



Institut National des Langues et Civilisations Orientales

**Master Traitement Automatique des Langues, spécialité Ingénierie
Multilingue**

Évolution et visualisation des émotions

Loïc DUMONET

Sous la direction de Natalia Grabar, chargée de recherche CNRS

Année universitaire 2013-2014



Remerciements

Je tiens à exprimer toute ma reconnaissance à ma directrice de stage, Natalia Grabar, chargée de recherche au CNRS, qui s'est rendue disponible et qui m'a donné les moyens nécessaires pour mener à bien ce projet dans les meilleures conditions. Grâce au suivi régulier de mes travaux et à ses précieux conseils, j'ai pu les mettre à profit pour la bonne réalisation de la mission qui fut très enrichissante. Je la remercie également pour sa participation à une tâche d'annotation et surtout pour le temps qu'elle a consacré à la relecture du mémoire ainsi qu'à ses instructions pertinentes qui m'ont permis de l'améliorer.

Ensuite, je souhaite adresser mes remerciements aux équipes professorale et de chercheurs de l'INaLCO, pour leurs enseignements sur ces deux dernières années, qui m'ont permis d'accomplir cette mission de façon efficace et autonome.

Enfin, je tiens à exprimer toute ma gratitude à ma conjointe qui s'est portée volontaire pour une tâche d'annotation et qui m'a soutenu sans relâche durant toute la durée du projet, ainsi qu'à ma famille et à mes proches amis qui m'ont sans cesse encouragé, ce qui alimentait ma motivation.

Résumé

Les forums médicaux permettent aux patients d'exprimer leurs ressentis par rapport à leur maladie ou leur traitement, et sont souvent teintés d'émotions. De ce fait, ils constituent une source précieuse d'informations pour les professionnels de la santé car offrent des indications sur la vie et le comportement des patients. Les discussions sur les forums sont souvent nuancées par la présence de marqueurs linguistiques comme des émotions et sentiments, des modifieurs d'intensité, des indicateurs d'incertitude ou encore la négation. Nous nous intéressons à l'expressions des émotions, à leurs valeurs contextuelles, ainsi qu'à leur évolution dans les messages de forums médicaux. Nous proposons également des mécanismes de visualisation de cette évolution des émotions. L'analyse des sentiments a connu un essor important depuis ces dernières années. Les données des microblogs comme les avis de consommateurs sur des produits, les tweets, les commentaires d'internautes sur des films, des jeux vidéo, ... constituent une source d'informations gigantesques pour les tâches d'analyse de sentiments ou d'opinions. Les travaux consacrés aux émotions sont bien plus complexes et moins nombreuses. En effet, l'analyse des sentiments concerne principalement trois classes polaires : positive, négative et neutre tandis que l'analyse des émotions concernent généralement au moins cinq ou six émotions. Cela devient rapidement compliqué dans les tâches de classification, car plus il y a de classes et plus il est difficile de répartir correctement les informations. Grâce à des traitements automatiques dédiés à ces caractéristiques sémantiques, nous proposons différentes visualisations des émotions et des incertitudes afin d'étudier leur évolution dans les interactions entre patients/professionnels et patients/patients. Pour y parvenir, nous mettons en place un système à base de règles, construites à partir de ces différents marqueurs, afin de détecter les émotions et leur polarité dans les divers types d'échanges. Grâce à ces règles, nous pouvons par la suite calculer l'intensité polaire des émotions pour chaque question et pour chaque réponse. Ce calcul permet de sélectionner la polarité émotionnelle dominante dans un type d'énoncé (question ou

réponse) afin d'étudier les caractéristiques propres aux patients (subjectivité, fréquence élevée de marqueurs émotionnels, . . .) et celles propres aux professionnels de santé (retenue émotionnelle, fréquence de marqueurs d'incertitude, . . .).

Mots-clés : Traitement automatique de la langue, forum de santé, émotions, évolution, visualisation

Table des matières

1	Introduction	8
1.1	Contexte	10
1.2	Travaux existants	11
1.2.1	Les différents systèmes	12
1.2.2	Les algorithmes de l'apprentissage supervisé souvent utilisés	14
1.2.3	La visualisation des sentiments et des émotions	15
1.2.4	Le domaine médical	16
1.2.5	Le matériel souvent exploités dans les travaux sur les émotions et les sentiments	19
1.2.5.1	Les corpus	19
1.2.5.2	Les lexiques	22
1.2.5.3	Les outils	26
1.3	Objectifs de notre travail	28
2	Méthodes et matériel	30
2.1	Matériel	30
2.1.1	Les corpus	30
2.1.1.1	Description et organisation des corpus	30
2.1.1.2	L'annotation linguistique et sémantique des corpus	31
2.1.2	Les ressources	32
2.1.2.1	Les ressources sémantiques	32
2.1.2.2	Les ressources linguistiques	33
2.2	Méthodes	35
2.2.1	Analyse des corpus	37
2.2.2	Ajustement des ressources	39
2.2.3	Réaccentuation des corpus	43
2.2.4	Sélection des connecteurs syntaxiques	44
2.2.5	Un système à base de règles	45
2.2.5.1	Les contextes	46
2.2.5.2	Les règles	46
2.2.5.2.1	Pré-traitement.	47

2.2.5.2.2	Traitement de la négation.	48
2.2.5.2.3	Traitement des modifieurs.	48
2.2.5.2.4	Traitement de l'incertitude.	49
2.2.5.2.5	Combinaison de marqueurs.	50
2.2.5.2.6	Traitement des connecteurs.	51
2.2.5.2.7	Incertitude comme forme pôle.	53
2.2.5.3	Calcul de la polarité émotionnelle dominante	54
2.2.6	L'évaluation du système	59
2.2.7	Les modèles de visualisation	62
2.2.7.1	Dynamique des émotions et des incertitudes	63
2.2.7.2	Graphes non orientés	64
2.2.7.3	Diagrammes circulaires	65
2.2.7.4	Constellations	66
2.2.7.5	Nuage de mots	66
2.3	Expériences	66
3	Résultats et Discussion	68
3.1	Évaluation du système	68
3.1.1	Évaluation par rapport aux annotateurs	69
3.1.2	Évaluation par rapport aux annotations communes	70
3.2	Distributions des polarités émotionnelles	71
3.2.1	Distribution polaire par couple question/réponse(s)	71
3.2.1.1	Échanges patients/professionnels	72
3.2.1.2	Échanges patients/patients	73
3.2.2	Distribution des types médicaux en fonction des polarités émotionnelles	74
3.2.2.1	Corpus QA	75
3.2.2.2	Corpus forum	76
3.2.2.3	Bilan	77
3.2.3	Distribution polaire en fonction du corpus	78
3.2.3.1	Corpus QA	79
3.2.3.2	Corpus forum	80
3.2.3.3	Bilan	81
3.3	Distribution des marqueurs par polarité émotionnelle (graphes non orientés)	81
3.3.1	Corpus QA	82
3.3.1.1	Patients	82
3.3.1.2	Professionnels de santé	83
3.3.2	Corpus forum	84
3.3.2.1	Patients qui posent des questions	84
3.3.2.2	Patients qui répondent aux questions	85
3.3.2.3	Bilan	86

3.4	Les diagrammes circulaires : distribution des informations sémantiques en fonction des polarités et des catégories d'émotions	86
3.4.1	Répartition des polarités émotionnelles	86
3.4.1.1	Corpus QA	87
3.4.1.2	Corpus forum	88
3.4.1.3	Bilan	89
3.4.2	Répartition des catégories émotionnelles	89
3.4.2.1	Émotions positives	89
3.4.2.1.1	Corpus QA	90
3.4.2.1.2	Corpus forum	91
3.4.2.2	Émotions négatives	92
3.4.2.2.1	Corpus QA	92
3.4.2.2.2	Corpus forum	93
3.4.2.3	Émotions neutres	94
3.4.2.3.1	Corpus QA	95
3.4.2.3.2	Corpus forum	96
3.4.2.4	Incertitudes	97
3.4.2.4.1	Corpus QA	98
3.4.2.4.2	Corpus forum	99
3.4.2.5	Bilan	100
3.5	Les "constellations" : représentation quantitative de l'expression des émotions par polarité	100
3.6	Les nuages de mots : modélisation du contenu des contextes émotionnels	104
4	Conclusion et perspectives	110
4.1	Conclusion	110
4.1.1	Particularités des discours et des corpus étudiés	110
4.1.1.1	Interactions patients/professionnels de santé dans le corpus Q/A	110
4.1.1.2	Interactions patients/patients dans le corpus forum	111
4.1.2	Les concepts médicaux	111
4.1.3	Efficacité du système à base de règles	111
4.1.4	Difficultés	112
4.1.4.1	Création des données de référence	112
4.1.4.2	Détection de l'ironie et du sarcasme	112
4.1.4.3	Orthographe et syntaxe particulières	112
4.1.4.4	Couverture et précision des ressources linguistiques et sémantiques	113
4.2	Perspectives	113

