

Avant-propos

Cet ouvrage porte sur l'analyse des opinions dans les interactions humain-humain et humain-agent. Il présente des méthodes issues de l'intelligence artificielle (apprentissage de modèles de comportements socio-émotionnels en combinant des méthodes symboliques et des méthodes issues de l'apprentissage automatique) et de l'informatique affective (analyse et synthèse de signaux socio-émotionnels).

Cet ouvrage reprend les travaux présentés dans mon « Habilitation à diriger des recherches (HDR) » soutenue le 29 mai 2017. De ce fait, il propose un point de vue personnel et n'a pas vocation à fournir un panorama exhaustif du domaine. Ce point de vue est empreint de la spécificité de mon parcours de chercheuse mené dans des contextes académiques et industriels.

Après mon doctorat sur la reconnaissance acoustique des émotions [CLA 07], j'ai poursuivi mes recherches en milieu R&D en tant que chercheuse au sein du centre de recherche *Thales Research and Technology*. J'ai élargi mon périmètre de recherche de l'analyse acoustique au traitement automatique des langues lors des travaux que j'ai menés sur l'analyse d'opinions pour des applications de gestion de la relation client, en tant que chercheuse à EDF Lab.

Je suis, depuis 2013, enseignante et chercheuse au sein du LTCI (Laboratoire traitement et communication et de l'information) de Telecom-ParisTech et mes travaux de recherche actuels poursuivent les travaux initiés à EDF Lab sur l'analyse des opinions et des sentiments dans les interactions écrites et orales et s'étendent au traitement des émotions et des attitudes sociales dans les interactions humain-agent. Plus précisément, mes travaux ont pour vocation d'examiner la modalité de parole (contenu verbal et prosodique) dans le contexte des interactions de l'humain avec un agent conversationnel animé et pour le développement d'agents socialement et émotionnellement compétents.

Je présente dans cet ouvrage les différentes études que j'ai menées sur ces sujets en distinguant les deux axes suivants :

- l'analyse des opinions dans les interactions humain-humain et humain-agent ;
- les stratégies d'interactions socio-émotionnelles et la génération de comportements socio-émotionnels en interaction humain-agent.

Les recherches qui y sont présentées se situent dans les domaines du traitement automatique du langage naturel, de l'apprentissage automatique appliqué, de l'informatique affective et de l'interaction humain-machine/robotique sociale. Elles interviennent dans un contexte pluridisciplinaire en étroite collaboration avec les domaines de recherche tels que la psychologie, la sociologie, et la linguistique. Ce contexte pluridisciplinaire s'est avéré fondamental, compte tenu des phénomènes étudiés allant des expressions d'opinions aux attitudes sociales.

Ce livre présente une synthèse de travaux majoritairement menés par les stagiaires, doctorants et post-doctorants que j'ai encadrés (par ordre chronologique) :

- stagiaires : Charlotte Danesi, Camille Dutrey, Rachel Bawden et Jessica Durand ;
- doctorants : Rémi Lavalley, Camille Dutrey, Caroline Langlet, Thomas Janssoone, Irina Maslowski, Valentin Barrière ;
- post-doctorants : Sabrina Campano, Guillaume Dubuisson-Duplessis, Brian Ravenet, Atef Ben-Youssef.

Ces travaux sont également le fruit de nombreuses collaborations – réalisées à travers, entre autres, le coencadrement de ces doctorants et post-doctorants.

Les travaux que je rassemble ici sur les méthodes d'analyse d'opinions dans les interactions ont été menés en collaboration avec Anne-Laure Guénet, Delphine Lagarde, Anne Peradotto et Alina Stoica Beck à EDF et avec Patrice Bellot et Marc El-Bèze (Laboratoire d'informatique d'Avignon).

Les travaux sur les données orales de centres d'appels et notamment sur l'analyse des disfluences présentés dans les chapitres 1 et 2 sont nés d'une collaboration avec le LIMSI-CNRS, le laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur de l'université Paris 11 (Sophie Rosset et Ioana Vasilescu) et avec le LPP - Laboratoire de phonétique et phonologie, Paris 3 (Martine Adda-Decker).

Les recherches menées sur l'interaction humain-agent avec mon arrivée au LTCI (présentés en chapitres 3, 4, 5) ont beaucoup bénéficié de la longue et fructueuse collaboration que j'ai pu avoir avec Catherine Pelachaud. Les travaux menés sur les champs aléatoires conditionnels présentés dans le chapitre 3 ont pu voir le jour grâce à l'expertise du LTCI dans le cadre de ma collaboration avec Slim Essid et ceux sur la

génération d'expressions faciales évoqués dans le chapitre 5 de l'expertise de Kévin Bailly (ISIR). Les travaux sur la génération de comportements prosodiques et sur les mesures d'alignement présentées en chapitres 4 et 5 ont bénéficié des échanges scientifiques que j'ai pu avoir avec Frédéric Landragin du Lattice.

Enfin, je tiens à remercier Björn Schuller, Dirk Heylen et Nicolas Sabouret, les rapporteurs de mon jury de mon « Habilitation à diriger des recherches », ainsi que Frédéric Bechet, Mohamed Chetouani et Patrick Paroubek, les examinateurs, pour leurs questions et les discussions motivantes, source d'inspiration pour la suite de mes recherches.

De la fouille d'opinions aux interactions humain-agent

L'analyse automatique des opinions dans les interactions est un domaine de recherche en plein essor, un essor poussé d'une part par les enjeux applicatifs du domaine et d'autre part par la multiplication des plateformes d'expression des citoyens et des médias sur le web. Ce foisonnement favorise le renouvellement des possibilités de critique et d'action *via* Internet [CAR 13] et les données sociales disponibles sont de plus en plus nombreuses et variées. L'enjeu pour le domaine du traitement automatique du langage naturel est l'analyse automatique de ces expressions d'opinion afin de participer à l'analyse des tendances sociétales. Les applications sont nombreuses : l'analyse des opinions des citoyens sur les candidats lors des élections, l'analyse des opinions des internautes sur un produit (également connue sous le terme d'e-réputation), l'identification des clients cibles pour des systèmes de recommandation, l'évaluation du succès d'une campagne de communication, etc.

En parallèle de ce phénomène lié au web social, la robotique sociale, et plus largement l'interaction humain-agent, est un domaine de l'interaction humain-machine pour lesquels l'intégration de l'analyse d'opinions dans l'interaction entre l'humain et l'agent virtuel présente un fort potentiel. Par exemple, les robots compagnons ont pour vocation d'apporter à leurs utilisateurs à la fois aide (notamment dans le cadre du marché de l'assistance et de la dépendance) et distraction (dans le cadre du marché du divertissement). Dans ce contexte, la connaissance de l'utilisateur et donc son profilage est une information cruciale pour l'établissement d'un lien social avec le robot. À partir de ce profil (et en particulier ici des préférences de l'utilisateur), le robot compagnon peut par exemple choisir les sujets à aborder dans ses interactions successives avec l'humain, ou encore lui recommander des produits, musiques, divertissements susceptibles de l'intéresser.

