Moteurs de personnalité pour agents dialogiques : étude d'un modèle pour les traits interpersonnels

J. P. Sansonnet[a] jps@limsi.fr
F. Bouchet[b] francois.bouchet@mcgill.ca
N. Sabouret[a] nicolas.sabouret@limsi.fr

[a]LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France

[b]SMART Laboratory, McGill University, 3700 McTavish, Montreal, Canada

Résumé

Nous présentons une approche flexible et déclarative pour l'implémentation informatique de traits de personnalité issus de la taxonomie FFM/NEOPI-R. *Elle repose sur le principe que les traits activent des opérateurs influençant le processus de décision rationnelle des agents. Notre approche est validée sur une étude de cas portant sur la sous-classe des traits interpersonnels, appliquée aux agents dialogiques.*

Mots-clés : *Modélisation cognitive, Psychologie computationnelle, Traits de personnalité.*

Abstract

We present a flexible and declarative approach to the implementation of personality traits from the FFM/NEOPI-R *taxonomy. This approach is based on the principle that personality traits influence the rational decision making process of the agents. The proposition is validated upon a case study involving the sub-class of interpersonal traits, applied to dialogical agents.*

Keywords: *Cognitive modeling, Computational psychology, Personality traits.*

1 Introduction

Dans leurs travaux pionniers, Rousseau et Hayes-Roth [15, 16] proposent de transposer aux agents artificiels le principe psychologique selon lequel les traits de personnalité ont une influence directe sur le comportement rationnel. Le domaine des systèmes multi-agents s'est intéressé à l'implémentation des *capacités cognitives* des humains dans des architectures logicielles. De ces recherches sont ressorties deux grandes familles d'architectures logicielles : SOAR/ACT-R et BDI. Par exemple, le modèle eBDI [9] implémente un modèle d'émotions dans un moteur BDI. CoJACK [6] ajoute un niveau destiné à la simulation de phénomènes physiologiques et cognitifs humains comme par exemple la limitation mémorielle. Tous ces travaux font apparaître la coexistence de modules de raisonnement à côté de modules psychologiques.

Plusieurs travaux ont proposé de modéliser une influence des traits de personnalité humaine sur le comportement d'agents conversationnels. Ainsi, Malatesta *et al.* [10] utilisent des traits pour engendrer l'expression des émotions dans des agents virtuels, en se fondant sur le modèle de déclenchement OCC [12]. De même, les travaux précurseurs de Rizzo *et al.* [14] ont montré que dans le cadre BDI, la phase de passage des désirs aux buts pouvait être influencée par la personnalité. Plus récemment, les travaux autour de l'architecture d'agent conversationnel animé GRETA [13] et son modèle de raisonnement FATIMA [5] proposent une opérationnalisation de l'influence des traits de personnalité.

Ces travaux se divisent en deux catégories :
– Les architectures multi-agents qui mettent l'accent sur la prise en compte des *capacités* physiologiques ou cognitives. Elles étudient l'impact des propriétés de l'agent sur l'action dans le monde (direction agent $\rightarrow$ monde) ;
– Les architectures d'agents virtuels qui mettent l'accent sur le déclenchement et l'expression corporelle (faciale, gestuelle) des *émotions*. Elles étudient l'impact des évènements du monde sur le modèle mental des agents (direction monde $\rightarrow$ agent).

Le travail présenté dans cet article se situe dans le cadre des architectures multi-agents dont la direction est agent $\rightarrow$ monde et notre originalité tient au fait que nous traitons les phénomènes psychologiques plutôt que les capacités de raisonnement rationnel. De plus, dans toutes les études précédentes de type agent $\rightarrow$ monde, les auteurs mettent l'accent sur des phénomènes ciblés sur des études de cas, ce qui pose un problème de généricité. Notre objectif est donc d'essayer de couvrir le domaine psychologique le plus complet possible, ce qui nous oblige à aborder les questions de gestion et d'intelligibilité du processus d'implémentation des phéno-

TABLE 1 – Schémas de TFS.

FFM	FACETS	Classe	+ POLE	- POLE
O	Fantasy	□	IDEALISTIC	PRACTICAL
O	Fantasy	□	SPIRITUALIST	MATERIALIST
O	Fantasy	□	CREATIVE	UNIMAGINATIVE
O	Aesthetics	□	AESTHETE	TASTELESS
O	Feelings	□	SENTIMENTAL	UNSENTIMENTAL
O	Actions	□	ADVENTUROUS	UNADVENTUROUS
O	Actions	□	ADAPTIVE	MALADAPTIVE
O	Ideas	□	CULTURED	UNEDUCATED
O	Ideas	★	BROADMINDED	BIGOTED
O	Values	★	PROGRESSIVE	CONSERVATIVE
O	Values	★	PERMISSIVE	AUTHORITARIAN
C	Competence	□	ASSURED	INSECURE
C	Competence	□	DECISIVE	HESITANT
C	Orderliness	□	METHODICAL	DISORGANIZED
C	Orderliness	□	TIDY	MESSY
C	Dutifulness	★	LOYAL	UNFAITHFUL
C	Dutifulness	★	VIRTUOUS	IMMORAL
C	Dutifulness	★	LAWABIDING	LAWLESS
C	Dutifulness	★	TRUTHFUL	UNTRUTHFUL
C	Dutifulness	★	FAIR	PREJUDICED
C	Self-discipline	□	CONCERNED	UNCONCERNED
C	Self-discipline	□	ATTENTIVE	INATTENTIVE
C	Self-discipline	□	MINDFUL	FORGETFUL
C	Achievement-striving	□	AMBITIOUS	UNAMBITIOUS
C	Achievement-striving	□	HARDWORKER	LAZY
C	Deliberation	□	STRATEGIC	TACTICAL
C	Deliberation	□	COMMONSENSE	UNSOUND
E	Warmth	★	FRIENDLY	COLD
E	Warmth	★	KEENINTERESTED	NONGOSSIPMONGER
E	Gregariousness	★	TALKATIVE	UNCOMMUNICATIVE
E	Gregariousness	★	COMPANIONABLE	SOLITARY
E	Assertiveness	★	CHARISMATIC	UNCHARISMATIC
E	Assertiveness	★	DOMINEERING	SUBMISSIVE
E	Assertiveness	★	INTIMIDATING	PLEADING
E	Assertiveness	★	SHOWY	DISCRET
E	Activity	□	LIVELY	LANGUID
E	Activity	□	ACTIVE	APATHETIC
E	Excitement-seeking	□	LUXURIOUS	ASCETIC
E	Positive-emotions	□	JOYFUL	BLASE
A	Trust	★	TRUSTFUL	SUSPICIOUS
A	Trust	★	UNGUARDED	SECRETIVE
A	Straightforwardness	★	HONEST	DISHONEST
A	Straightforwardness	★	FRANK	INSINCERE
A	Altruism	★	GENEROUS	GREEDY
A	Altruism	★	COOPERATIVE	ANTAGONISTIC
A	Altruism	★	SELFSACRIFYING	SELFISH
A	Compliance	★	AFFABLE	INSUFFERABLE
A	Compliance	★	CONCILIATORY	CONFRONTATIONAL
A	Compliance	★	TOLERANT	OFENSABLE
A	Compliance	★	FORGIVING	REPROACHING
A	Modesty	★	HUMBLE	ARROGANT
A	Tender-mindedness	★	COMFORTING	HARSH
A	Tender-mindedness	★	HARMLESS	HARMFUL
A	Tender-mindedness	★	EMPATHIC	INSENSITIVE
N	Anxiety	□	WORRIED	RELAXED
N	Anxiety	□	PESSIMIST	OPTIMIST
N	Anxiety	★	DEMANDING	SELF-ASSURED
N	Angry-Hostility	□	CHOLERIC	NONCHOLERIC
N	Angry-Hostility	□	IRRITABLE	CAREFREE
N	Angry-Hostility	★	MALICIOUS	MERCIFUL
N	Impulsiveness	□	INSTINCTIVE	DELIBERATE
N	Impulsiveness	□	FIERCE	PLACID
N	Impulsiveness	□	MOODY	STABLE
N	Impulsiveness	□	TEMPTABLE	UNTEMPTABLE
N	Depression	□	MELANCHOLIC	CHEERFUL
N	Depression	□	REMORSEFUL	UNPENITENT
N	Self-consciousness	★	SHY	OUTGOING
N	Vulnerability	□	EMOTIONAL	IMPASSIVE
N	Vulnerability	□	HOPELESS	HOPEFUL
N	Vulnerability	□	COWARDLY	COURAGEOUS

