
Traits de Personnalité Computationnels

Enrichissement de la taxonomie FFM/NEO PI-R avec des gloses WordNet liées à des adjectifs de personnalité

François Bouchet¹ Jean-Paul Sansonnet²

1. SMART Lab, McGill University 3700 McTavish, Montréal, QC H3A 1Y2, Canada
francois.bouchet@mcgill.ca

2. Laboratoire LIMSI-CNRS BP 133 F-91403 Orsay Cedex, France
jps@limsi.fr

RÉSUMÉ. Ce travail décrit le processus de classification d'un ensemble d'adjectifs décrivant des traits de personnalité dans la liste des facettes de la taxonomie NEO PI-R, en relation avec le Modèle des Cinq Facteurs (FFM– Five Factor Model). Le processus de classification est fondé sur des mots-la sémantique lexicale exprimée par les couples 'synset-gloses' attachés aux mots-adjectifs dans la base lexicale WordNet. Pour traiter par ordinateur l'arrangement des gloses dans les positions de l'arborescence définies par la taxonomie FFM/NEO PI-R, une phase de réarrangement en termes de 'schémas comportementaux' a été effectuée. Finalement, la classification est synthétisée sous forme d'une ressource XML, accessible librement sur le Web qui produit un outil logiciel manifestant une bonne couverture des traits de personnalité, en vue de leur intégration dans des agents conversationnels.

ABSTRACT. We classify a set of personality-trait adjectives within the facet list of the NEO PI-R taxonomy related to the Five Factor Model. This process is based on the lexical semantics expressed by the synset-gloss attached to the adjectives in the WordNet lexical base. In order to make the arrangement of the glosses within the positions of FFM/NEO PI-R taxonomy computationally treatable, a phase of rearrangement in terms of so-called behavioral schemes is performed. This classification is synthesized as an XML resource, freely accessible on the Web, which provides a computer based support with good coverage for the study and the computational implementation of psychological behaviors in conversational agents.

MOTS-CLÉS : Adjectifs de traits de personnalité, Taxonomies FFM et NEO PI-R, gloses WordNet.

KEYWORDS: Personality traits adjectives, FFM and NEO PI-R taxonomies, WordNet glosses.

DOI:10.3166/HSP.x.1-31 © 2011 Lavoisier, Paris

1. Introduction

1.1. Contexte

Ce travail se situe dans le cadre des études actuelles sur la modélisation cognitive pour des Agents Conversationnels Animés (ACA). La modélisation cognitive d'agents s'appuie sur les travaux en intelligence artificielle autour de la notion d'agent rationnel, où l'accent est mis sur le raisonnement concernant les actions et plans des agents artificiels. Dans le contexte particulier des agents conversationnels, les actions et les plans des agents sont essentiellement tournés vers la gestion de l'interaction multimodale avec des usagers humains. Alors que les agents purement rationnels visent à l'efficacité dans leur comportement, sans forcément émuler les modes de pensée humaine, le fait que les agents conversationnels aient affaire à des humains leur impose d'être « humainement crédibles ». Dès lors, les notions psychologiques, peu prises en compte dans la communauté en intelligence artificielle sur les agents rationnels, se trouvent portées au premier plan des préoccupations scientifiques.

Bien que nous les ayons présentées ci-dessus comme des notions séparées, les capacités de raisonnement rationnel et psychologique d'un agent fonctionnent en fait de manière très entremêlée (Ellsworth, Scherer, 2003 ; Frijda, 2006). Par exemple, certains facteurs associés à des notions psychologiques de haut niveau, comme les traits de personnalité (John, Robins, Pervin, 2008) (généralement stables et durables), pourraient être intégrés dans des architectures d'agents rationnels afin de produire des agents conversationnels à la fois rationnellement efficaces et humainement crédibles.

C'est dans cet objectif que nous avons proposé un cadre dédié à l'étude informatique de la nature des relations qu'entretiennent les fonctionnalités rationnelles et psychologiques d'un agent (Sansonnnet, Bouchet, 2010a) : l'architecture *R&B* (pour *Rational and Behavioral agents*¹, où *behavioral* prend le sens particulier de « comportement psychologique »). *R&B* est un cadre générique permettant la définition computationnelle et l'expérimentation de diverses hypothèses portant sur les relations entre raisonnement rationnel et raisonnement psychologique. Les travaux présentés dans cet article se placent au cœur des études qui visent à outiller des architectures d'ACA.

1.2. Objectifs

Au cours de travaux récents sur *R&B*, nous avons été amenés à proposer un modèle où les phénomènes psychologiques, comme les traits de personnalité, sont implémentés en termes d'opérateurs d'influence censés contrôler, voire altérer, le processus de raisonnement rationnel d'un agent. Pour le moment, le modèle ainsi proposé n'a été implémenté qu'avec un ensemble restreint de phénomènes psychologiques que nous avons choisis arbitrairement (par exemple, un agent est *lazy*, *gloomy* ou encore *cheerful*...) (Bouchet, Sansonnnet, 2009 ; Sansonnnet, Bouchet, 2011). Cette approche est suf-

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

fisante pour illustrer les principes et la faisabilité du cadre *R&B*, mais elle ne couvre pas de manière significative le domaine décrivant la psychologie d'une personne, en particulier en termes de traits de psychologie. Une illustration de ce problème est proposée dans le tableau 1 et discutée à la section 2.2.

Pour ces raisons, il existe un réel besoin de définir le domaine des phénomènes psychologiques traités d'une manière qui soit non seulement plus systématique, mais qui permette également de les rendre exploitables en termes informatiques, c'est-à-dire pouvant déboucher sur des implémentations effectives.

Pour ce faire, les informaticiens peuvent partir des travaux recensés dans la littérature en psychologie, qui offrent plusieurs taxonomies de la personnalité humaine (John, Robins, Pervin, 2008 ; Mischel *et al.*, 2008), mais il faut aussi compléter et adapter ces données pour pouvoir les intégrer dans des architectures informatiques telles que *R&B*. En effet, lors de la recherche de ces données, un double problème apparaît rapidement (*cf.* section 2.5 pour plus de détails) : a) le manque de disponibilité des données d'origine ; b) la trop forte généralité des taxonomies proposées. Cette généralité découle de la volonté des chercheurs d'enraciner les concepts au moyen d'outils statistiques, comme l'analyse factorielle largement utilisée pour montrer l'orthogonalité des catégories de traits proposées. Cet inconvénient a d'ailleurs été mentionné par les psychologues eux-mêmes, amenant certains à tenter de raffiner les taxonomies en introduisant dans les traits des sous-niveaux appelés des *facettes*. Malheureusement, même ainsi, les taxonomies obtenues demeurent encore trop générales pour une prise en compte au niveau informatique (on peut par exemple s'en rendre compte en lisant les gloses² associées aux facettes données dans le tableau 2).

Dans ces conditions, notre proposition consiste alors à :

1. raffiner le niveau des facettes afin de proposer une taxonomie à trois niveaux pour décrire la personnalité d'une personne. Plus exactement, ce troisième niveau ne vise pas tant à constituer une extension, au sens des études en psychologie, qu'un enrichissement à base d'exemples, qui se contente de fournir des instances pour le second niveau. Il est composé de *schémas psychologiques* (en abrégé *schémas*) qui regroupent des instanciations de facettes en notions décrivant des comportements opérationnels, donc plus directement tournées vers une implémentation logicielle.

2. éliciter les schémas psychologiques à partir d'un corpus d'adjectifs décrivant des traits de personnalité, qui sont à leur tour associés à des sens lexicaux et des gloses dans la base lexicale informatisée WordNet (Fellbaum, 1998). Si les chercheurs en psychologie ont depuis longtemps fait usage des ressources linguistiques (dictionnaires et thésaurus) pour construire des taxonomies (voir la section 2.4), WordNet offre trois avantages : a) c'est un outil logiciel ; b) il fait office de standard de fait en Traitement Automatique de la Langue (TAL) ; c) surtout, il propose un ensemble systématisé de sens lexicaux (appelés *synsets*) qui définissent une base solide pour la

2. Une glose est une petite phrase qui définit informellement (et non du point de vue informatique) la sémantique d'un symbole, comme par exemple les définitions de mots des dictionnaires – en anglais, les termes *gloss* et *short phrase* sont synonymes.

définition des schémas. Le processus d'élicitation des schémas a été conduit en deux phases principales :

- phase 1 : adjectifs de personnalité : sélection d'un sous-ensemble significatif d'adjectifs de personnalité, puis classification dans une taxonomie standard à deux niveaux ;
- phase 2 : pour chaque facette, réarrangement (clusterisation) des sens lexicaux concernés (synsets) en schémas comportementaux.

Plan de l'article. Dans la section 2, nous présentons les travaux relatifs à l'implémentation informatique de traits de personnalité dans le cadre d'agents autonomes et aussi d'agents conversationnels. Nous mettons par ailleurs en avant les approches fondées sur la linguistique, en particulier le domaine de la sémantique lexicale. Dans la section 3, nous exposons le processus de classification d'un ensemble de gloses WordNet associées à des adjectifs de personnalité ; la classification se fait dans le cadre de la taxonomie FFM/NEO PI-R (voir tableau 2). Enfin, dans la section 4, nous décrivons le résultat du processus d'élicitation des schémas comportementaux à partir de la ressource construite lors de l'étape précédente.

2. Contexte de la recherche

2.1. Définitions : agents rationnels, conversationnels et psychologiques

Agents rationnels. Dans le contexte de la recherche en Intelligence Artificielle (IA), plusieurs architectures cognitives d'agents sont fondées sur la notion de raisonnement pratique (« *practical reasoning* ») à la suite des travaux sur les modèles à base de règles comme SOAR (Laird *et al.*, 1987) et ACT-R (Anderson, 1996) ou encore, plus récemment, les architectures d'agents de type BDI (*Belief – Desire – Intention*) (Bratman, 1992 ; Rao, Georgeff, 1995). Dans la théorie de Bratman de la raison pratique (Bratman *et al.*, 1988), le comportement d'un agent est modélisé par la spécification de *croyances*, de *buts* et de *plans*. Ces architectures se sont montrées efficaces dans l'implémentation logicielle de systèmes multi-agents (SMA) capables d'agir dans des contextes opérationnels, comme par exemple la plateforme multi-agents JACK (Howden *et al.*, 2001).

