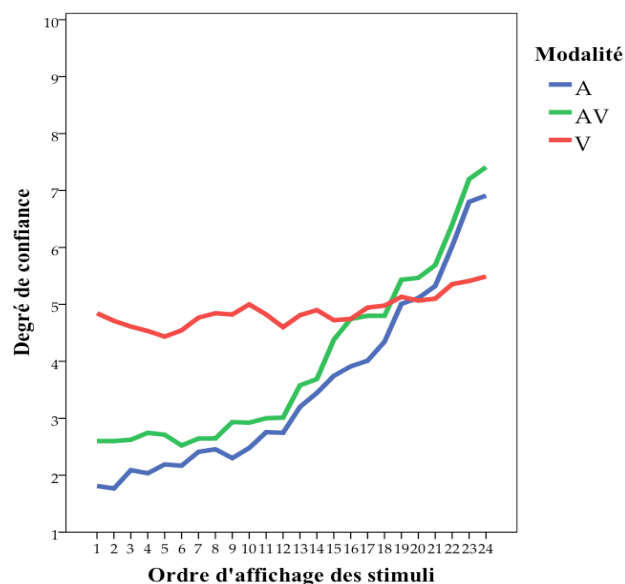
TAB. 3 : « Point de stabilité » du choix langue/culture par modalité de passation (A, V, AV), nationalité des juges (FRA vs. ITA) et sujet.

	A		V		AV	
	FRA	ITA	FRA	ITA	FRA	ITA
F-M	11,67	16,40	6,53	12,40	9,80	13,40
F-S	11,33	16,20	8,60	7,33	11,87	8,87
F-T	15,53	15,80	8	10,20	16	11,27
M-J	15,93	10,93	7,80	7,13	16,47	13,87
M-N	18,07	15,13	10,07	10,07	10,20	12,93
M-R	18	19,07	5,73	5,73	17,73	8,93
Total	15,09	15,59	7,79	8,61	13,68	11,54

Comme le montre le TAB. 3 les réponses des juges se stabilisent vers le 15^{ème} stimulus pour la modalité A et un peu plus précocement, autour du 12^{ème}/13^{ème} stimulus, pour la modalité AV. Cette stabilité correspond à l'apparition des stimuli pré-phonétiques ou

phonétiques, contrôlés prosodiquement.

1a – Juges Français



1b – Juges Italiens

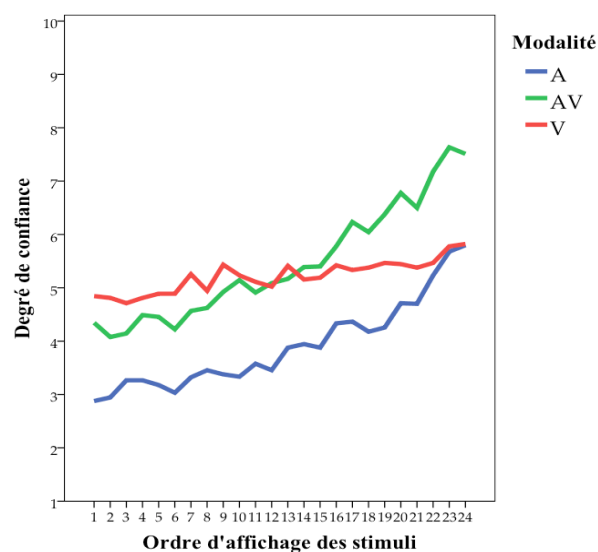


FIG. 1 : évolution du degré de confiance par rapport à l'ordre des stimuli (de 1 à 24) et à la modalité de passation du test (A, V, AV).

Par contre aucun point de stabilité significatif n'a été relevé dans la modalité V. Le choix de langue/culture, ainsi que le degré de confiance, restent stables pendant la présentation de la suite des stimuli. Chose qui va dans le sens de l'hypothèse prosodique : les informations visuelles seraient ainsi principalement liées à l'aspect global, les micro-événements n'apporteraient pas d'information dynamique.

Le point de stabilité diffère pour tout sujet : comme dit *supra* (cf. 3.2), les stimuli sélectionnés ne correspondent pas en terme de nature et de placement dans la suite d’affichage des stimuli.

4.4 Degré d’extraversion

Des patterns de réponse différents, par rapport à la modalité de passation du test et à la nationalité du juge, se sont avérés pour les sujets : F-M est toujours perçu comme introverti ; F-S est vu comme plus ou moins extraverti ; F-T comme extraverti en A, introverti en AV et V ; M-J comme introverti en A, comme extraverti en AV et V ; M-N comme extraverti en A et comme introverti en AV et V ; enfin, M-R comme extraverti en A et comme introverti en A et V (Analyse statistique : ANOVA à mesures répétées avec comme facteurs *inter-sujet* la modalité de passation du test ($F(8,21)=4.19$, $p<.001$) et la nationalité du juge ($F(4,90)=1.05$, $p=0.382$)). Les réponses « *a posteriori* » confirment ces tendances.

L’autre facteur qui semble influencer le jugement est l’ordre d’affichage des stimuli. Les juges français, plus que les juges italiens, montrent une tendance à noter comme introverti pendant les premiers stimuli, pour progresser relativement peu vers extraverti. Une analyse sur la régression linéaire de ces progressions montre qu’il existe une pente significative entre le début et la fin des stimuli (juges français ($F(1,64)=136.61$, $p<.001$) ; juges italiens ($F(1,64)=28.65$, $p<.001$)).

Enfin, nous n’avons pas constaté une interaction significative entre degré d’extraversion perçu et choix langue/culture.

5 Conclusions et perspectives

Ce travail préliminaire sur la discrimination de deux langues/cultures très proches, à partir de micro-événements – expressions du FoT – de 6 sujets français, a permis de considérer comme « valide » l’interrogation sur l’informativité langagière/culturelle de ces événements subtils, *a priori* non linguistiques.

La modalité de passation du test est l’un des facteurs qui a influencé le choix des juges. Ainsi, l’acoustique présentée seule ne s’est pas avérée suffisamment informative (en particulier pour les juges italiens), au point de ne pas leur permettre de faire le choix correct. Mais, la progression qualitative de l’information véhiculée et l’accumulation des stimuli contrôlés prosodiquement, ont permis aux juges de corriger leur choix. Les tendances de la modalité AV montrent cependant que (dans le cas des juges italiens) l’audio permet, en s’ajoutant à la vidéo, de faire basculer les résultats vers la bonne réponse. En particulier, dans ces deux modalités, l’ordre des stimuli a révélé un changement – au point de stabilité – dans le choix langue/culture. Cela précisément à l’apparition des stimuli donnés comme premier niveau informationnel selon l’hypothèse prosodique croissante de départ (supposée rapprocher ces éléments de faits de langue). D’autre part, les résultats de la modalité V nous montrent que les scores de reconnaissance des juges sont dus, vraisemblablement, à l’aspect des sujets. En visuel la réponse est précoce et correspond à un degré de confiance élevé. Les juges restent sur leur *a priori* de départ, correct ou non. D’ailleurs, aucun point de stabilité des réponses n’a été relevé, ce qui prouve donc que dès le départ les réponses sont stables (ou ne le sont jamais ce qui n’est pas le cas ici).

Enfin, le facteur d’extraversion, proposé *a priori* pour vérifier si ce trait « populaire stéréotypique » a une réalité tangible, ne s’est pas avéré pertinent. Cependant, il n’est pas possible d’en déduire que ce trait n’a pas eu d’influence sur la discrimination perceptive français vs. italien parce que aucune mesure « objective » sur la personnalité des sujets n’avait été menée en amont. Il serait donc intéressant, dans une expérience ultérieure, de mesurer *a priori* le degré d’extraversion des sujets et de les choisir dans un spectre large sur ce trait.

Pour compléter cette étude, et en tirer des conclusions plus solides, il est nécessaire de croiser symétriquement les juges et les sujets

dans chacune des deux langues/cultures. En outre, il faudrait reproduire cette expérience croisée sur des listes progressives de stimuli qui commencent précisément au premier niveau de contrôle prosodique tel qu'il a été validé dans cette expérience.

Ce type d'études des influences réciproques, en perception, des facteurs de langue/culture et de personnalité, pendant les interactions verbales, pointent sur l'importance de tenir compte de ces facteurs si l'on veut donner à un agent virtuel animé un comportement « culturellement » pertinent en situation, réaliste et écologique. Les « détails », non seulement dans la dynamique visuelle, mais également – c'est notre propos – dans la « périphérie » de la parole, commencent certainement à peine à nous montrer l'extrême importance que nous devons porter à sa pertinence cognitive et la nécessité de son réalisme virtuel. De plus, la présence ou l'absence de facteurs de langue/culture et de personnalité peuvent modifier et/ou activer des stratégies spécifiques dans l'agent virtuel pour la planification du discours et pour le choix des signaux de communication à utiliser [6]. Il s'agit de donner une capacité de perception « de haut niveau cognitif » à ces agents. Si l'agent a un petit modèle de l'utilisateur (i.e. s'il réussi à comprendre comme il est vraiment, d'où il vient, quel type de personnalité il a, etc.) il peut choisir, dans sa pré-planification de comment répondre à l'utilisateur, d'interagir d'une façon plutôt que d'une autre. Pour l'agent virtuel est une façon de décider : « en fonctionne de comme il/elle est, je lui dis ça ou pas ».

Références

- [1] Aubergé, V., Audibert, N., et Rilliard, A. (2006). « De E-Wiz à C-Clone. Recueil, modélisation et synthèse d'expressions authentiques ». In *Revue d'Intelligence Artificielle « Interactions émotionnelles »*, 20(4-5), pages 499-528.
- [2] Lévi-Strauss, C. (1958). *Anthropologie structurale*. Plon, Paris.
- [3] Loyau, F., et Aubergé, V. (2006). Expressions outside the talk turn: ethograms of the feeling of thinking. In *5th LREC*, pages 47-50.
- [4] Matsumoto, D. (s.d.). Subtle Expression Recognition Training (SubX) [méthode]. URL <http://www.humintell.com/subtle-expression-recognition-training> Dernière visite 17/10/2010.
- [5] Matsumoto, D., Consolacion, T., Yamada, H., Suzuki, R., Franklin, B., Paul, S., et al. (2002). American-Japanese cultural differences in judgements of emotional expressions of different intensities. In *Cognition & Emotion*, 16(6), 721-747.
- [6] Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., et De Carolis, B. (2005). GRETA. A Believable Embodied Conversational Agent. In Stock, O. and Zancarano, M., éditeurs, *Multimodal Intelligent Information Presentation*, pages 1–23. Kluwer, Belgique.
- [7] Rossi, M. (1977). L'intonation est la troisième articulation. In *Bulletin de la Société Linguistique de Paris*, LXXII, (1), pages 55-68.
- [8] Vanpé, A. et Aubergé, V. (2010a). Expressions des états mentaux/cognitifs et affectifs : Prosodie des productions vocales minimales – des grunt et burst à l'interjection. In *XXVIIIe JEP*.
- [9] Vanpé, A. et Aubergé, V. (2010b). Prosodie expressive audio-visuelle de l'interaction personne-machine. In *Technique et Science Informatiques, numéro spécial Agents Conversationnels Animés* 29.

NaviQuest : un outil pour naviguer dans une base de questions posées à un Agent Conversationnel

Ph. Suignard¹

Philippe.Suignard@edf.fr

¹EDF R&D

1, avenue du Général de Gaulle
92 141 Clamart Cédex – FRANCE

Résumé :

Cet article présente NaviQuest un outil mixant des techniques de visualisation et d'analyse textuelle pour naviguer dans une base de questions posées par les internautes à un agent conversationnel.

Mots-clés : Agent Conversationnel, Text mining, Visualisation

Abstract :

This paper presents a tool called NaviQuest, mixing visualization techniques and textual analysis, to browse a database of questions from users to a conversational agent.

Keywords: Conversational Agent, Text Mining, Vizualisation

1 Introduction

Présente depuis plusieurs années sur le site Bleu Ciel d'EDF¹, Laura est un agent-conversationnel qui répond aux questions posées par les internautes. Laura est réalisée par les sociétés Sémantia² pour la partie question-réponse et Cantoche³ pour l'animation du personnage. Sa base de connaissances métier est réalisée par EDF.

En plus des rapports mensuels fournis par Sémantia, nous avons voulu analyser finement les questions posées à Laura dans un but d'amélioration : savoir ce que disent les clients, comment ils le disent et comment ces demandes



évoluent avec le temps. Toutes ces questions constituent une source d'information importante pour connaître les préoccupations des clients, démarche qui peut se rapprocher du « crowdsourcing⁴ ».

Pour cela, nous avons voulu mixer des techniques de visualisation et d'analyse textuelle, ce qui constitue une préoccupation actuelle comme en témoigne la tenue du premier workshop⁵ sur le thème des « interfaces visuelles intelligentes pour le texte » ou les sessions « Text Visualization » des conférences InfoVis⁶ et « Text Analytics » des conférences Vast⁷. Tous ces efforts correspondent au besoin croissant d'analyser des données textuelles issues de mails, blogs, forum, twitter et autres.

En effet, plusieurs approches existent pour analyser ou fouiller ce type de données. D'un côté, il existe des logiciels d'analyse lexicométrique comme Lexico [8], initialement fait pour analyser des œuvres littéraires. De l'autre côté, se trouvent des logiciels fortement interactifs (pour naviguer dans les données afin de mieux les comprendre) comme Jigsaw [9], PosVis [11], ou encore Harvest [6].

⁴ - Le crowdsourcing est un néologisme créé en 2006, pouvant être traduit par « approvisionnement par la foule » <http://fr.wikipedia.org/wiki/Crowdsourcing>

⁵ - IVITA, First International Workshop on Intelligent Visual Interfaces for Text Analysis, Hong-Kong, Chine, Février 2010.

⁶ - IEEE Information Visualization Conference (IEEE InfoVis)

⁷ - IEEE Symposium on Visual Analytics Science and Technology (IEEE VAST)

¹ - <http://bleuciel.edf.com>

² - <http://www.semantia.com/>

³ - <http://www.cantoche.com/>

Et enfin, il existe des bibliothèques graphiques ou des techniques de visualisation, plus ou moins adaptées à chaque domaine, plus ou moins généralistes comme ManyEyes [10] ou Prefuse [7].

La suite du document présente le logiciel NaviQuest, inspiré par les logiciels ou techniques précédentes et adapté à nos besoins ainsi que ses principales fonctionnalités accompagnées de quelques résultats.

2 Architecture du logiciel

Pour développer NaviQuest, nous sommes repartis d'un logiciel développé dans le cadre du projet Callsurf [5] [2] qui permet d'enregistrer des conversations téléphoniques entre des clients et des conseillers en ligne, de les transcrire automatiquement, et d'offrir une interface de navigation dans ces conversations.

Bien évidemment, un dialogue entre un client et un conseiller est très différent d'un dialogue entre un internaute et Laura, mais ils ont pour dénominateur commun d'être constitués d'une suite de tours de parole. Donc ce qui avait été développé dans le cadre de Callsurf a pu être adapté, pour les dialogues avec Laura.

Le logiciel est développé en Java. Il s'appuie sur plusieurs bibliothèques Open Source :

- Lucene⁸ pour la partie moteur de recherche ;
- JFreeChart⁹ pour la représentation sous forme de courbe ou histogramme ;
- JTreeMap¹⁰ pour la représentation sous la forme de TreeMap.

Le logiciel se présente sous la forme de deux modules : le premier pour indexer les données et le second pour naviguer dans la base des questions-réponses.

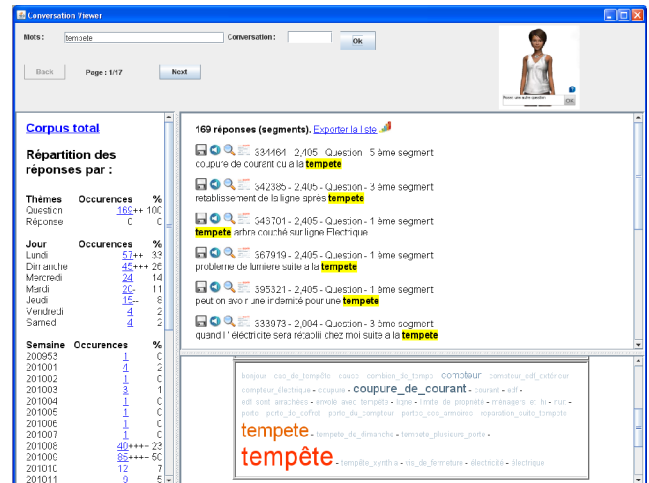


Fig. 1 : Copie d'écran de NaviQuest
Sur la fenêtre principale de l'application, se trouvent :

- En haut de l'écran, une barre de saisie pour rechercher un ou plusieurs mots ;
- A droite, la liste des questions posées à Laura contenant le ou les mots cherchés ;
- A gauche, la ventilation de ces questions selon les méta-données
- En bas à droite, un nuage de mots calculé à partir des questions posées à Laura et contenant le ou les mots cherchés.

3 Fonctionnalités du logiciel

L'indexation

Cette partie consiste à relire les fichiers de « log » de Laura et à les convertir en documents au sens de Lucene. Les questions sont indexées, les réponses ne le sont pas, elles sont juste sauvegardées dans la base Lucene. Cela signifie qu'on peut chercher un ou plusieurs mots dans les questions posées à Laura et qu'on peut voir la réponse fournie, mais qu'on ne peut pas chercher dans les réponses fournies. Un document sera constitué des éléments suivants :

- Le texte de la question posée par le client ou de la réponse fournie par Laura
- Nom du fichier (anecdote)
- Type du tour de parole : question ou réponse
- Rang de la question/réponse : 1 pour la 1^{ère}

⁸ - <http://lucene.apache.org>

⁹ - <http://www.jfree.org/jfreechart/>

¹⁰ - <http://jtreemap.sourceforge.net/>

question, puis 2, etc.

- Jour de la semaine : de lundi à dimanche
- Nombre de mots de la question
- Numéro de la semaine
- Date de la question/réponse

Recherche « full text »

L'application comprend un moteur de recherche « full text » permettant de trouver toutes les questions contenant un mot donné, comme « tempête », par exemple. Grâce aux requêtes floues autorisées par Lucene (`FuzzyQuery`) on s'affranchit, dans une certaine mesure, des nombreuses erreurs de saisies¹¹ ou faute d'orthographe. L'application autorise de saisir des suites de mots comme « service relevé client » grâce aux `MultiTermQuery` ou `QueryPhrase`. Il est également possible, avec des requêtes booléennes `BooleanQuery`, de faire des requêtes plus sophistiquées¹².

Après la recherche dans la base de données, le logiciel affiche la liste des questions correspondantes avec mise en évidence du ou des termes recherchés :

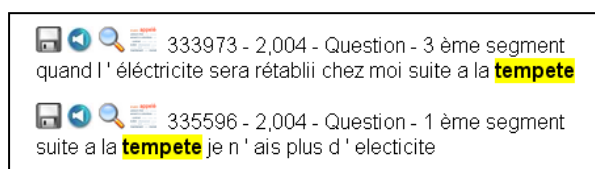


Fig. 2 : Surlignage de mot

En cliquant sur une des questions, on accède à l'ensemble du dialogue entre le client et Laura.

Ventilation des résultats de la requête

Au démarrage, le logiciel calcule la ventilation des questions selon toutes les méta-données disponibles (jour de la semaine, numéro de la semaine, etc.). Ensuite, pour chaque requête, il calcule la ventilation des réponses trouvées

¹¹ - Pour le mot « tempête » par exemple, on trouve : « tempoet », « tepete », « tempête », « tempêtes », « tempête », « tempête », « tempete ».