Les applications du domaine et les enjeux associés sont nombreux avec le développement des agents virtuels sur les sites Internet pour la gestion de la relation client et l'émergence d'assistants virtuels (Alexa d'Amazon, Siri d'Apple, Cortana de Microsoft, etc.). Si les assistants virtuels proposés par ces entreprises sont déjà commercialisés, la prise en compte de la composante sociale de l'interaction reste un enjeu crucial pour la fluidité et le naturel de l'interaction. Une autre application de la prise en compte des comportements socio-émotionnels dans l'interaction-humain-agent est celle des *Serious Games* pour lesquelles l'utilisateur peut s'entraîner avec un agent virtuel à faire face à différentes situations. Par exemple, dans [YOU 15], l'utilisateur peut s'entraîner à améliorer son comportement social lors d'entretiens d'embauches virtuels.

Les recherches sur l'analyse des opinions présentées dans cet ouvrage se positionnent dans ces deux situations d'interaction :

- 1) les interactions humain-humain collectées sur le web et à partir des données d'entreprise ;
- 2) les interactions humain-agent (agents conversationnels animés, robots).

Le domaine de l'analyse et de la détection des opinions est un domaine dans lequel la communauté en apprentissage automatique s'est lancé avec engouement, restant parfois déconnectée des domaines du traitement automatique des langues et de l'*Affective Computing*. Cet ouvrage s'appuie sur les recherches menées dans ces trois communautés (*Affective Computing*, apprentissage automatique et traitement automatique des langues). Il présente des méthodes de détection incluant à la fois des règles logico-sémantiques et des méthodes d'apprentissage automatique, choisissant les méthodes les plus adaptées en fonction de la problématique scientifique et de sa maturité. Les méthodes à base de règles et de connaissances constituent en effet une première étape nécessaire pour circonscrire des problèmes scientifiques peu explorés. À titre d'exemple, nous pouvons citer les règles logico-sémantiques que nous avons définies pour le développement d'un premier système de détection des opinions de l'utilisateur ancré dans le contexte de l'interaction humain-agent, car ce sujet n'avait pas encore été abordé dans la littérature.

Les recherches présentées s'articulent autour de questions de recherche communes aux différents chapitres. Dans ce chapitre d'introduction, nous exposons, pour chaque axe de recherche, son contexte scientifique et les questions de recherches associées :

- la première question de recherche porte sur les modèles théoriques d'opinions (Q1 présentée en section I.1) ;
- la seconde question de recherche porte sur les modèles computationnels d'opinions (Q2 présentée en section I.2) ;

– la troisième et la quatrième portent sur la génération d’agents conversationnels socio-affectifs (Q3 et Q4 en section I.3).

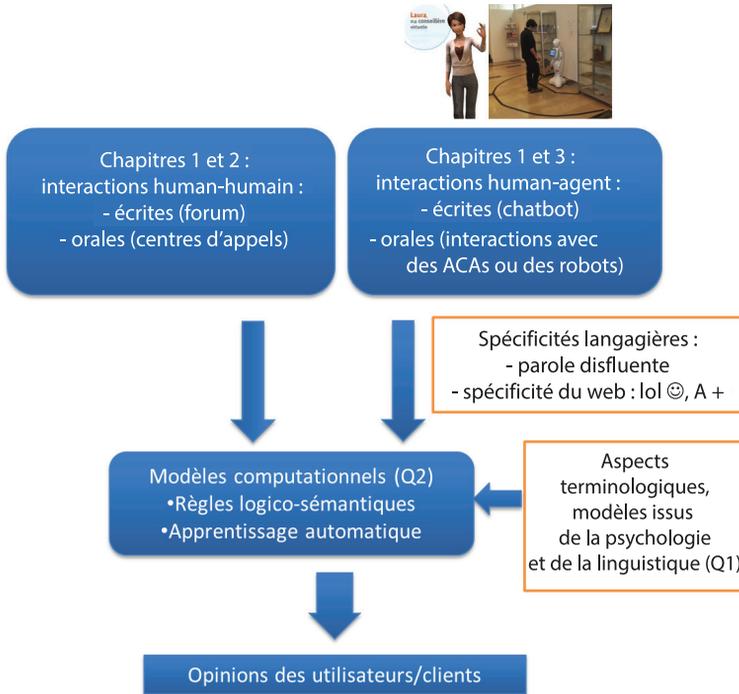


Figure I.1. *L'analyse des opinions dans les interactions telle que présentée dans ce livre*

Ces questions de recherches vont nous permettre, de faire ressortir les fils directeurs et les articulations existantes entre les chapitres. Ces questions de recherche permettent en effet d’articuler les deux volets de recherche présentés dans les chapitres de cet ouvrage :

1) le développement de systèmes de détection d’opinions dédiés à l’analyse d’interactions (écrites et orales¹), que ce soit des interactions humain-humain (chapitre 2) ou humain-agent (chapitre 3), comme représenté sur la figure I.1 ;

2) le développement d’agents virtuels capables d’exprimer des opinions et plus largement des comportements socio-émotionnels (chapitres 4 et 5), comme représenté sur la figure I.2.

1. C’est-à-dire concernant la détection des opinions contenues dans les textes et dans les transcriptions de la parole.

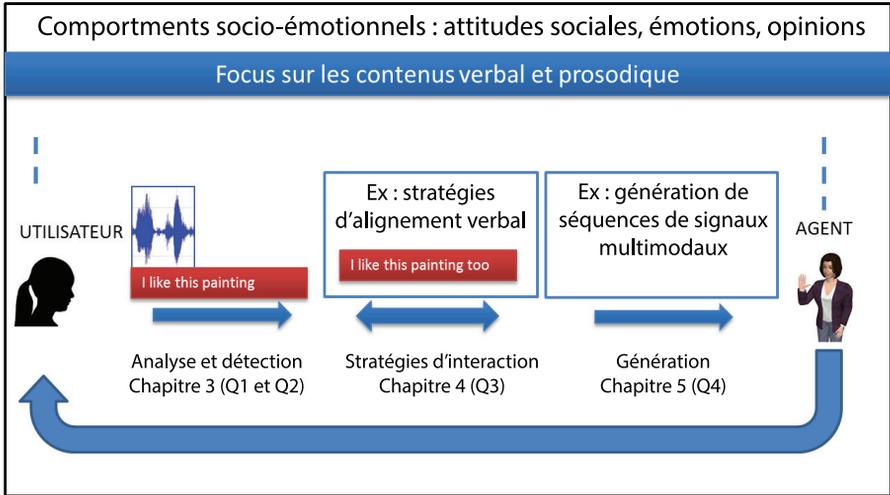


Figure 1.2. *L'interaction humain-agent et les comportements socio-émotionnels : les trois volets présentés dans cet ouvrage dans les chapitres 4 et 5 ainsi que les questions de recherche Q1, Q2, Q3 et Q4 associées*

1.1. Terminologies et modèles théoriques des opinions

Le problème de la détection des opinions est souvent ramené à un problème de classification positif/négatif. La question de la définition du phénomène et de ses différences avec les émotions ou les sentiments est une question importante et le choix du phénomène à étudier va différer en fonction de la problématique scientifique soulevée. Nous présentons dans cette section les travaux issus des différentes communautés et qui se sont attelés à préciser les aspects terminologiques et les théories sous-jacentes au phénomène d'opinion.