Agents conversationnels. Cependant, les agents logiciels sont de plus en plus en interaction avec des usagers humains, que ce soit en tant qu'agents conversationnels autonomes (Maes, 1994), (Cassell *et al.*, 1999) ou au sein de communautés mixtes (Traum *et al.*, 2003 ; Goodrich, Schultz, 2007). Dans ce nouveau contexte, il a été soutenu que les agents devaient être à la fois *compétents* (grâce à leurs capacités de raisonnement symbolique) et psychologiquement *pertinents*, de manière à augmenter a) leur facteur d'acceptabilité (Davis, 1989), surtout quand ils ont affaire à des personnes du grand public, et b) le facteur de performance du processus interactionnel du

couple humain/agent lui-même (par exemple, en compréhension ou en mémorisation dans des activités d'apprentissage (Baylor, Kim, 2004)).

Agents psychologiques. En fait, depuis que la notion d'agent « crédible » (*believable agent*) a été introduite au milieu des années 1990 (Bates, 1994 ; Maes, 1995 ; Hayes-Roth *et al.*, 1995), plusieurs tentatives d'implémentation de phénomènes psychologiques dans des architectures cognitives d'agents ont été entreprises. Par exemple, les travaux de Rousseau et Hayes-Roth (Rousseau, 1996 ; Rousseau, Hayes-Roth, 1996) ont établi une première base en donnant des exemples montrant comment les facteurs liés à la personnalité (aussi appelés traits de personnalité) peuvent être implémentés dans une architecture cognitive d'agent rationnel. Tous ces auteurs mettent en avant le fait que les agents devraient posséder à la fois un moteur de raisonnement rationnel et un moteur de raisonnement psychologique.

2.2. Intégration de traits de personnalité dans des agents rationnels

Comme mentionné ci-dessus, l'approche dominante actuellement, en intelligence artificielle et en systèmes multi-agents est celle des agents BDI de Bratman (1990) puis de Rao et Georgeff (1995), qui met aussi l'accent sur le raisonnement d'agents situés, à propos de leurs *actions* et de leurs *plans* (Allen *et al.*, 1991) et qui ont produit des architectures dédiées (Sansonnnet *et al.*, 1980). Les plateformes BDI opérationnelles sont nombreuses et utilisent divers supports formels comme la logique (KGP, MINERVA, AGENTSPEAK, CAN ; avec des règles : 2APL ; ou temporelle : MetateM, la famille Golog) ou simplement la programmation, par exemple en Java (JACK, JASON, JADEX, JADE). Toutes ces architectures sont fondées sur ce qu'il est convenu d'appeler le *cycle de délibération* de l'agent qui se décompose traditionnellement en cinq étapes :

1. Délibération : étant donné un agent muni d'un ensemble de « désirs » éventuellement non atteignables et/ou contradictoires, pendant cette étape, l'agent choisit parmi eux un ensemble de buts qu'il a pour intention de mettre en œuvre et qui doivent être atteignables et non contradictoires.
2. Planification : étant donné le but courant, l'agent génère un plan rationnel, c'est-à-dire une séquence d'actions sur l'état du monde qui doit amener dans une situation qui contient les conditions de ce but.
3. Décision : bien que le planificateur (étape précédente) fournisse un plan pour un but atteignable, ce plan peut contenir de nombreuses alternatives (actions ou sous-plans différents menant aux mêmes conditions) ou options (actions ou sous-plans ne modifiant pas les conditions visées). Ainsi les plans peuvent être sous-déterminés, ce qui laisse la possibilité d'introduire une phase décisionnelle sur les alternatives et options.
4. Exécution : l'agent exécute l'action courante. Dans la plupart des plateformes BDI opérationnelles, cette phase est implicite, c'est-à-dire considérée comme spécifique à l'implémentation. Dans l'architecture *R&B* elle est pleinement prise en compte.

5. Évaluation : l'agent reçoit du monde des informations relatives au résultat effectif de l'exécution de ses actions. Il a alors la possibilité d'évaluer ses performances et de réagir.

Ces étapes sont détaillées par exemple dans (Rao, Georgeff, 1995) ou (Sardina, 2007). Le cycle de délibération BDI joue un rôle central dans l'architecture *R&B* car les opérateurs d'influence s'appliquent aux cinq étapes de ce cycle, selon huit modes définis dans (Bouchet, Sansonnet, 2011).

Le passage de ces agents rationnels à des agents psychologiques s'est essentiellement fait par une intégration de notions psychologiques à des architectures d'agents rationnels déjà existantes. Les principales contributions dans ce domaine, présentées dans les deux paragraphes suivants, font ressortir deux raisons principales pour agir de la sorte : a) il s'agit de définir des modèles cognitifs plus complets intégrant des notions psychologiques ; b) il s'agit de définir des agents capables de soutenir des interactions avec les utilisateurs du grand public qui soient plus humaines (au sens de « *human-like* »), qui par exemple entrent en conversation avec des agents assistants sur des pages Internet (Sabouret, Sansonnet, 2001 ; 2003).

Gratch et Marsella (Gratch, Marsella, 2004) ont proposé un modèle d'émotions fondé sur une architecture traditionnelle de type SOAR qui n'a pas été sans modifications profondes au cœur même de l'architecture noyau, comme noté dans (Gratch, Marsella, 2004). La plupart des autres travaux se fondent essentiellement sur le modèle BDI et proposent des extensions pour le support de certains phénomènes psychologiques (Dias, Paiva, 2005 ; Lim *et al.*, 2008). Dans ces architectures, la psychologie des agents influence à la fois l'expression des émotions, via les modalités corporelles, mais a aussi un impact sur le processus de décision rationnelle des agents. Par exemple, le modèle eBDI de Jiang *et al.* (2007) implémente des émotions dans un cadre BDI. Ils donnent dans l'exposé de ce travail, une bonne introduction sur leur cheminement intellectuel et sur la nécessité d'implémenter des émotions dans un agent rationnel.

À partir de la plateforme multi-agents bien connue JACK (Howden *et al.*, 2001), qui implémente elle aussi la théorie BDI, Norling et Evertsz ont construit l'extension CoJACK (Norling, Ritter, 2004 ; Evertsz *et al.*, 2008) qui est une couche ayant pour objectif de simuler la physiologie humaine lorsqu'elle s'exprime sous forme de manques ou de contraintes, comme par exemple la durée qu'il faut pour prendre une décision (dans le cycle BDI traditionnel, la complexité d'un raisonnement n'est pas prise en compte comme un élément du raisonnement lui-même, ce que font par contre naturellement les humains), ou bien raisonner avec des limites en termes de mémoire (*e.g.* « perdre une croyance » si le niveau d'activation physiologique est bas ou « oublier l'étape suivante » dans une procédure), ou encore la remémoration des croyances fondée sur la logique floue, le focus attentionnel limité et l'utilisation de modérateurs altérant la cognition. Ici, comme pour la psychologie, il y a l'idée que des facteurs extérieurs contrôlent (et peuvent même altérer) le processus classique de raisonnement rationnel.

Tableau 1. Heuristique implémentant une réaction comportementale simple de l'agent lorsque l'utilisateur lui fait une remarque négative.

Heuristique : Traitement de remarques négatives sur l'agent.			
1 :	H_b : dislike-agent[DISLIKE _{user} [agent]]:	$\rightarrow$	{
2 :		$\rightarrow$	{ $Q[i, \text{MAP}[\text{energy}, \lambda x.x * 0.9]]$,
3 :			$Q[j, \text{MAP}[\text{confidence}, \lambda x.x * 0.9]]$,
4 :			$Q[k, \text{MAP}[\text{cooperation}, \lambda x.x * 0.9]]$,
5 :			}
6 :	$Q_i^+ \wedge Q_j.\text{value} < -0.5$	$\rightarrow$	ADD [$\mathcal{R}.\text{reply}$, TELL _a [energy, "tired"]]
7 :	$Q_j^+ \wedge Q_j.\text{value} < -0.5$	$\rightarrow$	ADD [$\mathcal{R}.\text{reply}$, TELL _a [confidence, "depressed"]]
8 :			}

Explications :

Cette heuristique est composée de trois règles gardées en 2 : , 6 : et 7 :

2 : l'agent exécute une séquence de trois requêtes internes qui modifient son état mental. 'energy' représente le niveau d'une variable dans le modèle mental de l'agent. $\lambda x.x * 0.9$ est une λ -expression décrémentant son argument x de 10%.

6 : si son énergie est très basse, on ajoute "Je me sens fatigué" à la liste des choses à dire.

7 : idem pour la variable 'confidence' ; le scheduling des règles assure qu'on puisse avoir les deux en même temps : "Je me sens fatigué et déprimé".

L'architecture *R&B* qui définit le cadre des travaux décrits dans cet article, partage avec les études listées dans les deux paragraphes précédents les deux motivations de modélisation cognitive et de *human-likeness*. Cependant les études listées portent sur des cas particuliers de phénomènes psychologiques, chaque phénomène donnant lieu à une étude de cas individuelle. Nous avons nous-mêmes développé des travaux similaires, présentant des études de cas (Bouchet, Sansonnet, 2009 ; Sansonnet, Bouchet, 2010b) qui ont pu être critiquées pour leur caractère *ad hoc*. Afin d'illustrer ce propos, nous donnons un exemple d'heuristique tiré et adapté de (Sansonnet, Bouchet, 2010b) qui traite un cas particulier de phénomène psychologique au moyen de règles qui influencent le processus de raisonnement rationnel d'un agent. Supposons que l'utilisateur envoie la remarque textuelle suivante « Je te déteste ! » à l'agent. Le module de traitement de la langue naturelle engendre alors la production d'une requête formelle ayant pour format DISLIKE_u[agent], pour laquelle nous avons implémenté une réaction possible qui est décrite dans le tableau 1. Même sans le détail du système formel support et des notations décrits dans (Sansonnet, Bouchet, 2010b), il est possible de faire les deux remarques suivantes :

- les heuristiques sont définies de manière opérationnelle. Grossièrement parlant, on peut dire qu'elles sont « codées en dur », quand bien même le langage est déclaratif ;
- la question se pose de la gestion de l'ensemble des heuristiques : dans leur idiosyncrasie (qui les construit ?), dans leur étendue et leur couverture (quel domaine psychologique est couvert ou non couvert ?), dans leur ordonnancement (*i.e.* lorsque plusieurs heuristiques sont actionnables au même moment), *etc.*

Ces critiques sont directement transposables aux travaux listés plus haut.