¹² - Mots obligatoires (`BooleanQuery.MUST`), mots à éliminer (`BooleanQuery.MUST_NOT`) ou mots requis (`BooleanQuery.SHOULD`).

selon ces mêmes méta-données, ce qui lui permet de comparer les deux ventilations, et d'indiquer grâce à des signes + et - une sur ou sous représentation par rapport à l'ensemble du corpus.

Semaine	Occurences	%
200953	1	0
201001	4	2
201002	1	0
201003	3	1
201004	1	0
201005	1	0
201006	1	0
201007	1	0
201008	40++++	23
201009	85++++	50
201010	12	7
201011	9	5
201012	5	2
201013	5	2

Fig. 3 : Ventilation des requêtes par semaines
Pour une recherche avec le mot « tempête », on constate une très forte sur représentation en semaine 8 et 9, d'où la présence de signes +.

Raffinement des recherches

L'application permet de raffiner une recherche en spécifiant une ou plusieurs méta-données. Sur l'exemple précédent, il est possible de rechercher le mot « tempête » en semaine 1 pour s'apercevoir qu'il s'agit de problèmes relatifs à une tempête habituelle, alors qu'en semaine 8 et 9, il s'agit de questions relatives à la tempête Xynthia qui a frappé le pays entre le 26 février et le 1^{er} mars, à cheval sur les semaines 8 et 9.

Analyses temporelles

Les mêmes meta-données vont permettre de mesurer la dynamique des questions et d'afficher leur occurrence selon les jours ou semaines, soit de manière absolue, soit relative au nombre de questions posées pendant une période donnée.

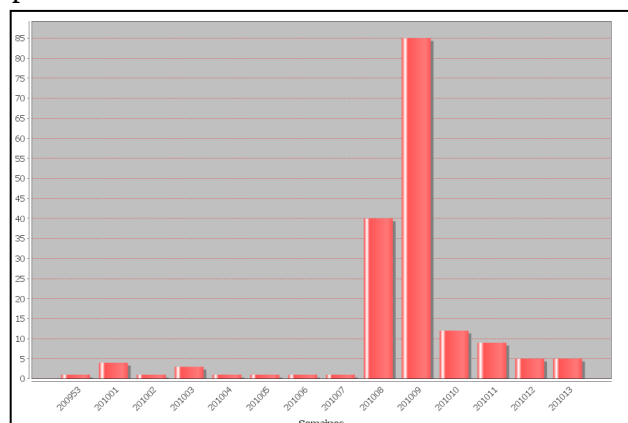


Fig. 4 : Distribution temporelle des questions

L'histogramme correspondant au mot « tempête » montre bien l'apparition soudaine de ce mot en semaine 8 et 9 puis sa lente décroissance. Cette analyse permet de voir qu'un mot a tendance à être cité de manière plus importante à certaines périodes.

Nuage de mot ou « Tag Cloud »

L'invention des « Tag Cloud » ou nuage de mots est attribuée, selon Wikipedia¹³, à D. Coupland [3], mais leur utilisation a été popularisée par le site Web de photos Flickr¹⁴. Pour chaque résultat de la requête, NaviQuest calcule un « nuage de mots » à partir de la liste des mots encadrant le ou les mots recherchés. La liste des mots trouvés (unigramme, bigramme ou trigramme) est visualisée sous la forme d'un nuage avec un code couleur et une police de caractère dépendant de la fréquence des mots : plus le mot ou la séquence de mot est fréquent¹⁵ plus sa police est grande et sa couleur est rouge.

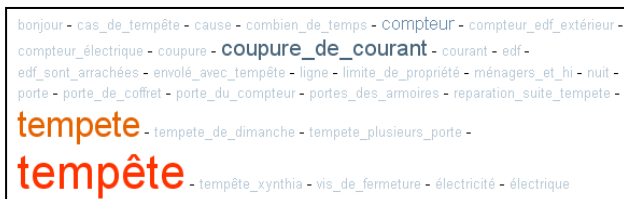


Fig. 5 : Nuage de mots associés à la requête « tempête »

Cette représentation a l'avantage de faire apparaître des associations de mots ou concepts, plus ou moins évidentes, comme ici : « coupure de courant », « porte » (de compteur, d'armoire, de coffret) arrachées ou envolées, « tempête de dimanche », « tempête Xynthia »...

Détection de nouveauté et TreeMap

Une manière d'explorer le corpus consiste à détecter les mots nouveaux ou qui évoluent fortement (à la hausse ou à la baisse) d'une semaine sur l'autre. L'algorithme utilisé est le

¹³ - http://en.wikipedia.org/wiki/Tag_cloud

¹⁴ - <http://www.flickr.com/>

¹⁵ - Afin de faire « ressortir » les bigramme et trigramme, un coefficient multiplicateur est attribué à leurs fréquences.

suivant : il calcule la liste de tous les mots (unigramme, bigramme ou trigramme) utilisés pour une semaine calendaire donnée. Ensuite, il calcule un score pour chacun d'eux : simple calcul de fréquence ou bien, pour les bigrammes, le rapport de vraisemblance RV [4] calculé de la manière suivante :

Pour tout couple (x_i, x_j) de mots qui se suivent, on calcule la table de contingence suivante :

- a_s : occurrences de (x_i, x_j) en semaine s
- b_s : occurrences de (x_i, x_i) en semaine s avec $i \neq j$
- c_s : occurrences de (x_k, x_j) en semaine s avec $k \neq i$
- d_s : occurrences de (x_k, x_i) en semaine s avec $k \neq i$ et $l \neq j$

$$RV_s(x_i, x_j) = a_s \cdot \log(a_s) + b_s \cdot \log(b_s) + c_s \cdot \log(c_s) + d_s \cdot \log(d_s) - (a_s + b_s) \cdot \log(a_s + b_s) - (a_s + c_s) \cdot \log(a_s + c_s) - (b_s + d_s) \cdot \log(b_s + d_s) - (c_s + d_s) \cdot \log(c_s + d_s) + (a_s + b_s + c_s + d_s) \cdot \log(a_s + b_s + c_s + d_s)$$

Enfin, l'algorithme compare le score de chaque mot ou suite de mots, d'une semaine à l'autre pour faire ressortir les mots qui apparaissent, ceux qui disparaissent, et ceux en forte variation (à la hausse ou à la baisse).

Les résultats sont affichés sous la forme de TreeMap [1], une technique qui consiste à représenter en 2D des informations de dimension supérieure.

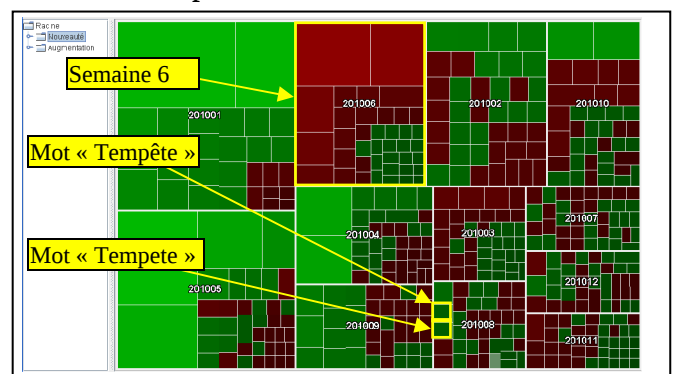


Fig. 6 : TreeMap des nouveaux mots

La figure montre les semaines et les mots qui présentent les variations les plus importantes (taille des rectangles) et leur sens de variation (verte pour une baisse et rouge pour une hausse).

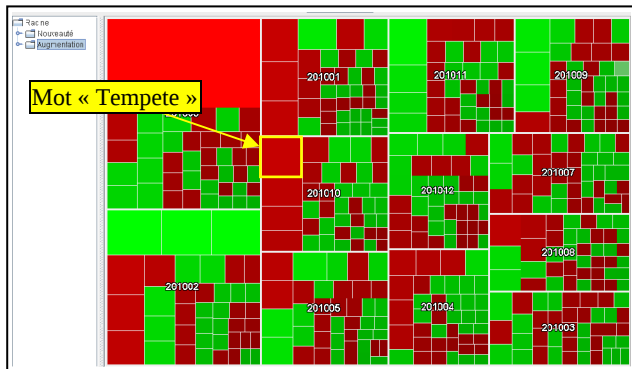


Fig. 7 : TreeMap des mots en augmentation

Les figures 6 et 7 font ressortir les fortes variations du mot « tempête », avec une brusque augmentation en semaine 8 et une forte chute en semaine 10.

La représentation sous forme de TreeMap présente l'avantage de mettre visuellement en évidence certains phénomènes, comme sur la figure 6, où l'on voit clairement qu'il y a eu beaucoup de disparitions de mots, car l'aire correspondant à cette semaine est globalement rouge. Cela constitue un indicateur pour aller ensuite regarder plus finement ce qui s'est passé cette semaine là.

4 Quelques résultats

Pour illustrer les possibilités du logiciel, voici quelques résultats obtenus :

- le lundi est le jour de la semaine où le nombre de questions est le plus élevé ;
- certaines questions très longues sont, en fait des mails, dans lesquels le client explique de manière détaillée sa demande ou sa problématique. On pourrait, dans ce cas, rediriger la question vers le traitement des mails ou bien lui faire une réponse adaptée en lui expliquant qu'il faut des questions courtes et que si il souhaite nous adresser un mail qu'il le fasse par le « bon » canal ;
- l'utilisation des évolutions de vocabulaire permet de détecter des questions qui apparaissent ponctuellement comme celles relatives à la « tempête ».

5 Conclusion

Afin de mieux comprendre les demandes qui lui sont faites et dans un but d'amélioration de Laura, l'agent conversationnel du site Bleu Ciel d'EDF, le logiciel NaviQuest a été développé pour naviguer dans la base des questions qui lui ont été posées. Ce logiciel mixe des techniques de visualisation et d'analyse textuelle. Il s'agit d'une première version du logiciel. Plusieurs évolutions sont envisageables :

- intégrer d'autres techniques de représentation ou visualisation pour fouiller le corpus de questions (évolution des réponses ou thèmes détectés)
- intégrer des informations de type analyse de sentiments ou d'opinions, qui constitueraient des méta-données supplémentaires pour fouiller le corpus.

Une étape importante consistera à prendre en compte les retours de la Direction Marketing d'EDF pour la définition précise et les évolutions d'un tel logiciel.

Références

- [1] B.B. Bederson, B. Shneiderman, M. Wattenberg. Ordered and Quantum Treemaps : Making Effective Use of 2D Space to Display Hierarchies. ACM Transactions on Graphics (TOG), 21, (4), Octobre 2002
- [2] L. Bozzi, P. Suignard, C. Waast-Richard. Segmentation et classification non supervisée de conversations téléphoniques automatiquement retranscrites, Actes de la conférence TALN'09 2009.
- [3] D. Coupland. Microserfs. Flamingo, 1996.
- [4] B. Daille, 1994, Approche Mixte pour l'Extraction Automatique de Terminologie, Thèse de Doctorat.
- [5] M. Garnier-Rizet, S. Guillermin-Lanne, F. Cailliau. CallSurf, Search by content, navigation and knowledge extraction on Call Center Conversational Speech, for marketing and strategic intelligence,

RIAO 2010.

- [6] D. Gotz, Z. When, J. Lu, P. Kissa, N. Cao, W.H. Qian, S.X. Liu. HARVEST : An Intelligent Visual Analytic Tool for the Masses, IVITA, First International Workshop on Intelligent Visual Interfaces for Text Analysis, Hong-Kong, Chine, Février 2010.
- [7] J. Heer. Prefuse : a software framework for interactive information visualization in Masters of Science, Computer Science Division, University of California, Berkeley (2004)
- [8] A. Salem *et al.* (2002). Manuel d'utilisation de Lexico. Université Paris 3. [http://www.cavi.univ-](http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/)
[paris3.fr/ilpga/ilpga/tal/lexicoWWW/](http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/)
- [9] J. Stasko, C. Görg, Z. Liu, and K. Singhal. Jigsaw: Supporting investigative analysis through interactive visualization. In Proceedings of IEEE, VAST '07, Sacramento, CA, Octobre 2007
- [10] F. B. Viégas, M. Wattenberg, F. Van Ham, J. Kriss, M. McKeon Many Eyes: A Site for Visualization at Internet Scale., Infovis, 2007.
- [11] Vuillemot, R., Clement, T., Plaisant, C., Kumar, A. (April 2009) What's Being Said Near "Martha" ? Exploring Name Entities in Literary Text Collections In Proceedings of IEEE VAST 2009.

Classification of Wordnet Personality Adjectives in the NEO PI-R Taxonomy

F. Bouchet
bouchet@limsi.fr

J-P. Sansonnet
jps@limsi.fr

LIMSI-CNRS BP 133 F-91403 Orsay cedex France

Abstract:

This work describes the process of classification of a set of personality-trait adjectives within the facet list of the NEO PI-R taxonomy related to the Five Factor Model. The classification process is not only based on adjective words but primarily on their lexical semantics as it is expressed by the synset-gloss attached to the adjectives in the Wordnet lexical base. This classification will provide a good coverage and support for the study and the computational implementation of psychological behaviors in conversational agents.

Keywords: Personality traits adjectives, FFM and NEO PI-R taxonomies, Wordnet.

1 Introduction

1.1 Context

In the Embodied Conversational Agents (ECA) community, research has focused on the expression of emotions through virtual characters modalities (face, gesture, speech...) but some authors have claimed that if an agent has a body it also has a mind [5]; by ‘mind’ they mean agents that can achieve practical reasoning upon the world, as required for example by non-player characters in games, by assistant agents in help systems or in tutoring systems. Authors have proposed mixed agent architectures exhibiting both rational reasoning modules (often based on BDI-agents) and psychological reasoning modules [4, 13]. In these architectures, the psychology of the agent influences both the expression of emotions through the body modalities but also has an impact on the decision making process.

Although presented here as separate notions, the rational and the psychological reasoning capacities of an agent actually work in quite an intricate manner [6, 9]. This is the reason why we have proposed an architecture dedicated to the study of the nature of their relationships: the Rational and Behavioral architecture (*R&B*)¹. *R&B* is a generic framework enabling the computational definition and the experimentation of various rational/psychological strategies. In recent work based on *R&B* [15], we have pro-

posed a model where personality traits are implemented in terms of influence operators altering the rational process of the assistant agent.

1.2 Motivation

For the time being, the *R&B* architecture has been implemented only with few arbitrary-chosen personality traits (e.g. a lazy-gloomy-cheerful... agent). This being insufficient to validate the genericity of the architecture, we need to implement a large class of psychological features, especially personality traits. In the literature about Psychology, various taxonomies have been proposed to cover the human personality (see Section 1.3 for some examples) but they are too generic for a computational implementation: 3 classes in Eysenck’s PEN system [7], 5 classes in FFM [10]. Some authors have tried to refine these classifications by adding sub classes, called *facets* (see Table 1) but the definitions attached to the facets are still too general to be directly implemented in terms of computational operators.

Our proposition is to enhance the precision of a standard taxonomy by refining the facets into sub classes, called *schemes*, with two main features: 1) a scheme represents a unique atomic behavior in the taxonomy; 2) a scheme is defined by a set of glosses associated with personality trait adjectives. To achieve this work we rely on two main standards: the FFF/NEO PI-R taxonomy for trait classes and facets and the Wordnet lexical data base for personality trait adjectives and glosses. This process was achieved in two phases: 1) the classification of a set of adjectives describing persons into the FFF/NEO PI-R taxonomy; 2) the process of clusterization of the schemes. In this paper, we present the first phase. The second phase is described in [16].

1.3 Works on personality adjectives

The Five Factor Model (FFM) is the most successful paradigm to day for personality traits representation. It is a convergent research from

1. <http://www.limsi.fr/~jps/research/rnb/rnb.htm>

many authors in Psychology for 20 years on. The FFM [10] is based on five large classes of psychological traits, often named Big Five model or O.C.E.A.N., by taking the first letter of the name of each class (they are listed in Table 1-col 1). Contrary to Cattell's classification into 16 factors [2] supported by Eysenck's personality questionnaires (EPQ) [7], FFM is based on the *lexical hypothesis* of Allport [1], stating that "personality vocabulary provides an extensive, yet finite, set of attributes that people speaking a given language have found important and useful in their daily interactions".

The NEO PI-R facets The FFM is a very generic classification so many authors have tried to refine this taxonomy by dividing the FFM classes into facets. Table 1, reproduced from [12], gives an attempt to align three of the facet lists proposed in the literature. The facet lists of Table 1 have many similarities although their facet number varies from 16 to 18 and 30. The **NEO PI-R** (Neuroticism Extraversion Openness Personality Inventory Revisited) facet list proposed by Costa and McCrae [3] (see Table 1-col 3) is a long standing model that provides the more precise facet list. For these reasons, we will rely on it for the work presented in section 2.

The Wordnet lexical data base In order to refine the facets into more precise categories (that we call schemes), and following the lexical approach, a first attempt was to define a scheme as a set of personality adjectives. However words can have several senses attached to them with quite distinct meanings. For example, the adjective 'kind' has three senses: Tolerant, Genial, Openhearted (see Table 4) that can be classified into two distinct FFM classes and three distinct NEO PI-R facets: *resp.* A-compliance, E-warmth, A-tendermindedness. This example shows that the word level is not sufficient to represent precisely the schemes. We have to work at the lexical semantics level.

The Wordnet lexical data base [8] comes handy when one has to treat a large amount of lexical data and one has to access the lexical semantics of words. In Wordnet, a word is attached to several so-called *synsets* that define a unique lexical sense described by a gloss (also called a short phrase — see Table 4 for examples). Moreover, because the Wordnet data base is freely accessible, it makes it easy to build a computer aided system for the classification process. Note that Wordnet has been used to support research on

affective computing, e.g. in the 'Wordnet affect' project of Strapparava *et al.* [21, 20]; however their work is dedicated to the recognition of affects in texts, not to the expression of psychological behaviors by rational agents and their classification is *ad hoc*.

It was decided to achieve a classification of a set of personality adjectives in terms of their Wordnet synsets/glosses while using the FFM / NEO PI-R taxonomy. This process has been carried out in two main stages, described in the next section:

Stage1: Adjective collecting. We collected a set of the personality trait adjectives (P_a) from Internet resources; then the adjectives were associated with their synset-gloss pairs in Wordnet; finally the synset-gloss pairs not related to personality description were discarded.

Stage 2: Classification. Remaining synset-pairs were manually classified within the FFM / NEO PI-R Taxonomy to generate a resource (XML file) of 60 psychological schemes containing 1,356 classed items.

2 Classification process

2.1 Stage 1: Personality adjectives selection

Using Wordnet, it is possible to extract a list of words according to their Part-of-Speech attribute (here, adjective) or related to a given word (e.g. all its synonym, antonym, meronym ...). However Wordnet lacks the possibility to extract a list of words related to a particular semantic domain (e.g. 'person personality'). Moreover, Wordnet aiming at covering exhaustively the English language, it does not provide information regarding the actual frequency of a given word or synset in everyday usage. Hence, the generation of a list of personality adjectives P_a was done in 4 sub steps S_{1-4} , described in the next paragraphs.