Bibliographie	115
Annexes	124
A Les corpus	124
A.1 Corpus Q/A	124
A.2 Corpus forum	124
B Les ressources	127
B.1 Ressources sémantiques	127
B.2 Ressources linguistiques	128
B.2.1 Les descripteurs	128
B.3 Les émotions	129
B.3.1 Les marqueurs émotionnels	129
B.3.2 L'association polarité/catégorie émotionnelle	130
C Ajustements des ressources	131
C.1 Ressources sémantiques	131
C.2 Ressources linguistiques	134
D Le système à base de règles	135
D.1 Pré-traitement des contextes émotionnels/incertitude : mise à jour des frontières contextuelles des marqueurs émotionnels et d'incertitude en fonction de la présence de connecteurs	136
D.2 Traitement de la négation	138
D.3 Traitement des modifieurs	140
D.4 Traitement l'incertitude	142
D.5 Traitement des connecteurs	143
E Calcul de la polarité émotionnelle dominante	144
E.1 Corpus Q/A	144
E.1.1 Détection de la polarité pour chaque marqueur émotionnel présent dans les questions et les réponses	144
E.1.2 Détection de la polarité émotionnelle dominante pour chaque question et réponse . .	144
E.2 Corpus forum	144
E.2.1 Détection de la polarité pour chaque marqueur émotionnel présent dans les questions et les réponses	144
E.2.2 Détection de la polarité émotionnelle dominante pour chaque question et réponse . .	144
E.2.3 Détection de la polarité émotionnelle dominante pour l'ensemble des réponses . . .	144

F	Création des données de référence	150
F.1	Corpus Q/A	151
F.2	Corpus forum	154
G	Modélisations de l'évolution des émotions	157
G.1	Distribution des types médicaux dans les énoncés en fonction des polarités émotionnelles dominantes (échelle : nombre d'occurrences)	157
G.2	Distribution des polarités émotionnelles dominantes dans les énoncés en fonction du corpus (échelle : nombre d'occurrences)	159
G.3	Distribution des marqueurs par polarité émotionnelle	160
	G.3.1 Corpus Q/A	160
	G.3.2 Corpus forum	163
G.4	Les "constellations" : représentation quantitative de l'expression des émotions par polarité et par catégorie émotionnelle	166
	G.4.1 Les émotions à polarité neutre	166
	G.4.2 Les émotions à polarité incertitude	166

Chapitre 1

Introduction

Les médias sociaux (forums, réseaux sociaux, ...) présents dans la sphère web offrent la possibilité à tout internaute de communiquer au sein d'une communauté sur des sujets précis. Ces médias sociaux permettent aux utilisateurs d'échanger des informations sur des thèmes de toute sorte : matériel informatique, médecine, jeux vidéo, événements culturels, ... Nous nous intéressons au domaine de la médecine et plus particulièrement aux forums de santé. Ces plateformes contiennent de nombreuses informations intéressantes sur les populations. Les forums médicaux fournissent divers renseignements sur ces populations comme leur état de santé, leurs savoirs techniques, leurs expériences personnelles avec la maladie, ... Ces plateformes de discussion en ligne proposent différents types d'échanges : des patients posent leur question et des professionnels y répondent (comme Dotissimo (Chauveau-Thoumelin & Grabar, 2014)) ou des patients posent leur question et ce sont d'autres patients qui y répondent (comme MaSanteNet (Abdaoui *et al.*, 2014)). Cependant, ces médias sociaux présentent plusieurs particularités, comme par exemple, l'expression très fréquente des émotions et des sentiments. Cela se justifie par cette facilité qu'ont les patients de pouvoir s'exprimer librement et avec moins de réticence que s'ils étaient face à face avec un médecin (Gauducheau, 2008). Dans le cadre de notre étude, nous proposons d'étudier ce type de matériel en nous concentrant sur ses différents aspects : le contenu conceptuel (termes médicaux comme les maladies, les pathologies, les médicaments, les traitements, les procédures ou diagnostics) et le contenu subjectif (comme des marqueurs d'incertitude, de négation, des modificateurs et les émotions), avec une attention particulière pour les émotions.

Si l'on distingue d'habitude les *sentiments* des *émotions*, la difficulté à les différencier est de taille. Selon Jacques Cosnier (Cosnier, 1994) :

- les émotions sont *des processus dynamiques qui ont un début et une fin et une durée relativement brève. Ces phénomènes "phasiques" sont causés par des événements précis et inattendus,*
- *Les "sentiments" tels que l'amour, la haine, l'angoisse, entre autres, se distinguent nettement des précédents par leurs causes plus complexes, par leur durée plus longue ("tonique"), et leur intensité plus basse.*

Ces définitions montrent bien le caractère évolutif des émotions.

Dans les travaux de TAL (Traitement Automatique de Langues), les travaux sur l'analyse des émotions sont moins nombreux que ceux consacrés à l'analyse de sentiments :

- L'analyse des sentiments ne considère que la polarité (*positive, négative, neutre*) dans la tâche de

classification. La majorité des études existants dans ce domaine, appelé aussi fouille d'opinions, consistent à classer un mot, une phrase ou un document dans une des polarités en fonction des termes qui les composent ;

- Dans l'analyse des émotions, l'objectif consiste à classer des énoncés (mots, phrases, documents) selon une liste de catégories émotionnelles *joie, peur, surprise,...* Aux émotions peuvent également être associées des polarités (aussi appelée *valence* ou *orientation sémantique*) :
 - positive (*attraction, joie, soulagement,...*),
 - négative (*colère, tristesse, peur, inquiétude,...*),
 - neutre (*étonnement, surprise, anticipation,...*).

Dans notre étude, nous adoptons la notion large d'émotions et nous considérons également les sentiments comme des marqueurs émotionnels.

Il existe également quelques travaux sur l'analyse des émotions dans le contexte médical. Parmi ceux-ci, mentionnons l'étude qui cherche à déterminer la pertinence des émotions pour la recherche d'informations par les patients (Battaïa, 2012) ; la détection des émotions exprimées par les patients qui sont liées à des entités médicales (Melzi *et al.*, 2014) ; l'identification de cibles des émotions dans les messages de forums médicaux (Bringay *et al.*, 2014) ; l'interaction des émotions par la présence de divers marqueurs (Chauveau-Thoumelin & Grabar, 2014) (*incertitude* forte et faible (Abdaoui *et al.*, 2014), *négation* et inversion de polarité émotionnelle (Maurel *et al.*, 2008), *modificateurs* d'amplification et d'atténuation).

Dans notre étude, nous proposons de nous concentrer sur la communication de patients (personne lambda atteinte de troubles de la santé) avec les professionnels de santé (médecins, spécialistes, étudiants en médecine), ainsi que sur la communication entre les patients. Plus particulièrement, nous proposons d'étudier plusieurs points :

- la détection de la portée de la négation,
- la détection d'unités linguistiques pertinentes dans l'altération des émotions (intensité et polarité),
- l'établissement de priorités dans l'application des règles concernant les marqueurs,
- l'analyse du rôle de l'incertitude dans les contextes émotionnels.

Pour atteindre nos objectifs, nous allons mettre en place un système à base règles. Finalement, nous allons modéliser le contenu des messages de forum (questions et réponses) afin de visualiser les émotions qui y sont exprimées. Nous supposons que la relation entre la question et ses réponses représente l'évolution des émotions des locuteurs : sa modélisation permet donc de visualiser cette évolution.

Dans la suite du rapport, nous présentons d'abord le contexte de notre étude (section 1.1), un état de l'art des travaux existants dans le domaine des sentiments, des émotions et des visualisations (section 1.2), nous précisons ensuite les objectifs de nos travaux (section 1.3). Ensuite nous présentons le matériel utilisé (section 1.2.5). Nous développons la méthodologie employée pour la mise en place du système à base de règles dans la section 2.2, et détaillons les diverses modélisations pour visualiser l'évolution des émotions dans la section 2.2.7. Enfin, nous discutons des résultats (section 3) avant de conclure (section 4.1) et de définir quelques perspectives pour de futurs travaux (section 4.2).

