Vers un modèle de détection des affects, appréciations et jugements dans le cadre d'interactions humain-agent

Caroline Langlet[1]
(1) Institut Mines-Télécom ; Télécom ParisTech ; CNRS LTCI, Paris
caroline.langlet@telecom-paristech.fr

Résumé. Cet article aborde la question de la détection des expressions d'attitude – i.e affect, d'appréciation et de jugement (Martin & White, 2005) – dans le contenu verbal de l'utilisateur au cours d'interactions en face-à-face avec un agent conversationnel animé. Il propose un positionnement en termes de modèles et de méthodes pour le développement d'un système de détection adapté aux buts communicationnels de l'agent et à une parole conversationnelle. Après une description du modèle théorique de référence choisi, l'article propose un modèle d'annotation des attitudes dédié l'exploration de ce phénomène dans un corpus d'interaction humain-agent. Il présente ensuite une première version de notre système. Cette première version se concentre sur la détection des expressions d'attitudes pouvant référer à ce qu'aime ou n'aime pas l'utilisateur. Le système est conçu selon une approche symbolique fondée sur un ensemble de règles sémantiques et de représentations logico-sémantiques des énoncés.

Abstract.

Toward a detection model of affects, appreciations, jugements within human-agent interactions
This article concerns the detection of attitudes – i.e affects, appreciations and jugements (Martin & White, 2005) – in the user's verbal content during a face-to-face interaction with an embodied conversational agent. It tackles the issue of the adaptation to the ECA's communicational goals and to the conversational speech. After a description of our theoretical model, it introduces an annotation model dedicated to the study of attitudes in a human-agent interaction corpus. Then, it describes a first version of our detection system, focusing on the attitude which can refer to a user's like or dislike. The system is rule-based and embeds logic and semantic representations of the sentences.

Mots-clés : Analyse de sentiments, interaction humain-agent, agents conversationnels animés, dialogue homme-machine.

Keywords: Sentiment analysis, human-agent interaction, embodied conversational agents, dialogue homme-machine.

1 Introduction

Dans le domaine croissant des agents conversationnels animées (ACA), de nombreuses applications permettent de faire interagir des agents virtuels avec des utilisateurs humains. Ces applications se fondent sur des scénarios variés où l'agent peut jouer différents rôles : assistant, tuteur ou encore compagnon. Pour de telles applications, la gestion de la composante affective de l'interaction est cruciale, tant du côté de la génération que de celui de la détection. Il est en effet nécessaire que l'agent puisse détecter les émotions, les sentiments et les attitudes sociales exprimés par l'utilisateur, afin de pouvoir produire des réactions affectives et sociales appropriées. Du côté de la détection, une majorité des travaux se concentrent sur l'analyse d'indices socio-affectifs non-verbaux (expressions faciales, indices acoustiques). Le contenu verbal reste quant à lui encore partiellement exploité. Deux études seulement présentent un système intégrant un module de détection des sentiments dans le contenu verbal pour les interactions humain-agent (Smith *et al.*, 2011; Yildirim *et al.*, 2011). Si ces méthodes proposent une première solution à cette question, les méthodes de détection qu'elles utilisent ne prennent néanmoins pas en compte la nature conversationnelle du contenu verbal qu'elles traitent.

Notre objectif est de développer un système de détection des sentiments exprimés par l'utilisateur au cours d'une interaction multi-modale, en anglais et en face-à-face avec un ACA. Si le rôle de l'ACA n'est pas défini a priori – notre modèle doit être adapté à des ACAs pouvant tenir différents rôles, comme celui d'assistant, de tuteur ou de compagnon – la prise en compte des spécificités énonciatrices d'une conversation humain-agent en face-à-face est néanmoins nécessaire pour

la conception de notre modèle de détection. Tout d'abord, le phénomène linguistique à détecter doit être circonscrit au regard des buts communicationnels de l'ACA et des modèles de relations sociales qu'il intègre. Dans un modèle socio-affectif de l'interaction, les expressions de sentiments pertinentes pour la détection sont celles qui permettront à l'ACA, à court-terme, de réagir de manière appropriée, à long-terme, de maintenir sa relation avec l'utilisateur. La nature de ces expressions sera donc différente de celles majoritairement visées en analyse de sentiments et en fouille d'opinions, où l'objectif est d'obtenir des informations sur des avis de consommateurs ou d'internautes. Cette première contrainte pose également la question du type d'analyse à fournir pour les expressions considérées : les résultats doivent être suffisamment précis pour être exploitables pour la création d'un modèle de préférences de l'utilisateur permettant le calcul de relations sociales ou la gestion de l'engagement de ce dernier dans la conversation. Ensuite, le système de détection doit pouvoir s'adapter à une parole spontanée et conversationnelle : il doit ainsi être en mesure d'en gérer des particularités tant au niveau syntaxique que pragmatique. Sur le plan syntaxique, il est nécessaire de prendre en compte les disfluences – hésitations, répétitions – conférant aux énoncés oraux une organisation syntaxique éloignée de la régularité de l'écrit. Sur le plan pragmatique, le caractère conversationnel du discours pris en charge implique une élaboration sémantique des expressions de sentiments pouvant s'effectuer sur plusieurs tours de parole et en collaboration avec l'agent.

Cet article s'intéresse plus particulièrement à deux de ces contraintes : l'adaptabilité aux buts communicationnels de l'agent et la prise en compte de la parole conversationnelle. Dans un premier temps, les méthodes développées en analyse de sentiments sont mises en regard des contraintes et des objectifs propres à notre cadre applicatif (Section 2). Ce panorama permettra de justifier la méthode et le modèle théorique de référence choisis : une méthode symbolique permettant d'intégrer une modélisation logico-sémantique des énoncés et s'appuyant sur le modèle théorique proposé par (Martin & White, 2005). Dans un deuxième temps, nous présentons une étude exploratoire des attitudes dans le contexte conversationnel de corpus Semaine (McKeown *et al.*, 2011) (Section 3). Enfin, sur la base des résultats de cette analyse, nous décrivons une première version du système reposant sur l'analyse conjointe des énoncés de l'agent et des énoncés de l'utilisateur pour la détection des attitudes de l'utilisateur 4 et 5.

2 Analyse de sentiments : modèles et méthodes face au contexte humain-agent

Le développement d'un système de détection des sentiments est, dans notre cadre applicatif, motivé par la volonté d'améliorer la relation sociale de l'ACA avec l'utilisateur. Dans cette perspective, les analyses fournies par le système doivent être suffisamment précises pour aider l'agent à se positionner et à réagir de manière appropriée face aux expressions détectées. Ainsi, le système doit tout d'abord être en mesure de détecter des expressions individualisées, dont la source est l'utilisateur. En effet, dans la majorité des cas, les sentiments ne pouvant être attribués à l'utilisateur ne seront que d'une utilité limitée pour le système. Ensuite, la catégorisation doit être suffisamment fine pour qu'il soit possible de cibler les phénomènes les plus appropriés aux besoins de l'agent. Enfin, le système doit intégrer une analyse précise des propriétés caractéristiques des expressions de sentiments : le calcul de la polarité doit être solide et la cible clairement identifiée afin que l'agent ne commette pas d'erreurs dans ses prises de décision et dans sa gestion de l'interaction. En vue de répondre à ces questions, cette section propose une synthèse des modèles linguistiques de référence et des méthodes utilisées en analyse de sentiments. L'objectif est de cibler dans l'existant les méthodes et modèles appropriés à nos objectifs et d'en définir les limites afin de proposer une adaptation pour l'analyse de sentiments dans le contexte conversationnel.