S1: Adjectives collecting. In order to work on actually used personality adjectives, we have collected a corpus of personality adjectives (C_{coll}), from ten different Internet sources explicitly claiming to provide "lists of adjectives describing personality traits" (the sources are summarized in Table 2). C_{coll} contains 1055 distinct adjectives, providing a first order approximation of the linguistic domain related to personality traits adjectives. Moreover, for each personality adjective a , we have associated a weight w_a

Table 1: Facets for the Big five trait domain: three approaches (alignment according to John et al. judgment, based on (John et al., 2008))

OCEAN Classes	Lexical facets ($N = 18$) (Saucier & Ostendorf, 1999)	NEO-PI-R facets ($N = 30$) (Costa & McCrae, 1992)	CPI-Big five facets ($N = 16$) (Soto & John, 2008)
Extraversion	Sociability	-	-
	Assertiveness	Gregariousness	Gregariousness
	-	Activity	-
	Activity/Adventurousness	Excitement seeking	(<i>O Adventurousness</i>) ^a
	-	-	Social confidence (vs Anxiety)
	Unrestraint (<i>A Warmth/Affection</i>) ^a	Positive emotions	-
Agreeableness	-	-	-
	Warmth/Affection	(<i>E Warmth</i>) ^a	-
	Modesty/Humility	Modesty	Modesty (vs Narcissism)
	-	Trust	Trust (vs Suspicion)
	-	Tender-Mindedness	Empathy/sympathy
	Generosity	-	Altruism
Conscientiousness	Gentleness	Compliance	-
	-	Straightforwardness	-
	Orderliness	Order	Orderliness
	Industriousness ^b	Achievement striving	-
	Reliability	Dutifulness	Industriousness ^b
	Decisiveness	-	-
Neuroticism	-	Self-discipline	Self-discipline
	(<i>O Perceptiveness</i>) ^a	Competence	-
	-	Deliberation	-
	Insecurity	Anxiety	Anxiety
	Emotionality	-	-
	Irritability	Angry hostility	Irritability
Openness	-	Depression	Depression
	-	-	Rumination-compulsiveness
	-	Self-consciousness	(<i>E Social confidence (vs Anxiety)</i>) ^a
	-	Vulnerability	-
	-	Impulsiveness	-
	Intellect	Ideas	Intellectualism
Openness	-	Aesthetics	Idealism
	Imagination/creativity	Fantasy	-
	-	-	Adventurousness
	-	Actions	-
	-	Feelings	-
	-	Values	-
Openness	Perceptiveness	-	-

a. Facet names in italics correspond to a secondary attachment of an already existing facet.

b. Despite a same name in two classifications, Industriousness doesn't have exactly the same meaning, explaining that the two aren't aligned.

(depending on the number of lists in which a given adjective appears) representing its usage frequency in resources (called *salience*).

S2: Representativity improvement with Wordnet.

To reinforce the weight of the most salient adjectives, we have used the functional relation synonym for each word in $\mathcal{C}_{\text{coll}}$ by adding 1 to the weight of any adjective in $\mathcal{C}_{\text{coll}}$ found to be synonym of another adjective from $\mathcal{C}_{\text{coll}}$ (i.e. $\forall a \in \mathcal{C}_{\text{coll}}, \text{ if } \exists a' \in \mathcal{C}_{\text{coll}} \text{ such as } a \in \text{synonym}(a') \text{ then } w_a = w_a + 1$). The maximum weight is thus theoretically of 11, although the

observed maximum weight is 9.

NOTE: We have also considered the possibility of adding to $\mathcal{C}_{\text{coll}}$, all the Wordnet synonyms of adjectives from $\mathcal{C}_{\text{coll}}$, but the synonym relation of Wordnet generates a lot of senses that exceed the personality trait domain. Actually, $\text{synonym}(\mathcal{C}_{\text{coll}})$ provides an extended list of 5,139 words of which 2,986 are distinct, which we considered too large to be processed.

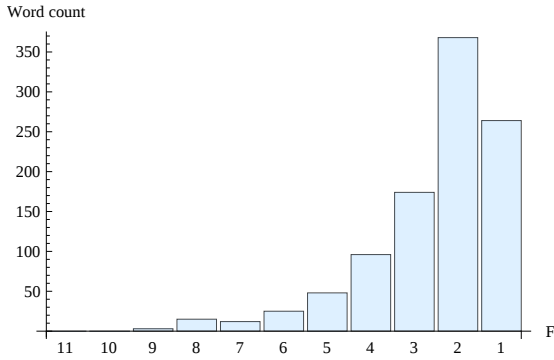
S3: Selection of Wordnet entries. To make $\mathcal{C}_{\text{coll}}$ technically compatible with Wordnet, words from $\mathcal{C}_{\text{coll}}$ absent from Wordnet ($\mathcal{C}_{\text{WN}}^{\text{abs}}$) or present

Table 2: Sources of personality adjectives for the corpus $\mathcal{C}_{\text{coll}}$

N°	Sources	Count
1	http://personal.georgiasouthern.edu/~jbjoy/Adjectives.html	315
2	http://www.keepandshare.com/doc/view.php?u=12894	182
3	http://www.lingolex.com/personalidad.htm	52
4	http://www.lesn.appstate.edu/fryeem/RE4030/character_trait_descriptive_adje.htm	183
5	http://www.mckinnonsc.vic.edu.au/la/lote/german/materials/describe/pers-adj.htm\#top	362
6 ^a	http://www.learnenglish.de/grammar/adjectivepersonality.htm\#positive	277
	http://www.examples-help.org.uk/parts-of-speech/personality-adjectives.htm	
7 ^b	http://www.esldesk.com/vocabulary/adjectives.htm	445
8	http://jobmob.co.il/blog/positive-personality-adjectives	130
9	http://www.nonstopenglish.com/exercise.asp?exid=440	20
10	http://www.scribd.com/doc/2212798/Adjective-List	80
Total		2046
Total of union (words different)		1055

a. List 6 is merged from two closely related sources.

b. List 7 is taken from a list of general adjectives, from which we selected those about persons.

Figure 1: Histogram of salience of personality adjectives in the corpus $\mathcal{C}_{\text{WN}}$

but not as adjectives ($\mathcal{C}_{\text{WN}}^{\text{-adj}}$) were deleted from $\mathcal{C}_{\text{coll}}$, with $|\mathcal{C}_{\text{WN}}^{\text{abs}}| = 29$ and $|\mathcal{C}_{\text{WN}}^{\text{-adj}}| = 21$. Note that $\forall a \in \mathcal{C}_{\text{WN}}^{\text{abs}} \cup \mathcal{C}_{\text{WN}}^{\text{-adj}}, w_a = 1$, indicating that Wordnet covers the most cited adjectives. We have thus obtained a list $\mathcal{C}_{\text{WN}}$ of 1,005 distinct adjective entries existing in Wordnet, such as $\mathcal{C}_{\text{coll}} = \mathcal{C}_{\text{WN}} \cup \mathcal{C}_{\text{WN}}^{\text{abs}} \cup \mathcal{C}_{\text{WN}}^{\text{-adj}}$, which can be sorted from the most ($w_a = 9$) to the least salient ($w_a = 1$). The histogram of salience of the $\mathcal{C}_{\text{WN}}$ adjectives is plotted in Figure 1 and adjectives a with $w_a \in [6..9]$ are given in Table 3.

S4: Selection of personality senses. Each adjective $a \in \mathcal{C}_{\text{WN}}$ has a Wordnet entry, which associates with a a set of lexical semantics senses (synsets). In Wordnet, each synset is unique and defines a unique sense with a

Table 3: Most salient $\mathcal{C}_{\text{WN}}$ adjectives $Total = 55$. Alphabetical order is used when same w_a

w_a	Adjectives	Count
9	friendly, kind, lively	3
8	ambitious, calm, cheerful, clumsy, crazy, energetic, excited, funny, gentle, helpful, proud, quiet, silly, talented, warm	15
7	bad, brave, bright, courageous, generous, good, happy, lazy, nervous, sensitive, successful, tough	12
6	anxious, careless, charming, cooperative, determined, dull, eager, enthusiastic, exuberant, fair, faithful, fantastic, jealous, loyal, mature, nice, placid, pleasant, sad, shy, thoughtful, unusual, wise, wonderful, zany	25

gloss. It is thus possible to obtain from the $\mathcal{C}_{\text{WN}}$ list of adjectives, a base $\mathcal{B}_{\text{all}}$ of triplets $\langle \text{adjective}, \text{synset}, \text{gloss} \rangle$. For the 1,005 adjectives in $\mathcal{C}_{\text{WN}}$, we obtained 2,924 entailing 2,924 entries in $\mathcal{B}_{\text{all}}$. However, all those synsets are not necessarily related to personality description (cf. entries with a star in the third column in Table 4) and they must be eliminated. To do that, a selection has to be made within $\mathcal{B}_{\text{all}}$ in order to keep only the triplets of $\mathcal{B}_{\text{all}}$ that could make sense to describe a personality trait or the behavior of a person.

This process of selection has been carried out by two independent annotators A_1 and A_2 in three successive rounds: during the first round, A_1

Table 4: $\mathcal{B}_{\text{all}}$ entries for adjectives from $\mathcal{C}_{\text{WN}}$ ranked with $w_a = 9$

w_a	Adjective	No	Synset	Gloss
9	friendly		Pally	characteristic of or befitting a friend
9		*	Allied	of or belonging to your own country's forces or those of an ally
9		*	Easy	easy to understand or use
9			Favorable	inclined to help or support; not antagonistic or hostile
9	kind		Tolerant	tolerant and forgiving under provocation
9			Genial	agreeable, conducive to comfort
9			Openhearted	having or showing a tender and considerate and helpful nature; used especially of persons and their behavior
9	lively		Vital	full of spirit
9		*	Eventful	filled with events or activity
9			Frothy	full of life and energy
9		*	Springy	elastic; rebounds readily
9			Alert	quick and energetic
9			Racy	full of zest or vigor

Table 5: Results of the 3-round selection process of personality senses from $\mathcal{B}_{\text{all}}$

Phase	Count	% $\mathcal{B}_{\text{all}}$
$\mathcal{C}_{\text{WN}}$ words (col 2 in Table 4)	1,005	
$\mathcal{B}_{\text{all}}$ initial triplets entries	2,924	100%
Round 1:		
triplets selected by A_1	1,771	60.6%
triplets selected by A_2	1,642	56.2%
A_1 / A_2 disagreements	449	15.4%
Round 2:		
disagreements left after A_1	241	8.2%
disagreements left after A_2	124	4.2%
Round 3:		
$\mathcal{B}_{\text{sel}}$ final triplets entries	1,715	58.7%

and A_2 were both working on the complete base, whereas during the second one, A_1 annotated only the words on which they disagreed during round 1 and then A_2 annotated only on the remaining disagreements. Finally, in the third round, A_1 and A_2 met for a discussion phase to take a decision regarding the remaining disagreements, the default preference being to keep the triplet (as it could always be discarded later on).

The results of this first filtering are summarized in Table 5. While only two annotators were involved, the quality of the filtering is quite good a) because the convergence on two simple predefined classes was strong; b) because in case of disagreement the decision was postponed to the next step, where better informed decision was possible. We have obtained a base $\mathcal{B}_{\text{sel}}$ of 1,715

triplets, containing 904 unique adjectives (each gloss being associated to at least one adjective, and each adjective to at least one gloss).

The final number of retained entries (58.7%) is a first result that confirms our claim about the necessity to work at the lexical semantics level (i.e. the synset-gloss) rather than the word level, as in previous classifications.

2.2 Stage 2: Facet association

Generation of the gloss-oriented base. In the next step, glosses appearing in $\mathcal{B}_{\text{sel}}$ must be classified. Unfortunately, it sometimes occurs that Wordnet glosses are associated with more than one pair (adjective, Synset): for instance, the gloss “causing or fraught with or showing anxiety” corresponds to (nervous, Queasy), (anxious, Nervous) and (uneasy, Queasy). In such specific cases, we have chosen to merge entries having the same gloss, providing an alternative base $\mathcal{B}'_{\text{sel}}$ with gloss-oriented pairs (gloss, (adjective, synset)*), such as $|\mathcal{B}'_{\text{sel}}| = 1356$.

Adding poles to the NEO PI-R taxonomy. We have seen in section 1.3 that while several facet taxonomies are proposed for FFM, the NEO PI-R taxonomy is prominent and provides a precision of 30 facets. In NEO PI-R, each facet encompasses both the concept and its antonym, e.g. the facet A-Modesty can stand for adjectives like ‘mild’ as well as ‘arrogant’. Hence, similarly to Goldberg’s works on 50 bipolar scales [11], each facet was divided into two poles: noted +

for the concept and - for its antonym(s), thus resulting in 60 positions noted +/-NEO PI-R.

A first attempt at classifying a subset of $\mathcal{B}'_{\text{sel}}$ entries over the 60 positions raised a major issue: despite the fact that the collected adjectives words were claimed by the resources to be personality-related, actually many of their corresponding synset-gloss pairs in Wordnet are also associated with other attributes of a person², which can be divided into five main categories:

- *physical descriptions*, e.g. “having a sturdy and well proportioned body”, (athletic, Muscular);
- *mental capacities* e.g. “characterized by quickness and ease in learning”, (bright, Smart), (smart, Bright);
- *judgmental qualities*, both physical or mental, e.g. “appealing to the emotions as well as the eye”, (lovely, Beautiful);
- *social roles*, e.g. “characterized by or indicating poverty”, (poor, Beggarly);
- *mind states* are transitory (like emotions), e.g. “in an aroused state”, (excited, Thrillful).
- *errors* are items that don’t fit in any of the facets of categories, i.e. if they should have been eliminated during the selection (see 2.1).

Annotation protocol. Consequently, $\mathcal{B}'_{\text{sel}}$ entries are manually classified either as one of the 60 +/-NEO PI-R poles or one of the five external categories listed above. The annotation of $\mathcal{B}'_{\text{sel}}$ has been carried out by two independent annotators A_1 and A_2 , following exactly the same protocol in three rounds as the one described in the fourth step of section 2.1 with the difference that in the third round, in the few cases where a strong disagreement remained between annotators, the gloss was discarded. The annotation was done using a specific software tool displayed in Figure 2.

2.3 Results and analysis

The results of the annotation process are given in Table 6. We see that after three rounds, the two annotators managed to agree on more than 90% of the positions. Note that if here the first disagreement rate between the two annotators is more important than in 2.1, it shouldn’t be misinterpreted: the number of possible options (60) is far more important here than in 2.1 (just 2). Roughly, three cases led to a disagreement:

1 characteristic of or befitting a friend
([Pity, kindly, 9])
Modesty + .

FIND BACK NEXT

NEO-PI-R facet:

Openness Fantasy Aesthetics Feelings Actions Ideas Values	Conscientiousness Competence Orderliness Dutifulness Achievement-striving Self-discipline Deliberation	Extraversion Warmth Gregariousness Assertiveness Activity Excitement-seeking Positive-emotions	Agreeableness Trust Straightforwardness Altruism Compliance Modesty Tender-mindedness	Neuroticism Anxiety Angry-Hostility Depression Self-consciousness Impulsiveness Vulnerability	Others Physicality Capacity Quality Role State	TODD ERROR
---	--	--	---	---	---	---------------

Value in the facet:

+ tendency to play down own achievements and be humble
=
- tendency to be proud and arrogant
?

Meta-infos:
Judgement

Other items linked to this facet by:

synset | word | gloss

GLOSSES = (37, 75, 153, 200, 201, 202, 203, 270, 271, 298, 427, 798, 803, 897)
(Majestic, lofty, Respectable, Respectable, Arrogant, Coy, Retiring, Unpretentious,
Mild, Modest, SelfRestraint, Circumspect, Unpretentious, Aloof, SelfLoving, Narcissistic, Psychopathology, Orotund)

Figure 2: Screen plot of the tool used to associate entries of $\mathcal{B}'_{\text{sel}}$ to +/-NEO PI-R facets

1. A mere error done by one of the two annotators. These errors disappeared at the end of the second round, when both annotators have had to recheck their own annotations by comparing them to the other annotator’s.
2. A disagreement regarding the interpretation of the glosses associated to the facets. They were sorted during rounds 2 and 3, by a discussion between the two annotators when a facet gloss was ambiguous. For instance, a strong degree of confusion has been observed between some facets like E-Assertiveness and A-Modesty, C-Achievement-Striving and C-Self-discipline or O-Ideas and O-Values.
3. A disagreement of interpretation regarding the gloss to annotate, especially when they are complex or composed: e.g. (“joyful and proud especially because of triumph or success”, ((exultant, Exultant), (prideful, Jubilant))) refers at the same time to a tendency to experience strongly positive emotions E-Positive-emotions and to a tendency to lack modesty A-Modesty. In such cases, annotators had to agree to favor one sense over another during round 3, or to ultimately discard the gloss.

Research in Psychology, especially on FFM or facets, involves multi-annotators (novices or peers) and is based on statically computed classes. Here our work is far less ambitious: 1) we already have 2×30 established classes to put the glosses into; 2) supposing a gloss is misplaced (the two annotators make the same error) this only jeopardize schemes having only

2. Again, this is in favor of the synset-gloss approach, which appears to be much more semantically detailed than the word level.

Table 6: Results of the 3-round classification of entries of $\mathcal{B}'_{\text{sel}}$ into the 60 NEO PI-R poles

Phase	Count	% $\mathcal{B}'_{\text{sel}}$
$\mathcal{B}'_{\text{sel}}$ entries	1,356	100%
Round 1:		
A_1 / A_2 facet disagreements	526	38.8%
A_1 / A_2 trait disagreements	376	27.7%
Round 2:		
facet disagreements after A_1	274	20.2%
trait disagreements after A_1	178	13.1%
facet disagreements after A_2	133	9.8%
trait disagreements after A_2	101	7.4%
Round 3:		
entries discarded	37	2.7%

one gloss (i.e. those with a small salience). We also discussed the possibility of allowing multiple membership of the glosses to the classes: 1) it could increase annotators agreement 2) it could handle some inherently ambiguous or polysemic glosses (a few Wordnet glosses can be rather vague). Finally, it was decided to provoke more annotator disagreement (which is better with two annotators) at the price of giving a single interpretation to vague glosses.

The final distribution of gloss-oriented pairs into the different categories (30 facets + external cases) is summarized by Figure 3. Although some facets like C-Dutifulness and A-Tender-mindedness appear to have gathered significantly more glosses, we can see that all NEO PI-R facets have had some associated pairs (with a minimum of 2 for O-Aesthetics). An XML file providing all the glosses, words and synsets associated to those categories is freely available on the web³.

3 Conclusion and further work

In this work, we have classified a large set of personality-trait adjectives within the facet list of the FFM / NEO PI-R taxonomy, which is based on the Wordnet lexical semantics given as synset-gloss couples. This classification provides a good coverage and makes available a resource for the study and the computational implementation of psychological behaviors in conversational agents. Further work will exploit the clusters of synset-gloss pairs to define, in a formal representation, a set of psychological be-

haviors to be implemented in the *R&B* framework.