1.1 Contexte

Le foisonnement des informations disponibles sur les plateformes sociales de l'Internet constitue une source de données particulièrement intéressante pour les chercheurs. Dans le cadre de notre étude, nous nous intéresserons aux forums de santé qui contiennent une richesse d'informations médicales (pathologie, traitements, effets secondaires, etc.) qui peuvent être très utiles et très instructives pour les professionnels de la santé comme pour les patients. De plus, nous pensons que les méthodes mises en place et testées sur les documents librement disponibles sur l'internet, peuvent aussi être efficacement exploitées avec les documents issus d'autres sources (littérature scientifique, littérature de vulgarisation, romans, etc.), également disponibles au format électronique. Cependant, le discours informel qui caractérise ces communautés sociales en ligne, (abréviations, fautes grammaticales et d'orthographe, argot, etc.), rend les tâches d'extraction d'informations assez délicates (difficultés de la reconnaissance automatique des mots quand ils sont mal écrits, perte d'informations, faible quantité de lexiques pour ce genre de discours). Ces échanges représentent pourtant une source d'informations très importante pour les professionnels de santé sur des questions comme celles relatives à la perception de la maladie, des traitements, des effets indésirables ou de toute autre question des patients. La possibilité de communiquer librement entre les patients, ou entre les patients et les professionnels de santé est également bénéfique aux patients. En effet, ce mode de communication sans cesse grandissant et offrant une certaine liberté de discours grâce à l'anonymat, met en confiance le patient (Gauducheau, 2008). L'anonymat permet en effet de s'exprimer sur des sujets délicats (problèmes dans la sexualité d'un couple, protubérances indésirables, état émotionnel ou psychologique inhérent à l'état de santé, etc.) aux personnes plutôt timides ou peu rassurées lors d'une conversation en face à face. C'est dans ce contexte, fortement subjectif et informel, que nous effectuons notre travail. Ce travail s'inscrit dans le cadre du projet *Parlons de nous*¹. De plus, notre travail est également lié au projet ANR TecSan *Ravel*² qui concerne la recherche et la visualisation des informations dans le dossier patient électronique. Les patients et les professionnels de santé doivent disposer d'outils capables de répondre à leurs besoins de la façon la plus pertinente et efficace possible (l'accès en temps réel aux informations précises sur la santé d'un patient afin d'améliorer la prise de décision des professionnels de santé ou encore faciliter la compréhension des patients sur leur état de santé). Il devient donc nécessaire et indispensable de mettre au point des systèmes capables de pallier aux difficultés liées aux gros volumes de données, engendrés par le développement des plateformes sociales communautaires. Ce type de systèmes devra également être en mesure de traiter les problèmes posés par l'hétérogénéité des données, typiques des forums et du web social en général. Notre travail fait suite aux expériences déjà réalisées dans l'équipe autour de la détection et l'annotation de la subjectivité (incertitude, émotions...) dans les documents biomédicaux (Grabar & Hamon, 2009; Périnet *et al.*, 2011; Chauveau-Thoumelin & Grabar, 2014). Dans la suite de cette section, nous présentons d'abord les travaux de l'état de l'art (section 1.2) et précisons ensuite les objectifs poursuivis (section 1.3).

1. Projet financé par la Maison des Sciences Humaines de Montpellier en 2013 puis par le réseau inter-MSH depuis 2014. Ce projet met en collaboration plusieurs laboratoires de recherche : <https://www.lirmm.fr/patient-mind/pmwiki/pmwiki.php?n=Site.Participants>

2. Projet financé par l'Agence Nationale de la Recherche depuis janvier 2012. Ce projet met en collaboration plusieurs laboratoires de recherche ainsi que des entreprises : [http://www.agence-nationale-recherche.fr/projet-anr/?tx_lwmsuivibilan_pi2\[CODE\]=ANR-11-TECS-0012](http://www.agence-nationale-recherche.fr/projet-anr/?tx_lwmsuivibilan_pi2[CODE]=ANR-11-TECS-0012)

1.2 Travaux existants

Depuis une décennie, les travaux sur l'analyse des sentiments et des émotions connaissent un investissement considérable. En effet, la détection et l'extraction de termes subjectifs dans les textes issus du web 2.0³ demeurent une tâche encore peu triviale dans le domaine du traitement automatique des langues (TAL). Après la présentation de quelques généralités de cette tâche, nous nous concentrons sur les systèmes souvent exploités (section 1.2.1), les algorithmes le plus souvent utilisés (section 1.2.2), les travaux sur la visualisation des sentiments et des émotions (section 1.2.3), quelques travaux effectués sur les données du domaine médical (section 1.2.4), et nous passons en revue le matériel disponible et souvent exploité (corpus, lexiques et outils) dans la section 1.2.5.

Nous considérerons les émotions (joie, peur, surprise, etc.) et les sentiments (crainte, angoisse, satisfaction, etc.), tels que décrits dans l'étude de (Blumenthal, 2009), comme faisant partie d'un même ensemble (Battaïa, 2012). De très nombreux travaux sur les sentiments et les émotions existants dans l'état de l'art sont consacrés à leur classification. La classification peut être effectuée à différents niveaux de granularité (Liu *et al.*, 2014) : au niveau du document pour proposer le calcul de la polarité globale (Li *et al.*, 2010), au niveau de la phrase (Glaser & Schütze, 2012) ou au niveau d'une unité lexicale (Sayeed *et al.*, 2012) pour déterminer la polarité d'un lexème et sa pondération (Li *et al.*, 2014). Il est possible de distinguer deux objectifs de classification, selon que les émotions ou les sentiments sont traités :

- en travaillant avec les sentiments, les chercheurs cherchent à regrouper au sein d'une même catégorie des sentiments positifs, négatifs (Feng *et al.*, 2013a) ou neutres (Athar & Teufel, 2012),
- en travaillant avec les émotions, les catégories recherchées sont plus précises, comme par exemple :
 - les six émotions de base d'(Ekman, 1992) (Vu *et al.*, 2014) : joie, tristesse, colère, peur, dégoût et surprise,
 - les huit émotions de base de (Plutchik, 1980) (Hasegawa *et al.*, 2013) : colère, anticipation, dégoût, peur, joie, tristesse, surprise et confiance.

De très nombreuses études se sont concentrées sur la construction et l'exploitation de corpus basés sur des tweets⁴ concernant une thématique précise et que l'on poste sur le site <http://www.twitter.com>, dont la récupération est effectuée avec des API dédiées de Twitter (Takase *et al.*, 2013)⁵, (Purver & Battersby, 2012)⁶. La granularité recherchée dans ce type de matériel se trouve au niveau du document, qui correspond alors à un tweet. Un autre type de matériel largement exploité est composé des avis de consommateurs ou des commentaires d'internautes liés à différents types d'entités :

- la politique (Bakliwal *et al.*, 2013),
- les catastrophes naturelles, comme le grand tremblement de terre qui a ravagé le Japon le 11 mars 2011, (Vo & Collier, 2013),
- l'état émotionnel et psychologique de la population observée sur un axe diachronique : il peut s'agir des intentions de vote lors d'une campagne électorale (Bermingham & Smeaton, 2011),

3. Le web 2.0 est aussi appelé le Web social. Il concerne les plateformes d'échanges et d'interaction entre internautes sur toutes sortes de thématiques. Twitter et Facebook sont deux plateformes très utilisées au niveau international

4. Messages de 140 caractères maximum

5. <https://dev.twitter.com/docs/api/1/get/search>

6. <https://dev.twitter.com/docs/streaming-api>

- les produits industriels comme les appareils électroniques (Ramteke *et al.*, 2013), ou les services comme les prestations dans le monde de la restauration (Brody & Elhadad, 2010), sur lesquels les consommateurs donnent leurs avis (Batra & Rao, 2010), permettent de filtrer les avantages et les inconvénients de ces derniers.

La classification poursuivie est alors polaire : les avis ou commentaires positifs, négatifs ou neutres sont recherchés. Ce type d'études a pour objectif de cibler les besoins des consommateurs afin de pouvoir proposer des solutions concrètes aux entreprises grâce aux outils TAL développés pour ces tâches (section 1.2.5.3).

Une grande majorité des travaux portant sur les sentiments et les émotions est réalisée avec le matériel en langue anglaise. Cependant, avec l'évolution du web 2.0, de plus en plus de matériel et de travaux portent sur d'autres langues (Volkova *et al.*, 2013), par exemple : le portugais (Souza & Vieira, 2013), l'italien (Basile & Nissim, 2013), l'espagnol (Moreno-Ortiz *et al.*, 2013), le tchèque (Habernal *et al.*, 2013), l'hindi (Mittal *et al.*, 2013), le chinois (Yu *et al.*, 2013) ou encore le japonais (Tokuhisa *et al.*, 2008). En ce qui concerne les travaux pour le français, quelques études ont été réalisées sur le rôle important des émoticônes dans les tweets (Paroubek, 2010a). En effet, les émoticônes ou smileys constituent une caractéristique importante dans la tâche consacrée aux émotions et aux sentiments dans les plateformes de microblogging⁷. Les émoticônes pondèrent en effet de façon significative les messages des internautes (Paroubek, 2010b; Gauducheau, 2008; Kouloumpis *et al.*, 2011) car ils sont une partie inhérente des messages et complètent les informations exprimées lexicalement. Il existe aussi des travaux sur : (1) la constitution d'un lexique sémantique de verbes de sentiments, d'émotions et d'états psychologiques en langue française (Mathieu, 2006); (2) sur l'analyse des sentiments dans les forums (Maurel *et al.*, 2008) qui calcule la relation entre le sentiment et sa cause ou son objet. L'objectif est alors d'extraire un maximum d'informations sur les sentiments (positifs ou négatifs) liés à des entités comme des lieux ou des produits; (3) sur la façon dont les émotions sont transmises par les patients dans les forums de santé afin d'en déduire si les émotions constituent un critère de recherche pertinent (Battaïa, 2012). Dans ce dernier travail, différentes formes d'expressions des émotions comme les smileys, la ponctuation et le lexique sont étudiées; (4) sur la détection des cibles des émotions exprimées par les patients dans les forums médicaux (Bringay *et al.*, 2014); (5) sur la détection de la subjectivité dans le discours médical en étudiant la présence de marqueurs d'incertitude (*si, possible, probable, etc.*) et d'émotions (*smileys, ponctuation et caractères répétés, etc.*). Cette étude montre les différents niveaux d'expression de la subjectivité propre aux patients et aux professionnels de santé (Chauveau-Thoumelin & Grabar, 2014).

