Agents conversationnels psychologiques

Modélisation des réactions rationnelles et comportementales des agents assistants conversationnels

François Bouchet [1], **Jean-Paul Sansonnet** [2]

1. *LIP6 – UPMC, BP 169, 4 Place Jussieu, 75252 Paris Cedex 05, France*
   francois.bouchet@lip6.fr
2. *LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France*
   jps@limsi.fr

RÉSUMÉ. *Plusieurs études ont récemment été menées sur l'attribution de compétences cognitives et de caractéristiques psychologiques à des agents artificiels. Cependant ces études reposent sur des approches procédurales, difficiles à analyser, et elles se focalisent sur des phénomènes particuliers au lieu de couvrir une partie significative du domaine psychologique des humains. Nous présentons ici une approche systématique de l'implémentation du principe stipulant que les traits de personnalité ont une influence potentielle et effective sur le processus de décision rationnelle d'agents cognitifs. L'apport de ce travail se situe au niveau de la couverture du domaine psychologique traité et de sa généricité par rapport aux modèles d'agents rationnels utilisés. Surtout, par sa nature déclarative, il facilite l'intelligibilité des relations associant les phénomènes psychologiques aux influences sur le cycle de délibération des agents.*

ABSTRACT. *Several studies have recently tried to provide artificial agents with cognitive capacities and psychological features. However these studies focused on procedural approaches, which are difficult to track, and have handled only small parts of the psychological domain. We present here a generic approach to the implementation of a principle stating that personality traits have an actual influence over the rational decision making process of cognitive agents. The contribution of this work concerns the coverage of the psychological domain treated, the genericity of the proposal with regard to the models of rational agents used, and above all, the improvement of the comprehensiveness of the relationships linking psychological phenomena with computational operators through influences on the deliberation cycle of the agents.*

MOTS-CLÉS : *architecture cognitive, psychologie computationnelle, traits de personnalité.*

KEYWORDS: *cognitive architecture, computational psychology, personality traits.*

DOI:10.3166/RIA.27.679-708

1. Introduction

Le principe selon lequel les traits de personnalité d'un humain (Cattell *et al.*, 1970 ; Goldberg, 1990) influencent son état mental (en particulier l'activation de ses émotions (Ortony *et al.*, 1988)) a été transposé à la modélisation d'agents artificiels de manière extensive, par exemple dans la communauté étudiant les personnages virtuels. L'idée que ce même principe s'applique aussi au processus de *raisonnement rationnel* des humains a été débattue en psychologie (Damasio, 1994), mais a été encore peu étudiée au niveau de la modélisation cognitive d'agents artificiels. En effet, depuis les travaux fondateurs de Rousseau et Hayes-Roth (Rousseau, 1996 ; Rousseau, Hayes-Roth, 1996) portant sur l'implémentation informatique de phénomènes de nature psychologique sur le raisonnement rationnel d'agents, les études qui ont suivi (détaillées à la section 2.2) souffrent de deux problèmes que nous désignerons par les termes de *couverture comportementale* et d'*intelligibilité des influences*.

La notion de couverture désigne en première acception l'extension du domaine des phénomènes psychologiques concernés par une implémentation informatique donnée. La plupart des travaux récents portent sur des études de cas particuliers (par exemple, elles traitent un ou deux traits de personnalité seulement). En psychologie, la question de la couverture est centrale et plusieurs larges taxonomies de traits de personnalité ont été proposées, comme le modèle 16PF de Catell (Cattell *et al.*, 1970) ou le modèle FFM (Goldberg, 1990). Cependant, du point de vue computationnel, ces modèles ont deux inconvénients : ils sont trop génériques et pas assez descriptifs des comportements effectifs des personnes qui portent les traits. En conséquence, nous étendrons la notion de couverture à celle de couverture comportementale (intuitivement, comment se traduit en pratique le fait d'être une personne ambitieuse, conciliante, pessimiste *etc.*).

La notion d'intelligibilité désigne la transparence du mécanisme d'implémentation des phénomènes psychologiques. L'essentiel des travaux portant sur les influences des traits psychologiques (détaillés à la section 2.2) ont recours à des implémentations procédurales ou à des heuristiques "enfouies dans le code" qui ne facilitent pas la discussion critique, en particulier avec les psychologues ; or celle-ci est indispensable car la nature de l'influence de la psychologie sur le raisonnement est encore mal connue. Il faut donc proposer une architecture flexible et observable, qui permette de tester et de suivre facilement des hypothèses sur les relations traits/raisonnement.

Dans cet article, nous proposons une architecture dédiée à l'implémentation informatique et à l'expérimentation des influences des traits de personnalité sur le raisonnement rationnel d'agents artificiels, en ayant pour objectifs de faciliter la couverture comportementale ainsi que l'intelligibilité des influences telles que définies ci-dessus. Dans les sections qui suivent, nous abordons la définition et la validation de cette architecture selon les points successifs qui décrivent le déroulement typique de notre méthodologie :

– **Section 2** : nous présentons tout d'abord un état de l'art qui couvre le contexte de cette étude. Il porte essentiellement sur les travaux existants sur les agents ration-

nels et psychologiques : d'abord nous présentons les principales architectures d'intelligence artificielle concernées ; ensuite nous abordons l'introduction de capacités psychologiques dans des agents rationnels ; puis nous nous focalisons sur les études portant sur les relations entre rationalité et psychologie dans un contexte d'interaction humain/agent.

– **Section 3** : partant de la taxonomie de traits classique FFM dans sa version étendue aux facettes de Costa et Mc Crae FFM/NEO PI-R, nous l'enrichissons de schémas qui décrivent les comportements effectifs des agents, répondant ainsi au besoin de couverture comportementale.

– **Section 4** : nous présentons les principes généraux et la structure d'une architecture, appelée Moteur de Personnalité (MP) dans laquelle les traits de personnalité sont liés à des opérateurs d'influence agissant sur le modèle d'agent rationnel. La liaison traits/opérateurs est réifiée de manière déclarative par une *matrice d'activation*, répondant ainsi au besoin d'intelligibilité des influences.

– **Section 5** : le principe d'influence est en théorie indépendant du modèle computationnel choisi pour supporter le processus de raisonnement rationnel de l'agent. Le choix d'un modèle particulier correspond donc à un premier niveau de distanciation de l'architecture définie précédemment. Nous nous plaçons dans le cadre d'agents BDI avec un modèle d'action proche d'INDIGOLOG. Il est alors possible d'éliciter pour ce modèle des classes d'opérateurs d'influence. Nous montrons que pour le cadre BDI choisi, cinq grandes classes d'opérateurs sont envisageables.

– **Section 6** : étant donné d'une part a) une taxonomie de schémas comportementaux et d'autre part b) un modèle de raisonnement rationnel et ses classes d'opérateurs associées, il est alors possible d'éliciter un ensemble d'opérateurs d'influence. Une première validation de notre approche est apportée par une étude de cas qui porte sur le trait FFM Conscientiousness. Cette étude apporte deux résultats : 1) la liste des opérateurs d'influence associés aux schémas comportementaux de ce trait ; 2) la matrice d'activation qui relie les schémas aux opérateurs.

– **Section 7** : la dernière section est consacrée à la présentation d'un exemple illustratif, montrant les effets des opérateurs d'influence sur une session d'interaction qui à l'origine est purement rationnelle. Pour ce faire, nous nous sommes appuyés sur des exemples de profils de personnalité issus de de la littérature du domaine.

2. Études sur les agents rationnels et psychologiques

2.1. Architectures d'agents rationnels

Dans notre contexte, le terme « agent rationnel » correspond aux définitions classiques en intelligence artificielle (IA) et systèmes multi-agents (SMA) du *raisonnement pratique* de Bratman déjà cité. Par exemple, la définition de Russell et Norvig (2009) (p. 38-39) ou celle de Wooldridge qui adapte le raisonnement pratique aux SMA : « Un agent est dit rationnel s'il choisit les actions qui satisfont son propre intérêt, étant donné les croyances qu'il a sur l'état courant du monde » (Wooldridge,

2000) (p.1) et il propose un modèle formel (Wooldridge, Jennings, 1995), fondé sur la théorie BDI de Rao et Georgeff (1995). En fait, le *raisonnement pratique* est souvent opposé au *raisonnement théorique* qui vise la consistance des représentations du monde et qui n'est pas discuté ici. De même, nous n'aborderons pas la notion d'agent rationnel telle qu'elle s'entend dans la théorie de la décision, où un agent maximise une fonction d'utilité. Ces deux derniers types de raisonnement rationnel nécessitent en effet une connaissance globale et totale du monde.

Les architectures d'agents rationnels sont souvent fondées sur SOAR (*State, Operator And Result*) qui fut le premier modèle populaire destiné à la modélisation des processus cognitifs au moyen de règles (Laird *et al.*, 1987). ACT-R (*Adaptive Control of Thought-Rational*) d'Anderson (1996) fonctionne sur les mêmes principes, avec une séparation claire entre les représentations déclaratives et procédurales. Faisant suite à SOAR et ACT-R, la théorie BDI (pour *belief*, *desire* et *intention*) de Bratman, Rao et Georgeff (cités ci-dessus) met en avant la primauté des désirs et des intentions dans l'action rationnelle. Les plateformes BDI sont nombreuses et utilisent des logiques pures (KGP, MINERVA, AGENTSPEAK, CAN), avec des règles (2APL) ou temporelles (MetateM, la famille Golog). Plusieurs implémentations de plateformes BDI développées notamment en Java ont été largement utilisées depuis plusieurs années (JASON, JADEX, JACK) – *cf.* (Mascardi *et al.*, 2005) pour une revue détaillée. Toutes ces architectures sont fondées sur le cycle de délibération BDI initial.

2.2. Introduction de capacités psychologiques dans des agents rationnels

Dans leurs travaux, Gratch et Marsella (2004) ont proposé un modèle d'émotions fondé sur l'architecture SOAR mais ce sont surtout les architectures de type BDI qui sont les plus utilisées, en particulier sous l'influence des chercheurs en SMA. En effet, ceux-ci ont exploré depuis le début la notion « d'agent cognitif » au moyen de théories destinées à modéliser les capacités de raisonnement des agents en s'appuyant généralement sur des architectures BDI. Remarquons que les termes *belief*, *desire* et *intention* ne sont pas réellement pris ici au sens psychologique ; ils correspondent de fait à des notions qui restent propres à l'action rationnelle (faits, buts, plan en cours).