2.1 Un modèle théorique adapté aux buts communicationnels de l'ACA

La conception d'un système de détection des sentiments implique généralement de s'appuyer sur une modélisation – même minimale – du phénomène tel qu'il s'exprime dans le discours. Concernant le développement d'un tel système dans le cadre d'interactions humain-agent, le modèle linguistique doit répondre à certains critères. Du point de vue de l'agent et selon son modèle de relations sociales, toutes les expressions verbales de sentiments ne seront pas pertinentes pour la détection. Dans certains scénarios, l'ACA peut avoir besoin d'informations concernant des expressions plus affectives (« I'm sad », par exemple), tandis que pour d'autres, l'intérêt portera davantage sur des expressions exprimant un jugement de valeur (« This painting is beautiful »). Ainsi, le modèle théorique doit fournir une typologie détaillée des sentiments dans le langage, proposant une hiérarchie complexe de catégories. Ensuite, il doit également modéliser les phénomènes d'expression de sentiments comme un processus évaluatif impliquant une source et une cible, la détection de ces deux éléments étant indispensable pour une exploitation efficace. Enfin, afin de faciliter le développement de ressources linguistiques (lexiques, patrons d'extraction), ce modèle doit s'accompagner d'une description linguistique précise des réalisations verbales de sentiments. Parmi les modèles présentés ci-dessous, le modèle proposé par (Martin &

White, 2005) nous est apparu comme le plus approprié.

Oppositions subjectif-objectif, positif-négatif De nombreux travaux en analyse de sentiments se concentrent sur une opposition objectif-subjectif pour classer des textes, des phrases (Wiebe & Riloff, 2005) ou des groupes syntaxiques (Wilson *et al.*, 2004). Cette opposition est également sollicitée lors de la constitution de ressources linguistiques, qu'elles soient lexicales ou syntaxiques. Alors que (Esuli & Sebastiani, 2005) constituent un lexique, dans (Riloff & Wiebe, 2003), les auteurs décrivent une méthode permettant d'apprendre automatiquement des patrons de phrases subjectives. A cette distinction subjective-objective, s'ajoute fréquemment une distinction relative à la polarité des expressions considérées comme subjectives. Cette distinction peut être établie en référence au modèle d'Osgood (Osgood *et al.*, 1975) ou des *Private States* (Quirk, 1985), dont il n'est retenu que cette notion d'axe de valence. Ainsi dans (Turney, 2002) et (Hu & Liu, 2004), les auteurs proposent de classer ou de résumer des revues d'internautes selon leur orientation sémantique. Dans certains cas, la granularité choisie peut être plus fine. Ainsi, dans (Nasukawa & Yi, 2003), les auteurs présentent un algorithme pour classer les phrases selon leur caractère favorable-défavorable, tandis que dans (Wilson *et al.*, 2005), le focus est mis sur l'attribution d'une polarité à des groupes syntaxiques. Dans un contexte humain-agent, une classification en termes de valence peut apporter des informations essentielles. Néanmoins, n'offrant pas de catégorisation précise des expressions de sentiments, cette approche ne peut suffire à elle seule.

Modèles psychologiques Afin de pouvoir classer plus précisément les énoncés ou les items lexicaux, certains travaux reprennent des classifications fournies par des modèles développés en psychologie cognitive. Ainsi (Neviarouskaya *et al.*, 2007) et (Neviarouskaya *et al.*, 2010a) reprennent les 9 classes d'émotions proposée par (Izard, 1977) : joie, dégoût, peur, colère, tristesse, surprise, honte, intérêt, culpabilité. D'autres travaux (par exemple (Ishizuka, 2012)) choisissent de s'appuyer sur la classification du modèle OCC (Ortony, Clore and Collins (Ortony *et al.*, 1990)). Si ces modèles offrent l'avantage d'une classification plus complexe et détaillée, ils restent néanmoins lacunaires quant à la description proprement linguistiques des expressions de sentiments. Leur objectif n'est pas de décrire des réalisations verbales mais bien d'appréhender des mécanismes psychologiques. Une telle description est cependant nécessaire pour le développement de notre système : elle sera un outil théorique indispensable à la conception de ressources linguistiques (lexiques, patrons d'extractions). Pour cette raison, nous avons décidé de fonder notre système de détection sur un modèle de référence plus orienté langage, celui décrit dans (Martin & White, 2005).

Affects, appréciations et jugements dans le langage Ce modèle, issu de la linguistique systémique fonctionnelle, fournit une description détaillée des réalisations verbales de sentiments. Son utilisation dans certains travaux d'analyse de sentiments (Neviarouskaya *et al.*, 2010b; Bloom *et al.*, 2007; Whitelaw *et al.*, 2005) a permis de démontrer son adaptabilité aux problèmes de modélisation des expressions de sentiments. Le modèle présente une hiérarchie complexe des *attitudes* divisées en trois sous-classes : les *affects*, qui réfèrent à des réactions émotionnelles ; les *jugements*, qui réfèrent à des évaluations axiologiques de comportements humains ; les *appréciations* qui expriment des évaluations d'artefacts ou d'événements naturels. Dans le cadre de ce modèle, les expressions d'attitudes sont décrites comme reposant sur trois éléments : la source, la personne évaluant ou expérimentant l'attitude, la cible, l'entité ou processus évalué, et enfin, l'expression linguistique permettant de référer à l'attitude en question. Du point de vue humain-agent, le modèle de (Martin & White, 2005) apparaît comme le plus approprié. Il cumule les avantages des modèles psychologiques – classification précise et modélisation des sources et des cibles – tout en gardant un point de vue linguistique sur le phénomène, facilitant ainsi la formalisation symboliques d'expressions pouvant intégrer un modèle de détection. De plus, le grand intérêt qu'il accorde à la modélisation des propriétés de ce phénomène – comme l'intensité ou l'engagement – garantit à terme la possibilité d'intégrer l'analyse de ces dernières par le module de détection.

2.2 Des méthodes de classification vers des analyses à grain fin

L'adaptabilité aux buts communicationnels de l'ACA doit déterminer la nature de l'analyse fournie par le module de détection : une analyse à grain fin capable (i) de distinguer différentes expressions d'attitudes au sein d'une même phrase, (ii) d'identifier la source et de la cible de chaque expression, (iii) de déterminer précisément la polarité. Enfin, le processus d'analyse devra être suffisamment modulaire pour être adaptable au contexte conversationnel. Cette section propose une présentation des différentes méthodes développées pour la détection de sentiments dans les textes et interroge la possibilité de leur adaptation aux interactions humain-agent. Parmi l'ensemble des méthodes proposées en analyse de sentiments, il est possible de distinguer celles optant pour un niveau de granularité haut et faisant une utilisation assez minimale de

modélisation linguistique, de celles s'intéressant à un niveau de granularité plus bas – celui de l'expression de sentiments – et intégrant des représentations symboliques des énoncés exprimant des sentiments.