Pour ces raisons, nous avons été amenés à proposer une architecture, dénotée *R&B*, qui a pour ambition de proposer un cadre plus large, idéalement générique, permettant de rendre compte de phénomènes psychologiques tels que ceux abordés par les études listées ci-dessus. L'architecture *R&B* tente de prendre en compte la question de la définition opérationnelle ainsi que la gestion de l'ensemble des heuristiques. Dans cet article, nous nous intéressons plus particulièrement à la couverture du domaine psychologique. Pour disposer d'une assise solide, préalable à l'étude de la question de la couverture, il est nécessaire de s'appuyer sur un système taxonomique pour la représentation de la personnalité humaine. Cette question est abordée dans la section suivante.

2.3. *Intégration de traits de personnalité dans des agents conversationnels*

Dans cette section, nous listons un certain nombre de travaux de chercheurs du domaine des agents conversationnels qui ont abordé, d'une manière ou d'une autre, la question de la personnalité de l'agent afin de l'intégrer, souvent partiellement, dans le modèle comportemental de l'agent. Chacune de ces études est contrastée avec l'approche que nous développons dans le cadre *R&B*.

Allbeck et Kress-Gazit (2010) viennent de proposer un cadre d'étude où des instructions élaborées peuvent être données par l'utilisateur, en particulier via la langue naturelle, pour contrôler un agent qui opère dans un environnement de simulation complexe. Dans ce cas, l'agent simule un robot, c'est-à-dire qu'il s'agit plutôt d'un agent autonome que d'un agent conversationnel au sens strict. Les instructions de l'utilisateur contraignent l'exécution des plans par l'agent. Ce travail met l'accent sur la dynamique des actions dans les plans des robots, dynamique qui facilite la mise en œuvre d'une implémentation de fonctions de contrôle et d'altération externes pour l'exécution des plans. Les contraintes qui sont alors appliquées sur les actions et plans sont d'une part, produites par une entité externe (ici l'utilisateur humain), et d'autre part, différentes des opérateurs d'influence que nous utilisons dans le cadre *R&B* (Sansonnnet, Bouchet, 2010a), au niveau de la nature même des influences exercées, bien qu'il y ait des similarités sur l'impact final sur les actions et les plans (*e.g.* empêcher une action d'être exécutée, changer l'ordre d'exécution des actions, *etc.*).

Les travaux de Rousseau et Hayes-Roth, mentionnés à la section 2.1, posent comme principe que les traits de personnalité s'expriment en fait comme des influences sur les actions et les plans. Une autre de leurs contributions majeures est l'idée que les traits de personnalité peuvent être associés à des facteurs intensifs (*i.e.* des poids) qui sont pris en compte dans le moteur de planification. Dans nos travaux cités ci-dessus, nous généralisons cette approche en systématisant la notion d'opérateur d'influence, dans un cadre plus générique (*R&B*). Mais ceci sort du cadre de cet article qui vise essentiellement à outiller ce processus.

Rizzo *et al.* (1997) ont prouvé que le caractère d'une personne peut être décrit en termes de ses buts et de ses plans, ceci avec une certaine précision. Pour ce faire, ils attribuent des comportements opératoires liés au personnage dans la façon dont il

poursuit ses buts. Les traits de personnalité sont utilisés pour choisir entre des buts multiples, lors de la phase de *désirs* $\Rightarrow$ *buts* (Sardina, 2007) du cycle BDI de l'agent (*i.e.* les traits influencent les désirs). Cependant une fois choisis, les buts sont planifiés et exécutés directement sans aucun contrôle supplémentaire, alors que dans le cadre *R&B*, les traits opèrent aussi sur l'exécution des plans (*i.e.* les traits influencent également les intentions). L'architecture *R&B* couvre les cinq étapes du cycle BDI alors que (Rizzo *et al.*, 1997) ne traitent que l'étape de délibération ; le spectre des influences possibles et donc beaucoup plus large et l'expression de la personnalité peut donc se faire sentir aussi le long des quatre autres étapes.

Malatesta *et al.* (2007) mettent en avant l'idée que les traits de personnalité peuvent être à l'origine de différentiation dans les comportements des agents. Dans leur étude, ils utilisent les traits pour créer différentes expressions dans les comportements psychologiques ("la psychologie influence la psychologie"), en particulier au niveau du module d'*appraisal* dans le modèle OCC de Ortony *et al.* (1988, 2003). Cependant la différence essentielle entre leur travail et le nôtre porte sur la *direction* des influences : 1) elle va du monde vers l'esprit (« *world to mind* ») pour (Malatesta *et al.*, 2007) *i.e.* la manière dont les agents évaluent à la fois les résultats de leurs actions et les événements externes ; 2) de l'esprit vers le monde pour l'architecture *R&B*. Les deux approches sont donc complémentaires.

L'architecture PMFserv (Silverman *et al.*, 2006) est dédiée à la création d'agents crédibles, au sens culturel du terme. Ceci est permis grâce à la mise en œuvre d'entités appelées « *performance moderators functions* » (PMFs). Les auteurs affirment que « la principale fonctionnalité de PMFserv est un modèle de raisonnement décisionnel fondé sur une utilité émotionnelle subjective contrainte par le stress et la physiologie ». Ce travail est partiellement fondé sur les travaux de Gillis et Hursh (1999), qui ont introduit la notion de « modérateurs de comportement ». La principale différence avec l'approche *R&B* est que, un peu sur le même principe que CoJACK, PMFserv met l'accent sur une différentiation comportementale fondée non tant sur les caractéristiques psychologiques des agents que sur leurs capacités *physiologiques* ; en effet, ces travaux font partie d'un projet dont l'objectif est de simuler les comportements de groupes de forces militaires.

Des travaux tels que (McRorie *et al.*, 2009) (Bevacqua *et al.*, 2010), fondés sur l'architecture d'agent conversationnel GRETA (Pelachaud, 2000), font également intervenir des modèles de personnalité pour l'expression des émotions (faciales, gestuelles, *etc.*). L'architecture FATIMA (Doce *et al.*, 2010) s'appuie sur ces travaux. Doce *et al.* affirment en substance que « dans l'ère de la globalisation, des notions comme l'individualité et la personnification (notre traduction) deviennent de plus en plus importantes. Les taxonomies comme FFM peuvent alors servir de base à des créations de personnages aux caractéristiques bien distinctes en utilisant les traits de personnalité pour influencer automatiquement les processus d'*appraisal* de *coping* ainsi que l'expression corporelle ». Il s'agit de travaux complémentaires à l'approche *R&B* dans la mesure où les influences de FATIMA portent essentiellement sur l'expression fine des émotions alors que nous traitons du processus décisionnel rationnel (les actions) ;

cependant le principe d'influence sous-jacent, issu de Rousseau et Hayes-Roth, est commun aux deux approches.

Tableau 2. Facettes NEO PI-R définies pour le domaine de traits FFM.

FFM facettes	Glose définissant les 30 facettes NEO PI-R
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 []
	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
Agreeability	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 [... shortened]
	Vulnerability general susceptibility to stress

2.4. Contribution des ressources linguistiques aux travaux sur la personnalité

2.4.1. Le modèle FFM

Historique. Le modèle des cinq facteurs (appelé ici FFM pour *Five Factor Model*) est aussi dénoté dans la littérature au moyen de l'acronyme OCEAN qui reprend la

première lettre de chaque classe (elles sont listées en colonne 1 du tableau 2). Lorsqu'on s'intéresse aux phénomènes psychologiques et plus particulièrement ceux qui concernent les traits de personnalité, le paradigme actuellement le plus saillant est FFM qui est issu d'une recherche convergente de plusieurs chercheurs en psychologie au cours des vingt dernières années. Ce paradigme a pris le pas sur des modèles qui ont eu du succès précédemment, comme la classification de Cattell en 16 facteurs (Cattell *et al.*, 1970), qui était le paradigme dominant dans les années 1980.

Les deux méthodologies. Les études des psychologues sur l'élicitation et la classification des traits de personnalité ont été sous-tendues par deux méthodologies complémentaires :

- l'approche à base de questionnaires : généralement composés de questions rapides (de type oui/non), comme ceux de Eysenck (EPQ – *Eysenck's Personality Questionnaire*), destinés à établir les traits de personnalité d'une personne (H. J. Eysenck, Eysenck, 1975 ; S. B. G. Eysenck *et al.*, 1985), et qui est à la base du modèle des 16 facteurs de Cattell.

- l'approche lexicale : fondée sur la langue naturelle et plus précisément sur les ressources lexicales comme celles qu'on trouve dans les dictionnaires (Baumgarten, 1933). L'*hypothèse lexicale* historique postule que l'essentiel des notions, socialement pertinentes, relatives aux traits de personnalité se sont peu à peu cristallisées, encodées en quelque sorte, dans la langue naturelle (Allport, 1937).

De nos jours, l'approche lexicale a été mise en avant par des auteurs comme Goldberg qui soutient que « le vocabulaire sur la personnalité (*e.g.* pour l'anglais, plus de 10 000 adjectifs et adverbes essentiellement) fournit un ensemble vaste mais fini d'attributs que les gens qui parlent une langue donnée ont trouvé suffisamment générique et utile pour leurs interactions conversationnelles de tous les jours » (Goldberg, 1981). En 1990, Goldberg a essayé de définir un petit ensemble de 475 adjectifs de personnalité les plus communs qui furent groupés en 131 ensembles de facteurs qui couvrent le domaine FFM (Goldberg, 1990). Ces travaux résultèrent en 1992 en un instrument de mesure à 50 items, appelé le « format transparent » (Goldberg, 1992), qui a convergé dans les années 90 avec les recherches d'autres auteurs, vers le modèle des cinq facteurs.

2.4.2. Les facettes NEO PI-R

Le modèle FFM est composé de cinq grandes classes considérées comme statistiquement indépendantes (cinq axes) de traits psychologiques qui sont décrits dans la colonne 1 du tableau 2. La première constatation, surtout si on se place du point de vue informatique (*i.e.* ayant en vue une implémentation effective des traits), c'est que la classification FFM est trop générale. Bien entendu, plusieurs auteurs en psychologie ont fait cette critique et ont essayé de raffiner cette taxonomie en la divisant en éléments appelés des *facettes* (Saucier, Ostendorf, 1999 ; Costa, McCrae, 1992 ; Soto, John, 2009).