1.2.1 Les différents systèmes

Il existe trois types d'approches automatiques en TAL :

- L'apprentissage supervisé, comme dans (Greene & Resnik, 2009) qui y intègre des connaissances linguistiques comme les relations syntaxiques et la sémantique lexicale, est basé sur un ensemble de données déjà étiquetées (corpus annoté);
- Les systèmes à base de règles (ou symbolique) utilisent des lexiques et/ou des règles (de grammaires, expressions régulières, etc.) comme dans les travaux de (Brody & Elhadad, 2010) dédiés à la détection des sentiments dans les avis de consommateurs;

7. Forums, réseaux sociaux comme Twitter et Facebook, avis de consommateurs sur des produits

- Puis, un peu plus rare mais présente, l’approche mixte ou combinée qui consiste en l’utilisation hybride des deux précédentes techniques.

Les systèmes d’apprentissage automatique et ceux à base de règles nécessitent la mise en place de caractéristiques plus ou moins linguistiques pour l’extraction des informations pertinentes en fonction de l’objectif fixé par les besoins de l’étude. Ainsi, nous trouverons des caractéristiques (ou critères) suivantes par exemple :

- les n-grammes (essentiellement unigrammes et bigrammes) (Balahur & Tanev, 2013; Smith & Lee, 2012),
- les étiquettes morphosyntaxiques des différentes parties du discours (adjectif, nom, adverbe, déterminant, etc.) (Kouloumpis *et al.*, 2011; Habernal *et al.*, 2013), permettant entre autre la désambiguïsation émotionnelle des lexèmes ainsi que l’orientation sentimentale des contextes (Lei *et al.*, 2014),
- la présence de certains types de descripteurs comme la négation (Wilson *et al.*, 2009; Paroubek, 2010a; Ikeda *et al.*, 2008),
- l’incertitude et la modulation d’intensité émotionnelle (Chauveau-Thoumelin & Grabar, 2014; Syssau & Font, 2005),
- les verbes de modalité (Liu *et al.*, 2014),
- les expressions conditionnelles (Narayanan *et al.*, 2009),
- les arbres de dépendances syntaxiques (Bringay *et al.*, 2014; Wu *et al.*, 2009; Sayeed *et al.*, 2012; Nakagawa *et al.*, 2010),
- les graphes de propagation (Kikuchi *et al.*, 2014; Takamura *et al.*, 2005; Brody & Elhadad, 2010; Hatzivassiloglou & McKeown, 1997) pour les caractéristiques linguistiques.

Les caractéristiques spécifiques aux plateformes de microblogging (émoticônes, hashtags des Tweets (Kuneman *et al.*, 2014), répétition de caractères (Brody & Diakopoulos, 2011)), permettent d’améliorer la classification elles sont combinées avec les caractéristiques linguistiques. L’utilisation et l’accumulation de plusieurs de ces paramètres permettent aux systèmes d’extraction d’informations, de détecter le plus précisément possible le contexte d’une unité linguistique (lexème, expression composée (Moreno-Ortiz *et al.*, 2013)). Ainsi, plus le système aura d’informations précises sur le contexte, plus il sera à même d’attribuer l’orientation sémantique à une unité linguistique, exprimée par l’auteur à travers un sentiment ou une émotion. Par exemple, par l’analyse des contextes de ces sentiments (Mitchell *et al.*, 2013) et en fonction des critères linguistiques exploités, le système attribuera directement une polarité, sinon calculera une pondération polaire correspondant à une émotion afin de proposer la classification la plus cohérente possible selon le niveau de granularité préalablement défini. Un exemple de système hybride est proposé dans (Maurel *et al.*, 2008) : il consiste en l’utilisation d’approches supervisées et symboliques afin d’extraire des corpus une plus grande quantité d’informations significatives en terme de sentiments et d’entités nommées (lieux et produits) et afin d’associer à ces derniers le ressenti des consommateurs.

En dehors des travaux qui utilisent ou complètent des lexiques ou dictionnaires déjà existants (Mihalcea & Strapparava, 2012; Mittal *et al.*, 2013), il existe aussi des travaux qui se consacrent à l’élaboration de nouveaux lexiques ou dictionnaires (Vu *et al.*, 2014; Mohammad & Yang, 2011), y compris sous forme d’ontologie (Ramteke *et al.*, 2013), afin de répondre à un besoin ou à une lacune de ressources dans un domaine d’étude ou bien dans une langue ciblée (Basile & Nissim, 2013). Nous récapitulons différents

lexiques ainsi que leur contenu et leur niveau de disponibilité dans la section 1.2.5.2. Notons aussi que des méthodes de bootstrapping⁸ (Qadir & Riloff, 2013; Velikovich *et al.*, 2010) permettent d'améliorer ou de créer de nouveaux lexiques spécifiques ou non (Chetviorkin & Loukachevitch, 2013) à un domaine ciblé ou à une langue ciblée.

1.2.2 Les algorithmes de l'apprentissage supervisé souvent utilisés

La classification des sentiments et des émotions est effectuée par l'intermédiaire de différents algorithmes de classification. Selon la méthodologie employée, l'utilisation des algorithmes sera adaptée au système d'apprentissage automatique mis en place. Dans la plupart des études, les auteurs évaluent leur système en le testant avec plusieurs algorithmes afin d'obtenir les meilleurs résultats possibles. Cela permet également de valider la ou les configurations optimales en fonction de la combinaison des caractéristiques choisies. Les algorithmes qui sont le plus souvent utilisés en classification des sentiments et émotions sont :

- le SVM (Support Vector Machine⁹),
- le Multinomial Naïve Bayes¹⁰,
- le maximum d'entropie, appelé aussi MaxEnt¹¹.

Dans les travaux de (Abdaoui *et al.*, 2014), le SVM donne les meilleurs résultats (sur une partie des tests seulement) avec une F-mesure supérieure à 94 % (contre un peu plus de 87 % pour le Naïve Bayes) en prenant en compte les unigrammes, les marqueurs d'incertitude, les émotions et les fautes d'orthographe, ce qui permet au final d'obtenir la configuration optimale. La tâche consistait à classer des énoncés de forum entre deux catégories : les énoncés écrits par les patients et ceux écrits par les professionnels de santé. Ensuite, les travaux de (Vo & Collier, 2013) ont consisté à classer les tweets concernant le tremblement de terre et le tsunami qui a ravagé le Japon le 11 mars 2011 dans le but d'analyser l'état émotionnel et affectif de la population tokyoïte à différentes périodes. Pour cela, les auteurs ont comparé les résultats de plusieurs algorithmes dont le Naïve Bayes, le SVM, le maximum d'entropie et le Multinomial Naïve Bayes, où ce dernier obtient les meilleurs résultats avec 64,7 % de F-mesure. Les caractéristiques utilisées sont les uni-, bi- et trigrammes de mots. Les travaux implémentant un SVM montrent que cet algorithme a su s'imposer dans la tâche de classification des sentiments et des émotions comme le soulignent les travaux de (Chetviorkin & Loukachevitch, 2013) consistant à classer des avis d'internautes en deux, trois ou cinq classes de sentiments (en fonction d'une échelle de sentiment définie par des annotateurs) ou ceux de (Jiang *et al.*, 2011) qui concernent une amélioration de la détection des émotions dans les textes de microblogs comme Twitter. Un autre algorithme ayant un certain succès dans l'analyse des sentiments et des émotions est le maximum d'entropie, comme par exemple son exploitation dans une tâche consacrée à l'analyse des sentiments dans les réseaux sociaux tchèques (Habernal *et al.*, 2013). Cette implémentation procure de meilleurs résultats que celle d'un SVM (mais à peine 0,02 de différence) et ce en y associant des critères tels que les unigrammes et bigrammes, les étiquettes morphosyntaxiques, les émoticônes et des n-grammes de caractères (permettant

8. Techniques d'amorçage à la construction de nouvelles données à partir d'une (très) faible quantité de données (appelées *seed terms* ou *termes germes*)

9. http://fr.wikipedia.org/wiki/Machine_%C3%A0_vecteurs_de_support

10. http://en.wikipedia.org/wiki/Naive_Bayes_classifier#Multinomial_naive_Bayes

11. http://en.wikipedia.org/wiki/Multinomial_logistic_regression

de récupérer les ponctuations et les émoticônes qui auraient été oubliés).