Méthodes statistiques et d'apprentissage automatique pour la classification des sentiments Un grand nombre de travaux proposés en analyse de sentiments se fondent sur des méthodes statistiques ou d'apprentissage automatique et ont pour objectif de classifier des items lexicaux, des textes ou des phrases. Les premiers travaux proposés ont ainsi majoritairement porté sur la classification d'items lexicaux. L'objectif est ainsi de distinguer des adjectifs objectifs d'adjectifs subjectifs. Ainsi dans (Hatzivassiloglou & McKeown, 1997), les auteurs cherchent à prédire l'orientation d'adjectifs conjoints, i.e. des adjectifs reliés par un conjonction de coordination, en exploitant un modèle de régression logistique. Rapidement ont également émergé des méthodes pour la classification de textes – généralement des revues de produits ou de films. Ainsi, dans (Pang & Lee, 2004), les auteurs comparent deux types de classifieurs pour déterminer automatiquement l'orientation positive ou négative des revues de films : SVM (Support Vector Machines) et NB (naive bayes). Si les systèmes de classification de textes subjectifs ont montré des résultats intéressants, de nombreux travaux ont néanmoins eu l'ambition de descendre au niveau de la phrase afin de ne plus traiter le texte de sa globalité. Là encore, différents algorithmes de classification par apprentissage automatique ont pu être utilisés : *boostrapping* (Wilson *et al.*, 2005; Riloff & Wiebe, 2003; Wilson *et al.*, 2004), *classifieur naïf bayésien* (Wiebe & Riloff, 2005), *rule learning* (Wilson *et al.*, 2004), *machine à vecteur de support* (Wilson *et al.*, 2004), *conditional random fields* (Breck *et al.*, 2007).

L'ensemble des méthodes présentées ici permettent de résoudre un certain nombre de difficultés liées à la détection de sentiments et d'opinions dans les textes. Dans le cadre de conversation humain-agent, elles peuvent permettre d'attribuer une polarité globale au tour de parole de l'utilisateur. Néanmoins, davantage développées pour des problématiques de fouille d'opinion dans des larges corpus, leur niveau de granularité est trop haut pour répondre à l'ensemble de nos problèmes. Tout d'abord, l'agent doit pouvoir distinguer les différentes expressions d'attitudes produites au sein d'un même tour de parole afin de se positionner clairement face à chacune d'elles. Ensuite, l'exploitation des données issues de la détection pour la création d'un modèle utilisateur exige un niveau de précision en termes de calcul de polarité que ne fournissent pas ces méthodes. Le calcul qu'elles effectuent ne prend généralement pas en compte le traitement des modificateurs de valence et lorsque cela est fait, celui-ci reste minimal. Enfin, la source et la cible sont rarement prises en charge.

Méthodes symboliques et hybrides pour l'analyse profonde des expressions de sentiments Les méthodes à grain fin semblent davantage répondre aux contraintes posées par notre cadre applicatif. Ces travaux intègrent généralement des représentations formelles des énoncés. Celles-ci sont exploitées soit par des algorithmes à base de règles, soit par des algorithmes hybrides associant analyse profonde de la phrase et méthode d'apprentissage. L'intérêt de telles méthodes est de pouvoir accorder une plus grande importance à l'analyse de propriétés intrinsèques des expressions de sentiments et d'opinions.

Tout d'abord, la prise en compte de la structure logico-sémantique des énoncés leur permet de gérer le principe de compositionnalité sémantique et d'améliorer ainsi sensiblement le calcul de la polarité. La prise en compte de ce principe, déterminant le sens d'un énoncé comme composé du sens de l'ensemble de ses constituants et de leurs relations hiérarchiques, permet de définir un certain nombre de règles de calcul de la polarité s'appuyant sur une représentation symbolique des structures logico-sémantiques. Ainsi les travaux présentés dans (Neviarouskaya *et al.*, 2010b) et (Moilanen & Pulman, 2007) proposent des méthodes de détection à base de règles, se fondant sur une analyse des relations de dépendance entre constituants et permettant un calcul fin de la polarité des expressions de sentiments. Ils modélisent ainsi des règles de propagation ou d'inversion pour résoudre des conflits de polarité à différents niveaux de la structure syntaxique. Cette approche est également adoptée dans (Shaikh *et al.*, 2009). Les auteurs y offrent une interprétation du modèle OCC (Ortony *et al.*, 1990). Sur la base de cette interprétation, les auteurs définissent des règles linguistiques pour la détection d'expressions référant à des émotions et le calcul précis de la polarité.

Au-delà d'un calcul plus fin de la polarité, ces méthodes ont également l'avantage d'être plus optimisées pour la détection des sources et des cibles des expressions d'opinions ou de sentiments. Ainsi, dans (Choi *et al.*, 2005), les auteurs proposent une méthode de détection des sources des opinions utilisant conjointement les CRF (*conditional random fields*) et des patrons d'extraction acquis via AutoSlog (Riloff, 1996). Ce type d'approche a également été exploité par (Yang & Cardie, 2013) pour la détection conjointe des sources et des cibles des expressions d'opinions. Là encore, les auteurs associent CRF et patrons syntaxiques.

Pour une première approche de la détection des attitudes dans le cadre de conversations humain-agent, nous avons fait le choix d'une méthode symbolique. Si les méthodes hybrides présentent des avantages en termes de coût de développement

et de rapidité de traitement, leur application à un contexte d'interaction humain-agent nécessite néanmoins une validation des ressources linguistiques qu'elles emploient dans le cadre d'une parole conversationnelle. Le développement d'une méthode symbolique permettra ainsi de pouvoir élaborer et valider un ensemble de règles linguistiques concernant tant le niveau logico-sémantique que pragmatique. Celles-ci pourront être exploitées à plus long-terme par une méthode hybride.

3 Les attitudes en contexte conversationnel : modèle d'annotation et exploration du corpus Semaine

Cette section décrit une étude exploratoire de corpus, visant à appréhender les attitudes telles qu'elles se manifestent dans le contexte d'une parole conversationnelle. Cette étude a pour objectif de servir de base au développement des règles linguistiques utilisées par le système. Dans un premier temps, nous présentons un modèle d'annotation (Section 3.1) s'attachant à décrire la relation entre les attitudes exprimées par l'utilisateur et les énoncés de l'agent. Dans un second temps, nous détaillons les résultats de l'application de ce modèle au corpus Semaine (Section 3.2).

3.1 Modéliser la relation des attitudes de l'utilisateur avec le contexte antérieur

Afin de pouvoir modéliser les expressions d'attitudes de l'utilisateur dans le contexte d'une parole conversationnelle, nous sommes partis du postulat développé en théorie de la conversation, considérant une partie des énoncés produits dans le cadre conversationnel comme des actes participatifs ou collectifs, fonctionnant comme des contributions au discours (Clark & Schaefer, 1989). En tant qu'énoncés à part entière, les expressions d'attitudes peuvent être considérées comme des formes de contribution intrinsèquement liées au déroulement de la conversation : motivées par ce qui a été énoncé au préalable, elles ont également une influence sur ce qui sera dit par la suite. En guise de première approche, le modèle d'annotation présenté ici s'intéresse aux liens énonciatifs que les attitudes de l'utilisateur entretiennent avec le contexte antérieur de la conversation. Plus particulièrement, il s'agit d'appréhender la manière dont l'agent peut influencer et déterminer les expressions d'attitude chez l'utilisateur. Pour cela, le modèle d'annotation considère à la fois le contenu verbal de l'utilisateur et celui de l'agent. Formellement, le modèle intègre des unités (« séquences d'élément textuels adjacents »), des relations (« rapports binaires entre deux unités ») et des schémas (« configurations textuelles complexes récurrentes impliquant unités et relations » (Widlöcher & Mathet, 2012)).