Récemment, les chercheurs en SMA se sont intéressés à l'intégration de notions plus proches de la psychologie dans des architectures cognitives d'agents, par exemple, en ajoutant une couche à des plateformes SMA déjà existantes, comme dans le cas de CoJACK (Norling, Ritter, 2004 ; Evertsz *et al.*, 2008) pour la plateforme JACK. CoJACK prend en compte des paramètres émulant certaines caractéristiques de la physiologie humaine comme : la durée d'apprentissage ; les limites mémorielles (*e.g.* « perdre une croyance » si l'activation est basse ou encore « oublier l'action suivante » dans une procédure) ; la remémoration vague de croyances (*fuzzy retrieval*) ; la limitation du champ d'attention ; ou encore l'utilisation d'opérateurs, appelés modérateurs, qui altèrent le raisonnement cognitif lui-même.

Des études ont aussi été menées afin d'ajouter des émotions aux architectures d'agents BDI, par exemple pour prendre en compte des notions comme la peur, l'an-

xiété ou encore la confiance en soi (Pereira *et al.*, 2008). Ceci a été effectué par l'ajout de paramètres supplémentaires comme les désirs fondamentaux, les capacités et les ressources. Une autre voie (Casali *et al.*, 2004) a consisté à ajouter des degrés aux logiques multivaluées représentant les croyances, désirs et intentions, en particulier en s'appuyant sur la logique de Łukasiewicz (1963).

De manière indépendante, un certain nombre d'études ont porté sur l'introduction de capacités psychologiques dans des agents conversationnels. Dès le milieu des années 1990, Maes (1994), Bates (1994) puis Hayes-Roth (Hayes-Roth *et al.*, 1995) et Rousseau (Rousseau, Hayes-Roth, 1996) ont introduit la notion de *believable agents*, qui pose le principe que les traits de personnalité ont une influence non négligeable sur la prise de décision dans l'action rationnelle (sans aller cependant jusqu'aux positions extrêmes de Damasio (1994)).

Plus récemment, les travaux de McRorie *et al.* (2009) ou Bevacqua *et al.* (2010), fondés sur l'architecture d'agent conversationnel GRETA (Pelachaud, 2000), font intervenir des modèles de personnalité pour l'expression des émotions (faciales, gestuelles *etc.*). S'appuyant sur ces modèles, avec l'architecture FATIMA, Doce *et al.* (2010) affirment qu'il est indispensable de construire des agents qui ont des apparences physiques bien distinctes mais aussi des comportements psychologiques bien différenciés : « in the era of globalization, concepts such as individualization and personalization become more and more important in virtual systems. The FFM [*cf.* section 3.1] can be a basis for the creation of distinguishable personalities by using the personality traits to automatically influence cognitive processes: appraisal, planning, coping and bodily expression ».

2.3. Relations entre rationalité et psychologie dans l'interaction humain/agent

La citation de Doce *et al.* ci-dessus, prend une importance particulière lorsqu'elle est appliquée dans un contexte d'interaction humain/agent, par exemple dans le cas d'agents assistants conversationnels. Ceci nous a amenés à effectuer des expérimentations afin de valider l'hypothèse de la « naturalité dialogique des agents » qui stipule que les humains (surtout dans le grand public) préfèrent interagir avec des agents dont le comportement observable est produit par des heuristiques qui privilégient la *pertinence anthropologique* (en anglais, « human-likeness ») à la *performance rationnelle* telle que produite par exemple par des outils de planification. Ceci nous a amené à développer un outil dédié aux expérimentations sur les agents conversationnels, dans le cadre de l'internet, appelé DIVA (Sansonnet, 2008). Cet outil s'est avéré simple et efficace d'emploi ce qui a permis de développer plusieurs applications expérimentales (Sansonnet, 2008 ; Miletto *et al.*, 2009). Cependant les premières expérimentations avec des usagers novices (Xuetao *et al.*, 2009a) ont fait apparaître certaines limitations des agents assistants lorsqu'ils sont fondés sur un moteur strictement rationnel et elles ont bien mis en lumière au moins trois phénomènes saillants qui gênent la perception de la naturalité dialogique des agents :

1. *Répétition des réactions coopératives de l'agent :* l'agent est toujours servile (*responsive*) quoi qu'il puisse se passer pendant la session, quel que soit le rôle qu'il endosse et ceci quoi que lui demande ou dise l'usager. Par exemple, l'agent ne peut cacher une information qu'il possède à l'usager ni s'empêcher d'exécuter une action qu'il peut faire si l'usager le lui demande.

2. *Répétition des schémas de réponses :* l'agent fournit la même information plusieurs fois de suite sans aucune variation linguistique.

3. *Répétition des réactions rationnelles :* si l'usager pose plusieurs fois de suite la même requête, à chaque fois l'agent réagit à la requête indépendamment de la session dialogique, c'est-à-dire comme si la requête n'avait jamais été posée auparavant.

Ces observations montrent l'importance que le grand public apporte à la perception de comportements jugés non naturels lors de leurs interactions dialogiques avec des agents conversationnels chargés de les assister. Dans la section suivante, nous proposons une architecture d'agent assistant psychologique qui tente d'intégrer à la fois le comportement rationnel, indispensable, et ce que nous nommons par opposition le comportement psychologique.

3. Schémas psychologiques

3.1. Les taxonomies FFM *et* FFM/NEO PI-R

Plusieurs familles de théories psychologiques décrivant la personnalité d'un individu ont été proposées depuis longtemps : la psychanalyse freudienne, les types et les traits, l'approche humaniste de Maslow et Rogers, la théorie socio-cognitive de Bandura, *etc.* Parmi celles-ci, les descriptions à base de traits, largement utilisées en informatique affective (*affective computing* de R. Picard au MIT (Picard, 2000)) ont montré qu'elles étaient bien adaptées pour les agents cognitifs. Nous suivrons donc cette approche de la personnalité, en utilisant la taxonomie FFM (Goldberg, 1981) qui prédomine dans les études computationnelles (John *et al.*, 2008).

Comme toutes les autres taxonomies de traits visant à décrire l'ensemble de la personnalité humaine, FFM présente cependant un inconvénient majeur d'un point de vue computationnel : elle est trop générale, puisqu'elle ne définit que cinq grandes classes de traits, souvent appelées O.C.E.A.N. (pour Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism). De nombreuses études scientifiques ont été menées avec FFM. Toutefois, dans le cadre d'une implémentation informatique des traits, il nous faut disposer d'une plus grande précision descriptive des mécanismes qui constituent par exemple, la notion de Conscientiousness ou de Agreeableness. Les chercheurs en psychologie sont depuis longtemps conscients de la généralité de la taxonomie FFM. C'est la raison pour laquelle certains d'entre eux (Costa, McCrae, 1992 ; Saucier, Ostendorf, 1999 ; Soto, John, 2009) ont fait des expériences psychologiques afin d'introduire un second niveau de raffinement à FFM à l'aide de *facettes bipolaires* (c'est-à-dire décrivant un concept et son antonyme) associées à chaque trait. Parmi ces études en psychologie, nous avons opté pour la taxonomie raffinée, définie

Tableau 1. Définition générale de la taxonomie FFM/NEO PI-R

Traits FFM	Facettes NEO PI-R	Glose du pôle + de la facette
Openness	Fantasy	receptivity to the inner world of imagination
	Aesthetics	appreciation of art and beauty
	Feelings	openness to inner feelings and emotions
	Actions	openness to new experiences on a practical level
	Ideas	intellectual curiosity
	Values	readiness to re-examine own values and those of authority figures
Conscientiousness	Competence	belief in own self efficacy
	Orderliness	personal organization
	Dutifulness	emphasis placed on importance of fulfilling moral obligations
	Achievement-striving	need for personal achievement and sense of direction
	Self-discipline	capacity to begin tasks and follow through to completion despite boredom or distractions
	Deliberation	tendency to think things through before acting or speaking
Extraversion	Warmth	interest in and friendliness towards others
	Gregariousness	preference for the company of others
	Assertiveness	social ascendancy and forcefulness of expression
	Activity	pace of living
	Excitement-seeking	need for environmental stimulation
	Positive-emotions	tendency to experience positive emotions
Agreability	Trust	belief in the sincerity and good intentions of others
	Straightforwardness	frankness in expression
	Altruism	active concern for the welfare of others
	Compliance	response to interpersonal conflict
	Modesty	tendency to play down own achievements and be humble
	Tender-mindedness	attitude of sympathy for others
Neuroticism	Anxiety	level of free floating anxiety
	Angry-Hostility	tendency to experience anger and related states such as frustration and bitterness
	Depression	tendency to experience feelings of guilt, sadness, despondency and loneliness
	Self-consciousness	shyness or social anxiety
	Impulsiveness	tendency to act on cravings and urges rather than reining them in and delaying gratification
	Vulnerability	general susceptibility to stress

par Costa et McCrae, FFM/NEO PI-R, en raison de sa popularité et du nombre supérieur de facettes qu'elle propose (6 facettes bipolaires (concept/antonyme) par traits, soit 30 au total). Le tableau 1 présente la taxonomie FFM/NEO PI-R complète, avec pour chaque facette sa définition intuitive sous forme de glose.

3.2. La taxonomie TFS

Malheureusement, comme le montre le tableau 1, la taxonomie FFM/NEO PI-R reste encore très générale : chaque facette est définie par une simple glose [1]. Ceci pose un problème au moment de l'implémentation informatique. En effet, il nous faudrait disposer de correspondances, si possible psychologiquement testées, entre les facettes et des comportements effectifs d'individus en action : pour chaque glose il faudrait disposer non pas d'une glose parfois assez abstraite et absconse, mais plutôt d'un ensemble de gloses qui décrivent concrètement des comportements précis (par exemple « évite les tâches pénibles », « vote en faveur des partis conservateurs », « console les personnes soudainement tristes », *etc.*).