1.1.1. Terminologies qui se recouvrent en fonction des communautés

Émotion, opinion, sentiment, humeur, attitude, positionnement interpersonnel, trait de personnalité, affect, jugement, appréciation, argumentation, etc. : la pluridisciplinarité du domaine a entraîné d'un côté l'utilisation de terminologies différentes pour désigner des phénomènes similaires et d'un autre côté, l'utilisation d'une même terminologie pour désigner des phénomènes différents. La communauté de l'*opinion mining* a tendance à utiliser des termes comme opinion, sentiment et affect qui réfèrent à des phénomènes différents. Cependant, les travaux existants présentent rarement de définition approfondie des phénomènes associés à ces termes.

La communauté de l'interaction humain-agent s'appuie, elle, sur des théories issues de la psychologie mais pour modéliser des phénomènes plutôt centrés sur l'émotion, l'affect et l'humeur et plus récemment sur les interactions sociales. Nous fournissons dans [CLA 16b] un état de l'art sur ces terminologies et les modèles théoriques dont nous résumons ici les principales références.

Scherer [SCH 05] propose une distinction des différents phénomènes. Il distingue notamment l'émotion des attitudes. L'émotion est définie comme un phénomène bref incluant une réaction physiologique suite à l'évaluation d'un événement majeur (*stimulus*) (comme c'est le cas de la peur, la tristesse, la joie, la colère). Les attitudes sont définies comme des dispositions envers des objets et des personnes (comme c'est le cas des préférences). Il définit également le positionnement interpersonnel (*interpersonal stance*) également appelé attitudes sociales comme un positionnement affectif vis-à-vis d'une autre personne dans une interaction, par exemple être poli, chaleureux ou méprisant.

Des études plus spécifiques au contenu verbal ont été menées par des linguistes. Martin et White [MAR 05] préfèrent aux termes sentiments ou opinions, comme Scherer, le terme d'attitudes. Ils définissent une attitude comme liée à des sentiments (*feelings*) et incluant les réactions émotionnelles, les jugements de comportement et l'évaluation des objets (« concerned with our feelings, including emotional reactions, judgements of behavior and evaluation of things » [MAR 05, p. 35]).

Ils distinguent trois types d'attitudes :

- les affects (réaction personnelle relative à un état émotionnel) ;
- les jugements (le fait d'assigner de qualités – par exemple la tenacité – à des individus en fonction de principes normatifs) ;
- les appréciations (l'évaluation d'un objet d'un produit ou d'un *process*).

[MUN 14] propose également pour la communauté *opinion mining* des définitions illustrées d'exemples textuels des phénomènes de sentiment et d'opinion. Les affects y sont définis comme précédant les émotions avant la prise de conscience de leur ressenti. Par conséquent, les affects n'ont pas d'expression dans le langage selon cet article. Les distinctions entre *émotions* et *sentiments* et entre *opinions* et *sentiments* y sont aussi précisées :

- les émotions diffèrent des sentiments de par leur durée (les émotions sont des phénomènes plus brefs) et de par la présence d'une cible (les émotions ne sont pas toujours reliées à un objet) ;
- les opinions sont des interprétations personnelles d'une information et ne sont pas nécessairement chargées émotionnellement comme les sentiments.

Les études présentées dans ce livre s'appuient sur le concept d'*attitude* tel que défini par Martin et White. La définition proposée permet en effet d'englober l'ensemble des phénomènes relatifs aux opinions en proposant des sous-catégories permettant de mieux circonscrire le phénomène d'intérêt en fonction du contexte scientifique.

1.1.2. Trois modèles théoriques issus de deux communautés

L'utilisation des modèles théoriques sous-jacents aux systèmes d'analyse des opinions et des émotions diffèrent également selon les communautés (*opinion mining* ou interaction humain-agent) et les applications. Nous avons recensé dans [CLA 16b] trois grandes familles de modèles théoriques pour définir les phénomènes liés aux opinions :

- les modèles dimensionnels ;
- les modèles catégoriels ;
- les modèles reposant sur la théorie de l'évaluation.

1.1.2.1. Modèles dimensionnels

En ce qui concerne la communauté de l'*opinion mining*, la détection de polarité (positif *versus* négatif) et d'intensité sont les tâches les plus fréquentes [WOL 13], [OSH 09]. La polarité et l'intensité sont deux dimensions permettant de décrire les opinions qui peuvent être reliées aux théories s'appuyant sur un **modèle dimensionnel** [RUS 80] des opinions/émotions. Ce mode de description représente les phénomènes socio-émotionnels sur des axes abstraits (par exemple sur les axes valence/activation).

La détection de polarité, en particulier, permet de simplifier le problème de l'analyse d'opinions en segmentant l'axe de polarité en deux ou trois classes (l'avis exprimé majoritairement dans un texte, est-il positif, négatif, ou neutre² ?) et est utilisée, par exemple, pour l'analyse des opinions associées à une marque (e-réputation) ou l'analyse des critiques de films.

Par exemple, le défi fouille de textes Deft07 [GRO 07] portait sur l'attribution d'une opinion (positive, négative ou éventuellement neutre) sur des corpus de critiques de livres, spectacles, jeux vidéo, relectures d'articles scientifiques et débats parlementaires.

L'analyse de la polarité des opinions ou de la valence des émotions est également utilisée dans le domaine de l'interaction humain-agent pour la gestion des émotions négatives dans l'interaction [SMI 11].

2. Selon le contexte, « neutre » peut signifier soit qu'aucun avis n'est exprimé dans le texte, soit que l'on y trouve à la fois des avis positifs et des avis négatifs.

1.1.2.2. *Modèles catégoriels*

D'autres travaux (par exemple [PER 13]) s'appuyant sur des **modèles catégoriels** issus de la psychologie [EKM 99, IZA 71, PIC 00, PLU 03, WHI 89] concernent la détection de catégories d'opinions ou d'émotions dans les données textuelles. L'approche catégorielle consiste en la dénomination des phénomènes socio-émotionnels par des items lexicaux adaptés et prédéfinis.

C'est la manière la plus intuitive pour décrire des phénomènes spécifiques, en utilisant des catégories issues du langage courant [CLA 07]. La définition de catégories consiste à tracer des frontières absolues dans l'espace perceptif. Chaque catégorie correspond alors à un prototype [KLE 90] auquel peut se rattacher d'autres manifestations similaires.