Sémantique des schémas détaillé à : http ://perso.limsi.fr/jps/research/rnb/toolkit/taxo-glosses/taxo.htm
Classes : □ Intrinsèque $N_{s\square} = 36$ ★ Interpersonnel $N_{s\bigstar} = 34$

mènes psychologiques dans les agents.

L'article est organisé comme suit. La section 2 présente une architecture dont la flexibilité et l'intelligibilité sont fondées sur la notion de moteurs de personnalité. La section 3 présente l'application de cette architecture aux traits interpersonnels, dans un cadre de modèle d'agents dialogiques. La section 4 développe un exemple illustratif, allant de la définition d'une personnalité d'agent à l'impact qu'elle peut avoir sur son comportement effectif.

2 Architecture

2.1 Domaine de personnalité couvert

La taxonomie TFS. Nos travaux s'appuient sur la taxonomie de traits de personnalité FFM/NEOPI-R [8] qui est l'une des plus répandues dans les études psychologiques. Elle est composée de deux niveaux de concepts :
– Le premier niveau, dénoté FFM (pour « Five Factor Model »), est composé de cinq classes de comportements psychologiques (appelés *traits*) qui sont listées dans la colonne 1 de la Table 1 : **O**penness, **C**onscientiousness, **E**xtraversion, **A**greeableness, **N**euroticism. – Au second niveau, chaque trait est décomposé en six sous-classes (appelés *facettes*) donnant ainsi 30 positions bipolaires [1] [4], qui sont listées dans colonne 2 de la Table 1. La sémantique de chaque facette est définie de manière intuitive au moyen d'une glose [2].

Toutefois, lorsqu'on s'intéresse à l'expression computationnelle des traits de personnalité, les gloses uniques définissant les facettes de FFM/NEOPI-R s'avèrent trop générales. C'est pourquoi on ne peut pas se limiter au deuxième niveau et nous avons proposé dans des travaux précédents [1] une taxonomie enrichie, appelée TFS (**T**raits-**F**acettes-**S**chémas), qui étend FFM/NEOPI-R avec un troisième niveau divisant les facettes en sous-cas opérationnellement distincts, appelés *schémas comportementaux* (ou *schémas*) [3]. Dans TFS, chaque schéma est défini non plus par une seule glose mais par un ensemble de gloses qui proviennent de deux sources : 1) elles sont tirées des synsets WordNet [7] qui sont associés à un ensemble de 1055 adjectifs de personnalité ; 2) puis complétées et alignées avec 300 éléments (appelés *q-items*) de questionnaires de personnalité de Goldberg [4]. Chaque schéma possède deux ensembles de gloses définissant un comportement opérationnel congruent : le pôle + correspond au concept du schéma et le pôle - à l'antonyme du concept. Quantitativement TFS est définie par : $N_{\text{facettes}} = 30$, $N_{\text{gloses}} = 766$, $N_{\text{schémas}} = 70$, $\overline{N}_{\text{gloses/facettes}} = 26$ et $\overline{N}_{\text{schémas/facettes}} = 2.3$.

Schémas intrinsèques et schémas interpersonnels. Lors de l'étude approfondie des gloses attachées aux schémas de la taxonomie TFS [17], nous avons pu identifier que les schémas se décomposent en deux grandes classes de taille comparable :

1. Chaque facette est munie d'un pôle positif (*resp.* négatif) noté $+$ (*resp.* $-$) associé au concept (*resp.* à l'antonyme du concept). Les facettes sont généralement désignées par leur pôle $+$.
2. En sémantique lexicale, une glose est un sens décrit par une phrase courte, comme celles qu'on trouve dans les dictionnaires ou celles des synsets de la base lexicale WordNet [7].
3. http ://perso.limsi.fr/jps/research/rnb/toolkit/taxo-glosses/taxo.htm
4. http ://ipip.ori.org/newNEOKey.htm

– *Les schémas intrinsèques* concernent les comportements liés à l'agent en tant qu'entité autonome, *i.e.* les schémas dans un monde où la seule entité intentionnelle est l'agent.
– *Les schémas interpersonnels* ne sont pertinents que dans un monde où l'agent est entouré d'autres agents.
Dans la Table 1, les schémas intrinsèques sont dénotés $\square$ ($N_{s\square} = 36$) et les schémas interpersonnels dénotés $\bigstar$ ($N_{s\bigstar} = 34$). Notons que les schémas intrinsèques sont parfois applicables au comportement social mais que l'inverse ne se produit pas.

Intuitivement, la distinction entre classes de schémas intrinsèques et interpersonnels correspond avec la distinction entre architectures mono-agent et multi-agents. C'est la raison pour laquelle nous avons été conduits à traiter ces deux classes dans deux études distinctes : dans [2] nous avons proposé une architecture mono agent, dédiée à la classe des traits intrinsèques ; dans la Section 2, nous nous appuyons sur une architecture de type multi-agents, pour proposer une approche des schémas interpersonnels qui sont fortement liés aux aspects dialogiques des agents.