C'est précisément à cette tâche que nous nous sommes consacrés lors de travaux précédents où avons été amenés à enrichir la taxonomie FFM/NEO PI-R, constituant ainsi la ressource TFS [2](Bouchet, Sansonnet, 2010), dans laquelle chaque facette est décomposée en concepts appelés *schémas comportementaux* (ou *schémas* en abrégé). Notons que TFS n'est pas tant une nouvelle taxonomie qu'un *enrichissement* de FFM/NEO PI-R : elle hérite donc des analyses ayant permis la constitution de celle-ci (analyses factorielles issues soit de tests de personnalité, soit d'analyses lexicales). La description complète de cette approche et la documentation de TFS sont disponibles en tant que ressource libre d'accès sur le web [3]. Chaque schéma y est défini par un couple d'ensembles de gloses : l'ensemble décrivant le concept du schéma et l'ensemble décrivant son antonyme. Ces ensembles ont été constitués par l'analyse de deux types de ressources croisées : tout d'abord nous avons eu recours à la base lexicale WordNet (Fellbaum, 1998) pour récolter et catégoriser les gloses définissant les synsets associés à 1 055 adjectifs de personnalité ; ensuite les catégories obtenues ont été alignées avec 300 gloses (appelées *q-items*) tirées de questionnaires utilisés en psychologie par Goldberg [4]. Chaque catégorie et son antonyme définissent alors un schéma comportemental bipolaire, souvent dénoté par le concept associé au pôle positif. La taxonomie TFS (dont la construction est détaillée dans (Bouchet, Sansonnet, 2010)) est constituée de 70 schémas bipolaires, contenant en tout 766 gloses distinctes.

La figure 1 donne un extrait pour deux des schémas contenus dans la facette Conscientiousness/Competence de la ressource.

---

1. Phrase courte, décrivant généralement un élément de sémantique lexicale, comme pour les définitions des sens d'un mot dans les dictionnaires ou pour les définitions des *synsets* de la base lexicale WordNet.
2. TFS pour **T**raits - **F**acettes - **S**chémas.
3. http://perso.limsi.fr/jps/research/rnb/toolkit/taxo-glosses/taxo.htm
4. http://ipip.ori.org/newNEOKey.htm

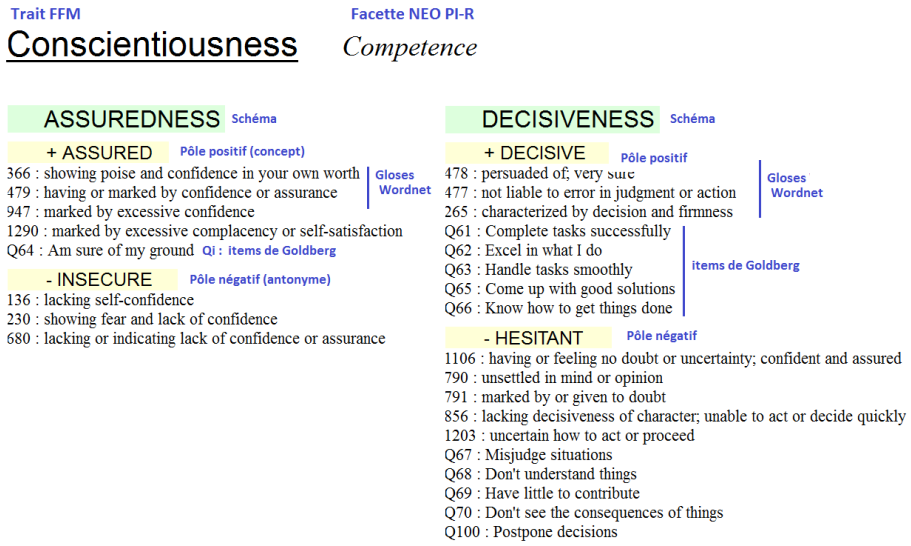

Figure 1. Extrait de la ressource Web définissant TFS*, présentant les deux schémas constitutifs de la facette* Compétence *du trait* Conscientiousness*:*
à gauche) les gloses associées aux pôles +/- du schéma ASSUREDNESS
à droite) les gloses associées aux pôles +/- du schéma DECISIVENESS

L'outil TFS permet non seulement de disposer de beaucoup plus de concepts pour décrire la personnalité d'un individu (70 schémas bipolaires *vs* 30 facettes bipolaires), mais surtout propose pour chacun d'entre eux plusieurs gloses, ce qui se révèle utile de deux manières complémentaires :

1) d'un point de vue psychologique : lors de la définition d'une personnalité, l'expert peut s'appuyer sur ces gloses pour attacher à un individu x non pas un trait global comme Agreeable mais plus particulièrement la facette Agreeable/+trust « l'individu fait confiance aux autres » et même de manière encore plus précise de distinguer entre Agreeable/+trust/+trustful « l'individu croit que les autres ne mentent pas » et Agreeable/+trust/+ungarded « l'individu ne se méfie pas des autres et leur donne des information intimes ». On constate bien qu'il s'agit de comportements en situation, plus facile à interpréter pour les implémenter dans le cadre d'une application donnée.

2) d'un point de vue computationnel : les gloses associées aux schémas apportent deux améliorations :
a) elles permettent d'éliciter plus facilement les opérateurs d'influence ainsi que la matrice d'activation schémas/opérateurs (ce point crucial sera développé dans l'étude de cas décrite dans la section 6) ;
b) elles offrent une meilleure garantie de couverture du domaine psychologique d'un trait donné. En effet, la décomposition du trait en sous-éléments décrit de manière

plus précise sa sémantique. On peut donc définir une personnalité dans son ensemble plutôt que seulement quelques aspects saillants de celle-ci.

Ainsi côté psychologique, il est permis de ne pas être d'accord avec une annotation particulière pour x, ou encore, côté computationnel, avec une association particulière entre un schéma et un opérateur, mais dans les deux cas, le processus est rendu plus précis et plus clair que dans certaines études similaires précédentes (*cf.* celles mentionnées dans la section 2) par l'architecture que nous proposons dans la section suivante.

4. Moteurs de personnalités

4.1. Architecture générale

Avant de décrire la structure formelle de l'architecture proposée, nous donnons une vue d'ensemble des notions qui sont utilisées ainsi que de leur articulations. La figure 2 présente le principe général de l'expression des traits de personnalité TFS comme influences sur le cycle de délibération BDI. Tout d'abord, elle met en exergue la séparation entre le domaine psychologique (en haut) et le domaine rationnel (en bas). Ensuite elle sépare :

– la phase d'élicitation des opérateurs d'influence (à gauche), réalisée une fois pour toute à partir de ressources existantes (*e.g.* littérature en psychologie, dictionnaires, base électronique WordNet, définition du cycle BDI) ;

– la phase d'implémentation d'une personnalité (à droite) réalisée à chaque nouvelle application, et qui dépend d'une part du profil de personnalité désiré (haut) et d'autre part du domaine applicatif visé (bas).

La construction de la taxonomie TFS est décrite en section 3.2 et l'élicitation des sites d'influence est effectuée à partir d'un modèle BDI et d'un modèle de plans qui sont décrits dans la section 5.1. La phase d'implémentation de la personnalité repose sur la notion de moteur de personnalité. Cette notion est introduite pour regrouper les différentes ressources nécessaires, et pour distinguer celles qui sont génériques (*i.e.* construites une fois pour toutes) de celles qui sont dépendantes de l'application. La structure d'un moteur de personnalité est décrite dans le paragraphe suivant. Ensuite la phase d'implémentation est illustrée à la section 6 par une étude de cas, suivie d'un exemple issu de la littérature.

4.2. Structure d'un moteur de personnalité

4.2.1. Composantes

Un moteur de personnalité MP est défini comme une structure à 4 composantes $\mathtt{MP} = \langle \mathtt{O}, \mathtt{W}, \Omega_{\mathtt{W}}, \mathtt{M} \rangle$ où :

– O est *ontologie de personnalité*, qui doit permettre la description précise de personnalités. Dans la suite, nous utilisons à fin d'exemple l'ensemble Σ de schémas bipolaires de TFS (*cf.* section 3), on aura donc $|\Sigma| = 70$. L'ensemble des posi-

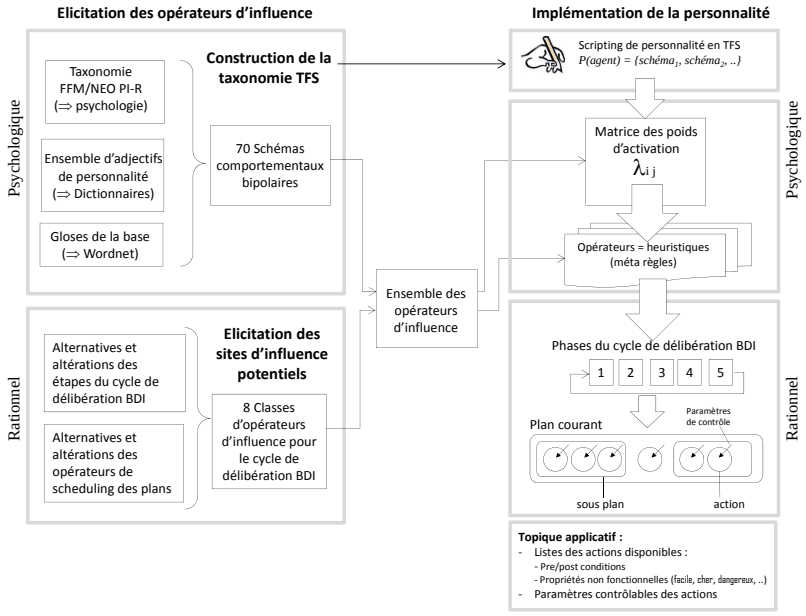

Figure 2. Principe général de l'expression des traits de personnalité TFS *comme influences sur le cycle de délibération BDI*

tions positives (*resp.* négatives) est noté $+\Sigma$ (*resp.* $-\Sigma$) et leur union est $\pm\Sigma$ tel que $\pm\Sigma = +\Sigma \cup -\Sigma$ et $|\pm\Sigma| = 140$;

– $\mathtt{W} = \langle \mathtt{Wk}, \mathtt{Wm}, \mathtt{Wc}, \mathtt{Wr} \rangle$ est un *modèle de monde d'agents* qui inclut : leur structure de connaissances rationnelles Wk (en anglais *knowledge base*) ; leur structure mentale Wm (*mind*) ; leurs protocole de communication externe Wc (*e.g.* type FIPA) ; leur processus de décision rationnelle Wr. Par exemple, Wr peut être un modèle fondé sur l'approche BDI ou un modèle encore plus spécifique comme celui défini à la section 5.1 ;

– $\Omega_{\mathtt{W}}$ est un ensemble d'*opérateurs d'influence* sur des règles de $\mathtt{Wr} \cup \mathtt{Wc} = \mathtt{Wrc}$;

– M est une *matrice d'activation* qui établit une relation sur $\pm\Sigma \times \Omega_{\mathtt{W}}$.