Plusieurs systèmes de facettes ont été proposés au cours des 20 dernières années. John *et al.* (John, Naumann, Soto, 2008) ont montré que les diverses listes de facettes pro-

posées sont assez similaires bien que le nombre de facettes varie par exemple de 16 dans (Saucier, Ostendorf, 1999) jusqu'à 30 dans l'approche NEO PI-R³ proposée par Costa et McCrae (1992). Actuellement, NEO PI-R est le modèle le plus reconnu et le mieux établi car il fournit une liste de facettes très précises, ce qui nous a conduit à le retenir comme base pour le travail décrit à la section 3. Les 30 facettes de la classification NEO PI-R sont listées dans la seconde colonne du tableau 2, en même temps que les gloses associées aux facettes.

2.5. Utilisation de la base de sémantique lexicale WordNet

Les taxonomies proposées par les psychologues sont sous-tendues par une des deux principales méthodologie employées : approche à questionnaires ou approche lexicale. Généralement, la taxonomie FFM/NEO PI-R que nous avons choisie est fondée sur l'approche lexicale et plus précisément sur la classification d'adjectifs (parfois aussi d'adverbes) qui décrivent la personnalité d'une personne (gentil, sérieux, ouvert, timide, ...) dans les positions de la taxonomie. Cela revient à dire que nous devons nous appuyer sur un ensemble d'adjectifs pour essayer d'exhiber puis d'implémenter des opérateurs d'influence, tels que définis dans le cadre *R&B* par exemple.

D'un point de vue informatique, si on considère les données effectives, issues de la psychologie, on fait alors face à deux types de problèmes :

- *Accès aux données* : les tables de classifications fournies dans les articles sont trop synthétiques pour être exploitées, et il n'est pas facile d'avoir accès aux fichiers de ressources d'origine.

- *Approche computationnelle de la sémantique* : plus important, afin de satisfaire les besoins liés à une implémentation informatique, il faut s'appuyer certes sur les mots, mais aussi sur les progrès récents en sémantique lexicale qui positionnent les *sens lexicaux* précis comme étant des citoyens de première classe (ce qui n'était pas le cas des études originelles en psychologie). Par exemple, le mot « *kind* » possède trois sens lexicaux : Tolerant, Genial, Openhearted (*cf.* tableau 3), qui peuvent être classés en fait dans deux catégories FFM distinctes, ou encore dans trois facettes NEO PI-R distinctes, respectivement : A-compliance, E-warmth, A-tendermindedness. Cet exemple montre bien que l'approche traditionnelle, fondée sur les mots, est insuffisante. Nous devons plutôt travailler au niveau sous-jacent de la sémantique lexicale attachée aux mots.

Il est pratique d'utiliser la base lexicale informatisée WordNet (Fellbaum, 1998) lorsqu'on a besoin de traiter systématiquement de larges ensembles de données lexicales, surtout si on a besoin de travailler au niveau des sens lexicaux. Dans WordNet, un mot est associé à plusieurs entités, appelées *synsets* qui définissent chacune un sens lexical unique, décrit au moyen d'une glose (aussi appelée phrase courte, *cf.* tableau 3 pour des exemples). La base informatisée WordNet a l'avantage supplémentaire d'être

3. NEO PI-R veut dire Neuroticism Extraversion Openness Personality Inventory-Revisited

librement accessible et largement partagée ce qui permet d’ancrer notre travail dans un domaine qui est un standard de fait. Enfin, le fait de travailler au niveau des synsets procure un avantage supplémentaire : il rend le processus de classification relativement indépendant de la langue d’origine des adjectifs de personnalité traités (ici l’anglais). En effet, non seulement on travaille au niveau des synsets qui sont des entités formelles, mais surtout, il existe d’autres bases informatisées comme EuroWordNet (Vossen, 1998) qui fournissent des équivalences entre synsets. Il est alors tout à fait envisageable que les travaux présentés ici puissent assez facilement être étendus au cas d’autres langues, en se servant des passerelles offerte par EuroWordNet.

WordNet a déjà servi de support à des études à cheval entre linguistique et psychologie, *e.g.* le projet *WordNet Affect* (Strapparava, Valitutti, 2004 ; Strapparava, Michalcea, 2008). Cependant, ces travaux sont orientés vers la reconnaissance des traces d’affects (*affect cues*) dans les textes et non vers l’expression de comportements psychologiques au moyen d’agents conversationnels ; de plus la classification choisie est essentiellement *ad hoc*. S’il s’agit bien de travaux complémentaires à ceux présentés ici, ils n’établissent pas une liaison entre les synsets des adjectifs WordNet et les taxonomies de traits (si possible FFM voire FFM/NEO PI-R), ce qui est notre besoin ici. Notons par ailleurs que dans notre cas, il est nécessaire de se restreindre à la classe des adjectifs décrivant la personnalité d’une personne, qui n’est pas réifiée dans WordNet, par exemple sous forme de la relation d’hyperonymie, justifiant le travail décrit dans la section suivante.

3. Processus de classification

Pour les raisons données ci-dessus, la taxonomie à deux niveaux FFM/NEO PI-R a été choisie pour servir de support au processus de classification d’un ensemble d’adjectifs de personnalité sous forme de paires synset-gloses. Ce processus a été réalisé en deux étapes principales, décrites dans la section qui suit :

1. **Sélection des adjectifs** : nous avons collecté un ensemble d’adjectifs décrivant des traits de personnalité (P_a) à partir de ressources disponibles sur Internet, et les avons associés avec toutes leurs paires synset-gloses fournies par WordNet. Dès lors, nous n’avons plus considéré les adjectifs eux-mêmes mais les paires synset-gloses, rejetant les paires non relatives à des traits de personnalité (*e.g.* les paires marquées * dans le tableau 3).

2. **Association aux facettes** : les paires synset-gloses restantes ont été classées manuellement dans la taxonomie FFM/NEO PI-R, permettant la constitution d’une ressource (fichier XML) de 1 356 éléments regroupés autour de 60 schémas psychologiques.

3.1. Étape 1 : Sélection des adjectifs de personnalité

WordNet permet l’extraction de listes de mots en fonction de leurs attributs, comme le *part-of-speech* (POS) (dans notre cas, POS = adjectif uniquement) ou bien qui sont

TABLE 3. *Synsets WordNet associés à trois adjectifs de personnalité très usités (entrées de $\mathcal{B}_{all}$ tirées de $\mathcal{C}_{WN}$ qui ont une fréquence $w_a = 9$). Dans la colonne 3, les étoiles indiquent les synsets qui ne sont pas relatifs à des traits de personnalité.*

w_a	Adjectif	→P	Synset	Glose (phrase courte de définition)
9	friendly		Pally	characteristic of or befitting a friend
9		*	Allied	of or belonging to your own country's forces or those of an ally
9		*	Easy	easy to understand or use
9			Favorable	inclined to help or support; not antagonistic or hostile
9	kind		Tolerant	tolerant and forgiving under provocation
9			Genial	agreeable, conducive to comfort
9			Openhearted	having or showing a tender and considerate and helpful nature; used especially of persons and their behavior
9	lively		Vital	full of spirit
9		*	Eventful	filled with events or activity
9			Frothy	full of life and energy
9		*	Springy	elastic; rebounds readily
9			Alert	quick and energetic
9			Racy	full of zest or vigor

en relation avec un mot donné (*e.g.* tous ses synonymes, antonymes, méronymes ...). Cependant, il n'est pas possible d'extraire une liste de mots appartenant à un domaine linguistique particulier (*e.g.* les « traits de personnalité »). De plus, dans la mesure où WordNet a pour vocation de couvrir la langue anglaise de manière exhaustive, il ne propose pas d'information de nature statistique, comme par exemple les fréquences d'usage quotidien des mots (ou plus précisément de leurs sens). Suite à ces considérations, nous avons procédé à une sélection d'un ensemble d'adjectifs P_a en quatre sous-étapes S_{1-4} décrites dans les paragraphes suivants.

3.1.1. S_1 : Recueil des adjectifs

Afin de travailler avec des adjectifs effectivement usités, nous avons collecté un corpus ($\mathcal{C}_{coll}$) à partir de dix sources Internet différentes affirmant offrir des « listes d'adjectifs décrivant les traits de personnalité des gens » (ces sources sont résumées dans le tableau 4). $\mathcal{C}_{coll}$ contient 1 055 adjectifs distincts, fournissant une approximation d'ordre un du domaine linguistique associé aux adjectifs de personnalité. De plus, pour chaque adjectif de personnalité a , nous avons associé un poids w_a indiquant le nombre de listes dans lequel il apparaît. Ceci permet d'obtenir une information quant à son usage fréquentiel dans les ressources, appelé *saillance*; notons bien qu'il ne

TABLE 4. Sources des adjectifs de personnalité pour le corpus $\mathcal{C}_{coll}$

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

^a La liste 6 est composée de l'union de deux sources proches.

^b La liste 7 est un sous-ensemble d'une liste décrivant des personnes d'une manière générale.

s'agit pas de son usage quotidien, même si on peut espérer qu'il existe une certaine corrélation.

 Tableau 5. Adjectifs de $\mathcal{C}_{WN}$ les plus saillants. Total = 55

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

3.1.2. S2 : Amélioration de la représentativité des adjectifs fondée sur WordNet

Afin de renforcer le poids des adjectifs les plus saillants dans les listes de ressources, nous avons utilisé la relation fonctionnelle synonym de WordNet : pour chaque élément $\mathcal{C}_{coll}$ on ajoute 1 au poids de tous les éléments de $\mathcal{C}_{coll}$ qui se trouvent être synonymes (i.e. $\forall a \in \mathcal{C}_{coll}$, si $\exists a' \in \mathcal{C}_{coll}$ tel que $a \in \text{synonym}(a')$ alors $w_a = w_a + 1$). Note : Nous avons aussi considéré la possibilité d'ajouter à l'ensemble $\mathcal{C}_{coll}$, tous les synonymes WordNet de $\mathcal{C}_{coll}$, mais la relation synonym de WordNet génère beaucoup de mots dont les sens sortent pour la plupart du domaine des traits de personnalité (sens avec *). En effet, $\text{synonym}(\mathcal{C}_{coll})$ produit une liste supplémentaire de 5 139 mots

dont 2 986 sont distincts entre eux et distincts de $\mathcal{C}_{\text{coll}}$, ce qui fait perdre l'intérêt même de la méthodologie employée en bruitant l'information réellement pertinente.