2.2 Moteurs de personnalité

Structure d'un moteur. On définit un moteur de personnalité $\mathtt{PE}$ comme un tuple : $\mathtt{PE} = \langle \mathtt{O}, \mathtt{W}, @, \Omega_{\mathtt{W}}, \mathtt{M} \rangle$ où :
– $\mathtt{O}$ est une *ontologie* qui doit permettre une description précise de profils de personnalité. Nous utilisons l'ensemble Σ des schémas bipolaires de TFS. Le sous-ensemble des positions positives (*resp.* négatives) est dénoté $+\Sigma$ (*resp.* $-\Sigma$), et leur union est $\pm\Sigma$ tel que $\pm\Sigma = +\Sigma \cup -\Sigma$;
– $\mathtt{W}$ est un modèle du *monde* des agents qui inclut : leur structure interne $\mathtt{W}_s$; leur protocole de communication $\mathtt{W}_c$; leur procédure de raisonnement rationnel $\mathtt{W}_r$. Par exemple, on peut considérer un modèle d'agent de type BDI ou bien un modèle plus spécifique comme celui défini Section 3.1 ;
– @ est un *topique* permettant d'instancier $\mathtt{W}$ dans un domaine applicatif particulier ;
– $\Omega_{\mathtt{W}}$ est un ensemble d'*opérateurs* d'influence, portant sur $\mathtt{W}_r \cup \mathtt{W}_c = \mathtt{W}_{rc}$;
– $\mathtt{M}$ est une *matrice* d'activation, établissant une relation sur $\pm\Sigma \times \Omega_{\mathtt{W}}$.

$\mathtt{O}$, $\mathtt{W}$ et @ sont considérés comme des ressources **données** par le contexte applicatif, alors que $\Omega_{\mathtt{W}}$ et $\mathtt{M}$ doivent être **élicités** à partir de ces ressources.

Elicitation des opérateurs d'influence. Étant donné un modèle d'agent $\mathtt{W}$, les opérateurs d'influence sont des méta-règles $\omega \in \Omega_{\mathtt{W}}$ qui contrôlent ou altèrent la partie non structurelle de $\mathtt{W}$, *i.e.* $\mathtt{W}_{rc}$.

La définition d'un algorithme qui prendrait un couple de ressources $\mathtt{W}$ et $\mathtt{O}$ et qui produirait *automatiquement* l'ensemble $\Omega_{\mathtt{W}}$ demeure une question scientifique ouverte. De plus, il faudrait aussi poser les questions de l'existence des éléments de $\Omega_{\mathtt{W}}$ (aucune influence trouvée) et de leur unicité (plusieurs ensembles distincts trouvés). De plus, il est nécessaire de s'appuyer sur une ontologie $\mathtt{O}$ qui soit issue des recherches en psychologie. Les éléments d'une telle ontologie ne sont pas facilement formalisables. Dès lors, il semble difficile de se passer d'une intervention humaine dans le processus d'élicitation des opérateurs d'influence.

Pour le moment, il nous faut donc recourir à une construction empirique de l'ensemble des opérateurs, ce qui les rend de facto concepteur-dépendants. C'est là qu'intervient la notion de moteur de personnalité. Elle a pour but essentiel de faciliter la gestion de leur diversité : *e.g.* des propositions distinctes $\mathtt{PE}_i$, fondées sur le même $\mathtt{W}$ et/ou $\mathtt{O}$, peuvent être testées et comparées de manière systématique.

Les opérateurs d'influence peuvent s'appliquer selon deux modes principaux :
1) Mode tout-ou-rien : l'heuristique est exécutée ou bloquée ;
2) Mode paramétrique : des paramètres sont passés à l'heuristique. Considérons deux cas fréquents :
– Un paramètre d'*intensité* est fourni, représentant le degré d'activité de l'opérateur. il est déterminé par les poids $\lambda_{i,j}$ d'une matrice d'activation (*cf.* § suivant) ;
– Un paramètre de *direction* est également fourni lorsque les opérateurs qui peuvent aussi fonctionner en mode inverse/antonyme (*e.g.* voir l'opérateur ω_{-safe} dans la Section 2.3).

Elicitation d'une matrice d'activation. Une fois choisi un ensemble de schémas $\sigma_i \in \pm\Sigma$ (ici ceux de TFS) ainsi qu'un ensemble d'opérateurs d'influence $\omega \in \Omega_{\mathtt{W}}$, les concepteurs d'un moteur particulier doivent alors décrire la manière dont les $\pm\sigma_i$ sont associés aux ω_i, c'est-à-dire qu'il faut définir quels schémas activent quels opérateurs. Pour les mêmes raisons que pour les opérateurs, cette relation est aussi concepteur-dépendante. Là encore, notre objectif est de faciliter la gestion de la diversité des relations schémas/opérateurs. C'est pourquoi nous avons

recours à une approche déclarative qui se fonde sur une matrice $\mathtt{M}$ contenant des éléments multi-valués, appelés *niveaux d'activation* $\lambda_{i,j}$ tels que $\mathtt{M} = \pm\Sigma \times \Omega_{\mathtt{W}}$. Les éléments $\lambda_{i,j}$ de $\mathtt{M}$ ont les valeurs et conventions suivantes :

- 2 active l'opérateur avec une force importante
- 1 active l'opérateur avec une force modérée
- 0 l'opérateur est désactivé
- -1 active l'opérateur antonyme (s'il existe) avec une force modérée
- -2 active l'opérateur antonyme (s'il existe) avec une force importante

2.3 Instanciation des moteurs

Une fois donné un moteur de personnalité particulier $\mathtt{PE}_0$, on dispose d'une structure symbolique qui peut être instanciée dans des situations effectives, variant selon deux points de vues : les topiques applicatifs et les profils de personnalité.

Topiques applicatifs. Soit $@_0$ un topique particulier fournissant un ensemble d'actions disponibles $\alpha_i \in \mathtt{A}(@_0)$. Le topique fournit alors pour chaque action des informations de nature application-dépendante nécessaires à la spécialisation des opérateurs d'influence. Par exemple, soit ω_{+safe} un opérateur qui trie un ensemble d'actions de la plus sûre à la plus dangereuse : $\omega_{+safe} \doteq Sort(\{\alpha_i\}, \prec_{danger})$. Pour être opérationnel, un opérateur ω_{+safe} a besoin que le topique $@_0$ lui fournisse une fonction de mesure $\mu_{danger} : \mathtt{A}(@_0) \longmapsto [0, 1]$. L'opérateur ω_{+safe} possède en outre un antonyme (ω_{-safe}) qui trie les actions en sens inverse.