1.2.3 La visualisation des sentiments et des émotions

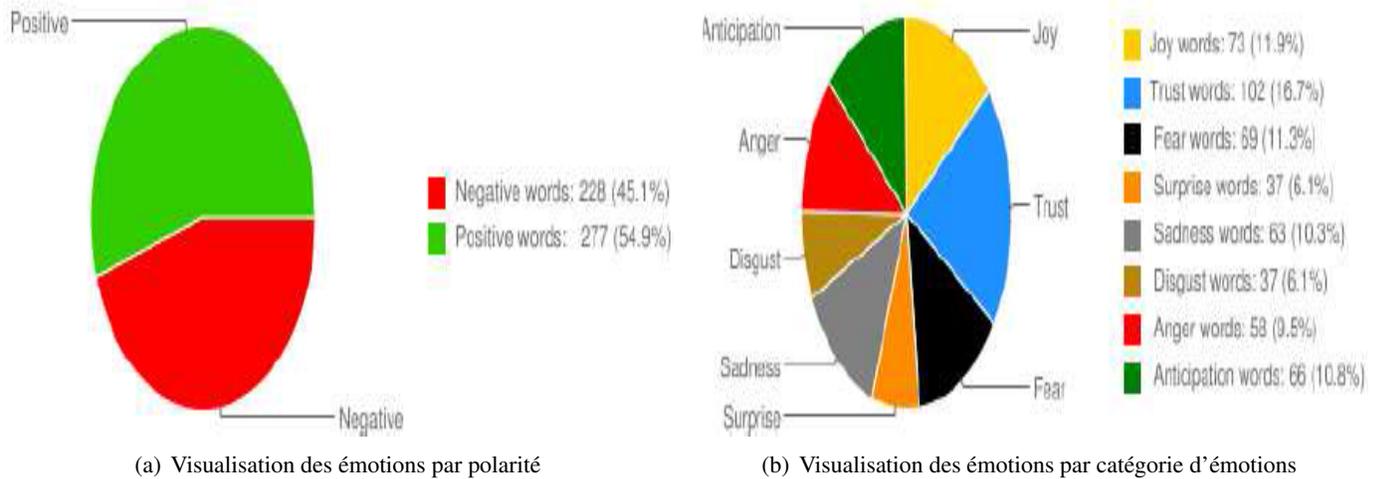
Les travaux portant sur la visualisation des émotions sont moins nombreux que ceux concernant la classification des sentiments et des émotions. La visualisation des émotions permet, grâce à une ou plusieurs représentations graphiques, d'interpréter l'état émotionnel, voire psychologique (Gauducheu, 2008), de la personne ayant exprimé cette émotion. Un récapitulatif des différents outils d'analyse des sentiments/émotions et de leur visualisation est disponible à la section 1.2.5.3 :

- L'étude de (Quan & Ren, 2014) a permis de mettre en place une interface proposant une représentation intuitive des émotions pour la langue chinoise dans des blogs, permettant alors de saisir les changements émotionnels. Leur système est également capable de reconnaître des émotions polysémiques ;
- Les travaux de (Cambria *et al.*, 2010), où les auteurs ont construit la base de connaissances SenticSpace qui mêle connaissances communes et affectives avec la possibilité de les visualiser tout en distinguant les concepts et les sentiments ;
- Une autre étude (Wecker *et al.*, 2014) a eu comme objectif de mettre en évidence les sentiments d'un document quelconque, comme un avis de consommateur, un blog, etc., tout en conservant le document original intact afin d'avoir un aperçu rapide et efficace de l'émotion manifestée dans un énoncé ;
- Le système de (Bista *et al.*, 2014) offre la possibilité de visualiser de différentes façons des données annotées issues des communautés en ligne. Les auteurs utilisent des graphes permettant de représenter des relations de similarité sémantique au travers de noeuds. Cela permet de représenter les relations entre des personnes et des concepts (comme les émotions) ;
- L'approche de (Mohammad & Yang, 2011) consistait à détecter et à quantifier les émotions dans les e-mails, afin d'étudier leur usage à travers différents types de e-mails : lettres d'amour, messages de haine et notes de suicide. Les auteurs ont mis au point différentes visualisations des émotions :
 - visualisation polaire (positif et négatif) (figure 1.1),
 - visualisation émotionnelle (les huit émotions de base décrites par (Plutchik, 1980)),
 - nuages de mots mettant en évidence les lexèmes les plus chargés en émotion (soit en fonction du type d'e-mail soit en fonction du genre de l'émetteur et du destinataire des e-mails de manière générale) (figure 1.1).

Les auteurs proposent également un type de visualisation en barres représentant le pourcentage d'émotion positive et négative exprimé par différentes personnes dans des e-mails envers une autre sur un lieu de travail. Puis ils terminent avec une nouvelle visualisation du même type que la précédente à l'exception qu'il s'agit cette fois de l'expression des émotions positives et négatives présentes dans des e-mails concernant deux personnes (émetteur et destinataire) selon une temporalité qui permet d'observer l'évolution des émotions de la personne émettrice envers celle recevant ces e-mails (figure 1.2).

- Les travaux de (Nalisnick & Baird, 2013) proposent une approche consistant à visualiser dynamiquement l'évolution des sentiments entre les personnages des pièces de Shakespeare. De cette manière, les auteurs ont pour objectif de déterminer la trajectoire émotionnelle des relations interpersonnelles

FIGURE 1.1 – Visualisation des émotions par polarité puis par catégorie d'émotions
suicide notes



plutôt que celle du texte dans sa globalité ou celle d'un personnage isolé. En utilisant les valences des unités lexicales chargées en émotion, ils vont pouvoir les additionner pour chaque instance de discours continue entre deux personnages afin d'observer l'évolution temporelle de leur relation au fil de la pièce (figure 1.3);

- L'approche proposée par (Bertrand *et al.*, 2013) permet de visualiser dans l'espace et le temps des émotions ressenties par les populations de New York, de Brooklyn et de Manhattan. Les auteurs effectuent ainsi un autre type de visualisation basé sur les coordonnées géographiques. En effet, on peut trouver ces coordonnées (mais pas systématiquement) dans les tweets, ce qui permet, dans le cadre de l'analyse des sentiments et des émotions, d'avoir un aperçu de l'état affectif dans lequel se trouve une population;
- Les travaux de (Basile & Nissim, 2013) exploitent également la méta-information géographique afin de justifier la qualité de l'approche proposée pour la détection des tweets en langue italienne. Les auteurs montrent clairement que les tweets sélectionnés par leur système proviennent effectivement d'Italie.

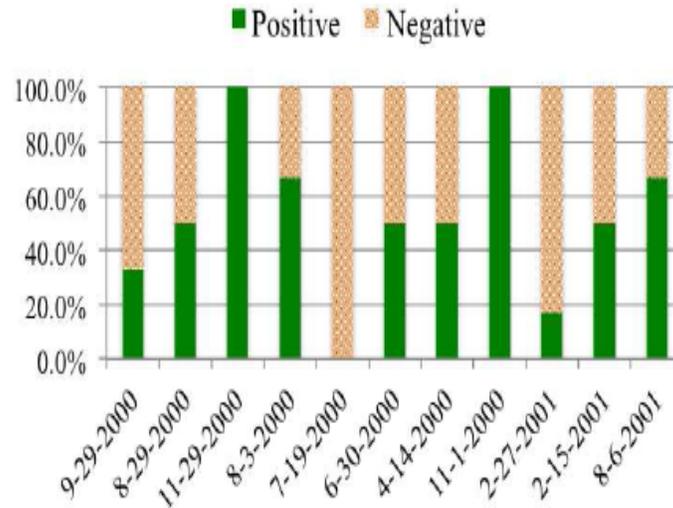
D'autres types de modélisation (figure 1.4) des émotions plus complexes existent comme par exemple la modélisation proposée dans les travaux de (Kempton *et al.*, 2014), détaillés également dans le tableau de la section 1.2.5.3.

1.2.4 Le domaine médical

En relation avec le domaine et le matériel de cette étude, que sont les forums de santé, il existe aussi quelques travaux sur la détection et l'extraction des émotions dans ce type de plateformes sociales :

- L'approche de (Melzi *et al.*, 2014) concerne les émotions ressenties par les patients en intégrant le lexique de (Mohammad & Turney, 2010). L'objectif consiste à extraire les émotions exprimées par

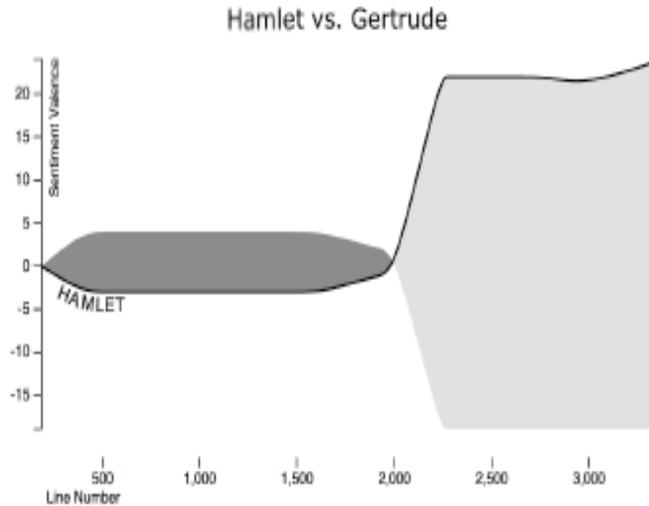
FIGURE 1.2 – Visualisation des émotions positives et négatives dans une dynamique temporelle