References

- [1] G. W. Allport. *Personality: A psychological interpretation*, 1937.
- [2] R. B. Cattell, H. W. Eber, and M. M. Tatsuoka. *Handbook for the sixteen personality factor questionnaire (16 PF)*. Champaign, Illinois, 1970.
- [3] P. T. Costa and R. R. McCrae. *The NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources, 1992.
- [4] Joao Dias and Ana Paiva. Feeling and reasoning: A computational model for emotional characters. In *Progress in Artificial Intelligence*, volume 3808 of *LNCS*, pages 127–140. Springer Berlin / Heidelberg, 2005.
- [5] Tiago Doce, Joao Dias, Rui Prada, and Ana Paiva. Creating individual agents through personality traits. In *Intelligent Virtual Agents (IVA 2010)*, volume 6356 of *LNAI*, pages 257–264, Philadelphia, PA, 2010. Springer-Verlag.
- [6] Phoebe C. Ellsworth and Klaus R. Scherer. Appraisal processes in emotion. In Richard J. Davidson, Klaus R. Scherer, and Hill H. Goldsmith, editors, *Handbook of affective sciences*, Series in affective science, pages 572–595. Oxford University Press, New York, NY, USA, 2003.
- [7] Sybil B. G. Eysenck, Hans Jorgen Eysenck, and Paul Barret. A revised version of the psychoticism scale. *Journal of Personality and Individual Differences*, 6(1):21–29, 1985.
- [8] Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
- [9] Nico H. Frijda. *The Laws of Emotion*. Psychology Press, 2006.
- [10] L. R. Goldberg. An alternative description of personality: The big-five factor structure. *Journal of Personality and Social Psychology*, 59:1216–1229, 1990.
- [11] L. R. Goldberg. The development of markers for the big-five factor structure. *Psychological Assessment*, 4:26–42, 1992.
- [12] Oliver P. John, Laura P. Naumann, and Christopher J. Soto. Paradigm shift to the

3. <http://fbouchet.vorty.net/projects/r&b/adjectives/neopiradj.xml>

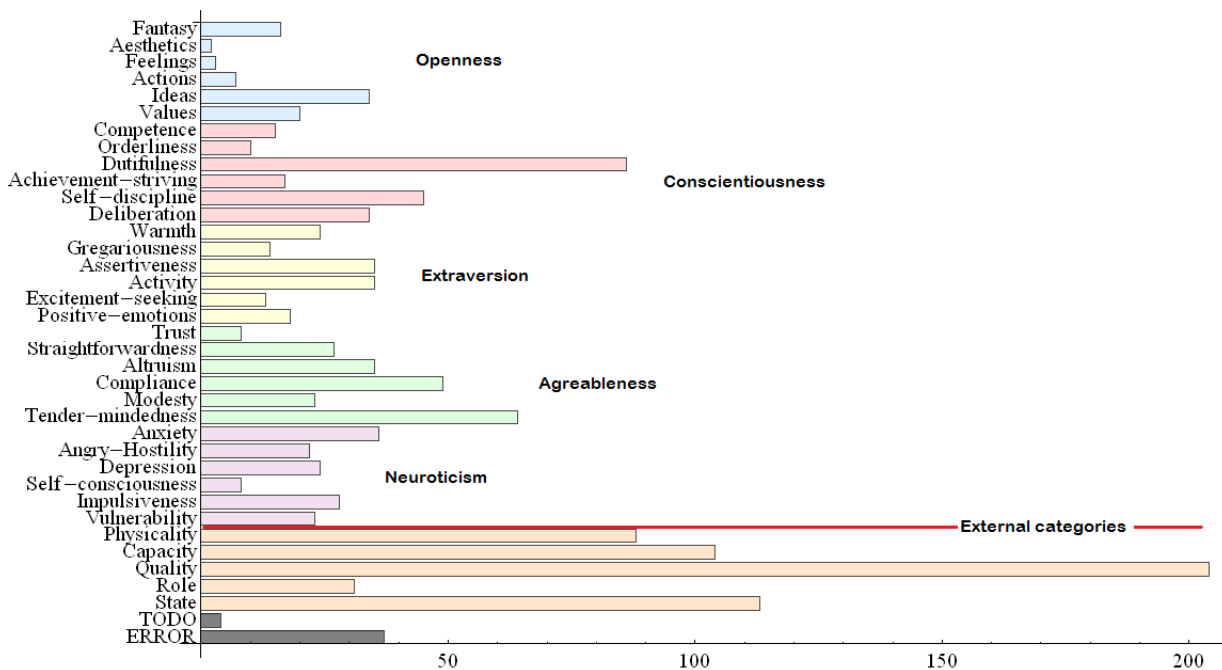


Figure 3: Number of synset-gloss pairs associated to each FFM / NEO PI-R facet

integrative big five trait taxonomy: History, measurements, and conceptual issues. In *Handbook of Personality: Theory and Research*. The Guilford Press, 3rd edition, 2008.

- [13] Mei Lim, Joao Dias, Ruth Aylett, and Ana Paiva. Improving adaptiveness in autonomous characters. In Helmut Prendinger, James Lester, and Mitsuru Ishizuka, editors, *Intelligent Virtual Agents*, volume 5208 of *Lecture Notes in Computer Science*, pages 348–355. Springer Berlin / Heidelberg, 2008.
- [14] N. Sabouret and J. P. Sansonnet. Learning collective behavior from local interactions. In *CEEMAS: Second International Workshop of Central and Eastern Europe on Multi-Agent Systems*, volume 2296 of *LNCS*, pages 273–282, Cracovia, Poland, 2002. Springer Verlag.
- [15] J. P. Sansonnet and F. Bouchet. Expression of behaviors in assistant agents as influences on rational execution of plans. In *Intelligent Virtual Agents (IVA 2010)*, volume 6356 of *LNAI*, pages 413–419, Philadelphia, PA, 2010. Springer-Verlag.
- [16] J. P. Sansonnet and F. Bouchet. Extraction of agent psychological behaviors from glosses of wordnet personality adjectives. In *Paper submitted at the Proc. of the 8th*

European Workshop on Multi-Agent Systems (EUMAS'10), Paris, France, 2010.

- [17] J. P. Sansonnet, N. Sabouret, and G. Pitel. An agent design and query language dedicated to natural language interaction. In *AAMAS:2002 First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, Bologna, Italy, 2002.
- [18] G. Saucier and F. Ostendorf. Hierarchical subcomponents of the big five personality factors: A cross-language replication. *Journal of Personality and Social Psychology*, 76(4):613–627, 1999.
- [19] Christopher J. Soto and Oliver P. John. Ten facet scales for the big five inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality*, 43(1):84–90, 2009.
- [20] Carlo Strapparava and Rada Mihalcea. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1556–1560, Fortaleza, Ceara, Brazil, 2008. ACM.
- [21] Carlo Strapparava and Alessandro Valentini. WordNet-Affect: an affective extension of WordNet. In *Proc. of the 4th Int. Conf. on Language Resources and Evaluation*, pages 1083–1086, 2004.

De la description de scénarios socio-émotionnels au narrateur virtuel expressif

R. Miletitch†

roman.miletitch@gmail.com

N. Sabouret†

Nicolas.Sabouret@lip6.fr

M. Ochs‡

magalie.ochs@gmail.com

†Laboratoire d'informatique de Paris 6
Université Pierre et Marie Curie
4 place Jussieu 75005 Paris, France

‡CNRS LTCI
Telecom-ParisTech
37 rue Dareau 75014 Paris, France

Résumé :

Raconter une histoire nécessite de relier entre eux un ensemble d'événements significatifs à l'aide d'énoncés permettant de maintenir la tension, de créer du suspense et de laisser du temps au développement des émotions. Lors de l'automatisation de ce processus, l'une des difficultés majeures est la génération de ces phrases brodées et le fait de leur donner suffisamment de cohérence avec l'histoire. Pour ce faire, nous proposons dans cet article d'utiliser un modèle de représentation de connaissances locales et cohérentes du monde. Nous présenterons une architecture sous forme d'ontologie de ces concepts. Mis en relations entre eux, ils permettront de décrire ce monde, et des algorithmes parcourront l'ontologie pour en extraire du sens de manière pseudo aléatoire, permettant la création de phrase. Ces différents éléments seront mis en commun lors du traitement d'une histoire prédéfinie. Lors du parcours de son squelette imposé, les phrases issues de nos algorithmes (Suspense, Relation sociale, Phrase seule, Événement, Phrase du Narrateur), une fois mises en forme, viendront enrichir l'histoire.

Mots-clés : Narration, Ontologie, Agent Conversationnel Animé, Emotions

Abstract:

Telling a story requires linking a series of significant events using statements to maintain tension, create suspense and to allow time for the development of emotions. When automating this process, one major challenge is the generation of these filling sentences and them being sufficiently consistent with the story. To this end we propose in this paper to use a knowledge representation model of a local coherent world. We will present an architecture in the form of an ontology of these concepts. By establishing relations between them, they will describe this world, and our algorithms will scan the ontology to drag sense out of it in a pseudo-random manner, allowing the creation of a sentence. These elements will be gathered during the process of a predefined storyline. During the scan of the scenario, an imposed skeleton, sentences from our algorithms (Suspense, Social Relations, one phrase, Event, Phrase Narrator), once edited, will enrich the story.

Keywords: Narration, Ontology, Animated Conversational Agent, Emotions

1 Introduction

Pour raconter une histoire, le narrateur ne doit pas seulement relater des événements, mais aussi faire ressentir des émotions tout au long

de la narration. Le fait de “broder” autour des événements, d'accaparer l'attention de l'interlocuteur, de stimuler sa visualisation et ses sensations, tout ceci amène à créer une histoire émouvante, intéressante et donc mémorable[1].

Il est important de garder à l'idée que lorsque l'on raconte une histoire, il ne s'agit pas juste de communication vocale, mais d'une communication à plus haute échelle, d'une création d'une ambiance par les sens, où tous les événements racontés sont mis en valeur les uns par rapport aux autres. Au-delà de ces sensations, ce qui motive le lecteur à poursuivre sa lecture, c'est l'intérêt que suscite l'histoire, et la tension qu'elle crée, à travers une répartition de cet intérêt[4].

En plus de ces sensations et de cet intérêt, une autre manière d'ancrer le public dans la scène du narrateur est de stimuler des émotions[2], et c'est là que le don d'empathie devient prépondérant. Nous pensons que le narrateur ne doit pas avoir un regard objectif sur la situation, mais doit plutôt exacerber ses émotions lors du conte pour en rendre la lecture intéressante. Signaler les émotions des personnages n'est pas suffisant, il est mieux pour le narrateur de les exprimer, de les manifester, à travers un choix de mot, à travers une gestualité ainsi que des expressions faciales.

Dans la section 2, nous allons présenter les travaux existant en "StoryTelling", ainsi qu'approfondir certaines thématiques abordées dans cette introduction. Ensuite dans la section 3, nous vous présenterons notre modèle, et les algorithmes mis en oeuvre. Enfin dans la section 4, nous exposerons les résultats obtenus.

2 Travaux existants

The virtual Story Teller [8] est un framework multi-agents qui permet la création dynamique

d'une histoire. Ses agents, guidés par leurs objectifs et contrôlés par un agent réalisateur construisent un scénario. Ce framework nous montre que l'aspect indéterminé de la création d'une histoire joue en sa faveur, effet que l'on retrouvera dans notre modèle. Deux agents s'occupent de traduire le scénario en histoire : le narrateur, et le présentateur. Cette séparation se retrouve dans nombre de systèmes, et c'est ce qui nous a fait opter pour deux classes d'algorithmes : une pour la construction de ces phrases, et l'autre pour leurs réalisations superficielles.

Cavazza traite dans un de ses articles [3] de l'intégration d'un générateur de langue naturelle au sein d'un planificateur émotionnel en s'inspirant de l'univers de Flaubert. Cette recherche tire aussi profit du fait que Flaubert avait l'habitude de créer ce que nous appellerions aujourd'hui une ontologie pour ses livres. Nous retrouvons ici l'intérêt pour les émotions qui serviront de base à notre modèle ainsi que l'utilisation d'une ontologie.

Jean Louis Dessales propose une modélisation de l'intérêt [4] issu de la communication d'événements, qui s'applique tout particulièrement au StoryTelling. Il définit l'intérêt d'un événement par la différence entre sa complexité d'occurrence, et sa complexité à être raconté. Au-delà de cela, il caractérise l'intérêt par l'aspect improbable d'un événement. Il suscite alors la curiosité, une certaine forme de suspense quant à sa raison d'être. C'est ce suspense qui sera la source d'un de nos algorithmes de création d'intérêt. Enfin, il indique que l'intérêt d'un événement est proportionnel à l'intensité émotionnelle contrastante qu'il suscite[5], peu importe le type d'émotion. Nos algorithmes se basent sur cette théorie pour savoir quelles émotions sont à développer dans notre scénario, et à quels moments.

Modèle OSSE. Notre travail a eu comme source un logiciel qui se fonde sur une modélisation dynamique de l'état socioémotionnel : OSSE [7]. Cette modélisation se concentre sur les facteurs de personnalité qui favorisent le déclenchement d'émotions, ainsi que l'influence que ces émotions ont sur le comportement et sur l'évolution des relations sociales entre les personnages mis en jeu. Le vocabulaire de OSSE est composé de 4 ensembles : $Voc_{OSSE} = O \cup V \cup R \cup P$, où O est l'ensemble des objet ou game characters, V l'ensemble des actions (verbes), R l'ensemble des rôles possibles pour les personnages,

et P l'ensemble des personnages. Ces rôles sont liés par des relations sociales : $RelSoc : R \times R \rightarrow \{Agreeability, Dominance, Solidarity, Familiarity\}$. Les personnages sont définis comme un set : $\{Name_p, Personality_p, Emotions_p, Roles_p, Attitudes_p\}$. Ce vocabulaire nous permet de définir des événements $e \in Evt$ par $e = \{Agent_e \in P, Verbe_e \in A, Patient_e \in O \cup P, date_e \in N, WhoWatching_e \subset P\}$. L' $Agent_e$ définit qui agit dans l'action, $Verbe_e$ définit le verbe, $Patient_e$ celui qui subit l'action. La $date_e$ permet d'organiser la liste des événements. $WhoWatching$ correspond à la liste des personnages de P qui observent l'action. Un scénario est un ensemble d'événements. Lors de l'avènement d'un triplet, l'état émotionnel ainsi que les relations sociales des observateurs de cet événement sont modifiés.

3 Modèle proposé

Comme dans le modèle OSSE, nous nous attachons à conserver une description du scénario à l'aide de triplets. Cette représentation, qui présente l'avantage d'être utilisable par des "game designers" non spécialistes de l'IA, pose le problème de la génération automatique de textes en langue naturelle par le narrateur virtuel expressif. Notre objectif est donc, en partant de cette suite d'événements, et à partir d'un vocabulaire restreint, d'enrichir une histoire de phrases cohérentes renforçant l'intérêt, l'impact émotionnel et la crédibilité de l'histoire chez les spectateurs sans pour autant la modifier fondamentalement.

Nous nous appuyons sur deux hypothèses :

- une modélisation simple et locale du monde mise en relation avec elle-même est suffisante pour renforcer l'intérêt, les sensations et les émotions
- la génération pseudo aléatoire d'éléments peut-être non pertinents ne gêne pas la narration. Au contraire, cela l'enrichit, car le lecteur/spectateur construira les liens absents.

3.1 Architecture globale

De la création d'événement au conte raconté, nous utiliserons deux modules. OSSE qui s'occupe de la modélisation socioémotionnelle et notre module composé d'une ontologie et d'algorithmes de parcours qui permettent d'en faire émerger du sens [11], et de créer la phrase.

Ontologie. Une ontologie est la spécification d'une conceptualisation d'un domaine de connaissance [6]. Elle constitue en soi un modèle de données représentatif d'un ensemble de concepts dans un domaine, ainsi que des relations entre ces concepts. Notre ontologie s'inspire de Cavazza [3], ainsi que des Graphes conceptuels de Sowa [10]. L'ontologie initiale définit un nouveau vocabulaire, en parallèle de celui d'OSSE, composée de concept et de relation entre ses concepts : $Voc_{Onto} = Con \cup Rel$, où chaque concept d'OSSE a son équivalent. Les concepts sont organisés de manière hiérarchique, et même si leur organisation est modifiable par le game designer, il en existe une de base, imposée, afin de guider le game designer. En plus des concepts, notre ontologie sera décrite par des relations r définies par $r : Con \times Con \rightarrow Rel$ permettant la mise en relation de deux concepts, par un lien d'un certain type de catégorie. L'utilisation d'une représentation de connaissance avec héritage permet de définir les propriétés et les relations entre concepts à un haut niveau, les sous-concepts héritant automatiquement des qualificatifs de leurs parents.

Le fait d'utiliser une ontologie plutôt que de scripter notre histoire nous permet d'avoir une mise en scène variée. Tout au long de notre démarche de création de langue, nous essaierons de garder le plus longtemps et le plus souvent possible cette idée de concepts et de lien entre concepts. Il est préférable d'avoir une liste de concept ordonnée plutôt qu'une chaîne de caractère ; une fois générée, elle ne nous serait en effet alors impossible de lui rajouter du sens, de jouer sur sa syntaxe [9]. Elle est un résultat fini, qui empêche tout rajout ultérieur. Voilà pourquoi nous gardons une organisation de phrase à partir d'item : $item = \{concept + typeGramatical + qualificatifs\}$.

Algorithme de génération. L'ontologie permet de décrire le monde, de le structurer. L'ordre est un moyen d'accès au sens, mais c'est le traitement de cette information qui permet de le faire émerger. Il s'agit là du travail des algorithmes de parcours que nous avons créé en nous inspirant des théories sur l'impact émotionnel et la modélisation de l'intérêt de Jean Louis Dessales [4]. Ces algorithmes sont au nombre de huit et chacun permet d'avoir un contrôle accru sur la tension des spectateurs tout au long de l'histoire. Deux sont particuliers : il s'agit des algorithmes d'introduction et de conclusion, afin de créer une ambiance au début, ainsi qu'un épilogue, qui permet un relâchement de la tension.

Cinq autres sont appelés à chaque date, et à partir de l'ensemble des triplets des événements de cette date, génère des liaisons entre ces triplets et l'ontologie :

- *Réalisation événement* traduit l'ensemble des événements d'une $date_t$, correspond à la structure de l'histoire créée par le game designer
- *Suspense* crée une phrase en amont d'accroche afin de créer une tension
- *Relation sociale* s'inspire des relations entre personnages renforçant le côté émotionnel de l'histoire
- *Stand Alone* crée des phrases seules, sans impact fort, afin de jouer sur la tension
- *Phrase narrateur* laisse l'occasion au narrateur de s'exprimer par commentaire sur les événements de la $date_t$, renforçant l'attention du spectateur, ainsi que le côté émotionnel
- *Filling* s'occupe de trouver des qualificatifs à rajouter aux items, pour varier l'expérience du spectateur

Ces algorithmes sont à la base de la tension, et de l'émotion générées dans notre histoire. Il est possible de modifier le paramètre narrateur, ce qui permet de modifier les histoires. En effet, l'histoire ne sera pas racontée de la même manière selon si l'on est le voleur qui se fait arrêté ou le policier qui effectue l'arrestation, par exemple. Enfin, un dernier algorithme, se chargeant de l'organisation, gère l'occurrence de ces algorithmes. Permettre à chacun de ces algorithmes de s'exprimer en permanence surchargerait l'histoire. L'algorithme d'organisation associe à chaque algorithme de génération une probabilité d'occurrence, et les appelle en fonction de cette dernière.

Algorithme de réalisation. Une fois les liaisons faites entre la liste des événements et les éléments de l'ontologie, il reste à réaliser une phrase de ces relations. Ce sont les algorithmes de réalisation superficielle qui transcrivent le sens extirpé par les algorithmes de parcours depuis l'ontologie en une phrase compréhensible par l'être humain et qui ajoutent des tags d'émotion et de comportement qui guideront l'interprétation du texte pour des agents conversationnels. L'objectif de ces deux catégories d'algorithme est d'interdire le plus possible des bouts de phrases écrits par le codeur et de favoriser une génération automatique de phrase de par l'agencement des mots et leurs liaisons. Dans les cas dont l'occurrence est unique, nous nous permettons d'en revenir à l'utilisation de phrase à trous (par exemple pour l'introduction), qui est alors remplie selon les éléments mis en jeux, les

concepts, ou encore les émotions du narrateur afin de lui donner un comportement plus proche de ce qui pourrait être attendu d'un narrateur humain.