Profils de personnalité. Étant donné un individu x, son profil de personnalité $\mathrm{P}(x)$ peut être spécifié par un ensemble de $|\Sigma|$ fonctions $\mathrm{p}(\sigma_i) : \Sigma \longrightarrow \{+, \asymp, -\}$ telles que :

- $\asymp$ signifie que vis-à-vis du schéma σ_i, le comportement de la personne x ne dévie pas de manière significative du comportement humain stéréotypique ;
- $+$ signifie que le comportement de x dévie significativement, dans le sens du pôle + ;
- $-$ signifie que le comportement de x dévie significativement, dans le sens du pôle -.

Notation : En général, on constate qu'un profil de personnalité ne dévie du comportement humain stéréotypique que par quelques schémas. C'est pourquoi il est aisé de définir $\mathrm{P}(x)$ par la liste de ses schémas dont la valeur est différente de $\asymp$.

3 Etude de cas

3.1 Le monde d'agents dialogiques TALKIE

Dans cette étude de cas, nous allons illustrer comment les moteurs de personnalités peuvent être définis, puis instanciés dans des situations réelles. Pour pouvoir mettre en œuvre le processus d'élicitation des opérateurs, il faut tout d'abord choisir un modèle d'application. Nous allons considérer ici un modèle multi-agents, appelé TALKIE, qui restreint les modèles classiques (*e.g.* KQML, ACL-FIPA) en vue d'une présentation intelligible de notre approche. Il n'est pas sans rappeler le modèle DIAL [11], auquel aurait été ajoutée une base psychologique. Comme cet article est focalisé sur les schémas interpersonnels (ceux notés ★ dans la Table 1), nous choisissons un modèle d'agents conversationnels communiquant par messages.

Modèle d'agent. Soit TALKIE un monde effectif, composé d'entités physiques et abstraites. On y accède via sa représentation sous forme de modèle symbolique $\mathcal{M}$. Une entité $e_i \in \mathcal{M}$ est définie dans $\mathcal{L}_M$, son langage de description associé, comme un ensemble de définitions, exprimées sous forme de règles ayant pour forme générale $\mathrm{D}_i = leftpart \mapsto rightpart$ telles que $\forall e_i \in \mathcal{L}_M;\ e_i = \{\mathrm{D}_i\}$.

Les agents $a_i \in \mathcal{A}$ représentent les entités dialogiques de $\mathcal{M}$ qui sont capables de raisonnement rationnel. Un agent $a_i \in \mathcal{A}$ est défini comme un 5-uple $\langle id, K, S, \Phi, \Psi \rangle$ où :
– id est le nom unique de l'agent ;
– Base de connaissances $K = k_i \in \mathcal{L}_k$ est un ensemble de propositions logiques sur $\mathcal{M}$;
– Base sociale S est l'ensemble des rôles endossés par l'agent dans TALKIE ;
– Base des caractéristiques ('features') Φ est l'ensemble des attributs de nature physique de l'agent (pour simplifier l'exposé, Φ ne sera plus rediscuté dans la suite) ;
– Base psychologique $\Psi = \Psi_T \cup \Psi_M$ est l'ensemble des traits Ψ_T et des humeurs [5] Ψ_M.

Structure des messages. Le système des agents TALKIE implémente les opérations $\mathrm{SEND}[t, a, \{b_i\}, m]$ qui permettent le transfert de message m lors du tour t entre un agent expéditeur a et un ou plusieurs agents receveurs

5. Les humeurs ('moods') sont des états mentaux plus dynamiques que les traits et moins dynamiques que les émotions de base ; dans cette étude qui concerne essentiellement les phénomènes à dynamique très lente, voire statique, nous posons : $\Psi = \Psi_T$.

TABLE 2 – Sémantique des niveaux d'activation pour les opérateurs de messagerie.

Niveaux		Opérateurs			Niveaux d'activation λ					
1	2	CODE	*Nom*	Définition générale	-2	-1	0	1	2	Intervalle
Proaction	Explicite	A	*Ask*	probabilité que l'agent utilise *Ask* ou *Propose*	-	-	nul	ask activé si nécessaire	ask activé même si non nécessaire	$[\![0, 2]\!]$
		P	*Propose*		-	-	nul	propose est activé si nécessaire	propose activé même si non nécessaire	
		D	*Dominance*	probabilité que l'agent utilise la force ou son antonyme	inférieur	supporter	nul	égal	supérieur	$[\![-2, 2]\!]$
		F	*Feeling*		agressif	froid	nul	poli	chaleureux	
		M	*Motivation*		motivations fausses	cache ses motivations	nul	motive sa demande	motive sa demande même si non nécessaire	
		I	*Incentive*		fait des menaces		nul	fait des promesses		
	Implicite	G	*Guess*	capacité à percevoir l'état mental des autres agents	perceptions faussées	ne perçoit rien	nul	perçoit ce qui est explicité	perçoit même ce qui est non explicité	
		C	*Conflict*	attitude de l'agent au sujet du risque de déclencher un conflit	aime les conflits	accepte les conflits	nul	n'aime pas les conflits	évite tous les conflits	
		S	*Sincerity*	sincérité de l'agent sur le contenu de son message	dit des faits qu'il croît faux	cache activement des faits	nul	franc	très/trop franc	
Réaction	Explicite	A+	*Reaction*	la réaction typique à *Ask* ou *Propose*. Elle dépend de l'évaluation globale par b des forces exprimées par a	toujours non	oui mais en protestant	nul	oui mais peut être conditionnel	toujours oui	$[\![-2, 2]\!]$
		A-								
		P+								
		P-								
	Implicite	B	*Bond*	réaction aux résultats de G (*e.g.* percevoir que a est triste, b donne : +) se-sentir-triste ; 0) indifférent ; -) se-sentir-heureux)	empathie inverse	aucune empathie	nul	empathie si nécessaire	empathie même si non nécessaire	
		N	*Negotiate*	réaction dans la gestion des conflits en cours et explicités	aggraver	se défendre	nul	chercher un agrément	toujours céder	

Le niveau $\lambda = 2$ (*resp.*-2) inclut le niveau $\lambda = 1$ (*resp.*-1), *i.e.* il est susceptible d'exhiber le comportement lié au niveau 1 (*resp.*-1) . nul = opérateur non activé

$\{b_i\}$. Dans la suite, nous restreignons cette définition aux interactions entre couples d'agents $a \Leftrightarrow b$ (a dénote le locuteur et b l'interlocuteur) ce qui se traduit par des opérations de la forme SEND$[t, a, b, m]$. Un message m dans une opération SEND contient quatre expressions, explicitement transmises par a à b :

$m = \langle$Réaction, Proaction, Forces, Contenu$\rangle$

Réaction est l'attitude qu'adopte a et qu'il exprime explicitement, en réaction à sa propre évaluation du message précédemment reçu de b au tour $t-1$. Les réactions sont organisées sur une échelle -/+, allant d'un désagrément total (noté No) à un agrément total (noté Yes). Le premier message du premier tour d'une session a une réaction vide (notée —).