Les composantes O et W sont considérées comme des ressources *données* tandis que les composantes $\Omega_{\mathtt{W}}$ et M doivent être élicitées à partir de O et W en fonction du cas à implémenter, selon le processus décrit à la section 6.2.

4.2.2. Opérateurs

Étant donné un modèle de raisonnement rationnel W, les opérateurs d'influence $\omega \in \Omega_{\mathtt{W}}$ sont des métarègles qui contrôlent ou altèrent la partie non structurelle de W, *i.e.* Wrc. Formellement, toute règle portant sur Wrc est un opérateur d'influence,

cependant seules seront considérées comme pertinentes les règles qui auront une interprétation possible en termes des schémas de $\mathtt{O}$. En conséquence, l'élicitation de tels opérateurs est un processus de $\mathtt{Wrc} \times \mathtt{O} \longmapsto \Omega_{\mathtt{W}}$ plutôt que $\mathtt{Wrc} \longmapsto \Omega_{\mathtt{W}}$.

La plupart des opérateurs sont activés en "tout ou rien" (la règle est appliquée ou pas), mais certains peuvent prendre des paramètres d'activation. On considérera deux cas fréquents :

– Une *intensité* ou force d'application est fournie ;

– Certains opérateurs peuvent fonctionner en mode inverse ou en mode antonyme et prennent alors une *direction*, notée par le préfixe +/- (+ est souvent omis).

4.2.3. Matrice d'activation

Étant donné un ensemble de schémas $\sigma \in \pm\Sigma$ et un ensemble d'opérateurs $\omega \in \Omega_{\mathtt{W}}$, le(s) concepteur(s) d'un moteur de personnalité particulier doit expliciter comment $\pm\sigma_i$ sont associés à ω_i, c'est-à-dire quels schémas activent quels opérateurs. Ces relations, qui peuvent varier d'un concepteur à l'autre, sont établies au moyen d'une matrice $\mathtt{M}$ dont les éléments sont appelés des *niveaux d'activation* $\lambda_{i,j}$, telle que $\mathtt{M} = \pm\Sigma \times \Omega_{\mathtt{W}}$ (*cf.* tableau 2).

Tableau 2. Valeurs des niveaux d'activation $\lambda_{i,j}$

Valeur	Opérateur actif	Direction	Intensité
2	oui	+	forte
1	oui	+	modérée
0	non	N/A	N/A
-1	oui	-	modérée
-2	oui	-	forte

4.3. Instanciation d'un moteur

Après avoir défini un moteur de personnalité particulier $\mathtt{MP}_0$, on dispose d'une structure symbolique qui peut alors être instanciée dans diverses situations effectives $\mathtt{MP}_0(\mathtt{T}_i, P(x_i))$ qui sont définies par deux paramètres indépendants :

– $\mathtt{T}_i$ est un *topique applicatif* ;

– $P(x_i)$ est un *profil de personnalité*.

4.3.1. Topique applicatif

On utilisera le terme 'topique' $\mathtt{T}_i$ pour désigner une instance de $\mathtt{W}$ qui a pour rôle de fournir : a) un ensemble d'actions disponibles $\alpha_i \in \mathtt{A}(\mathtt{T}_i)$ (où $\mathtt{A}$ est une API) ; b) des informations dépendantes de l'application définissant les modalités d'exécution des actions. Par exemple, posons l'opérateur $\omega_{+safe} \doteq Sort(\{\alpha_i\}, \prec_{danger})$ qui trie un ensemble d'actions de la plus sûre à la plus dangereuse. Il a besoin que $\mathtt{T}_i$ lui fournisse une fonction de mesure $\mu_{danger} : \mathtt{A}(\mathtt{T}_i) \longmapsto [0, 1]$.

4.3.2. Profil de personnalité

Étant donné un individu x_i sa personnalité, dénotée $P(x_i)$, fera office de profil de personnalité de l'agent dans la situation instanciée. Intuitivement, les profils de personnalité sont souvent définis par des ensembles de mots ; des adjectifs ou des adverbes qui décrivent le comportement d'une personne agissant seule ou en société (*e.g.* on dit de $Pierre$ qu'il est *attentif*, *travailleur* et *passif*). La recherche sur les taxonomies de traits a permis d'établir des profils techniquement plus précis. En particulier, l'usage de la taxonomie TFS permet de produire des profils bien fondés, selon la définition suivante :

Étant donné un individu x, sa personnalité $P(x)$ est définie par un ensemble de $|\Sigma|$ fonctions $\mathrm{p}(\sigma_i) : \Sigma \longrightarrow \{+, \asymp, -\}$ où:

- $\asymp$ signifie que vis-à-vis du schéma σ_i, le comportement de x n'est pas significativement déviant d'un comportement moyen ;
- $+$ signifie que le comportement de x est significativement déviant d'un comportement moyen, selon le pôle positif ;
- $-$ respectivement selon le pôle négatif.

Notation. Si on considère les 70 schémas de Σ, les personnes ont tendance à exhiber un comportement moyen dans la grande majorité des cas. En conséquence $P(x)$ est majoritairement composé d'éléments ayant pour valeur $\asymp$. C'est pourquoi $P(x)$ est préférablement donné comme l'ensemble des schémas différents de $\asymp$. Par exemple, pour $Pierre$ on aura $P(Pierre)$ = {+ATTENTIVE, +HARDWORKER, -PASSIVE}, ignorant ainsi les 67 autres schémas pour lesquels le comportement de $Pierre$ n'est pas particulièrement notable.

5. Support des influences

5.1. Modèle de raisonnement

5.1.1. Choix du modèle

En principe, chacun des modèles mentionnés en section 2.1 offre, selon des modalités qui leur sont propres, des opportunités d'expression des influences issues des traits psychologiques. Dans cette étude, nous appliquons ce principe dans le cadre BDI, en nous focalisant sur le modèle INDIGOLOG (Giacomo *et al.*, 2009) qui apporte à CAN les aspects situationnels dans un cadre déclaratif (en Prolog). Le modèle INDIGOLOG étant complexe, nous nous limitons à ses principaux opérateurs dans une version simplifiée (*cf.* tableau 3), suffisante toutefois pour illustrer notre propos.

5.1.2. Cycle BDI de l'agent

Le processus de raisonnement rationnel d'un agent de type BDI est typiquement un cycle composé de cinq étapes :

1. *Délibération* Étant donné un agent muni d'un ensemble Γ de désirs $\gamma_i \in \mathbb{G}$, éventuellement contradictoires, lors de la délibération l'agent sélectionne un sous-

ensemble d'éléments Γ^* dès lors appelés *buts* , tels que 1) l'agent a l'intention de les atteindre ; 2) ils ne sont pas mutuellement contradictoires.

2. *Planification* Étant donné le but en cours d'accomplissement γ^* dans Γ^*, l'agent génère un plan rationnel $\pi^* \in \mathbb{P}$ qui a γ^* dans ses postconditions : $\gamma^* \in post(\pi^*)$. Un plan est une expression symbolique bien formée dans le langage de plans $\mathcal{L}_\pi$, défini dans le paragraphe suivant. Il est composé de sous-plans $\pi_i \in \mathbb{P}$ et d'actions terminales $\alpha_i \in \mathbb{A}$.

3. *Décision* Bien que le planificateur propose une expression unique π^* afin d'accomplir le but courant γ^*, cette expression contient souvent des alternatives (sous-plans ou actions équivalents vis-à-vis du but) ainsi que des options (sous-plans ou actions indifférents vis-à-vis but). En conséquence les plans sont sous-déterminés et l'agent doit choisir entre les alternatives et décider s'il exécute ou non les options. En tout état de cause, le résultat de l'étape de décision est l'élection de l'action terminale suivante à exécuter α^* dans le plan π^*.

4. *Exécution* L'agent exécute l'action élue α^*. Dans la plupart des architectures BDI cette étape est implicite (c'est-à-dire vue comme une boîte noire). Un point original de notre architecture est que nous prenons en compte différents types de modalités avec lesquelles une action est exécutée. Ces modalités (*cf.* tableau 4) sont considérées comme des paramètres explicitement contrôlables de l'action. Ceci est possible grâce aux données fournies par $\mathtt{T}_i$. Cela permet d'offrir un support aux opérateurs d'influence qui portent sur la manière dont les actions sont exécutées (notions d'intensité et de gradualité).

5. *Évaluation* L'agent reçoit du topique applicatif des informations sur les conséquences effectives de l'exécution de α^*. Cela permet à l'agent d'évaluer ses performances et de les comparer à ses attentes pour en tirer des conséquences 1) en termes rationnels pour le prochaine itération du cycle BDI et 2) en termes psychologiques pour la mise à jour de son état mental $\mathtt{Wm}$.