Organisation grammaticale de la phrase. Nous avons décidé de séparer la phrase en item : $item \in I = \{concept \in Con, type_{gram} \in \{Sujet, Verbe, COD\}, qualificatifs\}$ où *qualificatifs* correspond à une liste de description de *concept* définissant le mot, et *type_{gram}* sa position dans la phrase. Une phrase est donc définie comme une liste de mots : $sentence = \{item_0, \dots, item_n\}$, et un paragraphe comme une liste de phrases : $paragraph = \{sentence_1, \dots, sentence_n\}$.

Nous partons des phrases simples des événements, qui forment déjà un paragraphe, auquel nous ajoutons la phrase simple de suspense, celle des relations sociales, celle Stand Alone, et lors de la réalisation de ces phrases, nous ajoutons la phrase du narrateur. À la suite de quoi, nous modifions leur mise en forme et nous ajoutons les qualificatifs.

3.2 Ontologie

Afin de permettre une manipulation plus aisée, notre ontologie se base sur une hiérarchie (FIG 1). Elle se remplit à l'aide de bibliothèque ou grâce à l'imagination de l'auteur qui permet ainsi de développer une hiérarchie plus complexe. Il ne s'agit pas juste de la remplir de connaissances, mais aussi de l'architecturer. Par exemple, nous décidons dans notre implémentation de séparer objets en gros objets, petits objets, et objets liquides. Ce choix est lié au choix de nos algorithmes et aux relations mises en jeu. Une fois cette architecture créée, l'auteur peut créer des liquides, définir des relations qui leur sont propres, mais aussi d'autres classes. De plus, les classes ou instances créées peuvent hériter de plusieurs concepts Parent. Il est possible d'avoir un gros objet récipient. Toute cette infrastructure permet à l'auteur de créer du sens, en reliant les classes entre elles par des relations. Par exemple, si nous permettons aux instances de Être de parler, alors tous nos personnages auront la possibilité de parler, et les algorithmes de génération en seront automatiquement informés.

Enfin, il existe une dernière distinction entre les concepts de l'ontologie. On considère par défaut que si le concept est une feuille, il s'agit d'une

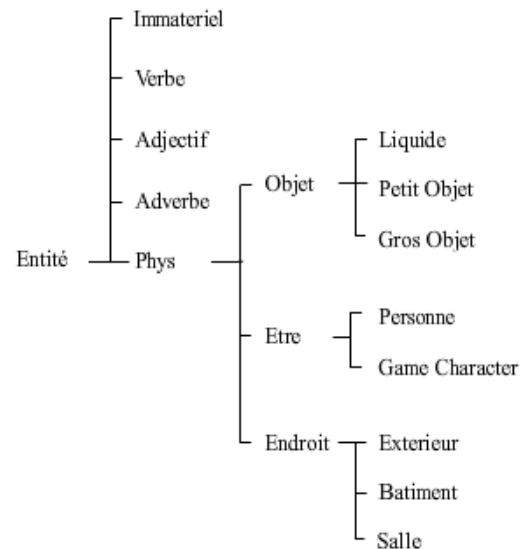


FIGURE 1 – Architecture de base de l'ontologie

instance, et que s'il s'agit d'un parent (hors Absurdité comme fils), alors on préférera l'imaginer comme une classe. Imaginons le concept *clef* héritant de *petit objet* et le concept *clef de la porte* héritant de *clef*. Le concept *clef* est alors parent, et le concept *clef de la porte* est alors une feuille. Lorsque nos algorithmes appelleront *clef*, ils auront tendance à écrire "une clef", mais lorsqu'ils appelleront *clef de la porte*, ils écriront "la clef de la porte" en y faisant référence directement puisqu'il s'agit d'une instance.

Les différentes relations. Pour que les algorithmes puissent faire émerger un sens de ces concepts organisés, il faut les mettre en rapport, et pour cela nous avons besoin des relations évoquées précédemment. Dans notre cas, ces relations sont binaires, et directionnelles. Une relation est définie par son type, et les concepts qu'elle relie (limitée à certains concepts, selon le type de la relation). Il existe de nombreux types : les types de placement qui permettent au designer de situer sa scène, et à cette description d'émerger dans l'histoire ; les types grammaticaux, définissant quel sujet et COD peuvent aller avec quels verbes ; les types qualificatifs (adjectif, adverbe) ; et les types plus sémantiques, comme *Obj_Vrb* liant un verbe avec un objet qui lui est souvent associé, comme *manger* et *fourchette*.

Notre but ici n'est pas de créer une liste exhaustive de type de relations, mais plus une architecture simple et pertinente, et de laisser libre cours à la création de différents types de relations dans le futur. De plus, les relations que nous avons

créées ici sont intimement liées au choix des algorithmes de génération de sens. Lors de la création de nouveau type de relations, il faudra les inclure dans la génération de sens de précédents algorithmes voir aussi de créer de nouveaux algorithmes.

En représentant notre monde par ses concepts, et en le décrivant par des relations, nous avons accès à un sens général. Ce sens permet de créer une masse de relation entre les concepts, décrivant plusieurs possibilités de notre monde, par exemple : “Le chapeau bleu est sur la commode” ou encore “Tuco a un grand chapeau”. Nos algorithmes créeront à la volée ce sens par association, et/ou choisiront parmi les associations préexistantes afin de créer un sens cohérent avec le scénario.

3.3 Algorithmes de Parcours

Deux des algorithmes de parcours sont particuliers : il s’agit des algorithmes d’introduction et de conclusion, afin de créer une ambiance au début, une attente, ainsi qu’un épilogue, qui permet un relâchement de la tension. Cinq autres s’occupent de gérer la tension au sein de l’histoire. Le premier de ces cinq : Réalisation événement, le plus simple, ne s’occupe que de récupérer la liste des événements, et de la traduire en une grammaire compréhensible par les algorithmes suivants. Il correspond au climax, point culminant du scénario, où la tension doit être au plus forte, avant de diminuer. L’algorithme Suspense part directement de la liste des événements, et essaye de construire une autre phrase qui va précéder les événements afin de construire une attente chez le spectateur, permettant de faire émerger une tension à travers l’attente de l’évènement. Son fonctionnement est explicité dans l’algorithme 1.

Algorithm 1 Algorithme Suspense

Require: *paragraph*

- 1: choose at random $stc = sentence$ in *paragraph*
 - 2: choose at random $cod' \in O \cup P$ so that $cod' \rightarrow Obj_Vrb \rightarrow stc.Verbe$ or $cod' = stc.COD$
 - 3: choose at random $Verbe' \in V$ so that $cod' \rightarrow Obj_Vrb \rightarrow Verbe'$
 - 4: create sentence :
 $Sente = \{stc.Sujet, Verbe', cod'\}$
 - 5: *paragraph.addSentence(Sente)*
-

Prenons un exemple, avec comme triplet issu de

notre paragraphe : {Tuco, to open, safe}. Nous devons instancier un nouveau *COD*, qui correspond soit au patient, soit à un concept lié au verbe par une relation *Obj_Vrb* ce qui dans notre cas donne : *safe* ou {*key*, *lockpick*}. Imaginons que l’on prenne *key*, nous en retirons une liste de verbe potentiel (grâce à la même relation) qui est {*to use*, *to watch*, *to play with*}, dont nous tirons *to play with*. En conservant le même agent, nous obtenons le triplet {Tuco, to play, key}, ce qui donne comme réalisation de phrase : “Tuco plays with the key. Tuco open the safe.”. On voit ici que le suspense est créé par une action triviale sur un objet qui est lourd en sens, et qui sera en effet utilisé lors de la phrase suivante (si ce n’est au niveau de l’écrit, du moins au niveau sémantique).

L’algorithme 2 Relation Sociale quant à lui part des personnages présents dans la scène, favorisant les acteurs, et selon leurs valeurs d’appréciation, de dominance, de familiarité et de solidarité génère une phrase. Plus que de décrire ces relations, nous avons voulu ici faire agir les personnages selon leur relation de manière simple et n’influençant pas le scénario (sourire, froncer les sourcils...).

Algorithm 2 Algorithme Relation Sociale

Require: *paragraph*

- 1: list all characters :
 $C = \{c \in P, c \in paragraph\}$
 - 2: choose at random $(c_1, c_2) \in C$
 - 3: get *SocRel* from (c_1, c_2)
 - 4: choose at random *SocRel.char* in {*Agreeability*, *Dominance*, *Solidarity*, *Familiarity*}
 - 5: **if** *SocRel.char* > *max* **then**
 - 6: *addSentence(SocRel.char is high)*
 - 7: **else if** *SocRel.char* < *min* **then**
 - 8: *addSentence(SocRel.char is low)*
 - 9: **end if**
-

Par exemple, imaginons un paragraphe formé des triplets {Tuco, to open, safe}, {Maria, to look, Tuco} et {Bank Man, to think, money}. La liste des personnages est {*Tuco*, *Maria*, *BankMan*}. Supposons que le hasard nous donne Maria et Tuco, nous nous intéressons donc à leur relation. Nous tirons au hasard une caractéristique parmi les quatre énoncées plus haut : {*Agreeability*, *Dominance*, *Solidarity*, *Familiarity*}. Une fois dominance choisie, il ne nous reste plus qu’à comparer sa valeur à un seuil minimal et maximal, qui diffèrent selon la caractéristique de la relation sociale

sur laquelle on se fonde. Dans notre cas, la dominance sera de 0.7 sur 1 dépassant le seuil de 0.6, en faveur de Maria. Nous créerons donc un triplet préassemblé, spécifique à ce seuil et cette caractéristique. Dans notre cas, cela sera le triplet : {Maria, to despise, Tuco}.

À ces algorithmes s'ajoute Stand-Alone, phrase seule, sans relation avec les événements précédents ou futurs, permettant de calmer le rythme, souvent trop rapide, car chaque événement OSSE est voulu comme ayant un fort impact émotionnel. Enfin, le dernier des cinq, Phrase Narrateur, permet à la fois au narrateur de transparaître, et de manifester ses émotions. Nous avons vu précédemment que l'importance du conteur dans le conte ne doit pas être négligée, voilà pourquoi nous avons voulu inclure cet algorithme. Dans ce cas, le narrateur émet un commentaire subjectif sur l'évènement qui vient de se passer, ce qui permet aussi de relâcher la tension après le climax.

Enfin, le dernier des huit algorithmes est un cas à part, il s'agit du Filling (Remplissage). Il parcourt les phrases créées précédemment, et cherche toutes opportunités dans l'ontologie, dans ses relations, et dans les propriétés d'OSSE pour ajouter des qualificatifs, tels que les adjectifs, des relations de positionnement, des traits de personnalité... Ces qualificatifs permettent de plonger l'histoire dans le réel. Ils donnent un historique aux objets qualifiés, et le spectateur ira à la recherche de cette histoire inventée à la volée. Les qualificatifs se rajoutent de manière aléatoire aux personnages, objets ou verbe, selon l'algorithme 3.

Algorithm 3 Algorithme Filling

Require: *paragraph*

```

1:  $\forall \text{sentence} \in \text{paragraph}$ 
2: choose at random  $k \in [0, 1000]$ 
3: if  $k > 0.35$  then
4:   Fill(sentence.getPersonne)
5:   execute Algorithme Filling again
6: else if  $k > 0.50$  then
7:   Fill(sentence.getPersonne)
8: else if  $k > 0.66$  then
9:   Fill(sentence.getObjet)
10: else if  $k > 0.83$  then
11:   Fill(sentence.Verbe)
12: end if
```

Par exemple, en partant du triplet {Tuco, to open, safe}, nous pourrions obtenir comme qualificatif *rusty* sur *safe*, *outlaw* sur Tuco et

quickly sur le verbe. Ce qui donnerait : “Tuco the outlaw quickly opens the rusty safe.”.

À cette suite d'algorithmes s'ajoute l'algorithme d'organisation qui gère l'appel des huit algorithmes que nous venons de décrire.

En plus de créer ces phrases, chaque algorithme ajoute un tag émotionnel, afin de permettre une narration plus chargée en émotion. Les phrases Stand Alone et relation sociales sont considérées comme neutres et le suspense est lié à l'émotion de peur ou d'espérance selon l'impact positif ou négatif des événements sur le narrateur. Pour ces derniers, le tag est défini par l'émotion perçue par le narrateur. Enfin, pour la phrase de narrateur, le tag correspond à une émotion dépendant de l'affection du narrateur pour le personnage en question, et de l'émotion que ce dernier ressent.

4 Résultats

4.1 Exemple

Nous avons utilisé plusieurs scénarios pour tester nos hypothèses. Le principal portait sur la thématique du Western, et se focalisait sur Clint le shérif, et Tuco le hors-la-loi. Le scénario débute par l'attaque de la banque par Tuco, et se finit par un duel. Nous allons ici montrer un exemple de phrase générée, en explicitant la part de chaque algorithme avant de montrer le résultat final. Chaque algorithme n'est pas automatiquement appelé lors d'une suite d'événements, mais nous avons choisi un exemple où c'est le cas afin de tous les montrer.

En entrée, nous avons les deux triplets suivants : {Tuco, to shoot, Clint} et {Clint, to shoot, Tuco}. Voici l'exemple disséqué de réalisation que notre programme a généré.

- Stand Alone : Clint thinks about himself.
- Relation Sociale : Tuco frowns to Clint.
- Suspense : Clint watches his gun.
- Évènement : Clint shoots Tuco. Tuco shoots Clint.
- Phrase narrateur : I wish I could have done something to help !
- Filling et Organisation : Clint peacefully thinks about himself near the saloon. Tuco frowns to Clint the sheriff. Clint watches his old gun. Then violent Clint and mean Tuco shoot each other. I wish I could have done something to help !

4.2 Scenarios

Les deux scénarios que nous avons choisi de créer pour tester nos algorithmes sont *Western* et *Policier*. Le premier a été déjà évoqué ; le second, traite d'un interrogatoire en cinq événements entre un policier et un voleur. Il a été présenté en détail dans un autre article [7]. Ces événements ont été créés à titre d'exemple d'évolution de relation sociale entre deux personnages. *Western* est notre scénario principal. En plus de Tuco et Clint, nous avons Maria, la shérif adjointe, un indien, un croque-mort, un prêtre et un guichetier de banque. Tous ces personnages possèdent une personnalité propre, définie par des mots clefs dans OSSE. Par exemple, le banquier est *dull* (ennuyeux), Tuco est *violent* et Clint *gentle* (doux). Comme nous l'avons vu dans le modèle OSSE, ils possèdent aussi des rôles : shérif, shérif adjoint, hors-la-loi, local ou stranger... Enfin, nous définissons leurs préférences (par exemple : Tuco a une vision moins négative du pistolet que les autres par exemple) ainsi que leurs relations sociales de départ (par exemple : Tuco et Clint ne s'aiment pas, Clint et Maria ont un fort sentiment de solidarité). Pour l'ontologie, une bibliothèque de concept spécifique à l'ambiance *Western* est rajoutée aux concepts standards définissant une bibliothèque standard minimale. Parmi ces concepts on retrouve *to shoot, saloon, sheriff, to duel, gun*.

Pour mieux voir l'effet de nos algorithmes, voici leur résultat sur un texte plus conséquent :

Avant : Tuco threatens Bank Man. Tuco could attack Bank Man. Bank Man gives the money. Tuco gets it. Tuco leaves the bank. Cherokee wounds Maria.

Après : Old Bank Man frowns to Tuco. Tuco plays with the gun and thinks about Bank Man. And after that, Tuco threatens and could slowly attack old Bank Man. I'm sure he would like to be as small as a mouse ! Strange Bank Man frowns to Tuco who loves helping. Tuco thinks about the money and Bank Man plays with it. Next, strange Bank Man gives the money and Tuco gets it. Look how proud he is ! Afterwards, Tuco the stranger leaves the bank. Show-off Maria the deputy sheriff dreads annoying Cherokee the stranger. Old Cherokee watches distressed Maria. Afterwards, annoying Cherokee wounds Maria the deputy sheriff. He shouldn't have done that, don't you agree ?

Le texte ainsi généré donne un aperçu des résultats que l'on peut attendre. Il nous reste à affiner la probabilité d'occurrence de chacun des algo-

rithmes ainsi que la variété des relations avant notre prochaine évaluation.

Planification de l'évaluation. Nous sommes en train d'organiser une évaluation pour tester l'influence de nos algorithmes sur les histoires racontées. Nous voulons comparer *sans algorithme* à *Suspense* seul, *Relation Sociale* seul, *Phrase du Narrateur* seul, *Filling* seul, ou *avec tous les algorithmes*. Cela nous permettra de voir l'impact global ainsi que celui de chacun des algorithmes majeurs. Les critères de jugements sont : *Surprising, Amusing, Boring, Repetitive, Captivating, Disturbing*. En plus de ces six critères, nous aurons des critères qui permettront de comparer les histoires entre elles : *Same* comparera la similarité, et *Preference* indiquera la préférence d'une histoire sur l'autre.

5 Conclusion

Nous venons de voir comment, à partir d'une ontologie, de ses relations, et d'algorithmes de parcours et de réalisation nous pouvons grossir le squelette d'une histoire, afin de lui donner une apparence que nous espérons plus réaliste, moins saccadée, et donc plus humaine. La création de suspense, sans pour autant créer une réelle intensité (qui est de la responsabilité du scénariste) lors de l'évolution de l'histoire, permet de jouer sur la tension le long de la narration. Le fait de manifester les relations sociales et leur évolution renforce l'immersion des spectateurs, ils se sentent plus concernés par ces personnages et par les péripéties qu'ils rencontrent. Enfin, les phrases du narrateur (pour ne parler que des principaux algorithmes) permettent d'humaniser le narrateur, et donnent une tout autre dimension à l'histoire. Notre future évaluation nous permettra de mettre à l'épreuve nos idées, et de nous guider dans notre recherche.

Les technologies employées ici permettent de rendre l'histoire plus riche et émotionnelle, sans être trop redondante, caractéristiques qui sont importante non seulement lors de sa narration, mais de manière beaucoup plus générale. Voilà pourquoi la représentation employée ici, ainsi que les algorithmes mis en oeuvre sont généralisables à tous les autres cas où un agent doit communiquer et où il ne possède qu'un squelette (établi ou généré automatiquement à chaque instanciation) de planification du discours (scénario, méthode d'argumentation, modèle basé sur des prédicats ...).

Perspectives. Notre objectif principal n'était pas de créer la meilleure ontologie possible ni de créer une liste exhaustive de relations, mais de montrer que notre méthode marchait et qu'une réalisation simple laissait voir les possibilités et les potentialités de cette architecture. Nous nous proposons ici de montrer des améliorations possibles.

Même si l'ontologie en tant que représentation des potentialités doit rester fixe, son interprétation doit être dynamique, et c'est en cela qu'il faudrait que les événements puissent créer à la volée des relations entre les concepts. Imaginons qu'un personnage se déplace, il faudrait que la liaison "Personnage dans Ville" soit détruite, et que l'on crée "Personnage dans Campagne".