Les actes illocutoires des énoncés de l'agent Des labels spécifiques ont été définis à la fois pour les attitudes exprimées par l'utilisateur et pour les énoncés de l'agent. Afin de prendre en compte la manière dont l'agent collabore et influence la production d'attitudes chez l'utilisateur, les énoncés de l'agent sont annotés selon l'acte illocutoire qu'ils performent. Pour cela, nous utilisons la classification établie par Searle (Searle, 1976) qui inclut cinq catégories : les actes assertifs qui engagent le locuteur sur la véracité de son propos (par exemple, « It's raining ») ; les actes directifs qui tentent d'obtenir quelque chose de l'interlocuteur (par exemple, « I order you to leave ») ; les actes commissifs qui engagent le locuteur sur des événements futurs (« I promise to pay you the money ») ; les actes expressifs qui expriment l'état mental du locuteur à propos de quelque chose (« I apologize for stepping on your toe ») ; les actes de déclaration dont la performance permet de faire correspondre le contenu propositionnel à un état de fait du monde (par exemple, « You're fired »). Pour étiqueter chaque énoncé de l'agent, nous utilisons l'unité *Enoncé Agent*, à laquelle nous ajoutons un trait spécifiant la nature de l'acte illocutoire. Un même tour de parole peut ainsi contenir plusieurs énoncés réalisant différents actes illocutoires.

Annotation des attitudes Le modèle prend compte les attitudes exprimées dans les énoncés de l'utilisateur et dans ceux de l'agent. Une expression d'attitude est composée de trois éléments qu'il est nécessaire de pouvoir annoter : l'indice linguistique exprimant le caractère évaluatif ou affectif de l'énoncé, la source et la cible. Des informations à propos du type d'attitude et de la polarité doivent également être spécifiées. Le type de l'attitude et la polarité (positive ou négative) sont indiqués grâce à une structure de traits associée soit au schéma utilisateur – décrit ci-dessous – soit à l'unité *Énoncé Agent*. Concernant le type d'attitude, la catégorie *affect* est conservée telle quelle, tandis que les catégories *appréciation* et *jugement* sont regroupées au sein de la catégorie *évaluation*. Les unités d'annotation *Source* et *Target* sont utilisées pour annoter les groupes syntaxiques référant à la source et à la cible des attitudes lorsqu'elles sont exprimées. Dans le but de vérifier l'influence de l'agent sur les expressions d'attitudes des l'utilisateur, la cible du côté utilisateur est reliée à la cible du côté agent lorsqu'elles réfèrent à la même entité. L'indice linguistique exprimant le caractère évaluatif ou affectif de

l'énoncé (*Indice Attitude*) est uniquement pris en compte dans le cas des attitudes exprimées par l'utilisateur. Concernant l'agent, le trait spécifiant le type d'attitude (*Type Attitude*) est suffisant pour indiquer si l'énoncé véhicule une expression d'attitude. L'unité d'annotation *Indice Attitude* est appliquée au niveau du syntagme et couvre à la fois les lexies référant à une attitude mais aussi les modifieurs de valence. Par exemple, dans une phrase comme « I don't really like my work », « don't really like » est annoté comme *Indice Attitude*, incluant le marqueur de négation « don't »

Relier les attitudes de l'utilisateur aux énoncés de l'agent A un niveau de granularité plus haut, notre modèle d'annotation se compose de deux types de schémas : des schémas utilisateur et des schémas agent. Chaque *Schéma Agent* est composé d'une unité *Enoncé Agent* et des éventuelles unités *Source* et *Cible* auxquelles elle peut être reliée. De même, un *Schéma Utilisateur* est composé de l'unité *Indice Attitude* et des unités *Source* et *Cible* auxquelles elle est éventuellement associée. Afin de pouvoir modéliser la relation entre les expressions d'attitudes de l'utilisateur et les énoncés de l'agent, chaque *Schéma Utilisateur* est relié au *Schéma Agent* qui le précède.

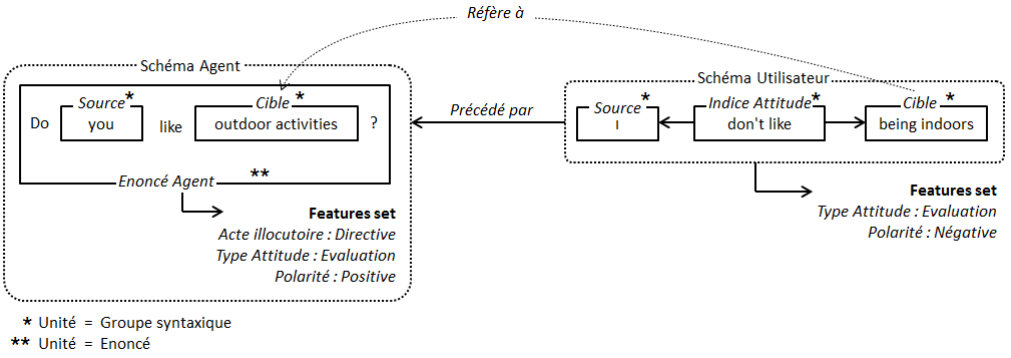

FIGURE 1 – Modèle d'annotation

3.2 Application sur un corpus d'interaction humain-agent

Corpus Semaine Le corpus Semaine (McKeown *et al.*, 2011) est composé de 65 transcriptions manuelles de sessions où un utilisateur humain interagit, en anglais, avec un opérateur humain jouant le rôle d'un agent conversationnel. Ces interactions sont fondées sur un scénario impliquant quatre personnages d'agent émotionnellement typés : Poppy, joyeuse et extravertie, Prudence, sensible et raisonnable, Spike, colérique et conflictuel, et enfin Obadiah, dépressif et maussade. Les énoncés que peuvent prononcer les opérateurs sont contraints par un script – néanmoins, certains énoncés dans le corpus s'écartent de ce script – ayant l'objectif de pousser l'utilisateur vers le même état émotionnel que celui du personnage joué par l'opérateur humain. Au total, pour notre étude, 16 sessions (4 pour chaque personnage) ont été annotées manuellement. Pour l'annotation nous avons utilisé la plate-forme Glozz (Widlöcher & Mathet, 2012). Dans la mesure où il s'agit d'une étude exploratoire et non de la constitution d'un corpus annoté pour l'apprentissage d'un algorithme supervisé, un seul annotateur a procédé à l'annotation. Les sessions annotées ont un nombre variable de tours de parole (132 pour la session la plus longue, 38 pour la session la plus courte). Parmi les 450 schémas agent annotés, 46% expriment un acte directif, 42% un acte expressif et 12% un acte représentatif. Aucun acte déclaratif ou commissif n'a été trouvé. Cela est probablement dû à la nature du scénario sur lequel est fondé le corpus Semaine : une conversation narrative où l'utilisateur est poussé à parler de sa vie. Parmi l'ensemble des schémas agent, 339 réfèrent à une attitude (187 à un affect, 152 à une évaluation). Du côté de l'utilisateur, 238 schémas ont été annotés : 44% sont notés comme affect, 56% comme évaluation (jugements et appréciations).