Tableau 3. Opérateurs procéduraux des plans

Nom	Op.	Description générale ($x_i = a_i \vert \pi_i$)
seq	;	$x_1; x_2$: Done(x_1) = précond pour exécuter x_2
alt	$\vert$	$x_1 \vert x_2$: un seul x est tiré au sort et exécuté
par	$\Vert$	$(x_1 \Vert x_2) \equiv (x_1; x_2) \vert (x_2; x_1)$ les x sont exhaustivement tirés au sort et exécutés
case	$\mapsto$	$\mathrm{guard}_1 \mapsto x_1, \mathrm{guard}_2 \mapsto x_2$ guard_i sont des préconds explicites pour que x_i soit exécutable. Si plusieurs gardes sont vraies, une est tirée au sort et son x est exécuté

5.1.3. Langage de plan

Le langage de plan $\mathcal{L}_\pi$ permet de définir des expressions composées de deux sortes d'entités : sous-plans et actions terminales. Les actions sont alors définies en termes des fonctions dépendantes de $\mathtt{T}_i$. Dans la suite, s'il n'y a pas ambiguïté, nous dirons

Tableau 4. Les huit classes d'opérateurs d'influence pour le cycle BDI

Classe (**C**ode) *Opérande* Etape BDI (n°) Description générale
Goal γ Délibération (1) Contrôle la gestion des buts : élicitation à partir des désirs, ordonnancement dans la pile des buts.
Preference $\alpha\|\pi$ Décision (3) Comparaison et tri en ordre total ou partiel d'ensembles de $\alpha\|\pi$ considérés comme *rationnellement* équivalents. Utilisation : implémentation des attitudes mentales à propos d'entités préférées/détestées comme décrit section 4.3–topique applicatif.
Filter $\alpha\|\pi\|\gamma$ Délibération (1), Décision (3) Décision de conserver ou de rejeter un élément $\gamma\|\alpha\|\pi$ dans la suite du cycle BDI. Alors que les préférences fonctionnent par comparaison d'éléments, les filtres implémentent les attitudes mentales intrinsèques de l'agent vis-à-vis d'un élément.
Scheduling π Exécution (4) Contrôle la phase de scheduling des plans, selon deux modes : les *politiques* d'ordonnancement portent sur les arguments des opérateurs (*e.g.* pour $S_p\|S_o\|S_d$ dans $CorpsDecl$) ; les *altérateurs* modifient la structure (*e.g.* remplacer un seq par un par est associé au schéma -DISORGANIZED du tableau 6 via l'opérateur S_{order}).
Modal $\varphi\|\alpha\|\pi$ Exécution (4) Contrôle l'exécution des actions selon deux modes, activables ou non en fonction des données fournies par le topique $\mathtt{T}_i$: la *gradualité* porte sur le degré d'accomplissement final d'une l'action ; l'*intensité* (*cf.* section 4.2) sur son déroulement.
Option α Exécution (4) Ajout d'actions, non nécessaires au succès du plan, avant ou après l'exécution de l'action courante. Elles modifient les pre/post conditions mais n'altèrent pas les conditions menant aux buts (*e.g.* appliquer l'option close-door après open-door).
e**X**pectation $\alpha\|\pi\|\gamma$ Evalualtion (5) Définition des attitudes mentales de l'agent au sujet des résultats *possibles* des actions exécutées selon trois mode principaux : *espérance*, *crainte* et *neutre*. Note : ils ont une attitude neutre vis-à-vis des résultats *nécessaires* de leurs actions.
Appraisal $\alpha\|\pi\|\gamma$ Evalualtion (5) Définition de l'impact sur le modèle mental de l'agent de la différence entre les résultats espérés/craints et les résultats constatés après l'exécution d'une action.

plans pour sous-plans, actions pour actions terminales, et fonctions pour fonctions dépendantes de $\mathtt{T}_i$.

Plans. Ce sont des expressions de la forme $\pi = \langle p_i, \gamma_i, pre, post, attr, corps \rangle$, où :

– p_i est un identifiant unique de plan. Ci-dessous les plans sont dénotés π ou p_i.

– pre est l'ensemble des préconditions rationnelles du plan (noté $pre(p_i)$), *i.e.* les faits qui *doivent* être vrais dans le monde pour que π soit applicable.

– $post$ est l'ensemble des postconditions rationnelles du plan (noté $post(p_i)$), *i.e.* les faits qui seront *nécessairement* ou *possiblement* vrais dans le monde après l'exécution de π.

– $\gamma_i \subseteq post(p_i)$ est l'ensemble des buts *associés au plan*. Notons que les agents n'ont pas forcément l'intention de rendre vraies toutes ces postconditions, mais seulement celles en intersection avec γ_i (*cf.* le bombardier stratégique (Bratman, 1987)).

– $attr$ est l'ensemble des attributs définissant le plan au regard des influences psychologiques.

– $corps$ est une expression définissant comment le plan est implémenté en termes de sous-plans et d'actions terminales, à l'aide d'opérateurs de composition définis dans le tableau 3.

Actions. Ce sont des expressions de la forme $\alpha = \langle a_i, \gamma_i, pre, post, attr, appel \rangle$ où $appel$ est un appel de fonction $f_i \in \mathbb{F}$, exécutable dans $\mathtt{T}_i$ et où les autres éléments sont définis comme pour les plans ci-dessus.

Fonctions. Fournies par le topique applicatif $\mathtt{T}_i$ par le biais d'une API $\mathbb{F}$, elles sont de la forme $\varphi = \langle f_i, args, params \rangle$, où :

– f_i est un identifiant unique.

– $args$ est l'ensemble des descriptions des arguments *rationnels* de la fonction f_i. Ils sont strictement opératoires.

– $params$ est l'ensemble des descriptions des arguments *modaux* de la fonction f_i. Ils correspondent aux propriétés non fonctionnelles (*e.g.* vitesse d'exécution, consommation de ressources, *etc.*). On posera $args \cap params = \emptyset$.

Postconditions. Notées $post(a_i)$ pour l'action a_i, on distingue :

– les postconditions nécessaires (notées $post_N$) sont forcément vraies après l'exécution de a_i, *e.g.* étant donnée a_0 associée à la fonction lancer-un-dé, alors $face = i$; $i \in [1..6]$ appartient à $post_N(a_0) \subseteq post(a_0)$;

– les postconditions possibles ($post_P$) ne sont que potentiellement vraies *e.g.* $face = 6 \in post_P(a_0) \subseteq post(a_0)$.

Corps de plan. Il a pour syntaxe :

Corps	$\rightarrow$ CorpsProc | CorpsDecl | a_i
CorpsProc	$\rightarrow$ ProcOp[Corps$\|p_i\|a_i, \ldots$]
CorpsDecl	$\rightarrow \langle S_p, S_o, S_d \rangle$
$S_p \| S_o \| S_d$	$\rightarrow \{p_i \| a_i, \ldots\}$

où :

– CorpsProc est composé d'opérateurs procéduraux (ProcOp) tels que définis dans le tableau 3 ;

– CorpsDecl est une expression déclarative composée de trois ensembles d'éléments $p_i|a_i$: S_p (éléments préférés), S_o (élements optionnels) et S_d (éléments à exécuter par défaut [5]).

Dans chacun des ensembles, les éléments sont considérés par le planificateur comme rationnellement équivalents vis-à-vis du but poursuivi. Par contre la distinction entre 'préférés', 'optionnels' et 'défauts' offre une opportunité aux opérateurs d'influence de s'exercer au moment de la sélection de l'action à exécuter.

5.2. Classes d'opérateurs

Les opérateurs d'influence peuvent être appliqués à quatre types d'opérandes : des buts (γ), des actions (α), des plans (π) ou des fonctions (φ). Étant donné un modèle particulier de raisonnement, on peut distinguer plusieurs grandes classes d'opérateurs. Pour le modèle de raisonnement défini à la section 5.1, il est possible d'exhiber huit classes dont les descriptions générales sont listées dans le tableau 4.

6. Étude de cas

6.1. Moteur de personnalité : $\mathtt{MP}_{\mathbf{C}}$

Afin de donner une première validation de l'approche proposée, nous détaillons dans cette section une étude de cas pour un moteur de personnalité particulier, $\mathtt{MP}_{\mathbf{C}}$, tel que $\mathtt{MP}_{\mathbf{C}} = \langle \mathtt{O}, \mathtt{W}, \Omega_{\mathtt{W}}, \mathtt{M}_{\mathbf{C}} \rangle$ où :

– $\mathtt{O}$ est TFS, restreinte pour cette étude au trait FFM Conscientiousness, choisi comme exemple car il est représentatif des phénomènes que l'on rencontre dans les quatre autres traits de FFM.

– $\mathtt{W} = \langle \mathtt{Ws}, \mathtt{Wc}, \mathtt{Wr} \rangle$ où $\mathtt{W}$ est le modèle du monde des agents qui inclut : leur structure interne $\mathtt{Ws}$; leur protocoles de communication externe $\mathtt{Wc}$; leur processus de décision rationnels $\mathtt{Wr}$. $\mathtt{W}$ n'est pas entièrement détaillé dans cet article (pour la partie $\mathtt{Ws}$ voir (Sansonnet, Bouchet, 2010) et pour la partie $\mathtt{Wc}$ (Sansonnet, Bouchet, 2013)). Le processus de décision rationnelle $\mathtt{Wr}$ est celui décrit section 5.1.

– $\Omega_{\mathtt{W}}$ est la liste des opérateurs d'influence associés au trait FFM Conscientiousness. La liste des traits élicités est décrite dans le tableau 5.

– $\mathtt{M}_{\mathbf{C}}$ est la sous-matrice d'activation reliant les schémas du trait Conscientiousness aux opérateurs associés tels que décrits dans le tableau 5 (*cf.* figure 1).

5. Ce sont des éléments $p_i|a_i$ tels que $pre(p_i|a_i) \neq \emptyset$. En effet pour être rationnellement corrects, les plans doivent à chaque étape posséder au moins un $p_i|a_i$ exécutable ; S_d permet d'éviter les blocages.

Tableau 5. 30 opérateurs pour Conscentiousness

$Classe_{nom}$	Glose (description générale) du pôle +
$G_{forget}+$	oublie γ (*e.g.* efface de la pile des buts)
G_{quick}	délibère rapidement (*e.g.* prend toujours le premier élément d'une liste d'alternatives)
$G_{deduct}+$	contrôle la 'profondeur' de déduction
$G_{prudent}$	prend en compte les conséquences de γ
$G_{nowaste}$	déteste les γ qui consomment des ressources
$G_{knowledge}+$	* possède une grande base de faits
$G_{reason}+$	* possède une grande base d'heuristiques
P_{easy}	préfère les $\gamma\|\pi$ faciles
P_{safe}	préfère les $\gamma\|\pi$ non dangereux
P_{clean}	préfère les $\gamma\|\pi$ propres/nets
$P_{ambitious}$	préfère les $\gamma\|\alpha\|\pi$ ambitieux
F_{lawful}	rejette les $\gamma\|\alpha\|\pi$ non légaux
F_{safe}	rejette les $\gamma\|\alpha\|\pi$ dangereux
S_{order}	exécute $\alpha\|\pi$ dans l'ordre donné
$S_{predictable}$	exécute $\alpha\|\pi$ toujours dans le même ordre
M_{focus}	exécute $\alpha\|\pi$ avec concentration mentale
$M_{precision}$	exécute $\alpha\|\pi$ avec précision
$M_{tenacity}$	persévère si difficultés/échecs lors de $\alpha\|\pi$
M_{quick}	exécute $\alpha\|\pi$ rapidement
M_{dopar}	contrôle la gradualité de $\alpha\|\pi$ (*cf.* tableau 4-M)
$M_{nowaste}$	ne gaspille pas de ressources lors de $\alpha\|\pi$
M_{clean}	exécute $\alpha\|\pi$ de manière 'propre/nette'
$O_{extra}+$	tendance à ajouter des actions optionnelles
O_{play}	ajoute des actions 'agréables' à $\alpha\|\pi$
$O_{cleanup}$	ajoute des actions de 'nettoyage' après $\alpha\|\pi$
$X_{success}$	attend normalement le succès de $\gamma\|\alpha\|\pi$
X_{hope}	espère fortement que $poss(\alpha\|\pi)$ advienne
$X_{choiceConf}$	sûr de ses choix lors l'élaboration de $\gamma\|\alpha\|\pi$
$A_{responsible}$	se sent responsable des résultats de $\gamma\|\alpha\|\pi$
$A_{success}$	a une sensibilité forte au succès