Proaction est l'attitude essentielle (*i.e.* la raison principale de l'envoi du message), exprimée explicitement par a envers b. Nous considérons deux classes de proactions, selon la direction de l'intention communicative de a :
– *Ask*, noté $a \xleftarrow{A\ Content} b$, où a envoie une demande à b au sujet du Contenu du message ;
– *Propose*, noté $a \xrightarrow{P\ Content} b$, où a envoie une proposition à b au sujet du Contenu du message.

Forces ce sont des modalités optionnelles des opérateurs de proaction (A|P = *Ask* | *Propose*), exprimées explicitement par a afin de contribuer au succès du message : un message de a est considéré réussi quand la réaction de b à ce message est positive et pertinente (au sens de Vanderveken ou Sperber) pour a. Nous considérons quatre types de forces, organisées en échelles bipolaires -/+ :
– *Dominance* a pour pôles -submissive et +dominant. Elle paramètre les opérateurs A|P *e.g.* A -submissive peut être vu comme 'implorer' et A +dominant comme 'exiger'.
– *Feelings* a pour pôles -aggressive et +affective. Elle paramètre A|P avec une modalité affective ;
– *Motivation* a pour pôles -hide et +open. Un agent a utilisant la force +open dans son message, est censé expliciter franchement les raisons rationnelles motivant l'envoi du message, tandis que -hide suggère que a peut essayer de cacher ces raisons, voire mentir à leur sujet ;
– *Incentive* a pour pôles -menace et +promising. Un agent a utilisant +promise essaye soit 1) de faire réussir le message en fournissant à b une raison rationnelle qui ferait que b aurait un intérêt à réagir positivement et pertinemment au message, ou encore 2) de fournir à b une récompense directe. Inversement, a peut essayer d'obtenir l'agrément de b par la -menace soit 1) en faisant état des conséquences négatives pour b si celui-ci traite le message négativement ou 2) en adressant des menaces directes.

Contenu est le corps du message, objet de la proaction. On considère cinq classes :
– *Croyance* est une proposition $k_i \in \mathcal{L}_k$;
– *Action* est une opération sur le monde. Par exemple, $A\ a(x)$ veut dire que a demande à b d'exécuter $a(x)$, alors que $P\ a(x)$ veut dire que a se propose d'exécuter $a(x)$;

– *Ressource* est une entité du monde qui peut être possédée par un agent et dont la possession peut être transférée à un autre agent (selon le type de la ressource, le transfert peut être un don, un partage ou une duplication) ;
– *Norme* décrit les droits et devoirs des agents à l'intérieur d'un collectif donné c_i ;
– *Émotion* regroupe deux sous-classes : un état mental intrinsèque à a (*e.g.* une humeur de a) ou bien une relation interpersonnelle (attitude affective de a envers b).

Avec ces définitions, la structure d'un message m est représentée par :

$$-|Yes|No \times A|P \times [D][F][M][I] \times k|a|r|n|e$$

où | sépare les alternatives, [] regroupe les forces optionnelles, k, a, r, n, e sont les cinq types de contenus et $\times$ le produit cartésien, définissant ainsi le domaine de messagerie. Un tour t est un couple $\langle \text{SEND}[t, a, b, m],\ \text{SEND}[t, b, a, m'] \rangle$ où m' est la réponse à m. Une session est une séquence de tours ; des sessions plus complexes peuvent inclure des sous-sessions (appelés fils) dans le cas de réactions conditionnelles.

3.2 Construction du moteur de TALKIE

Prinicipe d'élicitation des opérateurs. Si on considère le modèle du monde $\mathtt{W}$ = TALKIE, il est possible de lui associer un ensemble d'opérateurs d'influence Ω_{TALKIE}. Nous mettons l'accent ici, sur les opérations de construction des messages et d'envoi des messages, *i.e.* sur $\mathtt{W}_c$.

Une fois donné le modèle décrit à la Section 3.1, les concepteurs peuvent alors en étudier la structure afin de déterminer les éléments susceptibles d'être soumis à des influences. Il s'ensuit qu'on peut organiser les opérateurs selon une structure qui est le miroir de la structure du modèle : on obtient un premier niveau, où les opérateurs d'influence sur la passation de messages sont divisés en deux classes principales, *proaction* et *réaction* ; puis au second niveau où on distingue, pour chaque classe, des opérateurs *implicites* et *explicites*. On obtient ainsi :
– *Opérateurs de proaction explicite* qui sont exprimés dans le message ;
– *Opérateurs de proaction implicite* qui ne sont pas exprimés dans le message mais qui influencent sa construction ;
– *Opérateurs de réaction explicite* qui sont exprimés dans le message, en termes de réactions de type $Yes|No$;
– *Opérateurs de réaction implicite* qui sont le miroir de la proaction implicite.

Dans la mesure où nous avons utilisé 1) un modèle de communication agent simplifié et 2) la description des schémas de TFS, il a été possible aux concepteurs [6] d'exhiber 15 opérateurs pour ce modèle. La Table 2 donne la liste de ces opérateurs et leur associe sous forme abrégée, une sémantique pour chaque niveau d'activation λ appartenant à un intervalle de valeurs discrètes, tel que défini à la Section 2.2.

Exemple d'analyse de schéma. Si les niveaux (colonnes 1 et 2 de la table 2) se déduisent directement du modèle, les opérateurs doivent être déduits de la confrontation entre le modèle et les 34 schémas interpersonnels de TFS. Par exemple, considérons le schéma bipolaire $\mathbf{E}_{Assertiveness}$DOMINEERINGNESS. Il est défini par la liste de gloses suivantes[3] :

Id	Gloses (W = synsets wordnet Q = q-items de Goldberg)
	Pôle + DOMINEERING
W68	exerting force or influence
W169	characterized by action or force of personality
W518	tending to domineer
W617	having or showing a desire to control or dominate
W882	not making concessions
W946	expecting unquestioning obedience
W1145	exercising influence or control
W1151	forceful and definite in expression or action
W1016	severe and unremitting in making demands
Q141	Take charge
Q142	Try to lead others
Q144	Seek to influence others
Q145	Take control of things
	Pôle - SUBMISSIVE (antonyme)
W35	easily handled or managed
W699	showing humiliation or submissiveness
W916	evidencing little spirit or courage ; overly submissive
W917	very docile
W1025	quieted and brought under control
W1355	tending to give in or surrender or agree
W1356	inclined to yield to argument or influence or control
Q281	Am easily intimidated

Par exemple, l'analyse des gloses du pôle + suggère directement l'opérateur D (« tending to domineer ») (*cf.* Table 2) ainsi que les opérateurs I (« Seek to influence others ») C et N (« not making concessions »). Ceci est complété par l'analyse du pôle négatif.

Etablissement de la matrice d'activation. Étant donné l'ensemble $\pm\Sigma$ et l'ensemble Ω_{TALKIE} des opérateurs élicités dans l'étude de cas TALKIE, il est possible aux concepteurs [7] de définir une matrice d'activation $\mathtt{M}_{\text{TALKIE}}$ qui établit les relations entre les schémas et les opérateurs.