Quelle est l'influence des énoncés de l'agent sur l'expression des attitudes de l'utilisateur ? Au regard de l'annotation manuelle effectuée sur le corpus Semaine, il est possible d'avoir un premier aperçu de l'influence de l'agent sur les expressions d'attitudes de l'utilisateur. Tout d'abord, sur un plan pragmatique, la nature de l'acte illocutoire précédant les expressions d'attitude de l'utilisateur est un premier indicateur. En effet, les attitudes exprimées par l'utilisateur sont majoritairement précédés par des actes directifs de type interrogatif. Parmi ces questions, certaines concernent directement des attitudes de l'utilisateur. Ensuite, sur le plan du contenu propositionnel, il est à noter que, lorsqu'une attitude de

l'utilisateur est précédée par un énoncé de l'agent contenant lui aussi une expression d'attitude, les polarités concordent généralement. De plus, concernant les cibles, un quart des cibles évaluées par l'utilisateur réfèrent à une cible évaluée par l'agent.

4 Développement d'un système de détection adapté au modèle de relations sociales de l'ACA : focus sur les *likes* et les *dislikes*

Cette section décrit la première version de notre système de détection. Celle-ci a été décrite de manière exhaustive dans (Langlet & Clavel, 2015). Cette section et la suivante présentent donc une synthèse de son principe de fonctionnement et de son évaluation. Cette première version du système restreint volontairement le type d'attitudes à détecter. La délimitation est définie relativement à l'une des dimensions utilisées pour modéliser les relations sociales de l'ACA, le *liking*. La définition de ce concept est fondée sur la théorie de Heider – *Heider's Balance Theory* (Heider, 1958) qui envisage la manière dont les relations entre deux personnes, impliquant des entités impersonnelles, s'équilibrent relativement à la relation que chacune de ces deux personnes entretient avec ces entités individuelles. La théorie de Heider est utilisée par les modèles d'agents sociaux définissant des scénarios où la nature de la relation (relation de *liking*) entre l'agent et l'utilisateur est déterminée par leur goût (*liking*) pour d'autres entités (entités inanimées, processus ou événements). Afin de pouvoir fournir des informations nécessaires au calcul du liking, notre système est conçu pour détecter les expressions d'attitudes pouvant référer à ce que l'utilisateur aime ou n'aime pas (*likes* et *dislikes*). Ces expressions d'attitudes peuvent trouver leur place dans les trois catégories – affect, appréciation, jugement – définies dans le modèle de (Martin & White, 2005). Par exemple, si les phrases « This painting makes me sad » (« Cette peinture me rend triste ») et « this painting is a master-work » (« cette peinture est un chef-d'oeuvre ») appartiennent respectivement aux classes *affect* et *apprécition*, elles réfèrent également, chacune de manière différente, à ce qu'aime ou n'aime pas l'utilisateur.

Pour détecter les expressions d'attitudes exprimant un *like* ou un *dislike*, le système se fonde sur une analyse conjointe des énoncés de l'agent et de ceux de l'utilisateur. En effet, si l'analyse du corpus Semaine a montré un lien entre la forme et le contenu propositionnel des énoncés de l'agent et l'expression d'attitude de l'utilisateur, elle a également révélé l'importance de la prise en compte des énoncés de l'agent pour la détection des attitudes de l'utilisateur. Ainsi, un énoncé produit par l'utilisateur peut contenir l'expression d'une attitude sans qu'un indice lexical ne soit utilisé. Par exemple, dans le cadre d'un échange du type, Agent « Do you like this painting ? » (« Aimez-vous cette peinture ? ») – Utilisateur « Yes » (« Oui »), la seule analyse de l'énoncé de l'utilisateur ne permet pas d'y déceler l'expression d'une attitude. Afin de pouvoir gérer ces difficultés, pour chaque énoncé de l'utilisateur analysé, le système analyse parallèlement l'énoncé de l'agent auquel il répond. Du côté de l'agent, le système cherche des expressions d'attitude pouvant être confirmées et infirmées par l'utilisateur. Cette analyse ne peut se faire qu'automatiquement : dans la plate-forme ACA que nous utilisons, Greta (Poggi *et al.*, 2005), les énoncés de l'agent ne sont pas produits automatiquement, mais scriptés manuellement. Il n'est donc pas possible de s'appuyer sur des ressources de génération pour obtenir des informations sur la présence d'une expression d'attitude. Du côté de l'utilisateur, le système cherche des expressions de confirmations ou d'infirmations ainsi que des attitudes pleinement formulées.

4.1 Analyse de l'énoncé de l'agent

Dans l'énoncé de l'agent, le système cherche à détecter des attitudes exprimées sous la forme d'une affirmation ou d'une question fermée, pouvant référer à un *like* ou *dislike* et dont la source peut être l'agent ou l'utilisateur. Trois niveaux sont considérés : un niveau lexical, un niveau syntagmatique, et un niveau phrastique. Au niveau lexical, après une tokenisation et un POS-tagging, le système vérifie la présence d'un indice lexical d'attitude à l'aide du lexique WordNet-Affect (Valitutti, 2004). Trois parties de discours sont considérées : les noms, les adjectifs et les verbes. La définition des affects proposée par WordNet-Affect étant plus large que celle de Martin et White (Martin & White, 2005), le lexique intègre notamment des lexies pouvant référer à des appréciations et à des jugements. Afin néanmoins d'adapter cette ressource à nos objectifs de détection, nous avons procédé à une sélection des lexies les plus pertinentes. Parmi l'ensemble des synsets, nous avons conservé ceux pouvant référer aux concepts de *like* et de *dislike* et appartenant, dans la taxinomie de WordNet-Affect, à ces trois catégories : *positive-emotion*, *negative-emotion*, *neutral-emotion*. Une fois l'analyse lexicale effectuée, si un indice d'attitude est trouvé, le système lance l'analyse du niveau syntagmatique. Ce niveau est géré par une grammaire formelle implémentée sous forme d'automates via la plate-forme Unitex (Paumier, 2015). Les règles définies par la grammaire sont fondées sur les patrons permettant de reconnaître trois types de syntagmes : des syntagmes

verbaux, adjectivaux ou nominaux. Cette phase applique également un certain nombre de règles sémantiques permettant de calculer la polarité attribuée au syntagme lorsque celui-ci comporte une lexie étiquetée comme indice d'attitude. Le niveau phrastique permet de vérifier si la structure logico-sémantique de la phrase de l'agent correspond à l'expression d'une attitude référant à un *like* ou un *dislike* de modalité affirmative ou interrogative et ayant une forme attributive ou processive. Là encore, l'analyse emploie une grammaire formelle implémentée sous forme d'automates. Les figures 2 et 3 présentent une version simplifiée des règles supérieures des grammaires respectivement dédiées à la détection de ces deux formes syntaxiques d'expressions d'attitudes. Les non-terminaux $SyntVb$, $SyntNoun$ et $SyntAdj$ correspondent aux syntagmes verbaux, nominaux et adjectivaux détectés lors de la phase précédente de l'analyse. Leur trait sémantique att à une valeur égale à $true$ lorsqu'ils comportent une lexie référant à une attitude. Les arguments int et aff indiquent si l'expression est de modalité interrogative ou affirmative.