Les opérateurs ont aussi un pôle négatif (antonyme) *e.g.* $-O_{cleanup}$ = "qui ajoute des actions qui salissent ou mettent en désordre après $\alpha|\pi$ "

\+ Opérateur n'ayant pas d'antonyme

* Opérateur appartenant aussi au raisonnement rationnel

6.2. Processus de construction

6.2.1. Elicitation des opérateurs

Le processus de construction des opérateurs d'influence est mené en confrontant les ensembles de gloses associés aux schémas de O avec les classes d'influences engendrées par Wr. Par exemple, si on considère les gloses contenues dans le schéma ASSUREDNESS (pôles + et -) de la figure 1 et qu'on les confronte au tableau 4, on peut éliciter des opérateurs comme :

Pôle	Gloses de ASSUREDNESS	Opérateurs évoqués[a]
	G366: showing poise and confidence in your own worth	$+M_{tenacity}$ (**M**odal)
	G479: having or marked by confidence or assurance	$+M_{precision}$ (**M**odal)
+	G947: marked by excessive confidence	$+X_{hope}$ (e**X**pectation)
	G1290: marked by excessive complacency, self-satisfaction	
	Q64: Am sure of my ground	
	G136: lacking self-confidence	$-M_{tenacity}$
-	G230: showing fear and lack of confidence	$+P_{safe}$ (**P**reference)
	G680: lacking or indicating lack of confidence or assurance	$-M_{precision}$

[a] Pour chaque pôle, les opérateurs listés sont le résultat global de l'analyse de la liste des gloses du pôle.

Ces opérateurs ont un préfixe de direction (+/-) dénotant le concept et son antonyme. Par exemple, $+P_{safe}$ veut dire « qui recherche de préférence les actions non dangereuses (pour soi-même ou le monde) », tandis qu'à l'opposé, $-P_{safe}$ veut dire « qui recherche de préférence les actions dangereuses ». L'implémentation effective d'un opérateur comme P_{safe} par la classe **P**reference a été décrite dans la section 4.3.

Pour un même schéma de TFS, un même opérateur peut être évoqué plusieurs fois, par exemple lorsque les gloses sont très proches ; il peut être évoqué par le concept du schéma ou son antonyme ou les deux (dans ce cas souvent en mode inverse). A la fin du processus sur un ensemble des schémas TFS, on procède à l'union des opérateurs. Ainsi le tableau 5 liste les opérateurs élicités à partir des schémas du trait Conscientiousness de FFM.

6.2.2. Définition de la matrice d'activation

La matrice d'activation est alors construite en associant les schémas de O aux opérateurs au moyen de poids $\lambda_{i,j}$ définissant l'intensité d'activation de l'opérateur ainsi que sa direction (+/-). Pour le moteur $\mathtt{MP}_{\mathbf{C}}$, un exemple de sous-matrice d'activation $\mathtt{M}_{\mathbf{C}}$ est donné (sous forme d'un extrait limité aux quatre premières facettes du trait Conscientiousness) dans le tableau 6. Pour simplifier, nous avons restreint cette présentation au raisonnement d'agents autonomes et donc écarté quatre schémas portant sur les aspects interpersonnels [6].

6. Ceci est consistant avec le fait que Wc ne soit pas détaillé ici. Pour une étude plus complète sur les aspects interpersonnels *cf.* (Sansonnet, Bouchet, 2013).

Tableau 6. Extrait de la sous matrice d'activation M_C (facettes 1 – 4)

Facettes ⇒		Competence		Orderliness		Self-Discipline			Achievement -striving	
Schémas ⇒ Opérateurs ⇓	Agent générique (défaut)	⊕ ASSURED ⊖ INSECURE	⊕ DECISIVE ⊖ HESITANT	⊕ METHODICAL ⊖ DISORGANIZED	⊕ TIDY ⊖ MESSY	⊕ CONCERNED ⊖ UNCONCERNED	⊕ ATTENTIVE ⊖ INATTENTIVE	⊕ MINDFUL ⊖ FORGETFUL	⊕ AMBITIOUS ⊖ UNAMBITIOUS	⊕ HARDWORKER ⊖ LAZY
G_{forget}	1							0 2		0 2
G_{quick}	1		2 -2							1 -2
G_{deduct}	1									
$G_{prudent}$	1									
$G_{nowaste}$	1									
$G_{knowledge}$	1									
G_{reason}	1									
P_{easy}	1									0 -2
P_{safe}	1	0 2								
P_{clean}	1				2 -1					
$P_{ambitious}$	1								2 0	
F_{lawful}	1									
F_{safe}	1						2 0			
S_{order}	1			2 -1						
$S_{predictable}$	1			2 -1						
M_{focus}	1	2 -1	1 0	2 -1		2 0	2 0		2 0	2 -2
$M_{precision}$	1		1 0			2 0	2 0		2 0	2 -2
$M_{tenacity}$	1	2 -2	2 -1			2 0			2 -2	2 -1
M_{quick}	1		1 -2							1 -2
M_{dopar}	1			2 -2					0 2	2 -2
$M_{nowaste}$	1			2 0						
M_{clean}	1				2 -1					1 -2
O_{extra}	1			0 2						2 0
O_{play}	1									0 0
$O_{cleanup}$	1				2 0					2 0
$X_{success}$	1	2 -2							2 0	
X_{hope}	1	2 -2							2 0	
$X_{choiceConf}$	1	2 -2	2 -2							
$A_{responsible}$	1					2 0				2 -1
$A_{success}$	1								2 -2	1 -1

Pour ne pas surcharger la lecture, les valeurs $\lambda_{i,j}$ correspondant à un comportement considéré comme moyen/normal sont factorisées en tête dans la colonne « agent générique (défaut) » et représentés par des cellules vides.

6.2.3. Définition d'une personnalité cohérente

La personnalité $P(x)$ d'un individu x, telle que définie section 4.3, utilise la matrice M pour activer les opérateurs qui à leur tour influencent le cycle de délibération de l'agent. La cohérence des traits dans $P(x)$ est une question appartenant au domaine de la psychologie : par exemple, les traits +AMBITIOUS et -INSECURE apparaissent rarement ensemble tandis que -MESSY et -LAZY apparaissent fréquemment ensemble. Néanmoins, en théorie, pour un x donné, les psychologues peuvent choisir des traits qui activent un opérateur w_i de manière conflictuelle. Par exemple, si x est à la fois +AMBITIOUS et -LAZY, alors selon $\mathtt{M}_\mathbf{C}$, il active les opérateurs M_{focus} et $M_{precision}$ avec les poids +2 et -2. Dans d'autres cas, l'écart est moindre (*e.g.*-1 et -2) et peut être résolu automatiquement (*e.g.* le plus fort l'emporte). La question de la cohérence, essentielle en psychologie, trouve dans notre modèle une représentation formelle et un moyen d'expérimenter et d'observer les effets des choix.

6.3. Discussion

6.3.1. Pertinence et complétude des opérateurs

La méthode de construction des opérateurs garantit que les opérateurs élicités sont pertinents : émanant des schémas, ils sont activés au moins une fois de manière non triviale (*i.e.* $\forall \omega_i \exists j$ tel que $\lambda_{i,j} \neq$ agent-générique (i)).

La méthode de construction des opérateurs ne garantit pas que tous les opérateurs possibles soient bien élicités ; en pure théorie, ce n'est pas possible et c'est bien là que TFS joue un rôle crucial concernant la question de la couverture : fondée sur des ressources lexicales larges et reconnues, TFS couvre, pour ce que la littérature en connaît, les comportements associés aux traits et permet donc de limiter le risque de silence.

6.3.2. Validation des poids et du modèle

Les poids $\lambda_{i,j} \in \mathtt{M}_\mathbf{C}$ sont proposés par des annotateurs et à ce titre, ils sont susceptibles d'être discutés entre experts informaticiens et psychologues. Notre approche a le mérite de faire ressortir ce problème fondamental, bien souvent "enfoui dans le code" des approches procédurales citées à la section 2.1. Dans notre cas, le recours à une méthode déclarative via une matrice de poids plutôt que des règles procédurales, améliore l'intelligibilité et le suivi de l'association traits/comportements. De plus, elle facilite grandement le dialogue avec les psychologues devant *in fine* valider les choix.

Dans cet article, nous proposons une approche pour traiter le phénomène d'influence des traits sur les actions, tel qu'il est décrit dans la littérature. Il ne s'agit donc pas d'évaluer directement un modèle particulier (composé d'un modèle rationnel, des

opérateurs et des poids spécifiques associés) par une expérimentation. Notre objectif ici est double : montrer la faisabilité (il existe des points d'influence dans le raisonnement rationnel) et se placer au niveau de la couverture (potentiellement, la matrice d'activation couvre tout FFM).

7. Exemple situé dans l'étude de cas

7.1. Définition d'un profil de personnalité dans la taxonomie TFS

En nous plaçant dans un modèle d'agent tel que défini dans l'étude de cas de la section 6, nous considérons un exemple issu de la littérature du domaine, à savoir l'exemple du CyberCafé (Rousseau, Hayes-Roth, 1996), où quatre agents endossent le même rôle social de serveur de café (ou « waiter » w_i). Ces serveurs ont des profils psychologiques distincts $P(w_i)$ qui entraînent des comportements $B(w_i)$ différents. En reprenant directement les définitions fournies par Rousseau et Hayes-Roth, nous avons par exemple :

$P(w_1)$	realistic, insecure, introverted, passive, secretive
$B(w_1)$	Such a waiter does and says as little as he can
$P(w_2)$	imaginative, dominant, extroverted, active, open
$B(w_2)$	This waiter takes initiative, comes to the customer without being asked for, talks much
$w_{3,4}$	*etc.*

Si nous prenons par exemple le profil psychologique $P(w_1)$ du serveur w_1, il peut être assez facilement transposé en termes de schémas TFS :

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
}

où les éléments de $P(w_1)$ sont transposés dans l'ordre, séparés par des ';' dans P'(w_1). On peut noter que ce profil active essentiellement des pôles négatifs et qu'il y a quasiment une bijection entre les traits de TFS. Cela revient à dire que les traits de $P(w_1)$ correspondent en fait plus à des schémas qu'à des traits de FFM (ou des facettes de FFM/NEO PI-R) avec une seule exception, *introverted* qui est associé à l'ensemble du trait -**E**xtraversion, activant ainsi ses 12 schémas (en pôle -) ce qui ajoute de la précision à la définition initiale de Rousseau et Hayes-Roth.