- des patients ainsi que les entités médicales qui y sont liées. Pour réaliser cela, les auteurs utilisent différentes caractéristiques comme les unigrammes et bigrammes, les smileys :-), :-(, les mots d'émotion, les ponctuations !!, ??, les répétitions de caractères *loool*. Mais ce sera la combinaison des n-grammes et des mots d'émotion qui fournira les meilleurs résultats avec une F-mesure ¹² de 63,6 % ;
- Les travaux de (Bringay *et al.*, 2014) ont comme objectif l'identification de la cible des émotions exprimées par les patients sur leur maladie, les symptômes ressentis ou encore leur traitement. Cette approche permettra par exemple de déterminer le nombre de messages de patients exprimant leur peur vis-à-vis d'un médicament en particulier. Par l'intermédiaire du lexique de (Mohammad & Turney, 2010) pour l'annotation automatique des termes chargés d'émotion et le lexique *MESH* pour l'annotation des concepts médicaux, puis de la plateforme *FrameNet* (voir section 1.2.5.2) pour l'analyse sémantique, les auteurs ont ainsi testé une approche basée sur l'analyse syntaxique qui montre une performance moindre par rapport à l'approche basée sur une analyse des rôles sémantiques des constituants phrastiques (par exemple, un expérimentateur, une cible et éventuellement une circonstance) ;
 - Dans l'étude de (Battaia, 2012), il s'agit d'analyser les émotions dans différents forums de santé afin de déterminer dans quelle mesure elles peuvent être un critère de pertinence lors d'une recherche sur un forum. L'auteur essaie de découvrir quels types d'informations sont recherchés par les patients sur ce type de plateforme. De cette manière, l'auteur est arrivée à la conclusion que les personnes malades vont sur les forums dans l'objectif d'échanger des témoignages et des expériences de vie liées aux maladies les concernant afin de valider les informations récupérées sur l'Internet. En utilisant des grammaires destinées à repérer les adjectifs, les noms et les verbes porteurs d'émotion ainsi que les smileys et la répétition exagérée des ponctuations, permet de localiser les phrases contenant des affects. Ensuite, l'auteur vérifie la présence de termes médicaux afin d'étudier la relation entre émotion et terme médicale ;

12. Mesure statistique permettant d'évaluer un système sur la base du rappel et de la précision calculés en amont : http://fr.wikipedia.org/wiki/Pr%C3%A9cision_et_rappel

FIGURE 1.3 – Visualisation de l'évolution des sentiments entre deux personnages dans une œuvre littéraire de Shakespeare



- La tâche proposée par (Abdaoui *et al.*, 2014) consiste à distinguer les auteurs des messages dans les forums de santé en fonction du contenu des messages, comme l'expression des émotions par les patients et l'établissement d'un diagnostic provisoire par les professionnels de la santé. Afin d'y parvenir, les auteurs ont pris en considération plusieurs marqueurs. Parmi les caractéristiques des messages de patients on trouve ainsi de subjectivité comme les émotions (lexique d'émotion pour le français d'Augustyn *et al.*, 2008) et les fautes d'orthographe (en plus de l'utilisation de l'argot, d'émoticônes, etc.). Parmi les caractéristiques des messages des médecins on trouve les incertitudes (noms, verbes, adjectifs et adverbes) et la faible présence de fautes d'orthographe. Les résultats montrent que le système proposé est applicable à d'autres forums contenant des messages de patients et de professionnel de santé ;
- Les travaux de (Elhadad *et al.*, 2014) s'articulent autour de l'élaboration d'un lexique organisé selon trois catégories sémantiques : les traitements, les symptômes et effets secondaires puis les émotions. Ce travail propose ainsi une étude qui se consacre à la génération d'un lexique représentatif et caractéristique d'une communauté, comme les patients d'un forum traitant du cancer du sein, destiné à répertorier les différents lexèmes utilisés par ces internautes. Le système proposé est capable, à partir d'un ensemble de termes germes, d'extraire de nouveaux termes appartenant à ces catégories sémantiques. Ces *termes germes* sont choisis d'après un seuil d'apparition dans leur corpus afin d'amorcer la génération du lexique à partir des termes les plus pertinents. Les auteurs utilisent des vecteurs de contexte pour chaque terme candidat dont ils vont calculer le degré de similarité avec le vecteur représentatif de la catégorie sémantique, calculé selon l'ensemble des *termes germes* sélectionnés. Pour ceci, un seuil de présence dans le corpus est appliqué, la casse et la complexité des unités lexicales sont exploitées, ainsi que l'appartenance des unités à un lexique spécialisé comme *UMLS* et *SIDER* (voir section 1.2.5.2). Enfin, le modèle vectoriel est basé sur des caractéristiques syntaxiques, considérant le premier et deuxième mot précédant et succédant les *termes germes* ainsi que l'apparition d'un

FIGURE 1.4 – Visualisation de la distribution des émotions par catégorie en fonction du nombre de tweets traités (faisant varier le diamètre du cercle)

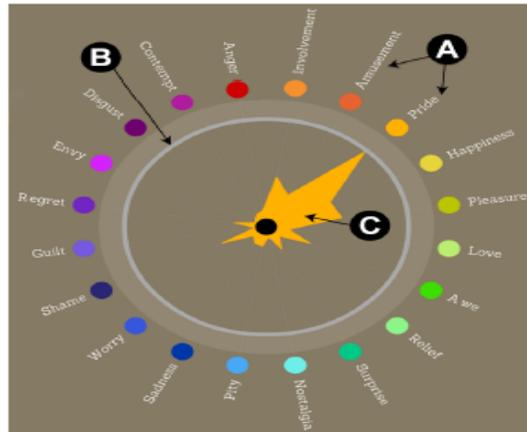


Figure 1: The emotion wheel. A - GEW emotion categories; B - Number of tweets visualized as the ring width; C - Emotion shape visualizing the emotional profile as a star plot.

mot dans une fenêtre de trois mots avant et après les *termes germes* ;

- L'étude de (Huh *et al.*, 2013) propose de classifier des textes issus de communautés sur la santé en ligne dont le but est de venir en aide aux modérateurs afin qu'ils puissent cibler les contenus nécessitant une aide médicale prioritaire. C'est en intégrant les unigrammes de mot, l'analyse des sentiments et la longueur des fils de discussion que les auteurs parviennent à sélectionner les messages prioritaires.

De manière générale, notons qu'un des attraits principaux de l'utilisation des forums dans le domaine de la santé réside dans la facilité et la liberté d'expression de façon anonyme par les patients de leurs maladies, symptômes, expériences, attentes ou espoirs. Ainsi, la communication des émotions médiatisée par ordinateur (Gauducheau, 2008) permet à ces internautes de ne pas subir le jugement éventuel d'autrui (Caplan & Turner, 2007) sous le couvert de l'anonymat (Hancock *et al.*, 2007).

1.2.5 Le matériel souvent exploités dans les travaux sur les émotions et les sentiments

Nous présentons dans cette section les corpus, les lexiques et les outils le plus souvent exploités et disponibles.

1.2.5.1 Les corpus

Les travaux dédiés à l'analyse des sentiments et des émotions ont permis la création de corpus de composition variée. En effet, de nombreuses recherches ont porté sur l'analyse des tweets, des forums et des plateformes de microblogging. Cette effervescence a entraîné la constitution d'une quantité importante de corpus d'entraînement, de développement et de test. Le contenu de ces corpus est assez varié : ils peuvent contenir les avis de consommateurs dans le but de repérer l'entité à l'origine de l'émotion, les fils de discussion des forums pour étudier la cible des émotions exprimées par les internautes.

Le tableau 1.1 présente une liste non exhaustive de corpus. Pour chaque corpus, nous indiquons sa disponibilité, les auteurs, la langue et en donnons une brève description (thématique, taille, etc).

<i>Corpus</i>	<i>Auteurs</i>	<i>Contenu</i>
Allodocteurs Corpus français	(Abdaoui <i>et al.</i> , 2014)	4 000 messages (2 000 par catégorie : patients/professionnels de santé) pour les données d'entraînement et 450 messages (même répartition) pour les données de test
Corpus d'avis et de critiques de consommateurs français	(Vincent & Winterstein, 2013)	14 000 textes issus de trois domaines différents : http://www.allocine.fr , http://www.amazon.fr et http://www.tripadvisor.fr . Corpus structuré en XML et annoté avec : identifiant de la critique, du produit, de l'auteur, descriptif du produit, note (de 1 à 5) associée à la critique (équivalent polaire positif et négatif), contenu de la critique
Czech dataset (disponible librement : http://liks.fav.zcu.cz/sentiment/) tchèque	(Habernal <i>et al.</i> , 2013)	91 381 critiques de film, annotées polairement (positif, négatif et neutre). 10 000 messages de Facebook, annotés polairement (positif, négatif, neutre et bipolaire). 145 307 avis de consommateurs, annotés polairement (positif, négatif et neutre)
Discursive relation tagged sentences (extrait du corpus ClueWeb09¹³) (disponible librement : http://nlg.csie.ntu.edu.tw/ntu-discourse/law-7.csv.gz) chinois	(Yu <i>et al.</i> , 2013)	7 638 phrases contenant 2 propositions, un marqueur du discours et 20 caractères chinois maximum pour une phrase. Annotation : polarité (positive, négative, neutre) des 2 propositions, de la phrase entière puis relation du discours (temporelle, comparaison, expansion, contingence)
Emoticon data set anglais	(Kouloumpis <i>et al.</i> , 2011) extrait du corpus de (Go <i>et al.</i> , 2009)	Prise en compte uniquement des données d'entraînement et des tweets positifs et négatifs : 230 811 positifs et 150 570 négatifs selon des émoticônes
Forum posts in health domain français	(Battaïa, 2012)	631 messages (8 fils de discussion) d'un forum généraliste modéré par un médecin (Forum Atoute). 582 messages de forums généralistes modérés (1 fil de Au féminin, 1 de Doctissimo, 1 de E-santé et 2 de Santé médecine). 621 messages de forums spécialisés modérés par un médecin (6 fils de Ligue Cancer et 14 de Traitement du pied bot à la naissance). 647 messages de forums spécialisés modérés (2 fils de Forum maladies Lysosomales, 2 de Renaloo, 3 de Solhand et 1 de Vivre sans Thyroïde)