3.1.3. S3 : Sélection des entrées pour WordNet

Le corpus $\mathcal{C}_{\text{coll}}$ est issu de listes de mots collectés hors de WordNet. Ainsi tous les éléments de $\mathcal{C}_{\text{coll}}$ ne sont pas présents dans la base WordNet même si celle-ci est vaste. Afin d'être techniquement compatible avec WordNet, les mots de $\mathcal{C}_{\text{coll}}$, absents de WordNet ($\mathcal{C}_{\text{WN}}^{\text{abs}}$) ainsi que ceux présents mais pas sous forme d'adjectifs ($\mathcal{C}_{\text{WN}}^{\neg\text{adj}}$), ont été rejetés de $\mathcal{C}_{\text{coll}}$. En pratique, ils sont très rares : on a $|\mathcal{C}_{\text{WN}}^{\text{abs}}| = 29$ et $|\mathcal{C}_{\text{WN}}^{\neg\text{adj}}| = 21$. Notons que $\forall a \in \mathcal{C}_{\text{WN}}^{\text{abs}} \cup \mathcal{C}_{\text{WN}}^{\neg\text{adj}}$ on a $w_a = 1$, ce qui confirme le fait que WordNet couvre effectivement les adjectifs les plus saillants. Nous avons obtenu ainsi une liste $\mathcal{C}_{\text{WN}}$ de 1 005 adjectifs distincts, traitable avec WordNet, telle que $\mathcal{C}_{\text{coll}} = \mathcal{C}_{\text{WN}} \cup \mathcal{C}_{\text{WN}}^{\text{abs}} \cup \mathcal{C}_{\text{WN}}^{\neg\text{adj}}$ et qui peut être triée du mot le plus saillant ($w_a = 9$) au moins saillant ($w_a = 1$). L'histogramme de saillance des adjectifs de $\mathcal{C}_{\text{WN}}$ est donné figure 1, et les adjectifs a tels que $w_a \in [6..9]$ sont listés dans le tableau 5.

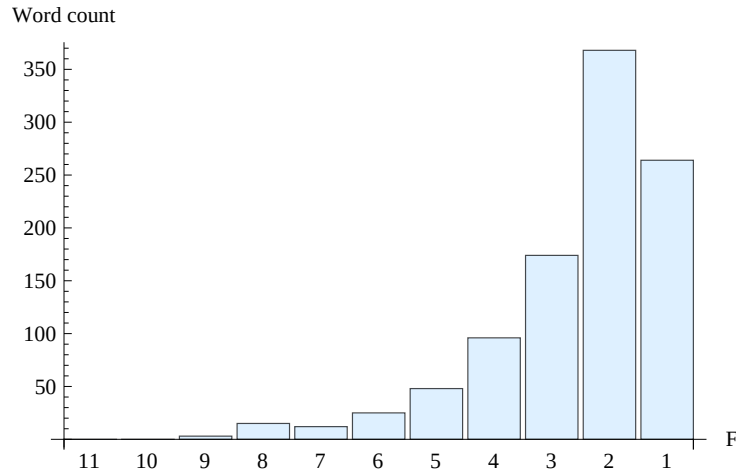


Figure 1. Histogramme de saillance des adjectifs de $\mathcal{C}_{\text{WN}}$ (F est w_a)

3.1.4. S4 : Sélection des sens liés à la personnalité

Chaque adjectif $a \in \mathcal{C}_{\text{WN}}$ possède une entrée WordNet (par construction) qui associe à a un ensemble de sens en sémantique lexicale (synsets). Dans WordNet, chaque synset est unique et définit un sens unique au moyen d'une glose. Il est donc possible d'obtenir à partir de $\mathcal{C}_{\text{WN}}$ une base $\mathcal{B}_{\text{all}}$ de triplets uniques $\langle \text{adjectif}, \text{synset}, \text{glose} \rangle$. Pour les 1 005 adjectifs de $\mathcal{C}_{\text{WN}}$, nous avons obtenu 2 924 triplets qui devraient constituer les 2 924 entrées de $\mathcal{B}_{\text{all}}$. Cependant, tous les synsets d'un adjectif ne sont pas nécessairement associés à la description de traits de personnalité (cf. les entrées avec

une * en colonne trois du tableau 3) et ils doivent être éliminés. C'est pourquoi, il ne faut conserver dans $\mathcal{B}_{all}$ que les entrées associées à des traits.

Tableau 6. Trois tours de sélection des sens liés aux traits dans $\mathcal{B}_{all}$

Phase	Opération	Compte	% $\mathcal{B}_{all}$
	$\mathcal{C}_{WN}$ mots (col 2 dans le Tableau 3)	1 005	
	$\mathcal{B}_{all}$ triplets initiaux	2 924	100%
Tour 1 :	triplets sélectionnés par A_1	1 771	60,6%
	triplets sélectionnés par A_2	1 642	56,2%
	conflits A_1/A_2	449	15,4%
Tour 2 :	conflits restants après A_1	241	8,2%
	conflits restants après A_2	124	4,2%
Tour 3 :	$\mathcal{B}_{sel}$ triplets finalement conservés	1 715	58,7%

Le processus de sélection a été effectué par deux annotateurs indépendants A_1 et A_2 en trois tours successifs : pendant le premier tour, A_1 et A_2 ont annoté la base en entier, alors que dans le deuxième tour A_1 a reconsidéré les mots sur lesquels il y avait un conflit A_1/A_2 puis A_2 a reconsidéré les conflits restants. Finalement, dans le troisième tour, A_1 et A_2 se sont rencontrés pour une discussion à propos des derniers conflits restants ; en cas de désaccord persistant, la décision par défaut consiste à conserver le triplet dans $\mathcal{B}_{all}$, car il peut toujours être rejeté lors d'une phase ultérieure.

Les résultats de ce premier processus de filtrage sont résumés dans le tableau 6. Bien que deux annotateurs seulement soient impliqués dans ce processus, la qualité du filtrage est plutôt bonne ($\kappa = 0.68$ après le tour 1). Ceci s'explique par le fait qu'il s'agit de choisir entre deux classes simples, et parce que en cas de conflit persistant, la décision est retardée à une étape future où une décision mieux informée est alors possible. Nous avons ainsi obtenu une base $\mathcal{B}_{sel}$ de 1 715 triplets, contenant 904 adjectifs distincts (où chaque glose est associée à au moins un adjectif et, inversement, où chaque adjectif est associé à une glose).

Le nombre d'entrées retenues au final (58.7% seulement des synsets d'origine) est déjà un premier résultat en soi qui confirme notre postulat sur la nécessité de travailler au niveau de la sémantique lexicale (*i.e.* au niveau synset-gloses) plutôt qu'au niveau des mots, comme cela était le cas dans les classifications en psychologie (Goldberg, 1981 ; 1990 ; 1992).

3.2. Étape 2 : Association aux facettes

3.2.1. Construction de la base $\mathcal{B}'_{sel}$ de triplets glose-adjectif-synset

L'objectif suivant consiste à classer les gloses présentes dans $\mathcal{B}_{sel}$. Malheureusement, il arrive parfois que les gloses WordNet soient associées à plus d'une paire (adjectif, Synset). Par exemple, la glose « *causing or fraught with or showing anxiety* » correspond à (nervous, Queasy), (anxious, Nervous) et (uneasy, Queasy). Dans de tels

cas, qui sont assez spéciaux, nous avons choisi de fusionner les entrées ayant la même glose. Ceci a débouché sur la constitution d'une seconde base, $\mathcal{B}'_{\text{sel}}$, qui est constituée de triplets (glose, (adjectif, synset)*), telle que $|\mathcal{B}'_{\text{sel}}| = 1\,356$.

3.2.2. Ajout de pôles à la taxonomie FFM/NEO PI-R

On a vu dans la section 2.4 qu'alors que plusieurs taxonomies à base de facettes ont été proposées pour FFM, la taxonomie NEO PI-R est actuellement une des plus influentes et fournit de plus une bonne précision à l'aide de 30 facettes. Dans NEO PI-R, chaque facette englobe à la fois le concept et son antonyme, *e.g.* la facette A-Modesty peut accueillir les adjectifs « *mild* » aussi bien que « *arrogant* ». Donc de manière similaire aux travaux de Goldberg sur les échelles bipolaires à 50 positions (Goldberg, 1992), chaque facette a été divisée en deux pôles, notés + pour le concept et - pour son antonyme principal, débouchant sur 60 positions notées FFM/NEO PI-R +/-.

Lors d'une première tentative pour répartir un sous-ensemble de $\mathcal{B}'_{\text{sel}}$ sur les 60 positions nous avons rencontré un problème majeur : malgré le fait que les adjectifs collectés sont considérés par leurs sources comme associés à la description de traits de personnalité, en fait, plusieurs de leurs paires synset-gloses dans WordNet sont associées à d'autres types d'attributs d'une personne⁴ qui peuvent être divisés en cinq grandes catégories, dites *externes* :

- *descriptions physiques*, *e.g.* « having a sturdy and well proportioned body », (athletic, Muscular) ;
- *capacités mentales* *e.g.* « characterized by quickness and ease in learning », (bright, Smart), (smart, Bright) ;
- *jugements de qualités*, soit physiques soit mentales, *e.g.* « appealing to the emotions as well as the eye », (lovely, Beautiful) ;
- *rôles sociaux*, *e.g.* « characterized by or indicating poverty », (poor, Beggarly) ;
- *états mentaux* qui sont transitoires (comme les émotions), *e.g.* « in an aroused state », (excited, Thrillful) ;
- *erreurs* correspondant à des éléments qui ne rentrent pas dans les facettes, *i.e.* qui auraient dû être éliminés pendant la phase de sélection (voir section 3.1).

3.2.3. Protocole d'annotation

En conséquence, les entrées $\mathcal{B}'_{\text{sel}}$ ont finalement été classées soit dans un des 60 pôles FFM/NEO PI-R +/- , soit dans une des cinq catégories externes listées ci-dessus. L'annotation de $\mathcal{B}'_{\text{sel}}$ a été réalisée par deux annotateurs indépendants A_1 and A_2 , selon le même protocole que celui décrit dans l'étape quatre de la section 3.1, à la différence près que dans le troisième tour, pour les rares cas où un conflit persistait entre les annotateurs, les gloses ont été rejetées. L'annotation a été réalisée grâce à un outil logiciel illustré à la figure 2.

4. Encore une fois, ceci est en faveur de notre approche synset-gloses, qui se trouve être bien plus précise que le niveau des mots.

1 characteristic of or befitting a friend
({Pally, friendly, 9})
Modesty + .