De manière similaire, le profil $P(w_2)$ du serveur w_2, peut être transposé :

P'(w_2) = {
imaginative ⇒ **O**+fantasy+CREATIVE;
dominant ⇒ **E**+assertiveness+DOMINEERING;
extroverted ⇒ **-E** = (ici les 12 schémas de **E** en pôle +) ;
active ⇒ **E**+activity-ACTIVE;
open ⇒ **+O** = (ici 11 schémas de **O** en pôle +) ;
}

Dans ce second exemple, la correspondance adjectif ↔ schéma est moins directe car deux adjectifs sont associés à des traits FFM. De plus, P'(w_2) active surtout des pôles positifs ; il est donc fort probable que P(w_1) et P(w_2) ont été construits à la main pour fournir des exemples de personnalités fortement contrastées, plutôt que d'être des cas pris dans le monde réel.

Ces deux exemples montrent que P'(w_1) et P'(w_2) offrent un positionnement plus systématique dans TFS ainsi qu'une définition comportementale plus précise car une définition comme B(w_1) est remplacée par les gloses associées aux schémas qu'elle active dans TFS. Par exemple, le seul schéma -PRACTICAL de P'(w_1) est défini par les gloses WordNet (N_i) ainsi que les Q-items de Goldberg (Q_i) suivants :

N618	guided by practical experience and observation rather than theory
N626	aware or expressing awareness of things as they really are
N788	freed from illusion
N1232	concerned with the world or worldly matters
N795	sensible and practical
Q6	Spend time reflecting on things
Q7	Seldom daydream
Q8	Do not have a good imagination
Q9	Seldom get lost in thought

de même pour -INSECURE, - COLD *etc.*

En résumé, TFS apporte une base solide pour le description de personnalité : non seulement elle couvre les huit classes qui sont présentes dans le CyberCafé, mais en plus elle offre une description des comportements effectifs beaucoup plus précise, ce qui justifie pleinement notre décision de l'utiliser comme support de l'élicitation des opérateurs d'influence.

7.2. Interactions rationnelles et psychologiques

Nous nous proposons maintenant de traiter un exemple de nature uniquement illustrative, puisque nous introduisons en particulier deux simplifications :

1. Il n'est pas traité exhaustivement car d'autres opérateurs d'influence que ceux présentés dans l'article interviennent aussi, et de plus l'implémentation des influences présentées ici n'est pas décrite techniquement. En effet, le traitement technique exhaustif, même pour un seul opérateur, sort du cadre imparti à l'article ; comme le montre le rapport technique consacré à l'implémentation effective des opérateurs de

la seule classe **P**reference de plus restreinte aux préférences atomiques [7] ;

2. Ci-dessous dans les interactions dialogiques, les phrases en langue naturelle n'ont pas été générées par le module de génération automatique d'un système de dialogue homme/machine. Il s'agit de transcriptions manuelles des structures de données produites par le système. Toutefois nous avons tenu à respecter les restrictions linguistiques introduites par un module de génération automatique. Les structures symboliques effectivement produites sont listées dans les colonnes de droite des tables de dialogues.

L'exemple développé illustre les comportements d'un agent serveur W endossant, dans la même situation d'interaction prise-de-commande, 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).

7.2.1. Description de la situation de l'heuristique : 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(cacahuètes) 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, pour lesquels l'agent serveur possède des faits dans sa base de connaissances K_W:

Plat	disponible	facile à servir	dangereux à servir	amusant à servir
pavé de boeuf	+	+	-	-
saucisse purée	-	+	-	-
soupe aux croutons	+	-	-	-
omelette flambée	+	-	+	+

7.2.2. Interaction entre agents rationnels

Étant donnée la situation décrite ci-dessus, voici une session d'interaction typique qui résulte de l'exécution de l'heuristique prise-de-commande par les agents C et W, considérés comme strictement rationnels :

7. consultable sur : http://perso.limsi.fr/jps/research/rnb/ranking/draft-ranking-V2.pdf.

Tour	Agent	Dialogue	Actions du serveur
1:	W	Quel plat désirez-vous ?	ouvrir()[a]
	C	Je voudrais une saucisse purée	consulter(disponibilité dans K_W)
2:	W	Il n'y en a plus	.
	C	Je voudrais un pavé de boeuf	noter(pavé de boeuf)
3:	W	Ensuite[b] Quel plat désirez-vous ?	.
	C	Aucun	clore()

[a] toutes les actions sauf ouvrir() 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 .

7.2.3. Instanciation avec P(W) = P'(w_1)

Pour cet exemple, nous nous restreignons aux opérateurs présentés dans le tableau 5. Ainsi, lorsque l'agent W endosse la personnalité P'(w_1) il active certains de ces opérateurs d'influence, en suivant le mode +/- défini dans la figure 2. Notons que cette figure ne porte que sur une partie des schémas associés à P'(w_1) ; par exemple les schémas liés à *introverted* peuvent eux aussi activer des opérateurs. Il s'ensuit que la session d'interaction rationnelle pour prise-de-commande est maintenant altérée par les opérateurs activés :

T	A	Dialogue	Actions	Opérateur(s) activé(s)
1:	W	Puis-je suggérer le pavé de boeuf ? [air dubitatif[a]]	ouvrir()	+P_{easy}, +P_{safe}, -$X_{success}$, -$X_{choiceConf}$
	C	Je voudrais une saucisse purée	consulter(K_W)	
2:	W	Désolé, il n'y en a plus [air triste]	.	+$A_{responsible}$, -$X_{success}$
	C	Je voudrais un pavé de boeuf	.[b]	-M_{focus}
3:	W	Ensuite, quel plat désirez-vous[c] ? [air dubitatif]	.	-$M_{tenacity}$,-$X_{success}$
	C	Aucun	clore()	

[a] [entre crochets] ajout d'expression physique (ici dû à -$X_{success}$, indiquant que W s'attend à échouer).
[b] W ne note pas la commande (dû à -M_{focus}).
[c] abandon de la proposition pavé de boeuf (dû à -$M_{tenacity}$).

On notera que le comportement du client n'a pas varié dans la mesure où nous avons conservé un caractère strictement rationnel pour l'agent C, qui ne réagit pas aux signaux interactionnels supplémentaires émis par W. Dans une situation d'interaction plus réaliste, les deux agents devraient être munis chacun d'une personnalité ce qui déboucherait sur une session plus complexe, ce que nous envisageons d'aborder dans des travaux ultérieurs.

7.2.4. Instanciation avec P(W) = P'(w_2)

Cette personnalité étant très différente, on obtient une interaction bien distincte :

T	A	Dialogue	Actions	Opérateur(s)
1:	W	Bonjour[a] [!!][b] Je vous conseille fortement l'omelette flambée ou la soupe aux croutons ; je suis sûr que vous allez aimer [!!]	ouvrir() ; **allumer**[b](bougie)[c]	+O_{play}, +$P_{ambitious}$, +$X_{success}$, +$X_{choiceConf}$, +O_{extra}, +M_{quick}
	C	Je voudrais une saucisse purée	consulter(K_W)	
2:	W	Il n'y en a plus !! Vous devriez suivre mes conseils ... [air sévère]	**poser**(cacahuètes)[c]	+M_{quick}, +$A_{responsible}$, +$A_{success}$
	C	Je voudrais un pavé de boeuf	**noter**(pavé-de-boeuf)	+M_{focus}
3:	W	Ensuite, je vous conseille fortement l'omelette flambée ou la soupe aux croutons[d] ; je suis sûr que vous allez aimer [!!]	.	+$M_{tenacity}$ plus ceux de 1:W
	C	Aucun		
	W	[Ah !] Je suis déçu [!!]	**éteindre**(bougie); clore()	+$A_{success}$, +M_{quick}

[a] expression liée à ouvrir().
[b] **en gras** = avec force/vivacité (ici dû à +M_{quick}).
[c] ajout d'action externe à l'heuristique (dû à +O_{extra}).
[c] répétition de 1:W (dû à +$M_{tenacity}$).

7.2.5. Analyse comparative des interactions

Nous pouvons faire trois remarques relatives aux différences existant entre les trois interactions décrites :

1. Par rapport à l'interaction strictement rationnelle, les deux instanciations avec une personnalité font apparaître : a) une forte variabilité de comportement ; b) une bien meilleure naturalité, au moins du côté de W.

2. Les deux instanciations avec une personnalité engendrent entre elles une différence de comportement notable pour W. Cet exemple n'étant qu'illustratif, nous n'avons pas effectué d'expérimentation en tant que telle, mais des personnes à qui nous avons présenté ces sessions ont trouvé l'agent lié à P'(w_1) timide, effacé, froid, ... et l'agent lié à P'(w_2) actif et intrusif (« il me dit quoi prendre », « il est pénible », ...) ou encore amical (« c'est sympa d'allumer la bougie »). En résumé, si l'exemple montre une différenciation comportementale claire, il reste à mener des expérimentations contrôlées pour vérifier si des sujets humains perçoivent et catégorisent correctement les traits de personnalité associés aux agents.

3. Les deux instanciations avec une personnalité font apparaître l'expression d'un nombre important d'opérateurs qui trouvent l'opportunité de s'exprimer au sein de l'heuristique prise de commande, et ceci selon diverses modalités : variation dans le

dialogue textuel ; expressions faciales et corporelles (si on dispose d'un agent conversationnel animé par exemple) ; actions sur le monde (modes ajout/suppression) et leur modalités d'exécution (force, vitesse, degré, *etc.*).