`Att(aff) -> Src(usr|agt),SyntVb(cop), SyntAdj(att:true), Target.` `Att(int) -> Aux, Source(usr), SyntVb(cop), SyntAdj(att:true), Target.`	*I am really happy to do that* *Are you really happy to do that?*
`Att(aff) -> Target, SyntVb(make), Src(usr|agt), SyntAdj(att:true).` `Att(int) -> Aux, Target, SyntVb(make), Src(usr|agt), SyntAdj(att:true).`	*This book makes me sad* *Does this book make you sad?*
`Att(aff) -> Src(usr|agt), SyntVb(have), SyntNoun(att:true).` `Att(int) -> Aux, Src(user), SyntVb(have), SyntNoun(att:true).`	*I had an awful week* *Did you have an awful week?*
`Att(aff) -> Target, SyntVb(cop), [SyntNoun|SyntAdj]`	*This book is amazing*
`Att(aff) -> SyntNoun(PronDem), SyntVb(cop),[SyntNoun(att:true)|SyntAdj(att:true)].`	*It is amazing to do that*
`Att(aff) -> SyntNoun(PronDem),SyntVb(cop),[SyntNoun(att:true)|SyntAdj(att:true)], "for"|"of", Target, InfClause.`	*It is silly of them to do that*
`Att(aff) -> Src(usr|agt), SyntVb(opinion), Target, "as"|"like", [SyntNoun(att:true)|SyntAdj(att:true)].` `Att(int)-> Aux, Src(usr), SyntVb(opinion),Target,"as"|"like", [SyntNoun(att:true)|SyntAdj(att:true)].`	*I consider this painting as beautiful* *Do you consider this book as beautiful?*

Règle de polarité

$\texttt{If Neg}_{\texttt{SyntVb}} \texttt{ == True : Att(pol:inv(Pol}_{\texttt{SyntAdj}}\texttt{|Pol}_{\texttt{SyntNoun}}\texttt{))}$
$\texttt{Else: Attitude(pol:Pol}_{\texttt{SyntAdj}}\texttt{|Pol}_{\texttt{SyntNoun}}\texttt{)}$

Dans ce type de phrase, la valeur attitudinale est portée par le syntagme attribut (nominal ou adjectival). Quand le verbe a une forme négative la polarité attribuée à l'expression est l'inverse de celle portée par le syntagme attitudinalement marqué.

FIGURE 2 – Règles du niveau phrastique prenant en compte des phrases de type attributif

Le niveau phrastique permet également de vérifier la nature de la source (agent ou utilisateur) et de catégoriser la cible. Pour la source, elle est assimilée à l'agent lorsque sa forme correspond à un pronom de la première personne ($Src(agt) \rightarrow$ "I"|"me") et à l'utilisateur lorsque forme correspond à un pronom de seconde personne ($Src(usr) \rightarrow$ "you"). Pour les cible, le système ne procédant pas à une résolution des anaphores, il n'est capable de leur assigner que des classes génériques. Les deux premières sont relatives aux deux membres de la conversation : l'agent et l'utilisateur. La troisième, appelée *other*, implique toutes les entités ou processus qui ne sont ni l'agent ni l'utilisateur. La dernière, appelée *unknown*, concerne des entités ou processus dont la classe même générique ne peut être connue. Enfin, lorsqu'une expression d'attitude est trouvée, le système produit une structure de traits de la forme suivante : $source = \{user, agent\}$, $polarity = \{neg, pos\}$, $targetType = \{user, agent, other, unknown\}$. A ces attributs sémantiques, est également ajouté un attribut syntaxique spécifiant la présence d'une négation dans l'expression d'attitude $negation = \{true, false\}$

`Att(aff) -> Src(usr|agt), SyntVb(att:true), Target.`	I like this book
`Att(int) -> Aux, Src(usr|agt), SyntVb(att:true), Target.`	Do you like this book?

Règle de polarité

$\texttt{Attitude(pol:Pol}_{\texttt{ChkVb}}\texttt{)}$

Dans ce type de phrase, la valeur attitudinale est portée par le syntagme verbal. Les règles de polarité sont donc appliquées au niveau syntagmatique de l'analyse (inversion de la valeur portée par le verbe en cas de négation). Au niveau phrastique, aucun modifieur n'entre en jeu, la polarité de l'expression est donc la même que celle du syntagme verbal.

FIGURE 3 – Règles du niveau phrastique prenant en compte des phrases de type verbal (processif)

4.2 Analyse des énoncés de l'utilisateur

4.2.1 Confirmation ou infirmation d'attitudes exprimées par l'agent

Lorsque l'agent exprime une attitude sous forme d'affirmation ou de question, l'utilisateur peut être amené à formuler : (i) une confirmation ou une infirmation simple, la première définissant le contenu propositionnel de l'agent comme vrai, la seconde comme faux ; (ii) une confirmation ou une infirmation modalisée, faisant porter sur le contenu propositionnel une modalité pouvant être aléthique, déontique, temporelle ou épistémique (Le Querler, 1996). Pour résumer de manière formelle le fonctionnement sémantique de ces deux types de segments, il est possible de dire qu'ils définissent une valeur pour un attribut $value_p$ prise dans l'ensemble $\{p, \neg p, \diamond p, \Box p, \diamond\neg p, \Box\neg p\}$. Pour le moment, le système ne détecte que les confirmations ou infirmations simples. Leurs versions modalisées seront intégrées dans une version ultérieure. Lorsque la réponse fait suite à un énoncé référant à une attitude, la valeur qu'elle attribue à $value_P$ va également entraîner une confirmation ou une infirmation de la valeur que l'énoncé de l'agent avait attribuée à $polarity$. En cas d'une infirmation, la valeur de l'attribut sera ainsi inversée. Il est également important de noter que, de même qu'une confirmation ne s'exprime pas nécessairement par un *yes* (ou ses synonymes), une infirmation ne l'est pas nécessairement par un *no* (ou ses synonymes). Ainsi, dans l'énoncé interro-négatif « Don't you love outdoors activities », le contenu propositionnel « you don't love outdoors activities » est affirmé par la réponse « no » et infirmé par « yes ». Afin de pouvoir déterminer la valeur des l'attribut $Value_p$ et par la suite celle de l'attribut $Polarity$, un typage de la réponse permettant de savoir si elle équivaut à un *yes* ou à un *no* est nécessaire. Ce typage est effectué par une simple vérification de la présence d'un «yes», d'un «no» ou d'un de leurs synonymes dans les cinq premiers mots de l'énoncé.

Une fois déterminé le type de la réponse (*yes*, *no*), le système lance le calcul de la valeur de l'attribut $Value_P$. Pour calculer cette valeur, les règles présentées ci-dessus s'appuient sur le type de la réponse et la valeur de l'attribut $negation$ définie au cours de l'analyse de l'énoncé de l'agent. Ainsi, lorsque la réponse équivaut à un *yes* : si $negation = false$ alors $value_P = p$, si $negation = true$ alors $value_P = \neg p$. En revanche, lorsque la réponse équivaut à un *no*, si $negation = false$ alors $value_P = \neg p$, si $negation = true$ alors $value_P = p$. A partir de la valeur attribuée à $value_P$, la valeur de $Polarity$ peut être définie. Lorsque p est nié, l'attribut est $polarity$ prend la valeur inverse de celle de la fiche de l'agent : $if value_P = \neg p, alors$: $polarity_{user} = reverse(polarity_{agent})$. La valeur sera en revanche conservée lorsque p est défini comme vrai par la réponse de l'utilisateur : $if Value_P = p, alors polarity_{user} = polarity_{agent}$.