13. <http://www.lemurproject.org/clueweb09/index.php#Specs>

Gender emotional axes corpus anglais	(Mohammad & Yang, 2011)	Love Letter corpus : 348 messages postés sur http://lovingyou.com . Hate Mail corpus : 279 extraits de mails de haine envoyés sur http://www.ratbags.com . Suicide Note corpus : 21 notes de suicides récupérées depuis http://www.well.com/art/suicidenotes.html#w
Hashtagged data set anglais	(Kouloumpis <i>et al.</i> , 2011) (Extrait du Edingurgh Twitter corpus de (Petrovic <i>et al.</i> , 2010))	+200 000 tweets contenant au moins un hashtag et annotés polairement (positif, négatif, neutre) selon les hashtags les plus fréquents (1 000 fois dans le corpus d'Edingurgh)
Health Forum Corpus anglais	(Bringay <i>et al.</i> , 2014)	345 phrases issues de messages du forum médical Spine-health. Annotations : les 6 émotions d'(Ekman, 1992) via le lexique de (Mohammad & Turney, 2010), entités médicales via MeSH, frames pertinentes de FrameNet (expressor pour l'émotion et topic et stimulus pour la cible médicale)
iSieve data set www.i-sieve.com anglais	(Kouloumpis <i>et al.</i> , 2011)	4 000 tweets annotés manuellement en relation avec une thématique accompagnés d'une annotation polaire : positive, négative ou neutre
News titles anglais	(Strapparava & Mihalcea, 2008)	250 titres pour le corpus de développement et 1 000 pour celui de test. Annotations : valence émotionnelle de 0 à 100, polarité émotionnelle de -100 (très négatif) à +100 (très positif) pour 6 émotions
Political tweet corpus anglais irlandais	(Bakliwal <i>et al.</i> , 2013)	2 624 tweets politiques relatifs aux cinq partis politiques les plus importants ainsi que leurs leaders. Annotation avec 6 catégories polaires (sentiments exprimé envers l'entité politique) : positive, négative, neutre et annotation si le tweet est sarcastique. Précision des parties du discours
Reviews anglais	(Ganu <i>et al.</i> , 2009)	+50 000 avis sur des restaurants de Citysearch New York. Métadonnées présentes : lieu, type de cuisine, "star rating" ¹⁴ , date et 3 400 phrases annotées (positive, neutre, négative ou conflit)
Tweets anglais et espagnol	(Etter <i>et al.</i> , 2013)	30 000 tweets espagnol et 10 000 en anglais, annotés avec 10 entités nommées (personne, organisation, lieu, etc.) via la plateforme AMT ¹⁵

14. Système d'évaluation basé sur un nombre d'étoile (cinq en général). Plus l'utilisateur coche d'étoiles et plus l'appréciation est meilleure

15. Amazon Mechanical Turk permet d'annoter des corpus par une communauté en contrepartie d'une rémunération : www.mturk.com/mturk/welcome

Twitta italien	(Basile & Nissim, 2013)	100 millions de tweets annotés avec des méta-informations : horodatage, coordonnées géographiques (si présentes), nom d'utilisateur, étiquettes morpho-syntaxiques et polarité (positive, négative, neutre)
--------------------------	-------------------------	---

TABLE 1.1: Quelques corpus à disponibilité variable construits pour l'analyse des sentiments et des émotions

1.2.5.2 Les lexiques

Avec l'augmentation des études sur les sentiments et les émotions, il devient donc indispensable de constituer les lexiques nécessaires afin de proposer une analyse correcte des corpus. Ainsi, plusieurs lexiques des termes affectifs ont vu le jour, tandis que d'autres sont enrichis afin de mieux satisfaire leur couverture. Lorsque de tels lexiques sont mis à disposition de la communauté des chercheurs, ils constituent des ressources extrêmement précieuses. En effet, ces ressources proposent un éventail d'informations lexicales, syntaxiques et sémantiques précises afin d'améliorer la détection et l'extraction des entités lexicales affectives dans les textes issus du web. Cependant, la nature subjective des textes de plateformes sociales sur l'Internet, comme les réseaux sociaux ou les forums, rend la tâche d'analyse des sentiments et des émotions complexe. La manifestation de la subjectivité, omniprésente dans ce genre d'énoncés, se présente sous différentes formes. Elle peut être exprimée explicitement par la terminologie émotionnelle à travers les adjectifs, les noms ou encore les adverbes ou implicitement par l'intermédiaire de formules rhétoriques comme l'ironie ou le sarcasme (González-Ibáñez *et al.*, 2011). Certains travaux dénoncent clairement la difficulté à mettre en place un système capable de détecter automatiquement les expressions sarcastiques ou ironiques (Haruechaiyasak *et al.*,). Ceux-ci jouent pourtant un rôle important dans l'expression des sentiments et des émotions, que ce soit par les patients dans les forums médicaux exprimant une critique à l'égard d'un professionnel de santé ou bien par des internautes échangeant sur des sujets moins spécialisés au sein d'une même communauté en ligne (jeux vidéo, matériel informatique, etc.). Nous présentons dans le tableau 1.2 quelques ressources lexicales existantes¹⁶. Nous indiquons les types d'informations contenues dans ces ressources : la polarité des termes affectifs, leur valence (ou pondération), leur étiquette morpho-syntaxique, la présence de synonymes, etc.

<i>Lexiques</i>	<i>Auteurs</i>	<i>Contenu</i>
Affective Norms for English Words http://csea.php.ufl.edu/media/requestform.html Anglais, espagnol, allemand	M.M. Bradley & P.J. Lang, 1999 Université de Floride (CSEA)	1 034 entrées auxquelles sont affectées une note (1 à 9) selon trois dimensions : le plaisir, l'excitation et la dominance

16. http://neuro.imm.dtu.dk/wiki/Sentiment_analysis
<http://sentiment.christopherpotts.net/lexicons.html#liwc>

AFINN http://www2.compute.dtu.dk/~faan/data/AFINN.zip Anglais	Finn Årup Nielsen, 2009-2011	2 477 mots et syntagmes auxquels a été attribuée une valence entre -5 (négatif) et +5 (positif)
BL Lexicon www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar Anglais	M. Hu & B.Liu, 2004	Entrées composées de termes positifs (2 006) et négatifs (4 783) contenant des formes avec des fautes d'orthographe, des variantes morphologiques, de l'argot et des formes propres aux plateformes de médias sociaux
CAO (database) http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/repository/cao.htm Japonais	(Ptaszynski <i>et al.</i> , 2010)	10 137 émoticônes répertoriés, qui sont utilisés dans les plateformes japonaises de communication en ligne. Ils définissent une émotion précise (joie, tendresse, colère, surprise, tristesse et excitation ; pour les plus représentée)
CLex contact : svitlana@jhu.edu Anglais	(Volkova <i>et al.</i> , 2012)	Lexique d'associations couleur-émotion-concept. 3 397 termes d'affect uniques représentant des sentiments, des émotions, des attitudes et des humeurs. 1 957 concepts uniques. Plus de 2 300 termes de couleur. Indications morpho-syntaxiques (noms, adjectifs, adverbes et verbes)
Dictionary of Affect in Language http://sail.usc.edu/dal_app.php Anglais	Whissell, 1989	8 742 entrées évaluées d'après trois dimensions : le degré de plaisir (plaisant, déplaisant, entre les deux), le degré d'activation (passif, actif, entre les deux), le degré d'imagination (difficile à imaginer, facile à imaginer, entre les deux)
FrameNet https://framenet.icsi.berkeley.edu/fndrupal/framenet_request_data Anglais	(Fillmore <i>et al.</i> , 2003)	Base de données contenant plus de 170 000 phrases annotées à la main. Lexique organisé en "frame sémantique" (type d'événement, de relation ou d'entité et les participants liés). Chaque concept (nourriture par exemple) est défini à l'aide d'"frame éléments" dont l'instance conceptuelle se nomme "unité lexicale" (poulet par exemple)
General Inquirer contact : rhuu@csail.mit.edu Anglais	P.J. Stone, D.C. Dunphy, M.S. Smith & D.M. Ogilvie, 1966 Harvard	Plusieurs dictionnaires : termes polaires (1 915 positifs, 2 291 négatifs, 7 582 objectifs (neutre)), termes d'intensité (1 902 intenses, 755 faibles), termes émotionnels, sentimentaux, termes liés aux institutions, lieux, animaux, etc.