6. Ici, les auteurs de l'article. Un procédé d'annotation standard doit être mis en place afin d'améliorer l'élicitation des opérateurs, même si l'existence d'un ensemble canonique est une question ouverte.

7. La définition des relations est concepteur-dépendante, de la même manière que pour les opérateurs.

TABLE 3 – Extraits de la matrice $\mathbb{M}_{\text{TALKIE}}$ (schémas utilisés dans les exemples de la Section 4.1).

			Proaction									Réaction						
		Code Opérateur	A	P	D	F	M	I	G	C	S	A+	A-	P+	P-	B	N	
		Intervalle de valeur	0 2	0 2	2-2	2-2	2-2	2-2	2-2	2-2	2-2					2-2	2-2	
		Agent générique	1	1	1	1	1	1	1	1	1	1	-1	1	-1	1	1	Serveurs
T	Facette	Schéma																w_1 w_2
O	fantasy	-PRACTICAL																*
O	fantasy	+IDEALISTIC		2		2	2	2		-1	2	2		2			0	*
O	fantasy	+CREATIVE		2			2				-2							*
C	competence	-INSECURE	2	0	-2		2	0		2			1		1		2	*
E	warmth	+FRIENDLY			-1	2			2			2		2		2		*
E	warmth	-COLD				-1			-1	-1	-1					-1		*
E	assertiv.	+DOMINEER.	2		2	-1	0	-1		-1	-1		-2		-2		-1	*
E	activity	+ACTIVE		2						-1		2		2			-1	*
E	activity	-APATHETIC	0	0	-2	0	0	0					1		1	-1	2	*
A	trust	-SECRETIVE				0	-1	0				-1	-2			0		*

-PRACTICAL est le pôle antonyme du schéma +IDEALISTIC *resp.* +FRIENDLY/-COLD, +ACTIVE/-APATHETIC.
Quand $\lambda_{i,j} = \emptyset$ alors $\lambda_{i,j} = \text{Agent générique}_j$ * indique que le serveur active le schéma.

La Table 3 donne un extrait d'une proposition de matrice pour $\mathbb{M}_{\text{TALKIE}}$. Alors que $\pm\Sigma$ contient 140 schémas, faute de place nous ne montrons dans la Table 3 que les 10 schémas effectivement utilisés dans les exemples développés à la Section 4.1. De plus, pour ne pas surcharger la lecture de la Table 3, les valeurs d'activation $\lambda_{i,j}$ associées au comportement humain stéréotypique sont factorisées dans la ligne du haut, appelée 'agent générique', et sont représentées par des cellules vides dans les schémas.

4 Exemple : le CyberCafé

4.1 Exemples de profils de personnalité

Comme exemple d'instanciation d'un moteur de personnalité pour TALKIE, nous considérons un exemple classique, tiré du CyberCafé de Rousseau et Hayes-Roth [16]. Dans le CyberCafé, plusieurs agents endossent le même rôle de serveur de café (waiter), notés w_i, mais avec des profils de personnalité distincts $P(w_i)$ impliquant des comportements psychologiques distincts qui sont décrits sous forme de plusieurs *adjectifs* et sont associés à une glose unique $G(w_i)$, tels que :

$P(w_1)$	*realistic, insecure, introverted, passive, secretive*
$G(w_1)$	« Such a waiter does and says as little as he can »
$P(w_2)$	*imaginative, dominant, extroverted, active, open*
$G(w_2)$	« This waiter takes initiative, comes to the customer without being asked for, talks much »
...	(il y a deux autres profils : w_3, w_4)

Nous reprenons les notations exactes de [16].

Considérant le profil psychologique $P(w_1)$ du serveur w_1, on peut transposer ses adjectifs en termes d'activation de schémas dans TFS (les adjectifs de $P(w_1)$ sont transposés dans l'ordre, séparés par des ';') :

P'(w_1) = {
realistic ⇒ **O**-fantasy-PRACTICAL ;
insecure ⇒ **C**-competence-INSECURE ;
introverted ⇒ **E***(-COLD, -NONGOSSIPMONGER, -SOLITARY, -UNCOMMUNICATIVE, -UNCHARISMATIC, -DISCRET, -SUBMISSIVE, -PLEADING, -LANGUID, -APATHETIC, -ASCETIC, -BLASE) ;
passive ⇒ **E**-activity-APATHETIC ;
secretive ⇒ **A**-trust-SECRETIVE ;
}.

Respectivement pour le serveur w_2 on obtient [8] :

P'(w_2) = { **O**fantasy+CREATIVE ;
Eassertiveness+DOMINEERING ; **E**warmth+FRIENDLY ;
Eactivity+ACTIVE ; **E**fantasy+IDEALISTIC }.

Analyse :

— *Différentiation des profils :* P'(w_1) active surtout des schémas en mode négatif alors que P'(w_2) n'active que des schémas positifs. De plus, P'(w_1) et P'(w_2) activent deux sous-ensembles de schémas (notés * dans la Table 3) dont l'intersection est vide. Cela suggère que les exemples du CyberCafé ont été *construits* afin de produire des profils de personnalité très contrastés.

— *Consistance des définitions :* La définition P'(w_1), fournie par le CyberCafé, n'est pas *consistante* c'est-à-dire que certains schémas sont activés en mode contradictoire *e.g.* les niveaux 1 et -2 sont présents en même temps

8. Ici, c'est une version simplifiée où on ne garde que le premier schéma lorsqu'une facette est activée en entier (idem pour un trait).

(idem pour P'(w_2)). En théorie, quand on écrit un profil, rien n'interdit de produire des activations conflictuelles pour un même opérateur. Notre approche, parce qu'elle est déclarative, a l'avantage de bien faire ressortir de tels effets (ici on montre que l'exemple de [16] est problématique) et rend même possible la vérification automatique d'existence de tels conflits.

Les valeurs d'activation du sous ensemble des opérateurs TALKIE associées aux schémas apparaissant dans P'(w_1) et P'(w_2) sont données dans la Table 3 ; elles ont été établies par les concepteurs de TALKIE.

4.2 Exemple de sessions dialogique

Nous considérons une situation d'interaction, appelée prise-de-commande, où un agent serveur W endosse successivement trois personnalités distinctes : P'(w_0) = $\emptyset$ (l'agent est strictement rationnel et n'active aucun opérateur d'influence) ; P'(w_1) ; P'(w_2). Cet exemple n'est pas traité exhaustivement car 1) d'autres opérateurs d'influence interviennent aussi et 2) l'implémentation des influences n'est pas décrite techniquement [9], seuls les effets sont montrés.