4.2.2 Détection d'attitudes pleinement formulées

Le processus de détection des attitudes pleinement formulées et référant à des *likes* ou *dislikes* est ici le même que celui appliqué pour la détection d'attitudes exprimées par l'agent, à l'exception que les expressions d'attitudes recherchées ne prennent pas de forme interrogative. Là encore, le processus s'applique en trois phases : lexicale, syntagmatique et phrastique. L'ensemble des patrons utilisés pour ces dernières phases correspondent aux patrons notés *affirm* dans les figures 2 et 3. Concernant la source, dans la mesure où le système ne cherche à détecter, dans l'énoncé de l'utilisateur, que des attitudes dont ce dernier est la source, une expression d'attitude n'est considérée comme pertinente que lorsque sa source correspond à un pronom de première personne ($Src(usr) \rightarrow$ "I"|"me"). Au terme de l'analyse de l'énoncé de l'utilisateur, le système fournit en sortie pour chaque attitude pertinente détectée – exprimées soit par confirmation-infirmation soit par formulation pleine – une structure de traits $source = user$, $polarity = \{neg, pos\}$, $targetType = \{user, agent, other, unknown\}$.

5 Première évaluation du système

5.1 Campagne d'annotation sur Mechanical Turk

Protocole Afin d'évaluer notre système, nous avons mené une campagne d'annotation via la plate-forme Amazon Mechanical Turk. 60 sous-ensembles du corpus Semaine, chacun composé de 10 paires d'énoncés (un énoncé de l'agent et la réponse de l'utilisateur), ont été annotés par 240 annotateurs anglophones natifs (4 pour chaque sous-ensemble du corpus). Pour chaque paire présentée aux annotateurs, une série de questions leur sont posées visant à vérifier la présence d'une expression référant à un *like* ou *dislike* de l'utilisateur. Tandis que la première question interroge sur la présence d'une telle expression, la seconde permet d'en spécifier le nombre d'occurrences dans la paire sélectionnée. Si l'annotateur détecte plusieurs expressions, les questions suivantes sont posées pour chaque d'elles. La troisième question interroge ensuite la

nature de la cible : seules quatre catégories – celles détectables par le système – sont proposées. La dernière question concerne la polarité des expressions détectées par l'annotateur (positive ou négative).

Mesures d'accord inter-annotateurs Concernant la réponse à la première question, nous mesurons le taux d'accord entre annotateurs sur la présence d'au moins une expression de référant à un *like* ou un *dislike*, grâce au Kappa de Fleiss (Fleiss, 1971) (Tableau 1). Nous mesurons la cohérence entre annotateurs concernant le nombre d'expressions détectées dans une paire en calculant le coefficient alpha de Cronbach (Cronbach, 1951). Pour la polarité et le type de cible, nous sélectionnons les paires où au moins deux annotateurs sont d'accord sur la présence d'une expression de *like* ou de *dislike* et nous considérons uniquement les annotations fournies par ces annotateurs. Nous mesurons ensuite le pourcentage d'accord. En fait, suite à la sélection des données montrant un consensus, nous obtenons un sous-ensemble d'annotations avec un taux non-fixe d'annotateurs : une simple mesure de pourcentage d'accord nous a donc semblé appropriée. Concernant la polarité, 41% des sous-corpus ont un pourcentage d'accord compris entre 50% et 75%, et 52% des sous-corpus ont un pourcentage d'accord supérieur à 75%. Concernant la catégorisation de la cible, 61% des sous-corpus ont un pourcentage d'accord supérieur à 50%.

	Fleiss' Kappa	**Cronbach's alpha**
Max	0.79	0.90
Min	0.00	0.29
Median	0.32	0.72
Average	0.25	0.59

TABLE 1 – Scores du Kappa de Fleiss et coefficients alpha de Cronbach pour chaque sous-ensemble annoté

5.2 Évaluation du système

Des 600 paires annotées, nous avons supprimé les paires où aucun consensus – une majorité d'annotateurs appliquant une même annotation – a été trouvé. 503 paires ont ainsi été conservées. Nous considérons la référence et notre système comme des annotateurs distincts. En effet, en analyse de sentiments, une vérité terrain construite à partir d'indices objectifs est difficile à obtenir : les observations faites par les annotateurs restent des interprétations subjectives. Sur la base de ce postulat, nous avons choisi d'évaluer l'accord entre notre système et la référence. Pour mesurer l'accord entre notre système et cette référence ainsi constituée, nous avons appliqué les trois mesures utilisées précédemment : kappa de Fleiss, pour la présence d'une expression référant à un *like* ou un *dislike* et la polarité, coefficient alpha de Cronbach, pour le nombre d'expression et pourcentage d'accord, pour le type de cible. L'accord entre les sorties du système et la référence est substantiel concernant la présence d'au moins une expression ($k = 0.61$). Le nombre d'expressions est lui-aussi correctement détecté par le système, le coefficient équivalant 0.67 (il est souvent admis que la valeur d'accord minimale acceptable est de 0.60). Le kappa de Fleiss obtenu pour la polarité est lui-aussi encourageant puisqu'il est égal à 0.844. Concernant le type de cible, on obtient un pourcentage d'accord de 53%. Le désaccord est souvent lié à un confusion entre les catégories *unknown* et *other*.

6 Conclusion et perspectives

Dans cet article, nous avons proposé une approche des attitudes – affects, appréciations, jugements – dans le cadre des conversations humain-agent. Afin de pouvoir s'adapter aux enjeux et contraintes de ce contexte spécifique, nous avons fourni un modèle d'annotation dédié à l'étude de ces expressions telles qu'elles se réalisent dans le contexte d'une parole conversationnelle. Appliqué au corpus Semaine, le modèle révèle un lien entre les attitudes exprimées par l'utilisateur et les énoncés de l'agent qui le précédent, tant le plan pragmatique que sur celui du contenu propositionnel. Sur la base de cette étude, nous proposons une première version de notre système se concentrant sur la détection des attitudes pouvant exprimer un *like* ou *dislike*. L'évaluation de cette première version montre des résultats encourageants. Après une analyse des résultats, il semble que les causes de désaccord entre le système et la référence soient en partie liées au manque de contexte des paires d'énoncés annotées. En effet, pour cette première version, le système ne considère que des paires d'énoncés. Si cela a permis de détecter un grand nombre d'expressions, il est néanmoins probable que cela ait créé une confusion interprétative dans certains cas, tant du côté du système que de celui des annotateurs humains. Cette confusion

peut ainsi être cause de désaccords. Dans l'exemple suivant, Agent : « good. Ah good » (« bien. Ah bien ») – Utilisateur : « my favorite emotion » (« mon émotion favorite »), si la source peut être identifiée, la détection de la cible est impossible pour une analyse se concentrant sur cette seule paire d'énoncés.

Plusieurs éléments peuvent être pris en compte pour améliorer ce résultat mais aussi les performances du système. Tout d'abord, sur le plan de l'organisation conversationnelle, il est important que le système puisse considérer un plus large contexte d'analyse. Tout d'abord, pour chaque tour de parole utilisateur analysé, il est important de pouvoir avoir accès à l'ensemble des tours de parole – agent et utilisateur – énoncés en amont de la conversation. Les règles sémantiques utilisées par le système devront modéliser la manière dont l'utilisateur et l'agent collaborent sur plusieurs tours de parole successifs à l'expression d'attitude. Ensuite, afin de pouvoir correctement gérer le prise en compte de ce contexte plus large, il sera nécessaire que chaque analyse effectuée sur un tour de parole spécifique soit rendue accessible par les analyses des tours de parole ultérieurs. Cette approche permettra une meilleure détection des expressions d'attitude mais aussi une meilleure identification de leur cible – notamment dans le cas de problème de référence anaphorique.