FIND
BACK
NEXT

NEO-PI-R facet:

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

Value in the facet:

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

Meta-Infos:

☐ Judgement

Other items linked to this facet by:

synset word gloss

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

Figure 2. Outil pour la classification des entrées $\mathcal{B}_{sel}^l$ dans FFM/NEO PI-R+/-.

3.3. Résultats et analyses

Les résultats du processus d’annotation sont donnés dans le tableau 7. On voit qu’après trois tours, les deux annotateurs sont arrivés à un accord sur 90% des positions. Il faut noter que le premier ratio de conflit est plus important ici que dans 3.1; cela ne devrait pas être mal interprété : le nombre de catégories possibles (60) est beaucoup plus important qu’en 3.1 (seulement 2). En fait, il y a trois cas de conflit :

1. Une simple erreur d’un annotateur. Ces erreurs disparaissent à la fin du deuxième tour quand les deux annotateurs ont à réviser leur propres annotations entrant en conflit (dès la fin de la première étape, les conflits sont soumis aux annotateurs qui doivent proposer une nouvelle annotation: garder leur opinion, adopter l’autre annotation, proposer une nouvelle annotation).
2. Un conflit à propos de l’interprétation des gloses ambiguës associées aux facettes. Ils sont en principe résolus pendant les tours 2 et 3 lors d’une discussion entre les annotateurs. Par exemple, on a pu observer un fort degré d’ambiguïté entre des facettes comme E-Assertiveness et A-Modesty, C-Achievement-Striving et C-Self-discipline ou encore O-Ideas et O-Values.

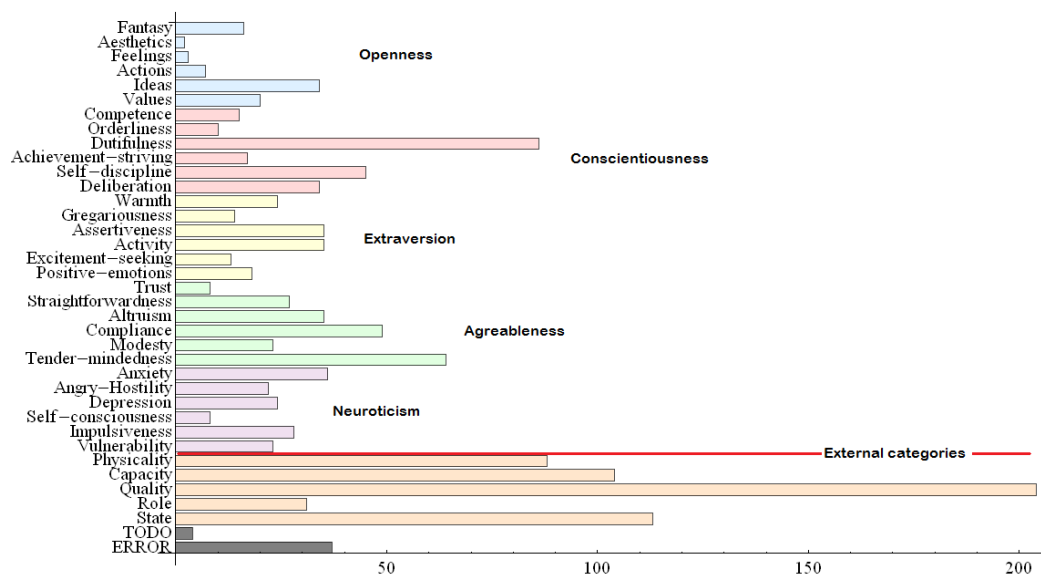


FIGURE 3. Histogramme du nombre de triplets par facette FFM/NEO PI-R

3. Un conflit d'interprétation relatif à une glose donnée, tout particulièrement quand les gloses sont complexes ou composées : *e.g.* ("joyful and proud especially because of triumph or success", ((exultant, Exultant), (prideful, Jubilant))) réfère en même temps à une tendance à ressentir des émotions positives E-Positive-emotions et à une tendance au manque de modestie A-Modesty. Dans de tels cas, les annotateurs doivent choisir ensemble une option, lors du tour 3, ou bien se résigner à rejeter la glose, puisque l'association d'une glose à plusieurs facettes n'était pas possible (*cf.* discussion ci-dessous).

Tableau 7. Résultats des trois tours de classification des triplets de $\mathcal{B}'_{sel}$

Phase	Opération	Nombre	% $\mathcal{B}'_{sel}$
	$\mathcal{B}'_{sel}$ entrées	1 356	100%
Tour 1:	conflits de facettes A_1/A_2	526	38,8%
	conflits de traits A_1/A_2	376	27,7%
Tour 2:	conflits de facettes après A_1	274	20,2%
	conflits de traits après A_1	178	13,1%
	conflits de facettes après A_2	133	9,8%
	conflits de traits après A_2	101	7,4%
Tour 3:	entrées rejetées	37	2,7%

Les recherches en psychologie, en particulier en ce qui concernent les facettes FFM, font généralement intervenir plusieurs annotateurs (novices ou pairs) et sont fondées sur des classes construites statistiquement. Notre travail ici est moins ambitieux : d'une part, nous disposons déjà comme cible pour les gloses d'une hiérarchie à deux niveaux de 2×30 positions qui a été validée par les recherches en psychologie, et, d'autre part, si une glose se retrouve mal placée (*i.e.* les deux annotateurs font indépendamment la même erreur), cela ne perturbe que les schémas composés de quelques gloses seulement (une ou deux), ce qui correspond de fait aux schémas associés à des adjectifs ayant une très faible saillance. Nous avons aussi envisagé la possibilité de classer certaines gloses de manière multiple, ce qui aurait permis d'augmenter notablement l'accord entre annotateurs. Cela aurait également permis de tenir compte de certaines gloses ambiguës ou même polysémiques (l'outil WordNet n'est pas parfait et quelques gloses WordNet peuvent s'avérer être assez vagues). En dernier ressort, cette option n'a toutefois pas été retenue car il a été décidé de ne pas éviter les conflits d'annotation (ce qui est mieux avec seulement deux annotateurs), quitte à réduire les gloses vagues à une seule interprétation.

La distribution finale des triplets glose-adjectif-synset dans les différentes catégories (30 facettes + les classes externes) est résumée dans la figure 3. Bien que certaines facettes comme C-Dutifulness ou A-Tender-mindedness se révèlent avoir été associées à plus de gloses que la moyenne, on peut voir que toutes les facettes NEO PI-R ont obtenu des triplets (avec un minimum de 2 pour O-Aesthetics).

Implémentation. Un fichier XML, issu de l'implémentation informatique de la classification, reprenant toutes les gloses, mots et synsets associés aux catégories FFM/NEO PI-R est accessible librement sur le Web (<http://www.limsi.fr/~jps/research/rnb/toolkit/taxo-glosses/taxo.htm>). La figure 4 présente un extrait de cette page.

4. Elicitation des schémas comportementaux

4.1. Définition des schémas comportementaux

4.1.1. La nécessité de raffiner les facettes en schémas

À partir du processus de classification des triplets de $\mathcal{B}'_{\text{sel}}$ dans la taxonomie FFM/NEO PI-R, nous avons obtenu une ressource informatisée dans laquelle chaque facette contient deux listes bipolaires disjointes de triplets. Les gloses contenues dans une facette donnée fournissent autant d'instanciations des notions psychologiques associées à la facette. À partir de là, on peut faire deux constatations:

1. Selon les principes du modèle *R&B*, les comportements doivent être implémentés en termes d'influences sur le processus de raisonnement rationnel de l'agent. La présentation des opérateurs d'influence et du modèle formel sous-jacent sort du cadre de l'article ; à ce sujet, on pourra consulter (Sansonnnet, Bouchet, 2010a).
2. Pour une facette donnée, si nous considérons l'ensemble des triplets contenus dans la liste bipolaire qui lui est associée nous nous pouvons l'observer d'un point

NEO-PI-R facets with asso... x

fouchet.vorty.net/projects/r&b/adjectives/neopiradj.xml

Openness

Catégorie o.c.e.a.n.

Fantasy

Facette NEO PI-R

+

Pôle positif (concept)

not sensible about practical matters; idealistic and unrealistic
[romantic, WildEyed, 5]

having the ability to produce or originate
[productive, Generative, 4]

concerned with sacred matters or religion or the church
[religious, Spiritual, 4] [spiritual, Religious, 3]

promoting construction or creation
[creative, Constructive, 3]

having the ability or power to create
[creative, Originative, 3]

(used of persons or artifacts) marked by independence and creativity in thought or action
[imaginative, Imaginative, 3] [inventive, Inventive, 3]

concerned with or affecting the spirit or soul
[spiritual, Unearthly, 3]

Glose Wordnet [mot, synset, saillance]

Multi synsets

abrégé

-

Pôle négatif (antonymes)

unimaginatively conventional
[conservative, ButtonDown, 3]

guided by practical experience and observation rather than theory
[practical, Hardheaded, 3] [hardheaded, Pragmatic, 2]

aware or expressing awareness of things as they really are
[realistic, Virtual, 3]

freed from illusion
[disillusioned, Disillusioned, 2]

sensible and practical
[earthy, DownToEarth, 2]

denying or questioning the tenets of especially a religion
[skeptical, Unbelieving, 2]

concerned with the world or worldly matters
[mundane, Terrestrial, 1]

deficient in originality or creativity; lacking powers of invention
[sterile, Uninspired, 1]

Figure 4. Extrait de la page Web contenant le fichier de classification des gloses.

de vue computationnel. En ce qui nous concerne, ce point de vue est dirigé par les principes du modèle *R&B* c'est-à-dire qu'il est implémenté au moyen d'opérateurs d'influence. On constate alors qu'au sein d'un pôle de facette, les comportements s'organisent en sous ensembles associés à des opérateurs d'influence distincts.

C'est la raison pour laquelle, pour chaque facette de la taxonomie FFM/NEO PI-R, il nous a fallu procéder à un réarrangement des triplets en termes de ce que nous appelons des *schémas comportementaux*. Les schémas sont des ensembles disjoints de triplets qui correspondent à une classe de comportement, sémantiquement cohérente du point de vue computationnel.

Par exemple, si on considère la position O-Fantasy dans la figure 4, ses gloses concernent trois comportements computationnels distincts : créativité, idéaux et spiritualité, ceci aussi bien dans le pôle positif que le pôle négatif. Cela engendre alors trois schémas, eux aussi bipolaires qui leur sont associés (voir la colonne 1 du tableau 8).