Enfin, un aspect qui n'est pas négligeable est la création d'une interface graphique pour l'ontologie qui simplifierait sa gestion, ce qui fait qu'elle pourrait grossir sans pour autant devenir illisible, permettant l'utilisation de bibliothèque de concepts et relations. Selon l'ambiance de l'histoire, selon les personnages mis en scène, selon l'ambiance voulue, le scénariste pourrait en effet faire appel à une bibliothèque particulière (film noir, science-fiction, roman historique...). L'intérêt de cette notion de librairie serait de créer une communauté autour de ce système, afin de rentabiliser dans le partage les efforts de chacun à la complexification de l'ontologie.

Références

- [1] C. Peterson A. McCabe. What makes a good story? *Journal of Psycholinguistic Research*, 13, 1984.
- [2] J. Bates. The role of emotion in believable agents. *Communications of the ACM*, 37 :122–125, 1994.
- [3] J.-L. Lugrin D. Pizz, F. Charles and M. Cavazza. Interactive storytelling with literary feelings. *ACII2007, Lisbon, Portugal*, 2007.
- [4] J.-L. Dessalles. Vers une modélisation de l'intérêt. *Actes des journées francophones 'Modèles formels de l'interaction'*, 2005.
- [5] J.-L. Dessalles. Le rôle de l'impact émotionnel dans la communication des événements. *Actes des journées francophones 'Modèles formels de l'interaction'*, pages 113–125, 2007.
- [6] T. R. Gruber. *Towards Principles for the Design of Ontologies Used for Knowledge Sharing in Formal Ontology in Conceptual Analysis and Knowledge Representation*. Kluwer Academic Publishers, 1993.
- [7] N. Sabouret M. Ochs and V. Corruble. Simulation of the dynamics of non-player characters emotions and social relations in games. *Transactions on Computational Intelligence and AI in Games*, 2009.
- [8] A. Nijholt M. Theune, S. Faas and D. Heylen. The virtual storyteller. *ACM SIG-GROUP Bulletin*, 2002.
- [9] F. De Rosis and F. Grasso. Affective natural language generation. *Affective interactions*, 2000.
- [10] J.F. Sowa. Conceptual structures : information processing in mind and machine. 1984.
- [11] M. Stone. Ontology and description in computational semantics. 2004.

Agents virtuels vendeurs: que veulent les consommateurs ?

M.S Ben Mimoun
m.slim_benmimoun@skema.edu

I.Poncin
i.poncin@skema.edu

Skema Business School, LSMRC
Université Lille Nord de France
Avenue Willy Brandt - 59777 – Euralille - FRANCE

Résumé

En dépit d'une littérature académique et professionnelle très optimiste les agents virtuels incarnés sur les sites Web commerciaux ne semblent pas tenir toutes leurs promesses. La mise à jour en décembre 2009, d'un benchmark conduit par OrangeLab en 2007 sur 36 agents incarnés présents sur des sites Web français, révèle que plus de 60 % d'entre eux avaient disparus. Ceci soulève une question primordiale: Pourquoi la plupart des agents virtuels incarnés précédemment existants ont disparu des sites Web commerciaux ? Ce travail apporte des éléments de réponse à cette question, à travers une série d'entrevues en profondeur auprès de consommateurs, il permet également d'identifier les principales attentes des consommateurs vis-à-vis des agents conversationnels incarnés dans le contexte d'une activité commerciale.

Mots clés : Agents virtuels vendeurs, sites web commerciaux, consommateurs.

Abstract

In spite of a very optimistic academic and professional literature, virtual embodied agents on commercial Web sites do not seem to keep all their promises. The update in December, 2009, of a benchmark led by OrangeLab in 2007 on 36 embodied agents present on French Web sites, reveals that more than 60 % of them had disappeared. This research deals with understanding the disappearance of virtual sales agents through a series of in-depth interviews with consumers. It also identifies the main consumers' expectations towards the interactive embodied agents in the context of a commercial web site.

Keywords: virtual salesmen Agents, commercial Web sites, consumer behavior.

1 Introduction

Le magasinage sur des sites de e-commerce est souvent considéré par les consommateurs comme étant, froid, impersonnel et distant. Une façon d'y

remédier et de donner une touche plus humaine aux sites Web commerciaux consiste en la présence d'agents incarnés pour aider les clients (Galanxhi et Nah, 2007).

Holzwarth, Janiszewski et Neumann (2006) établissent un parallèle entre l'effet de présence de vendeurs en magasin et la présence d'agents commerciaux sur des sites Web commerciaux. Selon ces auteurs, si les vendeurs humains permettent d'augmenter la satisfaction, d'améliorer l'attitude envers les produits et d'accroître l'intention d'achat, la présence d'agents commerciaux virtuels sur un site Web devrait conduire à des résultats semblables. Au départ des travaux de Moon (2000), ils montrent ainsi que la présence d'un avatar pendant l'interaction homme-machine rend cette dernière plus interpersonnelle et réciproque.

Ainsi, la littérature académique semble unanime sur le fait que l'utilisation d'un agent animé interactif sur des sites Web pourrait fournir une source de présence sociale et un moyen efficace pour avoir la confiance de l'utilisateur (Wood, Solomon et Englis 2005; Keeling et McGoldrick 2008). Les professionnels semblent également s'accorder sur le potentiel des agents conversationnels incarnés à créer un « lien » avec les clients. Par exemple Tania Maria, le PDG de Daughter's Nature indique que "l'agent SitePal a ajouté une touche humaine à leur magasin en ligne et leur clients apprécient la qualité de l'accueil que leur réserve l'agent. Tania

Maria précise que l'agent présente une aide importante dans la création d'un lien avec le consommateur et permet également de réduire le « gap » qui existe entre l'expérience vécue dans un magasin en ligne et celle vécue dans une boutique réelle. »

Toutefois, en dépit de l'enthousiasme de la littérature tant académique que professionnelle, les agents virtuels incarnés ne semblent pas tenir toutes leurs promesses. En effet, le « Journal du net » (2009) présente les agents virtuels comme un des plus grands « flops » marketing des dix dernières années. En outre, la mise à jour en décembre 2009, d'un benchmark conduit par OrangeLab en 2007 sur 36 agents incarnés présents sur des sites Web français, nous a révélé que plus de 60 % d'entre eux avaient disparus. Malgré le développement de l'activité de quelques-uns (par exemple, Laura d'EDF ou Anna d'IKEA) et la création de nouveaux agents sur d'autres sites Web, seulement 13 des 36 agents incarnés précédemment étudiés par OrangeLab étaient toujours actifs deux ans plus tard. La plupart ont totalement disparu du Web (20 sur 36) et le reste ont été remplacés par une vidéo interactive (3 sur 36).

Le résultat de ce benchmark soulève une question primordiale: Pourquoi la plupart des agents virtuels incarnés précédemment existants ont disparu des sites Web commerciaux ?

Ainsi, ce travail cherche à apporter des éléments de réponse à cette question. A travers une série d'entrevues en profondeur auprès de consommateurs, il vise à évaluer les agents incarnés conversationnels existants en les comparant à d'autres technologies concurrentes (notamment les vidéos interactives). Ce travail cherche également à identifier les principales attentes des consommateurs vis-à-vis des agents conversationnels incarnés dans le contexte d'une activité commerciale.

A l'issue d'une brève revue de la

littérature sur l'apport et les rôles des agents conversationnels incarnés, nous exposerons notre méthodologie de recherche et discuterons les principaux résultats. Nous concluons sur les implications managériales et les voies de recherche.

2 Revue de la littérature

Une revue de littérature portant sur cinquante-quatre articles scientifiques, traitant des agents virtuels, adoptant une approche expérimentale et publiés entre 1995 et 2009, révèle que trois disciplines majeures se sont intéressées à l'étude des agents virtuels : les Systèmes d'information (SI), l'Interaction Homme-machine (HCI) et le Marketing.

Elle indique également des différences critiques entre les articles des trois disciplines. En effet, la majorité des articles en SI (8 sur 9) s'intéressent seulement aux agents non-incarnés et se sont principalement concentrés sur les aspects fonctionnels de ce type d'agent. Comme le soulignaient déjà Qiu et Benbasat (2009), la littérature en IS a surtout pris en compte la conception technique et la valeur utilitaire que procure l'utilisation des agents virtuels, ignorant le potentiel d'influence sociale que peuvent exercer ce type d'agents sur leurs utilisateurs.

Au contraire, les articles en HCI, que nous avons passés en revue, se concentrent sur les agents incarnés et plus spécifiquement sur leurs aspects physiques. Finalement, les articles marketing se sont intéressés aux deux types d'agents incarnés et ont non-incarnés et aux deux facettes de la conception des agents (fonctionnelles et physiques). On notera, cependant, une prédominance de recherches centrées sur l'impact de la présence versus l'absence d'un agent incarné. Ainsi, plusieurs études ont traité des Agents Incarnés Conversationnel et de leurs impacts sur le comportement du consommateur et sur

l'expérience d'achat (Cassell et al. 2000; Holzwarth et al. 2006; Wang et al. 2007; Jeandrain et Diesbach 2008). Elles suggèrent que le simple ajout d'un tel agent améliore, la satisfaction des clients, l'attitude envers les produits ainsi que l'intention d'achat. Papadopoulou et al. (2000) indiquent que pour les sites de e-commerce les caractères, « personnalisé », « autonome », « adaptatif » et « proactif » des agents incarnés permettraient un niveau élevé d'interactivité et l'expressivité nécessaire pour une expérience client effective et satisfaisante menant à une relation de confiance avec l'entreprise. Dans ce même courant, en faveur de l'incarnation des agents sur les sites Web, Dehn et Vanmulken (2000) présentent l'augmentation de la motivation et du niveau d'engagement comme les principaux effets et arguments en faveur de la présence d'un agent sous forme incarnée sur un site Web. Gulz et Haake (2006) défendent également cette idée et considèrent, qu'indépendamment du contexte de son utilisation, l'effet le plus saillant de la présence d'un agent animé sur un site Web consiste en sa capacité à rendre l'expérience d'interaction avec le site plus impliquante. Assadi et al. (2008) précisent, pour le cas particulier des sites financiers, que l'utilisation d'agents virtuels incarnés permet d'augmenter l'engagement des clients envers leur banque. En outre, un impact sur la valeur hédonique du magasinage ou de la navigation a également été observé. Ainsi, Wang et al. (2007) ont montré empiriquement que la présence d'agents incarnés sur un site Web commercial exerce un effet sur les valeurs hédoniques et utilitaires liées à l'expérience de magasinage via le site et influence par la suite l'intention de revisiter ou d'acheter via le site. McGoldrick et al. (2008) établissent quant à eux un lien direct entre la valeur hédonique de l'expérience de navigation sur un site Web commercial muni d'un agent incarné et l'augmentation de la satisfaction des consommateurs.

Ainsi, plusieurs auteurs présentent l'utilisation d'un agent incarné sur un site de e-business comme une stratégie pouvant améliorer la confiance et la satisfaction de l'Internaute voire comme un moyen d'attirer de nouveaux clients en favorisant un bouche à oreille positif (Notebaert, 2005; Lemoine et Notebaert, 2009). Gauzente et Guilloux (2003) soulignent en outre que le recours aux agents virtuels incarnés peut permettre de développer « l'identité institutionnelle de l'entreprise ».

Finalement, la littérature marketing identifie plusieurs rôles que peuvent jouer les agents virtuels incarnés sur un site Web commercial. La diversité des rôles est ainsi soulignée (Holzwarth et al. 2006) allant de l'aide commercial, au guide ou en passant par le compagnon de conversation.

Tout en reprenant la typologie de Paraschiv (2002), Diesbach et Galan (2006) considèrent qu'il existe trois types d'agents qui correspondent à trois types de rôle que peut jouer un agent virtuel sur un site Web commercial : les agents de recommandation, les agents de recherche de prix et les agents de négociation. Se référant à des travaux portant sur les vendeurs « humains » (Beatty et al. 1996; Mittal et Lassar, 1996; Baron et al. 1996; Gremler et Gwinner, 2000), McGoldrick et al. (2008) identifient trois rôles possibles des vendeurs virtuels en ligne : le rôle d'ami (rôle social), le rôle d'acheteur personnel (agent de recommandation) et le rôle d'aide.

Ainsi à la lumière des différents rôles et apports possibles des agents virtuels incarnés (Gauzente et Guilloux, 2003; Notebaert, 2007; McGoldrick et al., 2008), il est important de souligner que souvent l'agent incarné ne pourra pas produire tous les effets que nous venons d'énumérer et l'entreprise devra attribuer un « rôle » précis à l'agent qu'elle intégrera à son site (Diesbach et Galan, 2006 ; McGoldrick et al. 2008). Le rôle de « simple guide à la navigation » étant le plus fréquent sur les

sites Web de e-business (Semeraro et al. 2008; Diesbach et al. 2007). Cette réduction du rôle attribué aux agents virtuels incarnés sur le Web commercial pourrait apporter une première explication au « gap » énorme entre une littérature fortement optimiste et une réalité beaucoup plus morose. Mais cette dernière explication reste à elle seule insuffisante.

Poursuivant notre but d'une meilleure compréhension de l'échec connu par un nombre important d'agents incarnés conversationnels, nous présentons dans ce qui suit notre méthodologie de recherche.

3 La méthodologie de recherche

A ce stade de notre recherche, l'objectif de l'étude qualitative, vingt interviews en profondeur, réalisée auprès de consommateurs est une compréhension de leurs évaluations des agents incarnés conversationnels existants et de leurs attentes face à ce type de technologie.

Elles ont également permis de confronter l'évaluation des agents virtuels existants à celle des vidéos interactives. Notre échantillon était composé de 50% d'hommes et 50% de femmes âgés de 20 à 60 possédant des niveaux différents, d'utilisation d'Internet, d'expérience d'achat en ligne et de jeu en ligne. Chaque entrevue a duré entre 40 et 90 minutes.

Afin de garantir une certaine validité de nos résultats, nous avons demandé à chacun des interviewés lors de l'entrevue de manipuler un agent incarné conversationnel (Anna sur le site d'IKEA) et une vidéo interactive commerciale (les Pixexperts de Pixmania)¹. Nous avons procédé à un enregistrement audio et vidéo des entrevues que nous avons retranscrites

intégralement.

4 Résultats

Nous structurons les résultats ressortis de l'analyse des entretiens individuels avec les consommateurs autour de trois thèmes principaux correspondant à nos objectifs de recherche : l'évaluation des agents incarnés conversationnels existants (avec une focalisation sur Anna d'IKEA) ; l'évaluation des vidéos interactives commerciales (évaluation des Pixexperts) ; et l'identification des caractéristiques idéales d'un agent virtuel vendeur.

Tout d'abord en ce qui concerne l'évaluation des agents conversationnels animés et en particulier Anna d'Ikea, trois idées principales semblent ressurgir :

Le principe d'un agent conversationnel est valorisé en soi, notamment dans la sensation d'accéder à un interlocuteur : « c'est une espèce de hotline virtuelle, c'est un peu plus convivial, ludique, il y a un petit personnage avec une icône comme si on était sur MSN en train de chatter avec quelqu'un » ; « le site Ikea c'est assez foisonnant, il y a plein d'articles avec plein de noms bizarres suédois et c'est pas toujours très clair donc je trouve que c'est une bonne idée ».

Des capacités conversationnelles trop limitées pour satisfaire et donner envie de réutiliser : « où puis-je trouver une armoire 2 portes vitrées ? » : Anna comprend la question comme une demande de localisation de magasins » ; « je cherche une glace » : Anna affiche « je n'ai pas la possibilité de répondre à votre demande » ; Si je veux une table basse....ah Regardons le message Catégorie jouets et jeux enfants C'est bizarre, ça n'a rien à voir ».

Une évaluation globale négative et une faible valeur ajoutée : « Anna a servi à pas grand-chose, puisque ...fauteuil je l'aurais vu direct, ça aurait mis – longtemps que de demander à Anna » ; Oui, c'est informatif, oui, ça donne l'information mais si on veut

¹ Les pixexperts de « Pixmania » sont une sorte de comparateurs multicritères auxquels on a associé de la vidéo, utilisant uniquement du filtrage de contenu et des questions fermés.

plus de détails...Si ça répond en général... La réponse ne va pas avec la question que j'ai demandée...vous aimeriez connaître la matière...voilà elle prend des mots, voilà l'exemple typique »

En ce qui concerne l'évaluation des Pixexperts, ces derniers semblent bénéficier d'une meilleure évaluation qu'Anna d'IKEA alors que l'analyse des entretiens permet d'identifier deux idées principales :

D'une part les PixExperts génèrent de la satisfaction réussissent à prendre naturellement le rôle d'un vendeur (virtuel) : « j'ai l'impression que c'est comme Anna mais plus 'conseiller' comme dans un magasin » ; « on imagine que ça va être affiné au plus juste par rapport à notre demande » ; « c'est comme si on était chez Darty ou Conforama, comme si on avait quelqu'un en face de nous, c'est intéressant car c'est comme la question qu'on poserait en magasin et en fonction de ce que je vais demander au vendeur, c'est la même chose, c'est interactif »

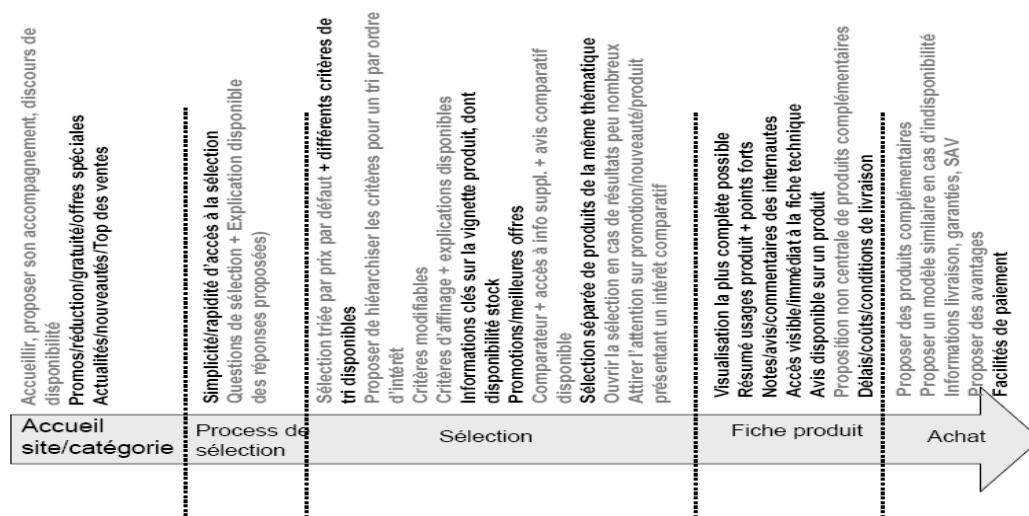
D'autre part la nature même des Pixexperts (vidéo interactive avec des question fermée) fait que les consommateurs leur associent un discours trop long : « il est un peu bavard », « si j'écoute tout j'ai l'impression que je vais rester des heures ! », « j'écoute pas tellement ce qu'il dit en

fait, je regarde plutôt l'écran, ce serait mieux qu'il parle pour expliquer quand on pointe sur quelque chose » ; ainsi qu'une démarche trop mécanique : « j'ai un peu l'impression qu'on va m'inciter dans un choix particulier, j'avais déjà vu cette fonction mais j'avais un a priori ».

Enfin, le troisième et dernier thème identifié suite à cette analyse porte sur les principales attentes des consommateurs face aux agents virtuels vendeurs. En effet, les consommateurs nous ont permis de dégager une série de caractéristiques que devra posséder le vendeur virtuel « idéal ». Ces caractéristiques concernent les cinq phases du processus d'achat sur le site Web : accueil du site, processus de sélection, sélection, lecture de la fiche produit, achat. En effet, à chaque phase de la navigation sur le site Web correspond une série d'attentes spécifiques du consommateur que nous résumons au niveau de la figure 1 et que nous discuterons au niveau de la section suivante.