L'établissement de telles frontières va fortement dépendre des données considérées. Dans le cas des corpus entièrement simulés : on part de prototype prédéfini que l'on cherche à illustrer. Toutes les manifestations émotionnelles contenues dans les corpus doivent donc fortement converger vers ce prototype. Pour les corpus spontanés, il s'agit de regrouper des expressions non contrôlées autour de prototype abstrait. La complexité de ce regroupement est accrue par la diversité des contextes d'émergence des émotions en parole spontanée.

Les classes considérées en analyse d'opinions dépendent ainsi fortement du contexte applicatif et des données considérées. Par exemple :

- la détection d'accord ou de désaccord [GAL 04] dans des enregistrements de réunions ;
- la détection de messages insultants sur le web [SPE 97] ;
- la détection de la subjectivité [TSY 12] ;
- la détection de la frustration lors de la conduite [BOR 10] ou pour des systèmes de soutien scolaire [LIT 06] ou pour des jeux sur ordinateur pour enfants [YIL 11] ;
- la représentation de l'émotion détectée dans le texte à travers des avatars [NEV 10b, ZHA 08].

1.1.2.3. *Modèles liés à la théorie de l'évaluation*

Des modèles plus riches, encore peu utilisés par la communauté de *opinion mining*, mais qui ont fait leurs preuves dans la communauté de l'interaction humain-agent, sont les **modèles liés à la théorie de l'évaluation**.

Le plus populaire de ces modèles au sein des deux communautés est le modèle d'Orthony Clore et Collins (OCC) qui modélise la structure cognitive des émotions. Il est utilisé par la communauté de *opinion mining* pour analyser les affects dans les

textes (*textual affect sensing* [SHA 09]) et est particulièrement populaire dans la communauté agent pour la génération des comportements émotionnels de l'agent, classifiant les événements, les objets et les actions afin de définir l'émotion à faire exprimer par l'agent [VAL 09].

Mais chaque communauté a aussi sa propre théorie de l'évaluation. En effet, il existe du côté de la communauté de l'*opinion mining* des travaux reposant sur une autre théorie de l'évaluation fournissant une définition de ce qui est appelé *attitude* ou évaluation dans le langage (Martin et White [MAR 05]). Cette théorie permet de représenter une opinion (ici appelée attitude) comme une évaluation par une source (par exemple la personne qui s'exprime) d'une cible (par exemple un service ou un produit) [BLO 07].

Du côté de la communauté agent, des alternatives au modèle OCC ont été utilisées : le modèle d'*appraisal* de Scherer qui décompose le processus d'évaluation en différentes étapes (par exemple : évaluation de la nouveauté) ou le modèle d'*appraisal* dynamique EMA [MAR 09].

Autre modèle théorique utilisé par une communauté connexe à celle de l'*opinion mining* : les modèles d'argumentation. Une formalisation sous forme de graphe des modèles d'argumentation issus de la philosophie [TOU 03a] a été proposée par [CAB 13] pour faire de l'*argument mining* dans les débats des réseaux sociaux. Cette formalisation présente l'avantage d'extraire des structures entre des opinions en liant par exemple des opinions correspondant à des réfutations (*rebuttal*) d'affirmations (*claim*).

Un premier effort de recensement des différentes terminologies et modèles théoriques avait été réalisé au sein du W3C avec la mise en place du EmotionML³ (*Emotion Markup Language*). L'objectif était de définir un langage commun pour l'annotation des émotions. Cet effort est actuellement étendu à la description des sentiments dans des sources de données liées (*Linked Data*) dans le cadre du W3C [SAN 16].

En cohérence avec notre choix de rassembler les phénomènes liés aux opinions autour du concept d'attitude, les travaux présentés dans ce livre s'appuient sur la théorie de l'évaluation dans le langage de Martin et White, car elle fournit une description des réalisations verbales des attitudes, facilitant la formalisation symbolique d'expressions pouvant intégrer un modèle de détection.

1.1.3. Question de recherche et articulation dans l'ouvrage

Dans [CLA 07], nous avons initié cette réflexion sur le modèle théorique à adopter dans l'objectif de construire un modèle computationnel dans un autre contexte, celui de l'analyse acoustique et la détection des émotions de type peur en situation

3. <https://www.w3.org/TR/emotionml/>.

anormale. Dans les travaux de recherche que nous présentons dans ce livre, cette réflexion s'est étendue aux comportements socio-émotionnels, incluant les phénomènes linguistiques liés aux opinions et aux sentiments.

Nous questionnons ici : (Q1) la pertinence des différents modèles théoriques pour la construction d'un modèle computationnel (linguistique et prosodique) en fonction de l'application (analyse des réseaux sociaux, gestion de la relation client, recommandation, interaction humain-agent et robotique sociale).

Cette question de recherche est abordée tout au long de ce livre :

- dans le chapitre 2, une première esquisse de réponse à cette question est proposée par la définition du concept de satisfaction du point de vue du marketing sur des données d'entreprise pour une application de gestion de relation client ;

- dans le chapitre 3, nous approfondissons cette question de la modélisation théorique en nous appuyant sur la théorie de l'*appraisal* de Martin et White [MAR 05] pour construire :

- des modèles de *like* et *dislike* de l'utilisateur dans l'interaction humain-agent, dans l'objectif de créer un profil d'utilisateur ;

- des modèles d'opinions de clients-utilisateurs envers l'interaction avec un *chatbot* ;

- dans le chapitre 4, c'est également le modèle de Martin et White qui est utilisé pour modéliser des appréciations des visiteurs dans un musée et proposer un agent capable de s'aligner sur les appréciations de l'utilisateur. Nous étendons également dans le chapitre 5 notre travail de recherche aux phénomènes liés aux attitudes sociales.

1.2. Modèles computationnels des opinions

La littérature de l'*opinion mining* et de l'analyse du sentiment, c'est-à-dire concernant la détection dans les textes ou dans les transcriptions de la parole, se décompose en trois grands types de méthodes : les méthodes à base de règles, les méthodes statistiques et une version hybride de ces deux types de méthodes.

Une opinion s'exprimant rarement par des phrases simples du type : « ce produit est mauvais », les enjeux pour ces méthodes sont multiples :

- le traitement de la négation et des intensifieurs [MOI 07, NEV 10a, TAB 11] pour pouvoir traiter des expressions telles que « ce film n'est pas vraiment ce que l'on appelle un bon film » ;

– l’identification des cibles et des sources [BLO 07] pour traiter des expressions telles que « je suis satisfait des contacts que j’ai eus avec EDF mais pas de ses services » ;

– la gestion des métaphores : par exemple, l’expression « réchauffement climatique » est connotée plus négativement que « changement climatique » [AHM 11, ZHA 09] et des anaphores⁴ [MOR 12] ;

– la gestion de la comparaison (par exemple « ce film était moins bon que le précédent »), de l’ironie (par exemple « je vous recommande chaudement d’aller voir ce film, pour peu que vous fussiez un adolescent attardé »), les expressions ironiques étant également très fréquentes et difficiles à détecter [REY 14] ;

– la prise en charge des spécificités structurelles et langagières des données (émoticônes, *hashtags* sur les *tweets*, etc.) [RUS 11] ;

– la prise en charge du contexte idiosyncratique (les opinions ne s’expriment pas de la même manière d’une personne à l’autre, en fonction de sa personnalité) ou du contexte social et politique dans lequel l’opinion a été exprimée.