4.1.2. Structure des schémas

Formellement, chaque schéma est une structure de la forme :

$$\sigma \doteq \langle S_i, T_\phi, \gamma_i, S_i^+[\tau_1, \dots, \tau_n], S_i^-[\tau_1, \dots, \tau_m] \rangle$$

où

S_i est un symbole, identifiant unique du schéma σ

T_ϕ la catégorie du schéma dans la taxonomie FFM/NEO PI-R: *Trait_{facette}*

γ_i la définition textuelle unique (ou glose) du schéma

S_i^+ est un symbole, identifiant unique du pôle positif (*resp.* négatif)

$[\tau_1, \dots, \tau_n]$ l'ensemble des triplets associés au schéma pour le pôle positif (*resp.* négatif)

Comme les facettes, les schémas sont organisés en deux pôles (positif et négatif), mais contrairement aux pôles des facettes qui sont placés sur une échelle à deux valeurs (le concept et son antonyme), les deux pôles des schémas sont placés sur une échelle à trois valeurs : tendance supérieure à la moyenne (+), la moyenne (=), tendance inférieure à la moyenne (-). En fait, la position moyenne (=) n'a pas de triplets qui lui sont associés, car elle dénote implicitement le « comportement moyen d'une personne moyenne ». Les pôles + et - dénotent explicitement une personnalité ayant une divergence de comportement par rapport à la moyenne. La définition de l'existence et de la définition d'une "personne psychologiquement moyenne" est une question qui demeure en discussion en psychologie. Dans le cadre d'une approche computationnelle, on peut toutefois se contenter d'une première approximation. Cette approximation peut être posée dans un premier temps en termes de sens commun ; par exemple, elle peut être fondée sur une ontologie définie par des annotateurs – c'est la méthode utilisée ici : *cf.* la remarque, ligne 1 dans le tableau 8. Dans un second temps, elle peut être posée sur une base statistique, par exemple, pour ensuite définir des schémas qui exhibent une différenciation par rapport à la définition fournie par l'ontologie.

4.1.3. Sémantique des schémas

Techniquement, la sémantique d'un schéma σ est définie par deux notions :

- la localisation dans FFM/NEO PI-R de σ qui hérite de la sémantique associée ;
- la liste des triplets que σ contient.

De plus, pour chaque pôle $\pm$ d'un schéma σ , nous donnons, à titre d'illustration, une définition informelle γ_i sous forme de phrase courte⁵ qui synthétise les triplets qu'il contient.

Exemple : les facettes de O-Fantasy, illustrées dans la figure 4, peuvent être divisées en trois schémas donnés dans le tableau 8 (les gloses des pôles des schémas sont abrégées par la simple donnée de leur nombre N_g).

Tableau 8. Clusterisation de la facette Openness-Fantasy (Ofa) en trois schémas bipolaires disjoints. Pour les pôles +/- du schéma CREATIVENESS, nous développons l'ensemble des triplets associés, réduits ici à leur glose ($\triangleright$).

Identifiants des pôle de schémas	Définition textuelle du schéma Remarque : + abrège 'more than average' (resp. - 'less than average')	N_g
1.IDEALISTICNESS	Ofa	
+ IDEALISTIC	+ inclined to like (or believe in) imaginary ideas or situations	1
- PRACTICAL	- inclined to like (or believe in) imaginary ideas or situations	5
2.SPIRITUALISTNESS	Ofa	
+ SPIRITUALIST	+ interested in spirit and soul matters	6
- MATERIALIST	- interested in spirit and soul matters, even denying tenets	2
3.CREATIVENESS	Ofa	
+ CREATIVE	+ having power of creativity in thought or action $\triangleright$ <i>having the ability to produce or originate</i> $\triangleright$ <i>promoting construction or creation</i> $\triangleright$ <i>having the ability or power to create</i> $\triangleright$ <i>[...] independence and creativity in thought or action</i>	4
- UNIMAGINATIVE	- having power of creativity in thought or action $\triangleright$ <i>unimaginatively conventional</i> $\triangleright$ <i>deficient in originality or creativity; lacking powers of invention</i> $\triangleright$ <i>unimaginative and conformist</i> $\triangleright$ <i>lacking spontaneity or originality or individuality</i>	4

4.2. Réorganisation des facettes en schémas comportementaux

Pendant le processus de réorganisation des facettes en schémas comportementaux, 766 gloses (en fait des triplets de $\mathcal{B}_{\text{sel}}^i$) ont été annotées et classées en 69 schémas distincts et disjoints, résultant en un ensemble total de $2 \times 69 = 138$ pôles. La figure 5

5. Elle a été produite par les annotateurs en fonction de la liste des triplets associés au schéma.

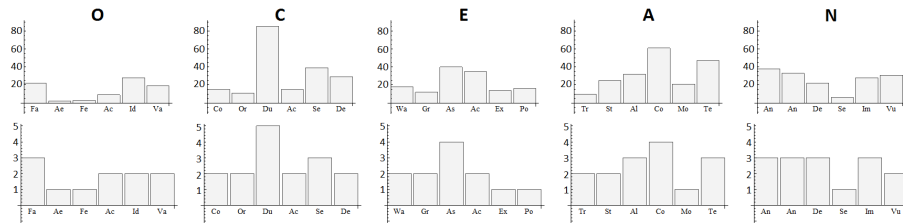


Figure 5. Distribution dans FFM/NEO PI-R des gloses (haut) et des schémas (bas), où: $N_{facette} = 30$, $N_{glose} = 766$, $N_{schma} = 69$, nombre moyen de gloses par facette = 26 et nombre moyen de schémas par facette = 2,3.

trace la distribution des gloses et des schémas dans la taxonomie FFM/NEO PI-R. Globalement, on observe une couverture bonne et homogène aussi bien en termes de gloses que de schémas. Il y a aussi une couverture bonne et homogène des gloses vis-à-vis des schémas : la distribution varie d'un minimum de 2 à un maximum de 40, avec une moyenne de 26 gloses par schéma. Cependant la couverture au niveau des pôles n'est pas complète : 11 pôles de schémas sur 138 n'ont pas de gloses attachées. Ceci est dû au fait que chaque fois qu'un nouveau schéma a été introduit (disons par son pôle positif), l'autre pôle a aussi été automatiquement créé, pour être complété plus tard (ou pas, dans 11 cas). Pour remplir ces positions il faudrait considérer d'étendre la liste des adjectifs sélectionnés, par exemple comme discuté en note de la section 3.1.2. Toutefois en première approximation, la sémantique des pôles vides peut être dérivée en prenant l'antonyme de celle de leur pôle opposé ; en effet par construction, au moins un des deux pôles d'un schéma n'est pas vide. Par exemple, dans le schéma LOYALNESS la sémantique du pôle négatif UNFAITHFUL (contenant 0 gloses) peut être inférée « en inversant » les 8 gloses de son pôle positif LOYAL, *e.g.* la glose + « *steadfast in affection or allegiance* » donne la glose – « *inconstant/wavering in affection or allegiance* ».

Implémentation. La classification des gloses en termes de schémas comportementaux a été implémentée informatiquement sous forme d'un second fichier XML qui est librement accessible sur le Web à la page⁶ du projet R&B. Cette ressource contient les gloses traitées lors des classifications présentées plus haut, réarrangées en 138 schémas bipolaires.

4.3. Discussion

Dans cet article nous avons exposé dans le détail un processus de construction d'un troisième niveau pour la taxonomie FFM/NEO PI-R à base de schémas comportementaux englobant des ensembles de gloses issues de WordNet. Dans la mesure où nous mettons à disposition de la communauté les données résultantes de ce travail,

6. <http://www.limsi.fr/jps/research/rnb/toolkit/taxo-glosses/taxo.htm>

il convient d'en tracer la portée mais aussi les limites. Nous discuterons d'abord des limites et ensuite des extensions déjà produites et envisageables.

Le processus décrit ici a été effectué par deux personnes (les auteurs de l'article) qui sont avant tout des experts en informatique. Dans les études en psychologie, on distingue en général les travaux requérant des novices (par exemple les gens qui remplissent des questionnaires tels que définis à la section 2.4) ou des experts (*e.g.* les comparaisons entre diverses listes de facettes citées à la section (John, Naumann, Soto, 2008) sont l'œuvre des trois auteurs : John, Naumann et Soto). Notre approche entre clairement dans la seconde catégorie, avec l'accent mis sur le fait que les experts doivent être des informaticiens, vu que les schémas sont ensuite utilisés pour éliciter des opérateurs d'influence (voir paragraphe ci-dessous). Cependant, il serait utile de confronter les annotations de plusieurs types d'experts, même si le processus est lourd. Nous avons néanmoins jugé bon de mener cette étude car d'une part elle correspond à un réel besoin comme outil support de l'architecture *R&B*, et d'autre part, les résultats obtenus nous paraissent valides, au moins dans le cadre restreint de leur utilisation dans *R&B*. En effet, il s'agit pour nous de couvrir le champ des traits de personnalité et aussi de « remplir » les facettes avec des exemples concrets de comportements. Dans le pire des cas, des gloses mal placées ne peuvent qu'avoir un effet faible lors de la phase d'élicitation des opérateurs.

À la suite du processus décrit ici, nous avons pu utiliser la ressource produite lors de travaux d'élicitation d'opérateurs d'influence pour l'architecture *R&B*, en particulier pour les traits *Agreeability* et *Conscientiousness* (Bouchet, Sansonnet, 2011). Ces travaux ont confirmé que la ressource que nous avons construite correspond bien à nos besoins et offre un substrat solide aux opérateurs d'influence. Cette première étude démontre tout l'intérêt de l'approche que nous avons proposée et développée, et elle encourage à l'étayer en la confrontant à d'autres annotateurs.

5. Conclusion et perspectives

Nous avons décrit ici le processus de construction d'une taxonomie de comportements psychologiques effectifs qui est fondée : 1) sur des taxonomies standard en psychologie (FFM comme base avec NEO PI-R pour les facettes) et 2) sur une ressource lexicale informatisée très utilisée (WordNet). Partant d'une présélection de 1 005 adjectifs de personnalité, nous les avons distribués sur les 60 facettes FFM/NEO PI-R, en prenant en compte la notion de polarité des concepts. Les gloses associées aux sens lexicaux de WordNet (les synsets) fournissent alors une information sur la manière dont les facettes doivent être effectivement implémentées. Enfin, nous avons montré que, bien qu'il y ait souvent une correspondance directe entre les facettes et les comportements effectifs, il est utile de réarranger, à l'intérieur de chaque facette, les gloses en ensembles sémantiquement cohérents du point de vue opérationnel, que nous avons appelé des schémas comportementaux.