8. Conclusion

La notion de moteur de personnalité est fondée sur une approche qui lie de manière déclarative et flexible (via les poids de la matrice d'activation) les traits d'une taxonomie à un ensemble d'opérateurs d'influence exhibés par un modèle de raisonnement donné. Cette approche doit permettre d'observer et de discuter l'impact des choix effectués, en accord avec les psychologues.

Nous avons appliqué notre approche d'un côté au modèle BDI pour déterminer huit grandes classes d'opérateurs, et de l'autre à la taxonomie TFS qui enrichit FFM de schémas comportementaux assurant la précision de la couverture du domaine psychologique. Nous envisageons d'appliquer l'approche à d'autres instances de BDI, en particulier à des systèmes actuellement opérationnels comme par exemple 2APL (Dastani, 2008) ou faisant l'objet de développements formels comme par exemple IndiGolog (Giacomo *et al.*, 2009). Nous envisageons aussi d'étendre l'approche à d'autres parties de la taxonomie FFM (par exemple les traits interpersonnels, *cf.* l'étude amorcée dans les travaux (Sansonnet, Bouchet, 2013)). De plus, des expérimentations portant sur les questions liées à la perception effective et la catégorisation correcte par les usagers des traits de personnalité exprimés dans les comportements des agents devront être menées.

L'approche développée ici est avant tout théorique, mais sa motivation première est d'ordre pratique. En effet, dans des travaux menés précédemment (Xuetao *et al.*, 2009b), nous avons pu montrer que les usagers de systèmes incorporant des agents conversationnels, endossant des rôles aussi variés que assistant, compagnon, tuteur *etc.* préfèrent les agents dont le comportement dialogique est 1) plus varié que le comportement monotone d'un agent rationnel strict (par exemple, la même question reçoit exactement la même réponse) et 2) surtout dont les variations peuvent être perçues comme émanant de l'expression d'une personnalité propre à l'agent. Ceci débouche sur un large champ applicatif pour les agents de médiation de systèmes et de services ; pour les agents utilisés dans les systèmes d'apprentissage humain ; pour les agents qui "coachent" des humains dans des environnements de type ambiant *etc.* Pour tous ces domaines applicatifs, nous avons déjà développé, ou bien nous sommes en train de développer, des applications intégrant des agents munis de personnalité.

Enfin dans les travaux futurs, nous aimerions étendre l'association de profils de personnalité à tous les agents concernés par une même situation interactionnelle. C'est une question difficile qui nécessitera certainement la définition de modèles de communication inter-agents appropriés car l'ajout simple d'une couche « psychologique » au-dessus de modèles multi-agents classiques (par exemple ACL-FIPA) risque de ne pas suffire.

Bibliographie

Anderson J. R. (1996). ACT: a simple theory of complex cognition. *American Psychologist*, vol. 51, n° 4, p. 355-365.

Bates J. (1994). The role of emotion in believable agents. *Commun. ACM*, vol. 37, n° 7, p. 122–125.

Bevacqua E., Sevin E. de, Pelachaud C., McRorie M., Sneddon I. (2010). Building credible agents: Behaviour influenced by personality and emotional traits. In *Proc. of the int. conf. on kansei engineering and emotion research (keer'10).* Paris, France.

Bouchet F., Sansonnet J. P. (2010). Classification of wordnet personality adjectives in the NEO PI-R taxonomy. In *The fourth workshop on animated conversational agents, WACA 2010*, p. 83–90. Lille, France.

Bratman M. E. (1987). *Intentions, plans, and practical reason*. Cambridge, MA, Harvard University Press.

Casali A., Godo L., Sierra C. (2004). Graded BDI models for agent architectures. In *Proceedings of CLIMA-V*, vol. 3487, p. 126–143. Lisbon, Portugal.

Cattell R. B., Eber H. W., Tatsuoka M. M. (1970). *Handbook for the sixteen personality factor questionnaire (16 PF)*. Champain, Illinois.

Costa P. T., McCrae R. R. (1992). *The neo pi-r professional manual*. Odessa, FL: Psychological Assessment Resources.

Damasio A. R. (1994). Descartes error: Emotion, reason and the human brain. New York: G.P., Putnam's Sons.

Dastani M. (2008). 2APL: a practical agent programming language. In *AAMAS '08:the seventh international joint conference on autonomous agents and multiagent systems*, vol. 16, p. 214-248. Estoril, Portugal, Springer-Verlag.

Doce T., Dias J., Prada R., Paiva A. (2010). Creating individual agents through personality traits. In *Intelligent virtual agents (IVA 2010)*, vol. 6356, p. 257-264. Philadelphia, PA, Springer-Verlag.

Evertsz R., Ritter F. E., Busetta P., Pedrotti M. (2008). Realistic behaviour variation in a BDI-based cognitive architecture. In *Proc. of SimTecT'08.* Melbourne, Australia.

Fellbaum C. (1998). *WordNet: an electronic lexical database*. MIT Press.

Giacomo G., Lespŕance Y., Levesque H. J., Sardina S. (2009). IndiGolog: A high-level programming language for embedded reasoning agents. In A. El Fallah Seghrouchni, J. Dix, M. Dastani, R. H. Bordini (Eds.), *Multi-agent programming: Languages, platforms and applications*, p. 31-72. Springer.

Goldberg L. R. (1981). Language and individual differences: The search for universal in personality lexicons. *Review of personality and social psychology*, vol. 2, p. 141–165.

Goldberg L. R. (1990). An alternative description of personality: The big-five factor structure. *Journal of Personality and Social Psychology*, vol. 59, p. 1216–1229.

Gratch J., Marsella S. (2004). A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research*, vol. 5, n° 4, p. 269-306.

Hayes-Roth B., Brownston L., van Gent R. (1995). Multiagent collaboration in directed improvisation. In *Proc. of the first int. conf. on multi-agent systems ICMAS 95*, p. 148-154. San Francisco, CA.

John O. P., Robins R. W., Pervin L. A. (Eds.). (2008). *Handbook of personality: Theory and research* (3rd éd.). The Guilford Press.

Laird J. E., Newell A., Rosenbloom P. S. (1987). Soar: an architecture for general intelligence. *Artif. Intell.*, vol. 33, n° 1, p. 1–64.

Lukasiewicz J. (1963). *Elements of mathematical logic*. Pergamon Press.

Maes P. (1994). Agents that reduce work and information overload. *Commun. ACM*, vol. 37, n° 7, p. 30–40.

Mascardi V., Demergasso D., Ancona D. (2005). Languages for programming BDI-style agents: an overview. In F. Corradini, F. De Paoli, E. Merelli, A. Omicini (Eds.), *Proceedings of WOA 2005*, p. 9–15. Camerino, MC, Italy, Pitagora Editrice Bologna.

McRorie M., Sneddon I., de Sevin E., Bevacqua E., Pelachaud C. (2009). A model of personality and emotional traits. In *Intelligent virtual agents (IVA 2009)*, vol. 5773, p. 27-33. Amsterdam, NL, Springer-Verlag.

Miletto E. M., Sansonnet J. P., Pimenta M. S., Bouchet F. (2009). Corpus-based design of a web 2.0 assisting agent. In L. Olsina, O. Pastor, D. Schwabe, G. Rossi, M. Winckler (Eds.), *Proc. of the 8th international workshop on Web-Oriented software technologies (IWWOST'2009)*, vol. 493, p. 52–63. San Sebastian, Spain, CEUR Workshop Proceedings.

Norling E., Ritter F. E. (2004). Towards supporting psychologically plausible variability in Agent-Based human modelling. In *Proceedings of the third international joint conference on autonomous agents and Multi-Agent systems.*

Ortony A., Clore G. L., Collins A. (1988). *The cognitive structure of emotions* (Cambridge University Press éd.). Cambridge, UK.

Pelachaud C. (2000). Some considerations about embodied agents. In *Proc. of achieving human-like behavior in interactive animated agents, the 4th international conference on autonomous agents.* Barcelona, Spain.

Pereira D., Oliveira E., Moreira N. (2008). Formal modelling of emotions in BDI agents. In F. Sadri, K. Satoh (Eds.), *Proceedings of CLIMA-VIII*, vol. 5056, p. 62–81. Porto, Portugal, Springer-Verlag.

Picard R. W. (2000). *Affective computing*. MIT Press.

Rao A. S., Georgeff M. P. (1995). BDI agents: From theory to practice. In *Proc. first int. conference on multi-agent systems (ICMAS-95)*, p. 312–319.

Rousseau D. (1996). Personality in computer characters. In *Aaai technical report ws-96-03*, p. 38-43.

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

Russell S., Norvig P. (2009). *Artificial intelligence: A modern approach* (3rd éd.). Prentice Hall Press.

Sansonnet J. P. (2008). *DIVA - DOM integrated virtual agent.* http://www.limsi.fr/~jps/online/diva/divahome/index.html.

Sansonnet J. P., Bouchet F. (2010). Joint handling of rational and behavioral reactions in assistant conversational agents. In H. Coehlo, R. Studer, M. Wooldridge (Eds.), *Proc. of the 19th european conference on artificial intelligence (ECAI 2010)*, p. 1049–1050. Lisbon, Portugal, IOS Press.

Sansonnet J.-P., Bouchet F. (2013, février). Managing personality influences in dialogical agents. In J. Filipe, A. Fred (Eds.), *Proceedings of the 5th international conference on agents and artificial intelligence (ICAART 2013)*, vol. 1, p. 89–98. Barcelona, Spain, SciTePress.

Saucier G., Ostendorf F. (1999). Hierarchical subcomponents of the big five personality factors: A cross-language replication. *Journal of Personality and Social Psychology*, vol. 76, n° 4, p. 613–627.

Soto C. J., John O. P. (2009). 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*, vol. 43, n° 1, p. 84–90.

Wooldridge M. (2000). *Reasoning about rational agents*. MIT Press.

Wooldridge M., Jennings N. R. (1995). Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, vol. 10, n° 2, p. 115-152.

Xuetao M., Bouchet F., Sansonnet J. P. (2009a). An ACA-based semantic space for processing domain knowledge in the assistance context. In F. Ko (Ed.), *Proc. of the 3rd international conference on new trends in information and service science (NISS 2009)*, p. 344–349. Beijing, China, CPS Publication.

Xuetao M., Bouchet F., Sansonnet J. P. (2009b). Impact of agent's answers variability on its believability and human-likeness and consequent chatbot improvements. In *Proc. of AISB 2009*, p. 31-36. Edinburgh, Scotland.