1.2.1. Méthodes à base de règles

Les méthodes à base de règles reposent sur l’occurrence de mots appartenant à des lexiques d’opinions [PEN 01] et sur la mise en place de règles ou de patrons d’extraction linguistiques intégrant ces lexiques et les différents niveaux d’analyse du texte, de l’analyse morpho-syntaxique (forme fléchie, lemme, catégories grammaticales) aux sorties des autres règles [MOI 07, OSH 09].

Avec ce type de méthodes qui intègrent des représentations formelles des énoncés, il est possible de s’attaquer aux problèmes classiques de l’*opinion mining* présentés dans le paragraphe ci-dessus. Par exemple, dans [MOI 07, NEV 10a], les auteurs s’attaquent à la gestion de la négation et des intensifieurs en s’appuyant sur une analyse de dépendance entre constituants avec des règles de propagation ou d’inversion permettant de gérer la polarité à différents niveaux de la structure syntaxique.

Les méthodes à base de règles permettent d’intégrer facilement des ressources et des connaissances [POR 14] et de rendre *computationnels* des modèles théoriques issus de la psychologie. Ainsi, [SHA 09] propose une interprétation du modèle OCC [ORT 90] à travers des règles linguistiques dédiées et [NEV 10a] utilise une approche compositionnelle pour distinguer les trois composantes, affect, jugement, et appréciation, du modèle de Martin et White [MAR 05]. Ces méthodes semblent également

4. La résolution d’anaphore et l’extraction de relation permettent d’identifier la cible de l’opinion lorsque cette cible est indiquée par un pronom personnel.

propices, de par la finesse des modèles construits, à la modélisation de spécificités langagières et du contexte d'interaction.

C'est ce que nous creuserons dans le chapitre 3 en tirant parti de ces méthodes pour la modélisation du contexte d'interaction humain-agent pour aider la détection des opinions de l'utilisateur. Le développement de ces méthodes est cependant très coûteux car elles nécessitent une expertise linguistique et une connaissance approfondie des données à étudier. Les modèles construits sont peu génériques et le passage à de nouvelles données et de nouvelles tâches requiert une adaptation du modèle par un expert.

1.2.2. Méthodes d'apprentissage

Le deuxième type correspond à des méthodes d'apprentissage. Ces méthodes permettent de générer des modèles plus robustes et d'apprendre automatiquement les caractéristiques linguistiques les plus pertinentes pour la classification d'opinions. Les méthodes d'apprentissage supervisé nécessitent la mise en place d'un volume suffisant de données annotées.

La tendance est actuellement à l'utilisation des méthodes de *deep learning* comme les réseaux de neurones récurrents. Ainsi, dans [SOC 13] l'utilisation de ces réseaux nécessite une base de données finement annotée à différents niveaux de la phrase (les différents nœuds de l'arbre syntaxique) afin de fournir la structure nécessaire à l'application d'un modèle récurrent.

De manière générale, la qualité des modèles appris va fortement dépendre de la qualité et la quantité des annotations obtenues. Toute la difficulté va résider dans la définition de protocoles d'annotations permettant de faire converger les annotations de plusieurs annotateurs, sachant que les opinions sont des phénomènes très subjectifs [WIE 05]. Les performances des modèles appris sont également évaluées en utilisant ces annotations comme référence, sachant qu'il est parfois difficile de trancher entre un désaccord entre le système et l'annotateur humain. Par ailleurs, les différentes classes d'opinions sont souvent mal distribuées dans les données avec une classe neutre plus fréquente [CAL 08]. Une réponse à cette difficulté à obtenir une grande quantité de données annotées est l'utilisation de méthodes d'apprentissage semi-supervisées [MAR 13b].

En entrée des méthodes d'apprentissage, les caractéristiques linguistiques utilisées pour représenter un document jouent un rôle important. Elles peuvent faire appel à des lexiques d'opinion, de modificateurs et de mots liés à la négation [KEN 06]. L'émergence de caractéristiques type *word embedding* *word2vec* semble également une tendance intéressante pour l'amélioration des performances des systèmes à base d'apprentissage automatique [POR 15, TAN 14]. À noter également que ces méthodes sont

particulièrement adaptées pour l'analyse multimodale de sentiments et facilitent l'utilisation conjointe des indices linguistiques, acoustiques et visuels [WOL 13].

Un premier pas vers ces méthodes a été fait dans le chapitre 2 avec l'utilisation de méthodes d'apprentissage automatique pour catégoriser des opinions et faire ressortir les expressions associées à ces opinions. Un système de détection des disfluences dans des interactions orales reposant sur les CRF (*Conditional Random Fields*) a également été développé. Dans le chapitre 3, ce sont les HCRF (*Hidden Conditional Random Fields*) qui seront investis pour l'analyse d'opinions dans les interactions humain-agent.

1.2.3. Méthodes hybrides

Le troisième type est une version hybride des méthodes à base de règles et des méthodes d'apprentissage automatique. L'hybridation consiste en l'utilisation des méthodes à base de règles pour améliorer les méthodes d'apprentissage automatique [POR 14], ou en l'utilisation des méthodes d'apprentissage pour l'extraction de motifs séquentiels qui pourront être utilisés dans les règles d'extraction des expressions subjectives [CHO 05].

Les méthodes hybrides tirent ainsi parti de la capacité de généralisation des méthodes d'apprentissage automatique et de la modélisation fine et flexible offerte par les règles sémantiques. Ces méthodes requièrent une quantité de données annotées raisonnable et, par là-même, constituent une alternative intéressante aux méthodes de type *deep learning*.

Dans le chapitre 2, nous montrons l'intérêt d'utiliser des règles linguistiques pour construire les caractéristiques en entrée de méthodes d'apprentissage pour l'analyse d'opinions dans les enquêtes de satisfaction. Dans ce même chapitre, nous montrons également la possibilité de piloter l'apprentissage des caractéristiques prosodiques et linguistiques pour la détection des disfluences par des modèles CRF (*Conditional Random Fields*).

1.2.4. Modèles computationnels et type de tâches

La question du choix de la méthode se pose également en fonction du niveau de granularité considéré pour faire de la classification d'opinions. L'approche la plus classique est de se positionner *au niveau du document* (une critique/un article/une conversation) et de procéder à une classification du document en catégories d'opinion [TAB 11].