Nous pensons que la taxonomie FFM/NEO PI-R, enrichie par les schémas que nous avons élicités, peut constituer une ressource utile pour les travaux qui visent à implé-

menter des notions psychologiques dans des agents logiciels, en particulier des agents conversationnels. Par exemple, dans des travaux comme ceux que nous avons menés sur des agents *lazy* ou *scatter-brained* (Bouchet, Sansonnet, 2009), où nous nous retrouvons face à un manque d'information de nature computationnelle sur le domaine des traits de personnalité. Ceci est aussi confirmé par le fait que certains schémas élicités dans (Bouchet, Sansonnet, 2011) correspondent directement à des phénomènes bien étudiés par d'autres chercheurs, comme par exemple *trust* (Ramchurn *et al.*, 2005) et *empathy*. (Paiva *et al.*, 2004).

Enfin, une fois implémentés, les schémas devront être évalués lors de travaux qui vont suivre, pour ce qui nous concerne, dans le domaine des agents conversationnels. Cela peut être effectué en faisant interagir des sujets humains avec des agents exprimant les influences de leurs traits de personnalité sur leur comportement rationnel (par exemple, selon les principes de l'architecture *R&B* (Sansonnet, Bouchet, 2010a)). De telles expérimentations devraient confirmer que les gens perçoivent ou non les différences psychologiques que les concepteurs souhaitent faire exprimer par les agents.

Article reçu le 30 mars 2011

Accepté après révisions le 17 novembre 2011

Bibliographie

- Allbeck J. M., Kress-Gazit H. (2010). Constraints-based complex behavior in rich environments. In *Intelligent virtual agents (IVA 2010)*, vol. 6356, p. 1-14. Philadelphia, PA, Springer-Verlag.
- Allen J. F., Kautz H. A., Pelavin R. N., Tenenbergh J. D. (1991). *Reasoning about plans*. Morgan Kaufmann Publishers, Inc.
- Allport G. W. (1937). *Personality: A psychological interpretation*. New York: Holt.
- Anderson J. R. (1996). ACT: a simple theory of complex cognition. *American Psychologist*, vol. 51, n° 4, p. 355-365.
- Bates J. (1994, juillet). The role of emotion in believable agents. *Commun. ACM*, vol. 37, n° 7, p. 122-125.
- Baumgarten F. (1933). Die charaktereigenschaften [The character traits]. *Beitrage zue Charakter-und Personlichkeitsforschung*, n° 1.
- Baylor A., Kim Y. (2004). Pedagogical agent design: The impact of agent realism, gender, ethnicity, and instructional role. In, p. 592-603. Maceio, Brasil, Springer.
- 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. (2009, décembre). A framework for modeling the relationships between the rational and behavioral reactions of assisting conversational agents. In *Proc. of the 7th european workshop on Multi-Agent systems (EUMAS'09)*. Agia Napa, Cyprus.

- Bouchet F., Sansonnet J. P. (2011). 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.
- Bratman M. E. (1992). What is intention? In P. R. Cohen, J. Morgan, M. E. Pollack (Eds.), 2nd éd., p. 15–32. Cambridge, MA, MIT Press.
- Bratman M. E., Israel D. J., Pollack M. E. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, vol. 4, n° 4, p. 349–355.
- Cassell J., Bickmore T., Billinghurst M., Campbell L., Chang K., Villiamsson H. (1999). Embodiment in conversational interfaces: Rea. In *CHI'99: proceedings of the SIGCHI conference on human factors in computing systems*, p. 520–527. New York, NY, USA, ACM Press.
- 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.
- Davis F. D. (1989, septembre). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, vol. 13, n° 3, p. 319–340.
- Dias J., Paiva A. (2005). Feeling and reasoning: A computational model for emotional characters. In *Progress in artificial intelligence*, vol. 3808, p. 127–140. Springer Berlin / Heidelberg.
- 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.
- Ellsworth P. C., Scherer K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, H. H. Goldsmith (Eds.), *Handbook of affective sciences*, p. 572–595. New York, NY, USA, Oxford University Press.
- 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.
- Eysenck H. J., Eysenck S. B. G. (1975). *Manual of the eysenck personality questionnaire*. London: Hodder and Stoughton.
- Eysenck S. B. G., Eysenck H. J., Barret P. (1985). A revised version of the psychoticism scale. *Journal of Personality and Individual Differences*, vol. 6, n° 1, p. 21–29.
- Fellbaum C. (1998). *WordNet: an electronic lexical database*. MIT Press.
- Frijda N. H. (2006). *The laws of emotion*. Psychology Press.
- Gillis P. D., Hursh S. R. (1999). Using behavior moderators to influence cgf command entity effectiveness and performance. In *Proceedings of the eighth conference on computer generated forces and behavioral representation*, p. 237–251. Orlando, FL.
- 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.

- Goldberg L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, vol. 4, p. 26-42.
- Goodrich M. A., Schultz A. C. (2007). Human-robot interaction: a survey. *Found. Trends Hum.-Comput. Interact.*, vol. 1, n° 3, p. 203-275.
- 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.
- Howden N., Rannquist R., Hodgson A., Lucas A. (2001). Intelligent agents - summary of an agent infrastructure. In *Proc. of the 5th international conference on autonomous agents*. Montreal.
- Jiang H., Vidal J. M., Huhns M. N. (2007). eBDI: an architecture for emotional agents. In *AAMAS '07: proceedings of the 6th international joint conference on autonomous agents and multiagent systems*, p. 1-3. New York, NY, USA, ACM.
- John O. P., Naumann L. P., Soto C. J. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurements, and conceptual issues. In *Handbook of personality: Theory and research*, 3rd éd.. The Guilford Press.
- 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.
- Lim M., Dias J., Aylett R., Paiva A. (2008). Improving adaptiveness in autonomous characters. In H. Prendinger, J. Lester, M. Ishizuka (Eds.), *Intelligent virtual agents*, vol. 5208, p. 348-355. Springer Berlin / Heidelberg.
- Maes P. (1994). Agents that reduce work and information overload. *Commun. ACM*, vol. 37, n° 7, p. 30-40.
- Maes P. (1995). Artificial life meets entertainment: lifelike autonomous agents. *Commun. ACM*, vol. 38, n° 11, p. 108-114.
- Malatesta L., Caridakis G., Raouzaïou A., Karpouzis K. (2007). Agent personality traits in virtual environments based on appraisal theory predictions. In *Artificial and ambient intelligence, language, speech and gesture for expressive characters, AISB 07*. Newcastle, UK.
- 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.
- Mischel W., Shoda Y., Ayduk O. (2008). *Introduction to personality* (Eighth éd.). John Wiley & Sons, Inc.
- 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. (2003). On making believable emotional agents believable. In R. Trappl, P. Petta, S. Payr (Eds.), *Emotions in humans and artifacts*, p. 189–211. Cambridge, MA, MIT Press.
- Ortony A., Clore G. L., Collins A. (1988). *The cognitive structure of emotions* (Cambridge University Press éd.). Cambridge, UK.
- Paiva A., Dias J., Sobral D., Aylett R., Sobrepepe P., Woods S. (2004). Caring for agents and agents that care: Building empathic relations with synthetic agents. In *Proceedings of the third international joint conference on autonomous agents and multiagent systems*, vol. 1, p. 201.
- Pelachaud C. (2000). Some considerations about embodied agents. In *Proc. of the workshop on "achieving human-like behavior in interactive animated agents", in the fourth international conference on autonomous agents*. Barcelona, Spain.
- Ramchurn S., Huynh D., Jennings N. (2005). Trust in multi-agent systems. *The Knowledge Engineering Review*, vol. 19, n° 1, p. 1-25.
- 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.
- Rizzo P., Veloso M. V., Miceli M., Cesta A. (1997). Personality-driven social behaviors in believable agents. In *AAAI symposium on socially intelligent agents*, p. 109-114.
- 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.
- Sabouret N., Sansonnet J. P. (2001, mai). Un modèle de requêtes sur le fonctionnement de composants actifs. In *Proc. MFI 01*, vol. 3, p. 419–436.
- Sabouret N., Sansonnet J. P. (2003). Un langage de description de composants actifs pour le Web sémantique. *Information - Interaction - Intelligence Review, Revue I3*, vol. 3, n° 1, p. 9–36.
- Sansonnet J. P., Bouchet F. (2010a). Expression of behaviors in assistant agents as influences on rational execution of plans. In *Intelligent virtual agents (IVA 2010)*, vol. 6356, p. 413-419. Philadelphia, PA, Springer-Verlag.
- Sansonnet J. P., Bouchet F. (2010b, août). 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. (2011, mai). Integrating psychological behaviors in the rational process of conversational assistant agents. In R. C. Murray, P. M. McCarthy (Eds.), *Proceedings of the 24th international FLAIRS conference (FLAIRS-24)*, p. 440–445. Palm Beach, Florida.
- Sansonnet J. P., Castan M., Percebois C. (1980). M3L: A list-directed architecture. In *ACM-IEEE international symposium on computer architecture ISCA-7*, p. 8. La Baule, France.
- Sardina S. (2007). *A course on BDI architectures*. <http://goanna.cs.rmit.edu.au/~ssardina/courses/Roma07PhDcourse/>. RMIT University, Melbourne, Australia.

- 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.
- Silverman B. G., Cornwell M., O'Brien K. (2006). Human behavior models for agents in simulators and games: Part I: enabling science with PMFserv. In *PRESENCE: teleoperators and virtual environments*, vol. 15, p. 139-162.
- 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.
- Strapparava C., Mihalcea R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on applied computing*, p. 1556–1560. Fortaleza, Ceara, Brazil, ACM.
- Strapparava C., Valitutti A. (2004). WordNet-Affect: an affective extension of WordNet. In *Proc. of the 4th international conference on language resources and evaluation*, p. 1083–1086. Lisbon, Portugal.
- Traum D., Rickel J., Gratch J., Marsella S. (2003). Negotiation over tasks in hybrid human-agent teams for simulation-based training. In *AAMAS '03: proceedings of the second international joint conference on autonomous agents and multiagent systems*, p. 441–448. New York, NY, USA, ACM.
- Vossen P. (1998). EuroWordNet: a multilingual database with lexical semantic networks. *Computational Linguistic*, vol. 25, n° 4, p. 628–630.