Description de prise-de-commande. Soit un monde avec deux agents capables d'actions dialogiques :
– Un client C qui s'assoit à la table d'un restaurant pour prendre un repas ;
– Un serveur W qui s'approche du client pour prendre sa commande.
L'agent serveur peut effectuer les trois actions optionnelles suivantes :
– noter(plat) écrire le nom d'un plat sur son carnet (pour mieux s'en souvenir) ;
– poser/prendre(chips) sur la table du client ;
– allumer/éteindre(bougie) qui est posée sur la table du client.
La carte du restaurant, posée sur la table du client, indique les quatre plats suivants : pavé de boeuf (pb) ; saucisse purée (sp) ; soupe aux croûtons (sc) ; omelette flambée (of). L'agent serveur possède des faits à leur sujet dans sa base de connaissances K_W. Nous utiliserons dans l'exemple : ¬disponible(sp) ; amusant-à-servir(of) ; amusant-à-servir(sc).

Interaction entre agents rationnels. On considère premièrement que les agents C et W sont strictement rationnels c'est-à-dire associés au profil P'(w_0). Voici une session d'interaction typique, notée S_{w_0,w_0} qui résulte de l'exécution de prise-de-commande :

T	A	Dialogue	Actions de W
1 :	W	Quel plat désirez-vous ?	ouvrirSession()[a]
	C	Je voudrais une saucisse purée	voir si disponible(sp) dans K_W
2 :	W	Il n'y en a plus	.
	C	Je voudrais un pavé de boeuf	noter(pb)
3 :	W	Ensuite[b] Quel plat désirez-vous ?	.
	C	Aucun	cloreSession()

T tour ; **A** agent [a] toutes les actions sauf ouvrirSession() sont exécutées après l'acte dialogique. [b] par rapport à la phrase 1 :W 'ensuite' indique que W exécute par ex. un 'stream-all' KQML .

Interactions entre agents rationnels et psychologiques. Deuxièmement l'agent C conserve le profil P'(w_0) mais l'agent W est associé au profil P'(w_1) : Cela produit alors la session S_{w_0,w_1}, altérée par les opérateurs d'influence de W comme le montre la session dans la Table 4-haut. Respectivement, W est associé à P'(w_2) donnant la session S_{w_0,w_2} dans la Table 4-bas.

Ici on s'est restreint aux opérateurs de la Table 2. De plus, l'agent C, fixé à P'(w_0), ne réagit pas aux signaux interactionnels supplémentaires émis par W. Dans une situation d'interaction plus réaliste, tous les agents devraient être munis chacun d'une personnalité ; nous envisageons d'aborder ce problème plus complexe dans des travaux ultérieurs.

On obtient une forte variabilité aussi bien si on oppose la session rationnelle S_{w_0,w_0} aux sessions $S_{w_0,w_{1|2}}$ ou bien entre S_{w_0,w_1} et S_{w_0,w_2}. De simples pré tests montrent que les sujets perçoivent bien cette variabilité. Il reste à évaluer s'ils caractérisent correctement les profils.

4.3 Discussion

Pertinence et complétude des opérateurs. Le processus d'élicitation des opérateurs assure que tous les opérateurs définis sont pertinents. Par exemple, dans l'étude de cas TALKIE, le fait que les opérateurs sont synthétisés à partir des gloses des schémas, assure qu'ils sont activés au moins une fois et de manière non stéréotypique [10] (*i.e.* $\forall \sigma \in \mathbb{M}_{\text{TALKIE}}, \exists j$ tel que $\lambda_{i,j} \neq$ agent générique (i)).

9. Pour un exemple d'implémentation complète d'un opérateur, voir le rapport technique consultable à : http://perso.limsi.fr/jps/research/rnb/ranking/draft-ranking-V2.pdf.

10. A l'exception de la première ligne de la Table 3 (O fantasy - practical) qui est similaire à la ligne agent générique dans la mesure où ce trait peut être vu comme la « raison pratique » de Bratman [3] qui peut alors être considéré comme un cas particulier d'agent psychologique.

TABLE 4 – Deux sessions d'interaction : en haut) P(W) = P'(w_1) ; en bas) P(W) = P'(w_2)

T	A	Dialogue	Actions de W	Schémas activés[a]
1 :	W	Quel plat désirez-vous ? [air froid[b]]	ouvrirSession()	+INSECURE, -COLD
	C	Je voudrais une saucisse purée	disponible(sp) ?	
2 :	W	Désolé, il n'y en a plus [air triste]	.	+INSECURE
	C	Je voudrais un pavé de boeuf	.[c]	-APATHETIC
3 :	W	Ensuite[d], quel plat désirez-vous ? [air froid]	.	+INSECURE, -COLD
	C	Aucun	cloreSession()	

[a] nous ne listons que ceux ayant un niveau d'activation fort. On se reportera à la able 3 pour voir les opérateurs activés par ces schémas, par ex. ligne 1 :W +COLD active -F, -G, -C, -S, -B (avec -F ⇒ [air froid]). [b] [entre crochets] ajout d'expressions (faciale, tonalité vocale *etc.*) ici dû à -COLD. [c] W ne note pas la commande (dû à -APATHETIC). [d] W pourrait clore sans relancer (dû à -APATHETIC) mais il respecte (dû à +INSECURE) les étapes de prise-de-commande.

1 :	W	Bonjour[a] [! !][b] Je vous conseille fortement l'omelette flambée ; je suis sûr que vous allez aimer[c] [! !]	ouvrirSession() ; **Allumer**[b](bougie)[d]	+FRIENDLY, +ACTIVE, +DOMINEERING, +CREATIVE
	C	Je voudrais une saucisse purée	disponible(sp) ?	
2 :	W	Il n'y en a plus ! ! Vous devriez suivre mes conseils ... [air sévère]	**poser**(chips)[d]	+ACTIVE, +DOMINEERING
	C	Je voudrais un pavé de boeuf	**noter**(pb)[e]	+ACTIVE
3 :	W	Ensuite, mon cher je vous conseille fortement l'omelette flambée ou la soupe aux croûtons[f] ; je suis sûr que vous allez aimer [! !]	.	+FRIENDLY, +ACTIVE, +DOMINEERING, +CREATIVE
	C	Aucun		
	W	[Ah !] Je suis déçu [! !] // ajout d'une incise	**éteindre**(bougie)[g] ; cloreSession()	+ACTIVE, +DOMINEERING

[a] expression due à +FRIENDLY liée à l'action ouvrirSession() (elle était inhibée dans le profil précédent). [b] [! !] ou **en gras** = exécuté avec force/vivacité, dû à +ACTIVE et +DOMINEERING. [c] incitation due à +CREATIVE et +DOMINEERING. [d] ajout d'action externe à l'heuristique dû à +ACTIVE. [e] ici non inhibé par -APATHETIC et conforté par +ACTIVE. [f] répétition de 1 :W mais avec en plus de la créativité (soupe). [g] indication de conflit potentiel dû à +DOMINEERING (si son activation > +FRIENDLY).