D'autres études s'intéressent à une sous-segmentation du document, par exemple au niveau de la phrase pour les documents qui contiennent plusieurs phrases ou au niveau du syntagme (voir par exemple les niveaux macro, meso et micro tels que définis dans [PAR 10]). Le document ayant été sous-segmenté, la méthodologie adoptée est la même que celle de la catégorisation de document.

Un autre type de tâche consiste à considérer le *niveau de l'expression d'opinions* : il s'agit alors de repérer des séquences de mots porteurs d'opinions et de leur attribuer une polarité. Différentes méthodes peuvent être utilisées pour cette tâche, de l'utilisation de métriques permettant d'extraire les termes discriminants d'une opinion [HAM 16] à l'utilisation des champs aléatoires conditionnels (*Conditional Random Fields* – CRF) [CHO 05].

Au-delà du repérage d'une expression ou de la catégorisation, certaines méthodes s'attaquent au problème de l'*extraction d'une structure* liée aux opinions. C'est le cas de la relation d'une opinion à sa cible (sur quoi porte l'opinion exprimée ?) et à sa source (par qui est-elle exprimée ?). Ainsi, dans [YAN 13], les auteurs proposent une méthode de détection des sources et des cibles des opinions reposant sur les CRF (*Conditional Random Fields*) et l'utilisation de patrons syntaxiques.

Les travaux présentés dans le chapitre 2 s'attaquent à la relation des opinions à leur cible en proposant une méthode basée sur des métriques permettant l'extraction des expressions d'opinions et des différentes thématiques dans les corpus étudiés. Dans le chapitre 3, nous présentons une méthode à base de règles qui utilise une formalisation des énoncés de l'agent et la structuration du scénario d'interaction en différentes thématiques permettant d'aider l'analyse des cibles des *likes* et des *dislikes* de l'utilisateur.

1.2.5. Modèles computationnels et interaction humain-agent

Si les travaux concernant l'analyse des opinions et des sentiments à partir du contenu verbal abondent dans le contexte de l'analyse des réseaux sociaux et des contenus Internet, ils sont encore très peu présents dans les applications de robotique sociale et plus largement de l'interaction humain-agent.

La majorité des travaux se concentre sur l'analyse d'indices non-verbaux (expressions faciales, indices acoustiques) pour l'analyse des comportements socio-émotionnels [SCH 11], le contenu verbal restant, quant à lui, encore partiellement exploité. Deux études seulement présentent un système intégrant un module de détection des sentiments à partir du contenu verbal pour les interactions humain-agent [PUL 10, SMI 11]. Elles s'appuient sur un module d'analyse de sentiments [MOI 07] initialement prévu pour l'analyse de textes hors contexte conversationnel. Ces études proposent une première solution pour l'intégration d'un système de détection

d'opinions centré sur le contenu verbal dans une interaction humain-agent et montrent ainsi l'importance de l'analyse de ces opinions dans l'interaction avec un agent conversationnel animé. Cependant, l'adaptation des méthodes d'analyse d'opinions au contexte conversationnel de l'interaction est un sujet qui n'a pas encore été abordé.

Ainsi, les systèmes proposés dans le chapitre 3 sont une première réponse à cette lacune. Les méthodes développées proposent d'utiliser le contexte d'interaction pour aider l'analyse des opinions de l'utilisateur.

1.2.6. Question de recherche et articulation dans l'ouvrage

Les différents travaux de recherche sur les méthodes d'analyse d'opinions que nous présentons ici questionnent la capacité des différentes méthodes à créer des modèles computationnels qui soient à la fois génériques et pertinents.

Nous étudions cette question de la généricité et de la pertinence sous deux angles.

(Q2a) Généricité et adaptabilité des modèles computationnels en fonction de l'application (la gestion de la relation client sur des données d'entreprise et l'interaction humain-agent).

Comment développer des méthodes de détection d'opinions qui permettent une modélisation fine des concepts d'opinions s'appuyant sur des modèles théoriques complexes et intégrant des connaissances spécifiques (connaissances métiers, connaissance du contexte d'interaction humain-agent) ? Nous nous interrogeons sur la manière de modéliser le contexte d'interaction humain-agent et sur l'adaptation des méthodes d'analyse d'opinions aux enjeux de l'interaction humain-agent.

Nous présenterons ces recherches selon l'angle du contexte applicatif en considérant deux types de méthodes d'analyse d'opinion :

– des méthodes à base de règles permettant de modéliser le concept de satisfaction en intégrant les connaissances métiers liées à EDF dans le chapitre 2 et d'autre part de modéliser les goûts de l'utilisateur en intégrant les connaissances thématiques du scénario d'interaction et les interactions problématiques dans le chapitre 3 ;

– des méthodes hybrides permettant l'utilisation de méthodes d'apprentissage automatique et le repérage de chaînes de mots relatives aux opinions dans le chapitre 2.

(Q2b) Généricité et adaptabilité des modèles computationnels en fonction des données.

Comment développer des méthodes de détection d'opinions qui permettent une modélisation à la fois des spécificités langagières et de la richesse des phénomènes

langagiers spontanés présents dans des données *in-the-wild*? Nous nous référons à cet enjeu de recherche en pleine expansion : le traitement des données *in-the-wild* [SCH 16], terme utilisé ici pour désigner des données collectées dans des cadres applicatifs réels. Un autre terme est parfois utilisé dans la littérature pour désigner ce type de corpus en texte : les corpus écologiques⁵.

L'angle des données a été étudié avec des méthodes à base de règles permettant de modéliser les spécificités langagières de l'oral conversationnel dans le chapitre 2 et des données de *chatbot* dans le chapitre 3. Des analyses de corpus ont permis la caractérisation des phénomènes langagiers présents dans des interactions écrites de forums et des interactions orales de centres d'appels (chapitre 2).

1.3. Interaction humain-agent et comportements socio-émotionnels

Le développement d'agents socialement et émotionnellement compétents dans une interaction avec un humain implique des recherches autour de trois volets présentés sur la figure I.2 [CLA 16a] :

- 1) l'analyse et la reconnaissance des comportements de l'utilisateur ;
- 2) le développement de stratégies d'interactions ;
- 3) la génération de comportements.

Le premier volet a été introduit dans la section I.2. Nous abordons dans cette section les deux autres volets en nous focalisant, comme cela a été fait pour l'analyse des comportements, sur la composante verbale, et, plus largement, la composante orale.

1.3.1. Stratégies d'interaction socio-émotionnelles

L'interaction humain-agent requiert la définition de stratégies socio-émotionnelles permettant à l'agent de réagir de façon socialement pertinente à ces comportements. La modulation de la réponse de l'agent en fonction des comportements socio-émotionnels de l'utilisateur aide l'agent à être plus efficace selon la tâche qui lui a été attribuée et à établir des relations plus riches, améliorant ainsi sa capacité à susciter de la sympathie [CAL 11b] et des sentiments de proximité (*rapport* [BIC 10a] en anglais) chez l'utilisateur.