Remerciements

L'auteur souhaite remercier l'équipe Greta pour sa contribution à la plateforme Greta-VIB. Ce travail a été développé dans le cadre du Labex SMART (ANR-11-LABX-65) supporté par l'ANR au sein du programme Investissements d'Avenir (ANR-11-IDEX-0004-02).

Références

BLOOM K., GARG N. & ARGAMON S. (2007). Extracting appraisal expressions. *HLT-NAACL*, p. 165–192.

BRECK E., CHOI Y. & CARDIE C. (2007). Identifying expressions of opinion in context. In S. S., M. H. & B. R. K., Eds., *International Joint Conference On Artifical Intelligence*, p. 2683–2688, San Francisco, CA : Morgan KoffMann Publishers.

CHOI Y., CARDIE C., RILOFF E. & PATWARDHAN S. (2005). Identifying sources of opinions with conditional random fields and extraction patterns. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, p. 355–362, Stroudsburg, PA, USA : Association for Computational Linguistics.

CLARK H. H. & SCHAEFER E. F. (1989). Contributing to discourse. *Cognitive Science*, **13**, 259–294.

CRONBACH L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**(3), 297–334.

ESULI A. & SEBASTIANI F. (2005). In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, CIKM '05, p. 617–624, New York, NY, USA : ACM.

FLEISS J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, **76**(5), 378–382.

HATZIVASSILOGLOU V. & MCKEOWN K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, ACL '98, p. 174–181, Stroudsburg, PA, USA : Association for Computational Linguistics.

HEIDER F. (1958). *The psychology of interpersonal relations*. Lawrence Erlbaum associates Inc.

HU M. & LIU B. (2004). Mining opinion features in customer reviews. In *Proceedings of the 19th National Conference on Artifical Intelligence*, AAAI'04, p. 755–760 : AAAI Press.

ISHIZUKA M. (2012). Textual affect sensing and affect communication. In *IEEE Transaction 11th International Conference on Cognitive Informatics and Cognitive Computing*, p. 2–3.

IZARD C. E. (1977). *Human Emotions*. New York, USA : Plenum Press.

LANGLET C. & CLAVEL C. (2015). Improving social relationships in face-to-face human-agent interactions : when the agent wants to know user's likes and dislikes. In *The Association for Computer Linguistics*, Beijin, China. to appear.

LE QUERLER N. (1996). *Typologie des modalités*. Presse Universitaire de Caen.

MARTIN J. R. & WHITE P. R. (2005). *The Language of Evaluation. Appraisal in English*. London and New York : Macmillan Basingstoke.

MCKEOWN G., VALSTAR M., COWIE R., PANTIC M. & SCHRODER M. (2011). The semaine database : Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *IEEE Transactions on Affective Computing*, **3**(1), 5–17.

MOILANEN K. & PULMAN S. (2007). Sentiment composition. In *Proceedings of Recent Advances in Natural Language Processing (RANLP 2007)*, p. 378–382.

NASUKAWA T. & YI J. (2003). Sentiment analysis : Capturing favorability using natural language processing. In *Proceedings of the 2Nd International Conference on Knowledge Capture*, K-CAP '03, p. 70–77, New York, NY, USA : ACM.

NEVIAROUSKAYA A., PRENDINGER H. & ISHIZUKA M. (2007). Textual affect sensing for sociable and expressive online communication. In *Proceedings of the 2Nd International Conference on Affective Computing and Intelligent Interaction*, ACII '07, p. 218–229, Berlin, Heidelberg : Springer-Verlag.

NEVIAROUSKAYA A., PRENDINGER H. & ISHIZUKA M. (2010a). Emoheart : Conveying emotions in second life based on affect sensing from text. *Adv. in Hum.-Comp. Int.*, **2010**.

NEVIAROUSKAYA A., PRENDINGER H. & ISHIZUKA M. (2010b). Recognition of affect, judgment, and appreciation in text. In *Proceedings of the 23rd International Conference on Computational Linguistics*, COLING '10, p. 806–814, Stroudsburg, PA, USA : Association for Computational Linguistics.

ORTONY A., CLORE G. & COLLINS A. (1990). *The Cognitive Structure of Emotions*. Cambridge, University Press.

OSGOOD C., MAI W. H. & MIRON M. S. (1975). *Cross-cultural Universals of Affective Meaning*. Urbana : University of Illinois Press.

PANG B. & LEE L. (2004). A sentimental education : Sentiment analysis using subjectivity. In *Proceedings of ACL*, p. 271–278.

PAUMIER S. (2015). *Unitex user manual*. Université de Paris-Est Marne-la-Vallée.

POGGI I., PELACHAUD C., DE ROSIS F., CAROFIGLIO V. & DE CAROLIS B. (2005). Greta. a believable embodied conversational agent. In *Multimodal intelligent information presentation*, p. 3–25. Springer.

QUIRK R. (1985). *A Comprehensive grammar of the English language*. General Grammar Series. Longman.

RILOFF E. (1996). An empirical study of automated dictionary construction for information extraction in three domains. *Artificial Intelligence*, **85**, 101–134.

RILOFF E. & WIEBE J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, EMNLP '03, p. 105–112, Stroudsburg, PA, USA : Association for Computational Linguistics.

SEARLE J. R. (1976). A classification of illocutionary acts. *Language in society*, **5**(01), 1–23.

SHAIKH M., PRENDINGER H. & ISHIZUKA M. (2009). A linguistic interpretation of the occ emotion model for affect sensing from text. In *Affective Information Processing*, p. 378–382 : Springer London.

SMITH C., CROOK N., DOBNIK S. & CHARLTON D. (2011). Interaction strategies for an affective conversational agent. In *Presence : Teleoperators and Virtual Environments*, volume 20, p. 395–411 : MIT Press.

TURNEY P. D. (2002). Thumbs up or thumbs down ? : Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, p. 417–424, Stroudsburg, PA, USA : Association for Computational Linguistics.

VALITUTTI R. (2004). Wordnet-affect : an affective extension of wordnet. In *In Proceedings of the 4th International Conference on Language Resources and Evaluation*, p. 1083–1086.

WHITELAW C., GARG N. & ARGAMON S. (2005). Using appraisal taxonomies for sentiment analysis. *Proceedings of CIKM-05, the ACM SIGIR Conference on Information and Knowledge Management*.

WIDLÖCHER A. & MATHET Y. (2012). The glozz platform : A corpus annotation and mining tool. In *Proceedings of the 2012 ACM Symposium on Document Engineering*, p. 171–180, Paris, France.

WIEBE J. & RILOFF E. (2005). Creating subjective and objective sentence classifiers from unannotated texts. In *Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing*, CICLing'05, p. 486–497, Berlin, Heidelberg : Springer-Verlag.

WILSON T., WIEBE J. & HOFFMANN P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, p. 347–354, Stroudsburg, PA, USA : Association for Computational Linguistics.

WILSON T., WIEBE J. & HWA R. (2004). Just how mad are you ? finding strong and weak opinion clauses. In *Proceedings of the 19th National Conference on Artifical Intelligence*, AAAI'04, p. 761–767 : AAAI Press.

YANG B. & CARDIE C. (2013). Joint inference for fine-grained opinion extraction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, p. 1640–1649, Sofia, Bulgaria : Association for Computational Linguistics.

YILDIRIM S., NARAYANAN S. & POTAMIANOS A. (2011). Detecting emotional state of a child in a conversational computer game. *Computer Speech & Language*, **25**(1), 29–44.