Il est important de noter que les éléments présentés en noir au niveau de la figure 1 sont des attributs qui existent aujourd'hui au niveau des sites de e-commerce dotés d'agents virtuels incarnés alors que ceux en bleu correspondent à des attentes non satisfaites par la génération actuelle de vendeur virtuels.

Figure 1 : Les caractéristiques du vendeur virtuel idéal



5 Discussion et conclusion

L'analyse de nos résultats, en lien avec la littérature, met en lumière plusieurs explications permettant d'expliquer l'échec connu par un nombre important d'agents conversationnels animés sur le Web commercial.

La première explication est liée à la forme de la majorité des ACA. La littérature a souligné l'importance de la forme des ACA, l'apparence humaine apparaissant comme un bon moyen de créer une relation interpersonnelle avec l'agent. Cette apparence humaine conduit donc à des attentes élevées au niveau de la performance. Ainsi, les entretiens mettent en lumière que le fait qu'Anna possède l'apparence d'une hot-liner et qu'elle utilise un langage naturel conduit certains utilisateurs à l'assimiler à une personne humaine travaillant dans un call-center. Dès lors, ils en attendent une performance identique. L'incapacité d'Anna à comprendre parfaitement les requêtes des consommateurs malgré son apparence « humaine » est dès lors très décevante pour certains utilisateurs. Cette explication trouve un écho dans la littérature en HCI notamment par les travaux de Groom et al. (2009) ou de Nowak et Rauh (2008).

La deuxième explication est liée aux attentes en termes de compétences et d'expertise des agents. Le langage naturel favorise l'impression d'une conversation avec un humain toutefois la qualité/pertinence des réponses obtenues n'est malheureusement pas encore à la hauteur des attentes suscitées. L'expertise de l'interlocuteur, élément qui apparaît comme clé dans la description d'un vendeur idéal, est alors mise en doute par le consommateur. Ce manque de pertinence dans certaines des réponses explique sans doute la disparition de certains agents. Au contraire, le filtrage de contenu itératif employé par les Pixexperts accroît de manière importante la pertinence des réponses obtenues et donc la satisfaction de l'utilisateur. En effet,

contrairement à Anna par exemple, chacun des Pixexpert indique clairement son domaine d'expertise et argumente chacune de ses propositions. Ceci conduit tout d'abord à une réduction des attentes vis-à-vis de ce type d'agent (risque de déception plus faible) et ensuite à une meilleure reconnaissance de l'expertise de chacun des Pixexperts dans son domaine précis (assimilé à un vrai vendeur expert de son domaine). Dès lors, malgré le fait qu'ils ne soient pas dotés d'une forte interactivité ni de capacité à dialoguer en langage naturel, ces derniers semblent être mieux appréciés qu'Anna, pourtant agent considéré par les professionnels et les académiques comme très performant.

Toutefois, malgré leurs points positifs, ces vidéos ne parviennent pas à satisfaire complètement leurs utilisateurs et ne correspondent pas vraiment au vendeur virtuel idéal. En effet, les attentes des consommateurs vis-à-vis du vendeur virtuel idéal portent sur deux éléments majeurs : la sélection du produit et la finalisation de l'achat. Ces éléments correspondent aux maillons faibles des agents virtuels existants. Pour la sélection, les consommateurs semblent avoir trois exigences majeures : le caractère comparatif, l'argumentation et l'affinage des critères de sélection. Concernant la finalisation de l'achat, les consommateurs considèrent que l'agent devra être capable de présenter en même temps des produits complémentaires et des produits de substitution. En outre, pour être mieux accepté par les consommateurs, les agents virtuels vendeurs devront faire preuve d'une certaine empathie (accueil, reconnaissance du client et de ces goûts), d'une plus grande capacité d'écoute (capacité à identifier les vrais besoins des consommateurs et de proposer les solutions correspondantes) et surtout d'une capacité à argumenter les produits proposés aux consommateurs. Finalement, alors que bon nombre de professionnels et d'académiques ont accordé beaucoup d'importance aux aspects physiques des agents, aucune des personnes interrogées n'a exprimé d'attentes

relatives à l'aspect physique de l'agent. Ceci peut être interprété positivement comme synonyme d'une satisfaction par rapport à l'apparence des agents existants ou comme non pertinent pour le consommateur.

Bibliographie :

Assadi D., Notebaert J-F., Attuel-Mendes L., (2008), La qualité de l'interactivité dans la communication du service client sur les sites des banques : du courriel au web 2.0, 3ème Journées Internationales sur la Communication Marketing, Nancy,

Baron, S., Harris, K. et Davies, B. J. (1996), Oral Participation in Retail Service Delivery: A Comparison of the Roles of Contact Personnel and Customers, *European Journal of Marketing*, 30, 9, 75-90.

Beatty, S., Mayer, M., Coleman, J. Reynolds, K. et Lee, J. (1996), Customer-Sales Associate Retail Relationships, *Journal of Retailing*, 72, 3, 223-247.

Cassell, J., M. Stone, et H. Yan. (2000), Coordination and context-dependence in the generation of embodied conversation, *Proceedings of the first international conference on Natural language generation*, 171-178.

Dehn D. M. et S. Vanmulken (2000), The impact of animated interface agents: a review of empirical research, *Int. J. Human-Computer Studies*, 52, 1-22.

Diesbach, P.B., Galan, J.P.(2006), L'agent virtuel incarné dans la distribution en ligne: cadre théorique et revue de littérature, 6ème journée de E-marketing, Nantes, p. 24, Septembre.

Diesbach, B.P., Chandon, J.L. et Galan, J.P. (2007 a), Effets de la présence et de la congruence d'un agent virtuel incarné sur le pouvoir de rétention du site web, Actes du XXIIIème Congrès International de

l'Association Française du Marketing, Aix les Bains, CDRom.

Galanxhi H. et Nah F. (2007), Deception in cyberspace: A comparison of text-only vs. avatar-supported medium, *Int. J. Human-Computer Studies*, 65, 770-783.

Gauzente C. et Guilloux V. (2003), Marketing relationnel sur Internet : Quelle place pour les agents d'interface ?, 2ème journée de recherche sur le E-Marketing AFM-AIM, septembre, Nantes

Gremler, D. D. et Gwinner, K. P. (2000), *Customer-Employee Rapport in Service Relationships*, *Journal of Service Research*, 3, 1, 82-104.

Groom,V.,et al. (2009) ,Evaluating the effects of behavioural realism in embodied agents. *International Journal of Human Computer Studies*,doi:10.1016/j.ijhcs.2009.07.001

Gulz A. et , Haake M. (2006), Design of animated pedagogical agents—A look at their look” *Int. J. Human-Computer Studies*, 64, 322-339.

Holzwarth, M., Janiszewski, C. et Neumann, M. M. (2006), The Influence of Avatars on Online Consumer Shopping Behavior, *Journal of Marketing*, 70, 4, 19-36.

Jeandrain, A. et Diesbach, B. (2008), Immersion In An Online Merchant Environment: Are Consumers Ready To Feel Their Presence In Such Environment?, *Proceedings of the 37th EMAC Conference*, May, Brighton.

Journaldunet (2009), disponible à <http://www.journaldunet.com/>

Keeling, K. and McGoldrick, P. (2008), "Relationship with a Byte ? Attractions, Interaction and Intention for Avatar use a

retail website”, *Advances in Consumer Research* 35, 84-89.

Lemoine J. et Notebaert J. (2009), Agents virtuels, confiance envers les sites web et intentions comportementales des internautes, Colloque de l’AIM, Marrakech, octobre.

McGoldrick Peter J., Kathleen A. Keeling et Susan F. Beatty (2008), A typology of roles for avatars in online retailing, *Journal of Marketing Management*, 24, 3-4, 433-461.

Mittal, B. et Lassar, W.M. (1996), The role of personalization in service encounters, *Journal of Retailing*, 72, 1, 95-109.

Moon, Y. (2000), “Intimates exchanges: using computers to elicit self-disclosure from Consumers”, *Journal of Consumer Research*, 26 (4), 323-339.

Notebaert J.-F. (2005), L’interface homme-machine en commerce électronique : vers une création de lien social comme outil de positionnement stratégique, *Revue Française du Marketing*, 5, 205, 71-90.

Notebaert J-F (2007), La création d’une relation via Internet : un changement d’attitude des internautes propre à orienter la stratégie des entreprises, Colloque de l’ACFAS, Trois-Rivières, Québec, Canada

Nowak, K. and Rauh, C. (2008), “Choose your ‘buddy icon’ carefully: The influence of avatar androgyny, anthropomorphism and credibility in online interactions”, *Computers in Human Behavior*, 24, 1473-1493.

Papadopoulou P., A. Andreou, P. Kanellis, D. Martakos (2000), Enabling Trust Formation Within Agent-mediated Virtual Environments, *Virtual Reality*, 5, 149-159.

Paraschiv C. (2002) Gestion des agents intelligents sur l’Internet : l’apport conjoint du marketing et de l’économie, Thèse pour l’obtention d’un doctorat en sciences économiques, Ecole Normale Supérieure de Cachan, GRID, décembre.

Semeraro, G., Andersen, V., Andersen, H., de Gemmis, M. and Lops, P. (2008), “User profiling and virtual agents: a case study on e-commerce services”, *Univ Access Inf Soc*, 7, 179-194.

Wang, L., Baker, J., Wagner, J. et Wakefield, K. (2007), Can a Retail Web Site Be Social?, *Journal of Marketing*, 71, 143-157.

Wood, N., Solomon, M. and Englis, B. (2005), “Personalization of Online Avatars: Is the Messenger as Important as the Message?”, *International Journal of Internet Marketing and Advertising*, 2(1/2), 143-61.

Business-Driven Negotiation Strategies for Proactive Embodied Conversational Agents in E-Commerce

Sameh Abdel-Naby Maxime Morge Bruno Beauflis
sameh.abdel-naby@lifl.fr maxime.morge@lifl.fr bruno.beaufils@lifl.fr

Laboratoire d'Informatique Fondamentale de Lille
Université des Sciences et Technologies de Lille
59655 Villeneuve d'ascq cédex – France

Abstract

A considerable number of the research conducted in the area of Software Agents focus on the enhancement and the proper provision of online Embodied Conversational Agents (ECAs). However, the ability of these agents to transform an ordinary visitor of an e-commerce who needs assistance to an actual buyer is yet of no notable weight. In this paper, we adopt a sell-based approach for online customer service, so that the conversational agent is proactive since he initiates an offer without waiting for assistance requests. We introduce new negotiation strategies for embodied conversational agents which reflect real-life sales approaches, such as bundling two or more items into one deal, or providing a promotion on a certain item for a certain time, or the commonly used value-added sales.

Keywords: Embodied Conversational Agents, Negotiation, E-Commerce.

1 Introduction

Within the last twelve years e-commerce has succeeded to pursue a massive number of shoppers to change their idea of buying [5]. Several existing businesses have taken an advantage of this boom by adding a virtual presence to their physical one by means of an e-commerce website, moreover, new companies that exist only through the web have also appeared, *bricks and clicks* businesses, (e.g., Amazon). Although the online presence of companies is cost-efficient, yet the lack of a persuading salesman affects the transformation ratio (sales vs. visits). Then, several companies have started to embody a virtual assistant to aid potential online shoppers.

In Computer Science, several research efforts were made to study, analyze, and better shape the processes of assisting customers while being present in an e-commerce space [8]. In Artificial Intelligence, a considerable amount of the research conducted in the area of Software Agents [14] focus on the enhancement and the proper provision of online Embodied Conversational Agents (ECAs) [3].

Whether these agents sell, assist, or just recommend, it is now clear that such autonomous

agents are capable of engaging in verbal and non-verbal dialogues with e-commerce's customers. However, the ability of these agents to transform an ordinary visitor of an e-commerce who needs assistance to an actual buyer is yet of no notable weight.

In the context of ECA, we adopt a sell-based approach for online customer service, so that the conversational agent is proactive since he initiates an offer without waiting for assistance requests. Among several other advantages, making an agent proactive is necessary when it is important for the concerned online seller to maximize their benefits by driving specific conversations to certain goals. For example, turning a normal assistance dialogue to a sale or a proper product recommendation.

However, since most of the currently available ECAs are designed to ask questions and wait for answers, one of the major challenges we are facing these days is about the reversibility of the current options. Meaning, to reach a proper ECA proactivity / sales attitude, the questions an agent should ask to collect sales data should be placed in nowadays agents' answers, and the vice-versa. Consequently, for the ECAs of existing literature; the current design approach of agents' answers generation mechanism must be adjusted for a conversational agent who is in the process of asking questions too (proactive) and not just giving answers.

In addition, since the existing negotiation strategies for software agents are not adapted, we get inspiration from marketing literature in order to provide selling strategies for an ECA to convince a (human) buyer to buy a product or bundled products in one conversation.

This paper is structured as follows. Section 2 gives a formal description of the context we expect our negotiation strategies to be employed within. In section 3 we introduce the business-driven negotiation strategies. Section 4 gives an overview of the related research efforts. We then

conclude this paper and provide a summary of our future work in section 5.

2 The Online Negotiation Context

We assume in this paper that an e-commerce website is representing a specific company or a group of united companies. For each company, or a union of companies, there is a set of **Products** wherein each of its elements represents one sellable item; **Products** = $\{p_1, \dots, p_n\}$. We also refer to all of the companies represented online by means of an e-commerce websites as a set **Companies** = $\{c_1, \dots, c_k\}$. The total number of potential buyers that each may acquire an item $p \in \mathbf{Products}$ can also be referred to as the set **Surfers**.

Since an online buyer's visit to a specific website refers to his willingness to eventually acquire an item, then the interaction involving this buyer, the provisioning company, and the list of available products is a negotiation session. In the context of our research, we formally represent an **Online Negotiation Session (ONS)** as a tuple,

$$\langle T, s, Products, Companies \rangle$$

where

- T is the specific time a buyer started to browse the concerned website;
- s is the potential buyer, $s \in \mathbf{Surfers}$;
- $Products$ is the set of items a buyer is showing interest in.
- $Companies$ is the set of product provisioning entities, (e.g., the product manufacturer and its delivery partner).

Figure 1 shows a possible online negotiation session that is traditionally found in e-commerce websites that has no embodied conversational agents assisting or selling to potential buyers - surfers. From the set **Companies** we show a possible union between two companies, (e.g., a retailer and a delivery service company), which is likely to happen online. In the situation depicted in the same figure, we show that the entire negotiation is linked to a single product $p \in Products$, but in real-life situation, the shopping-cart may include multiple products, which is the dotted line linking two Ps. However, the potential buyer is commonly one.

However, the recent involvement of ECAs in most of the current online interactions between

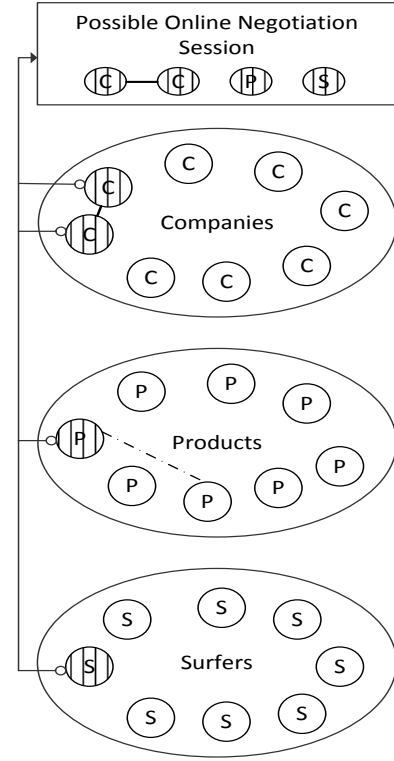


Figure 1: Possible Online Negotiation Session Involving Two United Companies

potential online buyers and companies' websites has made it clear that a mediator could exist. The existence of this mediator will lead us to widening the dimensions of the Online Negotiation Process (ONP) to also include this mediating agent.

Assuming that all mediating agents belong to the set **Agents**, we then formally represent the Mediated Online Negotiation Session (MONS) as a tuple

$$\langle T, a, s, Products, Companies \rangle$$

where,

- $\langle T, s, Products, Companies \rangle$ is an ONS as defined previously;
- a is the agent, $a \in \mathbf{Agents}$, mediating between $s \in \mathbf{Surfers}$ and one or more of **Companies**.

In figure 2, we show a possible online negotiation session while an ECA is mediating in-between the sellers and the buyers. Different from figure 1, in this figure we assume that one company of the set **Companies** is attempting to sell $\{p_1, p_2, p_3\} \in \mathbf{Products}$, therefore,

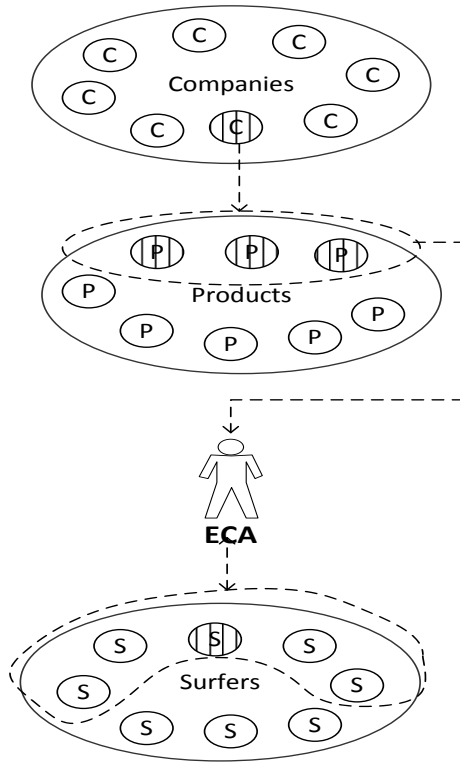


Figure 2: ECA Mediating in Possible Online Negotiation Session

$c \in \mathbf{Companies}$ is also delivering what it offers to online surfers.

Following the depictions of figure 2, the ECA is delegated by the company to sell or assist on a specific set of products online surfers are expected to show interest in. Therefore, the involvement of an ECA in any online negotiation session is always optional, depending on surfer's willingness to rely on a virtual assistant or seller, and whether the e-business imposes certain regulations to push online sales.

3 Negotiation Strategies

In this section, we introduce three different business-driven strategies for an ECA to convince a potential online buyer - *surfer* - to take a certain action, (e.g., buy). We start by the Bundling Strategy, and then the Promotion Strategy, and at the end we introduce the Value-Added Strategy.

3.1 The Bundling Strategy

Finding the best strategy to bundle a product is an attractive research issue that is being tack-

led by many scholars from different fields. For example, from economics we see some studies linking Product Bundling to market monopoly and other challenges in [1]. From computer Science, we see scholars relating the diversity of online shopping behaviors to the available bundling strategies in [15].

In the context of our research, we consider bundling as the action taken by a mediating entity to combine a set of products in order to facilitate reaching an agreement between two conflicting parties.

For example, considering a commonly encountered ONS that includes an e-commerce company called Taxilux, which is using DKL, as sub-company for delivering its products to the online shoppers. And, a surfer John that is showing an interest in a Leather Shoe. Therefore, this ONS will look similar to,

$$\langle T, John, LeatherShoe, \{Taxilux, DKL\} \rangle$$

In the above example, the involvement of a mediating entity - *ECA* - between the website's surfer John and Taxilux may make it possible to convert John's interest in the Leather Shoe to an actual buying desire. Assuming that ECA's approach here to finalizing this trade will be *Bundling*, then this MONS - *Mediated Online Negotiation Session* - will have {Leather Shoe, Green Socks, Delivery} in its **Products** set. Assuming that the Delivery Service have an actual cost, the ECA here have combined three different products into one bundle.