5. Le terme de « corpus écologique » dans [WIS 10] fait référence au recueil de données textuelles produites le plus spontanément possible par le locuteur, par opposition à des exemples construits par les linguistes pour les besoins d'une étude, produisant des méthodes, outils et théories déconnectés du monde réel.

Le concept d'*engagement* joue également un rôle crucial en interaction humain-agent : la mise en place d'agents engageants est primordiale, que ce soit pour les interactions humain-agent ponctuelles ou celles dédiées à la construction de relations long-terme entre l'utilisateur et l'agent.

De nombreuses applications de l'interaction humain-agent comme la formation [DME 13] ou la santé [BIC 10b, GRI 14] ou les musées [CAM 15a, KOP 05] montrent l'importance du paradigme d'engagement. Celui-ci est abordé sous des angles différents dans la littérature de l'interaction humain-agent. Une manière intéressante de la structurer est la distinction entre l'engagement attentionnel (*attentional engagement*) et l'engagement émotionnel (*emotional engagement*) [PET 09], même s'il est parfois difficile de les distinguer complètement, étant donné que l'attention est influencée par les émotions.

Si la pertinence de la réponse de l'agent à l'utilisateur est clairement un facteur d'engagement de l'utilisateur [NOV 10], la mise en place de stratégie d'interactions socio-émotionnelles peut permettre de pallier les limites des capacités de compréhension des systèmes de dialogue. C'est ce qui avait été fait dans le projet européen Semaine avec le choix d'un scénario d'interaction où l'agent était proactif dans l'interaction – c'était lui qui posait des questions – et était doté d'une palette de *feedbacks* et de *backchannels* émotionnels [SCH 12]. Ces *feedbacks* et *backchannels* sont un exemple typique de stratégie d'interaction socio-émotionnelle : ils permettent de maintenir l'engagement de l'utilisateur en dotant l'agent de comportement d'écoute [LAM 11]. Ainsi, [DME 13] propose un agent générant des *feedbacks* pour aider les étudiants à rester engagés dans l'interaction et dans leur tâche. Dans [TRU 10], un modèle à base de règles est utilisé pour planifier les *backchannels* de l'agent en fonction de la prosodie de l'utilisateur.

Un autre exemple de stratégies d'interaction socio-émotionnelles davantage axé sur le contenu verbal se trouve dans les stratégies de politesse qui pourvoient l'agent d'une intelligence sociale [WAN 08] et lui permettent d'être perçu comme plus engagé dans l'interaction [DEJ 08, GLA 14]. fournit ainsi un modèle centré sur le contenu verbal permettant à l'agent de s'aligner sur le niveau de politesse et de formalité de l'utilisateur.

Ce dernier article fait référence au concept de l'alignement et montre l'intérêt des stratégies d'alignement dans l'interaction socio-émotionnelle entre l'agent et l'utilisateur. Le concept d'alignement, tel que défini dans [PIC 04], fait référence à la tendance qu'ont les personnes à s'aligner sur leur manière de parler lorsqu'elles parlent entre elles et lorsqu'elles parlent à une machine [BRA 10]. Le processus d'alignement est vu comme un signe d'empathie et de compétence interpersonnelle [PFE 08], et peut par conséquent être considéré comme un comportement permettant d'améliorer la perception de la compétence sociale de l'agent.

Si l'on considère le processus d'alignement au sens large, différentes terminologies peuvent être trouvées dans la littérature faisant référence à des notions connexes à l'alignement. Les notions diffèrent dans la manière dont elles intègrent les aspects temporels et dynamique. Par exemple, la notion de *mimicry* est définie comme l'imitation directe de ce que l'utilisateur produit [BER 13] alors que la synchronie est définie comme l'adaptation réciproque et dynamique des structures temporelles des comportements entre deux interactants [DEL 12].

Les processus d'alignement ont été largement explorés dans des études linguistiques dédiées à l'observation de corpus [TRU 12]. Cependant, ces dernières années, nous avons pu constater un gain d'intérêt pour l'implémentation des processus d'alignement dans l'interaction humain-machine. Dans le dialogue humain-machine, celle-ci concerne principalement l'alignement sur les niveaux lexical et syntaxique [BUS 09], alors que dans l'interaction en face à face (*face-to-face*) entre l'utilisateur et l'agent, la majeure partie des implémentations s'attaque à un processus d'alignement – processus d'alignement, pris au sens large et incluant par exemple l'imitation (*mimicry*) [HES 99] et la résonance socio-émotionnelle (*social/emotional resonance*) [GRA 13] – reposant sur le contenu non-verbal.

Les méthodes sous-jacentes à la mise en place de stratégies socio-émotionnelles d'interactions sont actuellement des méthodes à base de règles et de patrons linguistiques (voir l'utilisation de l'AIML *Markup Language* intégrant les émotions de l'utilisateur [SKO 11]), même si la tendance actuelle pousse l'émergence des méthodes reposant sur l'apprentissage. Ainsi, [BUI 10] utilise les processus de décision de Markov (POMPD – *Partially Observable Markov Decision Processes*) qui prend comme entrée l'information sur l'état émotionnel de l'utilisateur et [KHO 15] utilise des méthodes d'apprentissage par renforcement pour gérer les tours de parole.

Dans le chapitre 4, nous proposons un nouveau type de stratégies d'interaction qui permet de s'aligner sur les appréciations de l'utilisateur en paramétrant le contenu verbal de l'agent en fonction de celui de l'utilisateur. Ce type d'alignement étant encore très peu représenté dans la littérature, nous avons choisi de mettre en place une méthodologie reposant sur l'observation de corpus et sur la littérature en analyse conversationnelle pour construire notre modèle.

1.3.2. Génération d'attitudes sociales

Les études menées pour générer chez les agents artificiels des comportements socio-émotionnels (phénomènes liés aux opinions tels que décrits dans la section I.1, émotions, attitudes sociales, personnalité, etc.) se situent essentiellement au niveau de l'animation visuelle (geste, expressions faciales) [OCH 13, XU 14]. La génération

de parole expressive avec les systèmes de TTS (*Text To Speech*) [CHA 13] et la génération du contenu verbal [WAL 14] ont également été abordés dans le contexte de la génération de comportement pour un ACA mais de manière plus marginale.

Une première façon de construire les modèles d'expressions est le développement de règles élaborées à partir de la synthèse des observations réalisées dans la littérature en psychologie [BRU 15].

Une seconde façon de procéder est de s'appuyer sur les analyses quantitatives faites à partir de l'observation de corpus. Dans ce cas, les modèles construits commencent à intégrer la variabilité intrinsèque aux observations et permettent de réduire le travail d'analyse manuelle à fournir pour chaque nouveau contexte applicatif, lorsque celui-ci ne dispose pas d'études préalables dans les domaines de la psychologie.