De plus, toutes les lignes de la Table 3 sont distinctes, impliquant que tous les schémas sont des concepts distincts, ayant des influences distinctes (ceci s'applique à tout $\pm\Sigma$).
Inversement, le processus d'élicitation n'assure pas que tous les opérateurs possibles sont découverts (risque de silence) ; nous avons déjà mentionné que ce n'est pas actuellement atteignable. Le fait que cette question reste ouverte apporte un argument en faveur de notre approche : pour éviter le silence des opérateurs, nous utilisons une taxinomie couvrant largement le domaine FFM/NEOPI-R ; de plus la version enrichie TFS, fondée sur un ensemble de ressources lexicales largement utilisées (*e.g.* WordNet) couvre, selon la littérature, les comportements effectifs qui sont associés aux traits.

Validation des valeurs de la matrice d'activation. Les valeurs $\lambda_{i,j} \in \mathbb{M}_{\text{TALKIE}}$ sont définies par des annotateurs, idéalement experts en informatique et en psychologie. Cela engendre potentiellement : 1) des différences quantitatives entre annotateurs qui peuvent être en partie contrôlées avec des outils statistiques classiques utilisés pour l'annotation en groupe (*cf.* le κ de Cohen) ; 2) des controverses qualitatives, en particulier entre informaticiens et psychologues.
Dans notre cas, le recourt à une méthode déclarative, utilisant une matrice de niveaux d'activation plutôt que des règles procédurales, augmente l'intelligibilité et le suivi ('tracking') des modes d'association traits/comportements. De plus, l'approche déclarative clarifie les discussions avec les psychologues, qui *in fine* doivent valider les valeurs dans $\mathbb{M}_{\text{TALKIE}}$.

Évaluation du modèle. Dans cet article, nous proposons une approche du principe, posé dans la littérature, que les traits de personnalité influencent les actions et les plans des agents. Notre propos n'est pas de fournir une évaluation directe via une expérimentation, d'un modèle d'agent particulier. Notre objectif est double :
– Présenter une preuve de concept du principe d'influence : il existe effectivement des « éléments décisionnels sous-déterminés » dans le processus de décision rationnel des agents qui peuvent être exploités pour exprimer des influences ;
– Proposer une méthode qui soit 1) générique *i.e.* non pas taillée pour quelques facteurs psychologiques mais couvrant tout le domaine des traits (ou un large sous-domaine comme ici celui les traits interpersonnels) ; et 2) déclarative *i.e.* qui utilise des valeurs explicites et non des règles codées.
Une conséquence directe est que les Tables 2 et 3 peuvent être vues comme des instances de cette approche, et ce sont donc elles qui doivent

être évaluées via des expérimentations psychologiques, dépassant le cadre de l'article.

Implémentation du modèle. Notre approche a été partiellement implémentée : elle été appliquée à plusieurs études qui sont listées sur la page Web du projet [11], voir en particulier[9].

5 Conclusion

Nous avons présenté une approche déclarative pour la prise en compte des traits dans les agents. Elle a trois avantage sur les approches procédurales : 1) elle circonscrit et elle réifie les parties du problème qui sont concepteur-dépendantes : ontologie de traits ; ensemble d'opérateurs d'influence ; matrice d'activation. 2) elle définit un processus générique pour construire les ressources et pour implémenter, de manière déclarative (matrice d'activation) les influences. 3) elle propose un cadre flexible où les choix d'influences peuvent être suivis et observés.

Nous envisageons d'étendre ce travail dans deux directions principales : en posant comme ressource des modèles d'agents BDI plus généraux afin de montrer l'indépendance de l'approche vis-à-vis du cadre de rationalité choisi ; en développant un cadre d'expérimentation, avec des psychologues, où des utilisateurs interagiront avec les agents et seront interrogés sur la perception effective qu'ils ont de la personnalité des agents.

Références

[1] F. Bouchet and J. P. Sansonnet. Classification of wordnet personality adjectives in the NEO PI-R taxonomy. In *Fourth Workshop on Animated Conversational Agents, WACA 2010*, pages 83–90, Lille, France, 2010.

[2] F. Bouchet and J. P. Sansonnet. Influence of personality traits on the rational process of cognitive agents. In *The 2011 IEEE/WIC /ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Lyon, France, 2011.

[3] Michael E Bratman. *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.

[4] P. T. Costa and R. R. McCrae. *The NEO PI-R professional manual*. Odessa, FL : Psychological Assessment Resources, 1992.

11. http://perso.limsi.fr/jps/research/rnb/rnb.htm

[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] Rick Evertsz, Franck E. Ritter, Paolo Busetta, and Matteo Pedrotti. Realistic behaviour variation in a BDI-based cognitive architecture. In *Proc. of SimTecT'08*, Melbourne, Australia, 2008.

[7] Christiane Fellbaum. *WordNet : An Electronic Lexical Database*. MIT Press, 1998.

[8] Lewis R. Goldberg. Language and individual differences : The search for universal in personality lexicons. *Review of personality and social psychology*, 2 :141–165, 1981.

[9] Hong Jiang, Jose M. Vidal, and Michael N. Huhns. eBDI : an architecture for emotional agents. In *AAMAS '07 : Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–3, New York, NY, USA, 2007. ACM.

[10] Lori Malatesta, George Caridakis, Amaryllis Raouzaiou, and Kostas Karpouzis. Agent personality traits in virtual environments based on appraisal theory predictions. In *Artificial and Ambient Intelligence, Language, Speech and Gesture for Expressive Characters,achie AISB'07*, Newcastle, UK, 2007.

[11] Maxime Morge. Collective decision making process to compose divergent interests and perspectives. *Special issue of Artificial Intelligence and Law on Argumentation in Artificial Intelligence and Law*, pages 75–92, 2006.

[12] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge, UK, cambridge university press edition, 1988.

[13] C. Pelachaud. Some considerations about embodied agents. In *Int. Conf. on Autonomous Agents*, Barcelona, 2000.

[14] P. Rizzo, M. V. Veloso, M. Miceli, and A. Cesta. Personality-driven social behaviors in believable agents. In *AAAI Symposium on Socially Intelligent Agents*, pages 109–114, 1997.

[15] D. Rousseau. Personality in computer characters. In *AAAI Technical Report WS-96-03*, pages 38–43, 1996.

[16] D. Rousseau and B. Hayes-Roth. Personality in synthetic characters. In *Technical report, KSL 96-21*. Knowledge Systems Laboratory, Stanford University, 1996.

[17] J. P. Sansonnet and F. Bouchet. Extraction of agent psychological behaviors from glosses of wordnet personality adjectives. In *Proc. of the 8th European Workshop on Multi-Agent Systems (EUMAS'10)*, Paris, France, 2010.