In our research, we consider it possible for an ECA to apply a bundling approach in a specific situation under two conditions. These conditions are:

- If bundling a product p_1 with p_2 of *Products* will make it possible to acquire any $p \in Products$ for no cost.

For example, assuming that it is a company's business rule to grant their online shoppers a free delivery if their total amount exceeds 10€, and the ECA is designed to handle such rule, in figure 3, *relation C* reflects this condition. Here, the ECA is combining the Shoe and the Socks so that John can qualify for a free delivery.

- If combining two or more products of the set *Products* will decrease the cost of buying at least one of them separately.

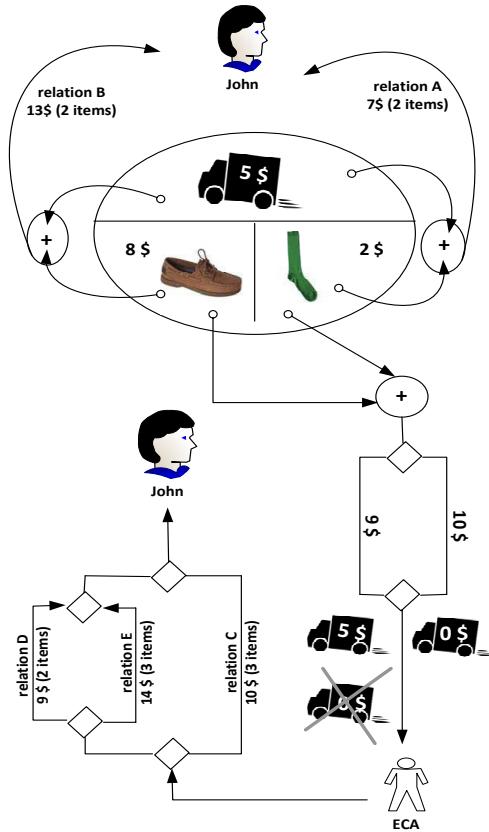


Figure 3: ECA and the use of bundling to generate new business offers

For example, in *relation D* of figure 3 the shoe can cost less than 8€ if it gets bundled with the socks. The reason some of the companies are relying on this approach goes back to the fact that the total net profit of the bundled items can be higher than selling each separately, even after applying a slight discount on one of them.

From figure 3 we can observe that the bundling option for the Embodied Conversational Agent (ECA) of this example have made it possible to generate three different offers for John, which may also lead to three different negotiation session and increase the sales potentiality. Although some of the generated bundling offers are not mutually beneficial, (i.e., *relation E*), yet an ECA can be designed to take advantage of such situations to generate as much profit as possible for the e-business and its partners only.

3.2 The Promotion Strategy

In real-life trading situations, it is quite common to walk into a store and find plenty of promotions on a range of products or services, (e.g., 5-

Year warranty on all TVs). Promoting a specific product takes place according to several business circumstances, which some of them can be:

1. the unexpected delay in selling a particular set of products with respect to the cost of their storage;
2. the sudden change of agreements made with business partners, (e.g., courier or maintenance companies), which will be allowing possible business extensions that did not exist earlier;
3. the necessity to increase the cash flow of the offering company in particular timeframe.

In our research, we look at product promotion as the action taken by a mediating entity to discount a certain product with respect to predefined condition in order to persuade a buyer's decision to conclude a specific deal.

For example, assuming that **KEX** is an electrodomeotics retailer that is having a deal with **Philcos** - TV manufacturer - to receive 100 sets every month for 10% discount per set. Since the TV market was not that alive in this specific month, **KEX** have decided to give a 5% discount on **Philcos** TVs for buyers of this month. In such situation, all involved parties are benefiting from the promotion since the buyer is taking 5% off, **KEX** is making 5% less of their profit but still there is a profit, and **Philcos** keep on manufacturing its TVs.

Online, an example like the above will be happening on the e-commerce website of KEX through banners that advertise such promotion, or newsletters, and then it is up to surfers of this website to chose whether to click on the banner or no. Therefore, an Online Negotiation Session (ONS) for such commonly encountered situation will look similar to,

$$\langle T, Bill, TV, \{KEX, Philcos, DKL\} \rangle$$

However, involving an ECA into the above **ONS** will make the **Products** set increase according to the value previously defined by KEX to push this certain product into the market. Therefore, the newly emerged **MONS** will have {TV, -5%} in its set of **Products**.

In order for us to introduce the condition under which an ECA would apply the promotion negotiation strategy, we will first give a brief explanation on how a product price is calculated. Considering the fact that for each item of the set

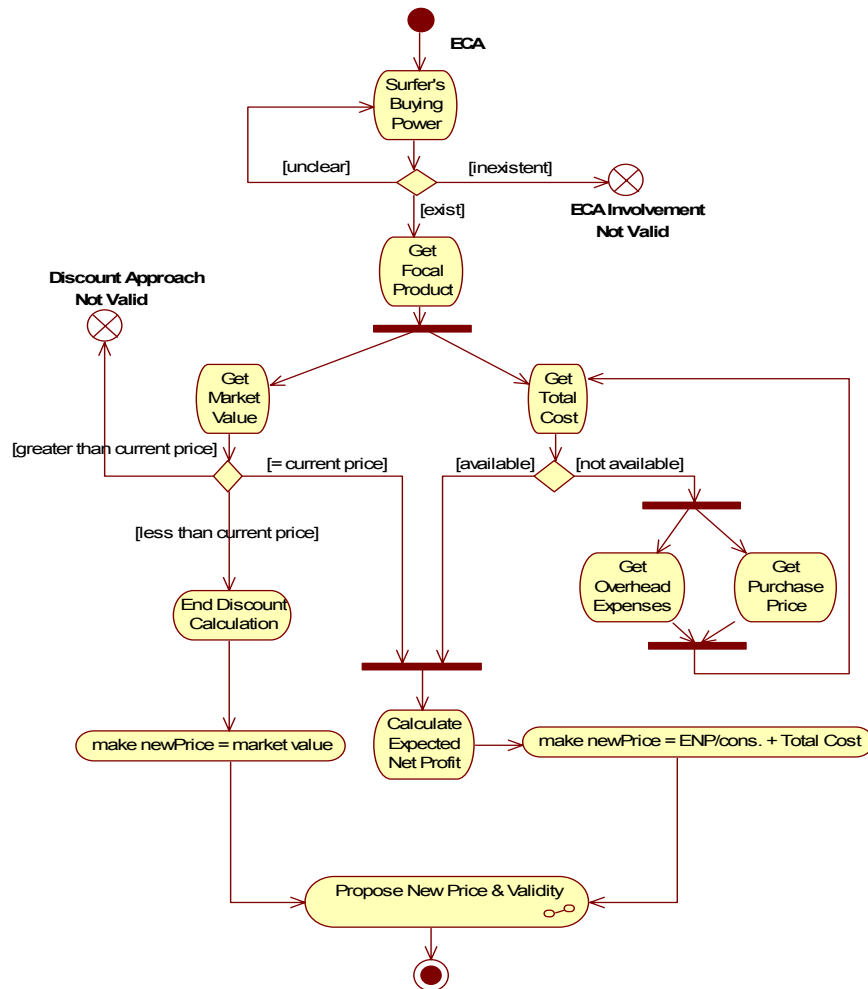


Figure 4: ECA's activities to obtain a new promotional price for a potential buyer

Products of any company there must be a *Market Value* and an *Estimated Overhead Expenses*.

In our research, we look at *market value* as the maximum price given by a company to a product without exceeding the price of buying the same product or similar from competitors. And, in our research, we look at a product's *estimated overhead expenses* as the average cost of all hidden services that are made by the seller in order to prepare an offer to a potential buyer.

For example, the market value of a used BMW X5 cannot exceed or equal the price of buying a new one and, the price of a new BMW X5 cannot also be very much far from the price of similar cars offered by competitors.

We then consider that a *Product Price* is defined by summing the actual purchase price, estimated overhead expenses per unit, and the desired net profit. Therefore, a product discount is valid if the discounted price is less than the mar-

ket value and greater than or equal to the total sum of the *actual purchase price* and *estimated overhead expenses*.

In figure 4, using a UML activity diagram we summarize the steps we expect an ECA to take in order to examine the possibility to obtain a product promotion, (i.e., a discount), for a potential buyer. Assuming that an external function exists, (e.g., a triggering engine), that is responsible of calculating the buying power - *potentiality to purchase* - of a surfer according to his/her behavior through a specific e-business. Once a sufficient buying power regarding a specific product is observed by this triggering engine, it is now the ECA turn to approach the potential buyer.

As long as the buying power of a certain product is *inexistent* or *unclear*, (i.e., less than a predefined rate), there will be no need for the ECA to approach the surfer. Eventually, the ECA will start calculating the new promo-

tional price by retrieving the *Market Value* and the *Total Cost* of the focal product. For example, the market value of $p_1 \in \text{Products}$ equals to 20€ since it is the maximum price a buyer can find for it in the market. And 10€ is the actual total cost of p_1 after summing its purchasing cost to its estimated overhead expenses, (i.e., delivery, storage, ... etc).

During the ECA's retrieval process of the market value a brief sub-operation is expected to occur in order to position the Current Price of p_1 with respect to competitors. If the current price is greater than market value, we then assume that it will be feasible for the ECA to make the new price of p_1 equal to the current market value. And, if the current price of p_1 is less than its market value the overall Promotion Strategy will not be of great use since higher prices elsewhere will auto-promote it.

However, if the market value of the concerned product is equal to its current price, then it is still valid for an ECA to apply this strategy. Following to that, a step wherein the **Expected Net Profit (ENP)** is calculated by retrieving the difference between p_1 's market value - *also current price in this case* - and its actual total cost.

Here, the *newPrice* will then be the ENP divided by a **constant** value - *defined by each company separately* - and then added to the total cost. For example, assuming that p_1 's market value is equal to its current price that is 20€, and its actual total cost is equal to 10€, then p_1 's ENP is equal to 10€.

Further on, assuming that KEX has its own business rule which says that ECAs can deduct 50% of a product's profit to generate its promotional price, then the constant will be equal to 2. Therefore, as an example, the equation allowing the ECA here to generate a *newPrice* for p_1 will be: $10/2 + 10$, that is 15 €.

3.3 The Value-added Strategy

In [11], Reilly presents the value-added sales as an approach a seller may employ to be proactive while negotiating a deal. Reilly also confirms the confusion found in literature when it comes to defining value-added sales in general. However, it is commonly found that buyers perceive value-added items as those acquired for free on top of their purchases. On the other hand, sellers whether they are service providers or retailers, a value-added item for them must have low

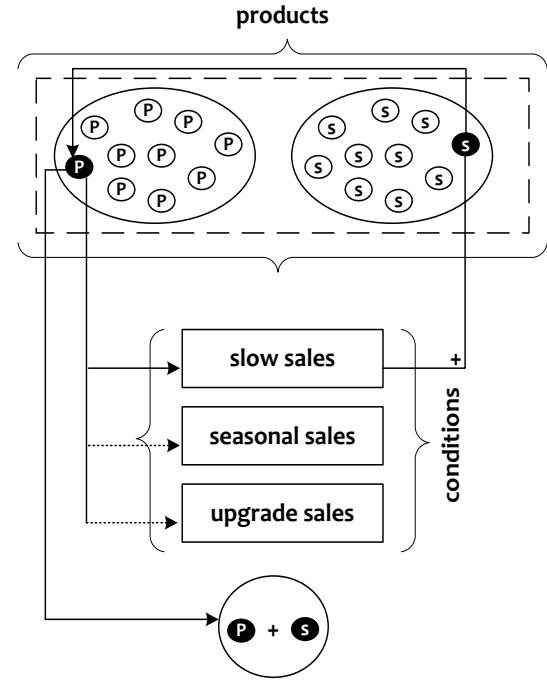


Figure 5: A value-added sale: products, conditions and services.

cost and low profit, and once attached to another item with greater profit both are then better sold to customers or pushed to market.

For example, a free memory card on top of any purchased digital camera is a value-added item. However, when the cost of this card is covered by the profit generated from selling the camera itself it is then looked at as a marketing strategy. Discounting the price of the camera itself is another marketing strategy that does not involve any value-added services but lead to the same goal.

In our research, an agent's decision whether to make use of the value-added sales approach or not is linked to the validity of at least one of the following conditions:

- **slow sales**: if the last time of selling an item that is similar to the negotiated one has reached a certain undesirable period. For example, if an online shopper is showing an interest in a TV that a similar one was last sold three months back, which is undesirable non-selling period for the seller. Then the seller agent may then be instructed to make use of the value-added approach.
- **seasonal sales**: if the product is designed to fulfill the needs of a specific season that is approaching an end within a period that is predefined by the selling entity. For example, if

the negotiated product is an air-condition and, as predefined by the seller, summer sales end in two weeks.

- **upgrade sales:** if the negotiated product is about to be replaced with an alternative one that is of the same money value but yet with higher functionality. For example, expanding a car's warranty to four or five years few months before introducing a new shape of the same model.

As it was previously assumed that each $c \in Companies$ is linked to a set of *products* that are all offered to online companies. In this section, we assume that two different types of products exists - *Products Subsets*. A subset represents the core products - *Products* - of a company and, the other subset - *Services* - represents the secondary ones. As described in figure 5, a value-added sale for us will have to combine an item or more of the *Products* set with an item or more from the *Services* set; while keeping the total cost of the sale equal to sum of prices from the *products* set only.

For example, assuming that a potential online buyer is showing an interest in a digital camera that costs 200€ and a TV that costs 300€. Assuming that both items are under one or two of the conditions we mentioned above. In such scenario, a seller agent will then pick one or two items of the *Services* set, for instance; 2 years warranty & free delivery, add them to the camera and TV, then offer them all for 500€.

4 Related Work

Rahwan et al. [10] propose an analysis grid of strategies for agents engaged in negotiations. According to this grid, the factors which influence our strategy are: the goals (an optimal profit here), the domain (the products), the negotiation protocol (a one-shot conversation), the abilities of agents (sell services here), the values (cooperativeness of the agent).

Few concrete strategies of agents engaged in negotiations have been proposed. Jeffrey S. Rosenschein and Gilad Zlotkin have proposed a monotonic concession strategy for bilateral negotiations in [12]. In this protocol, each agent starts from the deal that is best for him and either concedes or stand stills in each round. A (monotonic) concession means that an agent proposes a new deal that is better for the other agent. Sierra et al. [13] propose different strategies based on arguments such as threats, rewards

or appeals (e.g. to authority). It is clear that these strategies are adapted for automated (bargaining) negotiation amongst software agents in the context of business-to-business. By contrast, we have proposed selling strategies for an ECA to convince a (human) buyer to buy a (bundle of) product(s) in a one shot conversation.

Rahwan et al. distinguish in [9] different approaches for automated negotiation, including game-theoretic approaches [12] (which usually assume complete information and unlimited computation capabilities), heuristic-based approaches [4] (which try to cope with these limitations) and argumentation-based approaches [2] (which allow for more sophisticated forms of interaction). By adopting a game-theoretic approach, our negotiation model is simple. It should be interesting to adopt an argumentation-based approach. In this way, agents may deal naturally with new information in order to mutually influence their behaviors. By arguing (even if it is internally), parties can take into account the information given by the buyer, and then let the agents make some points to convince the buyer.

Authors of [6] and [7] highlights the significant role Embodied Conversational Agents (ECA) play in nowadays e-commerce websites, and how large is the potential for these ECAs to replace real-life human interactions, which also include business-related ones. However, in [6] the main focus was on introducing and using the Virtual Reality Modeling Language (VRML), which is one of the available tools used to manage 3D environments while considering players of this environment as ECAs.

5 Conclusions & Future Work

In this paper, we introduced the idea of applying business inspired selling approaches to the negotiation strategies Embodied Conversational Agents (ECAs) can employ to achieve specific deals online. In order for us to present this notion, we briefly presented a formalization for the context we imagine these kind of negotiations to take place. We looked at ECAs as the mediating entities between the selling players, (i.e., companies), and the potential buyers, (i.e., certain websites' surfers). We then presented our business-driven negotiation strategies that reflect real-life product bundling, promotions, and value-added sales.

As an extension to our work, we aim for build-

ing a simulation environment that reflects the correct behavior of websites' surfers, and also reflects the needs of e-commerce websites. With that simulation in hands, we will look for evaluating these negotiation strategies in more practical way, and tuning them up and down until a proper design schema is reached.

References

- [1] William James Adams and Janet L. Yellen. Commodity bundling and the burden of monopoly. *The Quarterly Journal of Economics*, 90(3):475–498, August 1976.
- [2] Leila Amgoud, Yannis Dimopoulos, and Pavlos Moraitis. A unified and general framework for argumentation-based negotiation. In *Proc. of AAMAS*, pages 963–970, Honolulu, Hawaii, 2007.
- [3] Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill. *Embodied conversational agents*. MIT Press, Cambridge, USA, April 2000.
- [4] Peyman Faratin, Carles Sierra, and Nick R. Jennings. Negotiation decision functions for autonomous agents. *International Journal of Robotics and Autonomous Systems*, 24:3–4, 1998.
- [5] R. Hof, H. Green, and L. Himmelstein. Now it's YOUR WEB. *BusinessWeek*, pages 68–75, October 1998.
- [6] Helen McBreen. Embodied conversational agents in e-commerce applications. In Gerhard Weiss et al., editor, *Socially Intelligent Agents*, volume 3 of *Multiagent Systems, Artificial Societies, and Simulated Organizations*, pages 267–274. Springer US, 2002.
- [7] Hazel Morton, Helen McBreen, and Mervyn Jack. Experimental evaluation of the use of ecas in ecommerce applications. *From brows to trust: evaluating embodied conversational agents*, pages 293–321, 2004.
- [8] Luigi Palopoli, Domenico Rosaci, and Domenico Ursino. Agents' roles in B2C e-commerce. *AI Communications*, 19(2):95–126, 2006.
- [9] I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. Argumentation-based negotiation. *The Knowledge Engineering Review*, 18(4):343–375, 2003.
- [10] Iyad Rahwan, Peter McBurney, and Liz Sonenberg. Towards a theory of negotiation strategy (a preliminary report). In *Proc. of the AAMAS Workshop on Game Theoretic and Decision Theoretic Agents (GTDT)*, pages 1–8, Melbourne, Australia, 2003.
- [11] Tom Reilly. *Value-Added Sales Management*. McGraw-Hill, April 1993.
- [12] Jeffrey S. Rosenschein and Gilad Zlotkin. *Rules of encounter: designing conventions for automated negotiation among computers*. MIT Press, Cambridge, MA, USA, 1994.
- [13] Carles Sierra, Nick R. Jennings, Pablo Noriega, and Simon Parsons. A framework for argumentation-based negotiation. In *Proc. of ATAL*, volume 1365 of *LNCS*, pages 177–192, Provident, RI, 1998. Springer.
- [14] Michael Wooldridge and Nicholas R. Jennings. Intelligent agents: Theory and practice. *Knowledge Engineering Review*, 10:115–152, 1995.
- [15] Tzyy-Ching Yang and Hsiangchu Lai. Comparison of product bundling strategies on different online shopping behaviors. *Electronic Commerce Research and Applications*, 5(4):295 – 304, 2006.