La méthodologie adoptée est d'analyser sur des corpus d'enregistrements d'humain des corrélations entre des labels correspondant à des comportements socio-émotionnels et des indices linguistiques, acoustiques et visuels. Nous pouvons trouver ainsi des illustrations de cette méthodologie pour l'étude des caractéristiques prosodiques qui influent sur la perception de la dominance [TUS 00] ou du charisme [ROS 09].

L'utilisation des méthodes d'apprentissage participe également à l'introduction d'une variabilité dans la génération de comportements, la génération de comportements systématiques ayant tendance à entraver le naturel et la fluidité de l'interaction. La tendance est d'utiliser des méthodes de type *sequence mining* pour identifier des séquences de signaux ou d'indices qui caractérisent un comportement socio-émotionnel, comme les *Generalized Sequence Patterns* [CHO 14, MAR 11].

Le défi est la synchronisation des modalités entre elles lors de la génération des comportements expressifs [MAR 13a]. Les méthodes d'apprentissage automatique ont pour vocation d'apprendre comment générer des gestes ou des expressions faciales à partir de la parole par exemple [DIN 13].

Nous proposons dans le chapitre 5 deux types de méthodes :

- une nouvelle méthode de *sequence mining*, la chaîne de traitement SMART (*Social Multimodal Association with Timing*) qui permet d'extraire automatiquement, à partir d'enregistrements audio-visuels, des règles d'association temporelle entre les signaux sociaux et de les injecter dans les modèles de génération de comportements de l'agent ;

- une méthode à base de règles modélisant au sein de l'agent le processus cognitif mis en œuvre par l'humain pour synchroniser ses gestes sur sa parole.

I.3.3. Questions de recherche et articulation dans l'ouvrage.

Si dans la section précédente, nous présentons, entre autres, la question de recherche liée à la détection des comportements de l'utilisateur, la question de recherche que j'aborde dans ce paragraphe porte sur les comportements de l'agent conversationnel animé. Nous positionnons les travaux présentés dans ce livre sur deux volets.

La premier volet porte sur les stratégies d'interaction socio-émotionnelles centrées sur le contenu verbal. Notre motivation principale est d'utiliser ces stratégies pour tenter d'influer sur les facteurs de l'engagement de l'utilisateur, l'engagement de l'utilisateur dans l'interaction étant un défi majeur de l'interaction humain-agent, comme nous l'avons vu dans l'état de l'art.

Dans la lignée des travaux émergeant sur l'implémentation de modèle d'alignement dans l'interaction humain-agent, la question de recherche associée à ce premier volet examine : (Q3) le rôle des stratégies d'alignement dans l'engagement de l'utilisateur, en s'attaquant à l'alignement au niveau des opinions et de leur expression.

Le développement d'un agent virtuel doté des capacités d'exprimer des opinions en réaction aux opinions exprimées par l'utilisateur selon un modèle d'alignement verbal de l'agent sur l'utilisateur est présenté dans le chapitre 4.

Le deuxième volet, également abordé dans le chapitre 5, porte sur la génération des comportements socio-émotionnels. Nous nous sommes focalisées sur le contenu prosodique et ses relations avec les autres modalités (expressions faciales, gestes), d'une part, et sur des comportements socio-émotionnels de type attitudes sociales.

La question abordée ici est la suivante : (Q4) comment paramétrer la prosodie dans un contexte multimodal pour la génération de comportements multimodaux chez un agent virtuel ?

L'état de l'art montre que les travaux s'attaquant à cette question sont encore rares. La méthodologie investie pour répondre à cette question est l'utilisation de méthodes de *sequence mining* intégrant l'information temporelle.

I.4. Plan de l'ouvrage

L'ouvrage commence dans le chapitre 1 par une étude comparative des différents corpus humain-humain et humain-agent collectés, annotés et étudiés dans les chapitres suivants.

Le chapitre 2 présente les travaux de recherche sur l'analyse des opinions dans les interactions humain-humain.

Notre première contribution a été de constituer des corpus *in-the-wild* (centres d'appels, forums) riches en expressions spontanées, à partir des données d'entreprise qui sont présentés dans le chapitre 1.

Notre seconde contribution concerne le développement de modèles d'opinions reposant sur des grammaires constituées de lexiques et de règles linguistiques et sur des méthodes d'apprentissage hybrides permettant d'intégrer des connaissances linguistiques au sein d'algorithmes d'apprentissage automatique.

Enfin, la troisième contribution porte sur la caractérisation des structures énonciatives propres à la communication virtuelle écrite et ses spécificités rédactionnelles et sur la caractérisation des phénomènes de parole spontanée (disfluences). Pour ce dernier aspect, nous avons travaillé sur la modélisation jointe des paramètres acoustiques et des marqueurs linguistiques au sein des CRF (*Conditional Random Fields*) pour la détection des disfluences.

Le chapitre 3 aborde le contexte de l'interaction humain-agent. Nos contributions portent sur deux aspects.

D'une part, nous avons proposé une modélisation fine du phénomène des opinions, en le circonscrivant aux goûts de l'utilisateur (*likes* et *dislikes*) et aux opinions des utilisateurs envers l'interaction (pour la détection d'interactions problématiques).

D'autre part, la méthode de détection des opinions proposée ancre son fonctionnement dans le contexte de l'interaction, avec la prise en compte à la fois du contexte dialogique (paires adjacentes et énoncés précédents de l'utilisateur), des modalités communicatives de l'agent et de la structure de *Topic* donnée par le scénario d'interaction.

Concernant nos travaux sur les stratégies d'interactions socio-émotionnelles de l'agent face à un utilisateur humain (chapitre 4), la stratégie d'interaction proposée permet de décider de l'alignement de l'agent sur l'appréciation de l'utilisateur et d'instancier les paramètres verbaux de réalisation de cet alignement, et constitue ainsi une première proposition de système d'alignement sur les appréciations dans une interaction humain-agent.

Concernant la génération chez l'agent d'énoncés présentée dans le chapitre 5, la méthodologie proposée permet l'extraction automatique de séquences de signaux sociaux (prosodie et expressions faciales) caractéristiques d'une attitude sociale, en travaillant directement à partir de corpus d'enregistrements, sans passer par l'étape d'annotation manuelle des signaux sociaux.

Nous récapitulons ci-dessous les différentes questions de recherche abordées au sein de ces différents chapitres :

(Q1) pertinence des différents modèles théoriques pour la construction d'un modèle computationnel (linguistique et prosodique) en fonction de l'application ;

(Q2a) généricité et adaptabilité des modèles computationnels en fonction de l'application ;

(Q2b) généricité et adaptabilité des modèles computationnels en fonction des données ;

(Q3) le rôle des stratégies d'alignement dans l'engagement de l'utilisateur ;

(Q4) la modélisation de la prosodie dans un contexte multimodal pour la génération de comportements multimodaux chez un agent virtuel.