HowNet http://www.keenage.com/html/e_index.html Chinois	Z. Dong & Q. Dong, 1988	Plus de 50 000 concepts définis avec des relations sémantiques entre eux (hyperonymie, synonymie, etc.)
Lexique sémantique de verbes français pour les émotions Français	(Mathieu, 2006)	600 verbes (décrivant les sentiments, les émotions et les états psychologiques) organisés en 33 classes sémantiques. Trois niveaux de catégorisation : positif, neutre, négatif. Trois types de liens entre les classes : le sens, l'intensité et l'antonymie
Linguistic Inquiry and Word Count Anglais, allemand, espagnol, italien, néerlandais	L.W. Pennebaker, R.J. Booth & M.E. Francis	À l'aide de dictionnaires spécialisés, extraction de mots appartenant à différentes dimensions linguistiques comme les émotions positives et négatives
Medical Subject Headings http://mesh.inserm.fr/mesh/ Anglais, français	National Library of Medicine, 1954	Thésaurus biomédical contenant 27 149 descripteurs (termes et concepts) organisés en 16 catégories thématiques (anatomie, organismes, etc.), elles-mêmes organisées hiérarchiquement (jusqu'à 11 niveaux)
Micro Opinion Lexicon http://www.unipv.it/wnop/micrownop.tgz Anglais	(Cerini <i>et al.</i> , 2007)	Sous-lexique de WordNet, composé de 1 105 synsets (contre plus de 100 000 dans WordNet). Ces synsets sélectionnés expriment une forme d'opinion et sont représentatifs de la distribution des parties du discours dans WordNet
Multi-Perspective Question Answering (OpinionFinder's Subjectivity Lexicon) http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/ Anglais	(Wilson <i>et al.</i> , 2005)	8 221 entrées associées à diverses informations : intensité subjective (faible, forte), polarité (positive, négative, neutre), étiquette morpho-syntaxique, indique si la forme est une racine (stem)
Multi WordNet http://hlt.distributor.fbk.eu/index.php Italien	(Pianta <i>et al.</i> , 2002)	Version italienne de WordNet (alignée avec WordNet version anglaise) avec la possibilité d'aligner les termes avec d'autres langues. Informations sur les relations lexicales entre mots, relations sémantiques entre concepts, champs sémantiques (domaine)
NRC word-emotion association (EmoLex) http://www.saifmohammad.com/emoform/emotionForm.html Anglais	(Mohammad & Turney, 2010)	24 200 signifiés pour 14 200 signifiants. Chaque signifié est associé à une émotion par annotation humaine (AMT ¹⁷). Indication de polarité (positive, négative) et association à une des huit émotions de base (Plutchik, 1980)

17. Amazon Mechanical Turk : plateforme de crowdsourcing d'annotation humaine

<p>OpLexicon http://ontolp.inf.pucrs.br/Recursos/downloads-OpLexicon.php Portugais</p>	<p>(Souza <i>et al.</i>, 2011)</p>	<p>7 077 entrées composées de hashtag, émoticônes, verbes, adjectifs avec polarisation : 1 (positif), 0 (neutre), -1 (négatif)</p>
<p>Roget's Thesaurus http://www.thesaurus.com/Roget-Alpha-Index.html, http://www.roget.org/ Anglais</p>	<p>P.M. Roget, 1852</p>	<p>Thésaurus composé de 6 classes primaires, contenant des subdivisions organisées en section (http://en.wikipedia.org/wiki/Wikipedia:Outline_of_Roget%27s_Thesaurus). Les unités lexicales (connotations, sens) sont reliées sémantiquement à leur concept</p>
<p>Sentiment140 www.umiacs.umd.edu/~saif/WebDocs/Sentiment140-Lexicon-v0.1.zip Anglais</p>	<p>(Mohammad <i>et al.</i>, 2013)</p>	<p>Lexique de plus d'un million d'entrées basé sur des indices de sentiments et d'émotions (émoticônes, ponctuations, répétition de caractères, etc.), produits uniquement à partir d'un corpus de tweets annotés soit positif soit négatif selon l'émoticône</p>
<p>SentiWordNet http://sentiwordnet.isti.cnr.it/download.php Anglais</p>	<p>(Esuli & Sebastiani, 2006) (version 1.0) (Baccianella <i>et al.</i>, 2010) (version 3.0)</p>	<p>Formes lexicales présentes dans WordNet. À chaque terme est attribué un score entre 0 et 1 pour chacune des polarités : positive, négative et objective (neutre)</p>
<p>SIDER ftp://sideeffects.embl.de/SIDER/ Anglais</p>	<p>(Kuhn <i>et al.</i>, 2010)</p>	<p>Lexique médical connectant 996 médicaments à 4 192 effets secondaires. Contient des informations sur la fréquence des effets secondaires (fréquent, moins fréquent, rare) liés à un médicament ou à un placebo</p>
<p>Taboada and Grieve's Turney adjective list disponible via Yahoo! sentimentAI group Anglais</p>	<p>(Taboada & Grieve, 2004)</p>	<p>Adjectifs classés en trois catégories et évalués d'après leurs contextes : affectif (émotion positive ou négative), jugement (moralement positif ou négatif) et appréciation (esthétisme positif ou négatif)</p>
<p>VerbNet http://verbs.colorado.edu/~mpalmer/projects/verbnet/downloads.html Anglais</p>	<p>(Schuler, 2005)</p>	<p>Verbes organisés en classes hiérarchisées avec des indications sur leurs propriétés syntaxiques et sémantiques. Compatibilité avec WordNet</p>

WordNet http://wordnet.princeton.edu/download/current-version/ Anglais et certaines autres langues ¹⁸	Princeton University	Noms, verbes, adjectifs et adverbes. Synsets (ensemble de synonymes) représentant des concepts liés entre eux par des relations sémantiques. Plus de 100 000 synsets
WordNet Affect http://wdomains.fbkeu/download.html Anglais	(Strapparava & Valitutti, 2004)	Extension de WordNet en associant des termes affectifs aux concepts dans WordNet. Possibilité d'assigner une ou plusieurs étiquettes affectives pour un groupe de synsets. Différentes labélisations possibles : émotion, humeur, attitude, états (cognitif, physique), etc. Distinction de la valence émotionnelle : étiquettes positive, négative, neutre et ambiguë

TABLE 1.2: Lexiques de sentiments, d'émotions et médicaux

1.2.5.3 Les outils

Dans le tableau 1.3, nous présentons quelques outils développés pour des tâches d'analyse de sentiments et d'émotions. Certains de ces outils proposent également des modules de visualisations des émotions : nuages de mots, graphiques en bâtons, représentations circulaires, etc. Ces outils rendent plus facile l'interprétation des émotions exprimées dans les textes issus du web. De plus, certaines de ces visualisations permettent d'aborder l'aspect dynamique d'évolution temporelle des émotions au fil d'un texte ou d'un fil de discussion par exemple.

<i>Outils</i>	<i>Auteurs</i>	<i>Objectif de la plateforme</i>
Calico http://woops.crashdump.net/calico/index.php?page=home&cwid=0f3af3263eac90ba5ee3c39e94209ea4f82d26cf	Laboratoires GREYC UMR 6072 et STEF	Crawle des forums (en français) et récupère les fils de discussion, offre la possibilité de travail collaboratif, propose plusieurs modules de visualisation
ConceptNet http://conceptnet5.media.mit.edu/downloads/current/	R. Speer, C. Havasi, J. Chaidez, J. Venezuela & Y.-L. Kuo, début des années 2000	Réseau sémantique contenant des connaissances du monde afin de faciliter "l'assimilation sémantique" par les ordinateurs. Les noeuds sont les concepts (unités lexicales simples ou composées) liés par des relations sémantiques. Disponible en anglais et japonais
EmotionWatch	(Kempton <i>et al.</i> , 2014)	Reconnaissance automatique des émotions dans les médias sociaux en proposant un résumé visuel des réactions du public corrélées à des événements sous la forme d'une roue des émotions. Anglais

18. http://fr.wikipedia.org/wiki/WordNet#Dans_d.27autres_langues

Étiqueteur morpho-syntaxique de Stanford http://nlp.stanford.edu/software/lex-parser.shtml#Download	Stanford University (D. Klein, C. Manning)	Analyseur en dépendances syntaxiques. Anglais, chinois, arabe, allemand, italien, bulgare et portugais
NOOJ (anciennement In-text) http://www.nooj4nlp.net/pages/download.html	M. Silberztein, 2002	Environnement de développement linguistique avec des dictionnaires et grammaires (couverture large). Permet d'analyser des corpus de gros volume en temps réel, la construction de concordances et l'extraction d'entités nommées. Une douzaine de langues
Ogmios http://search.cpan.org/~thhamon/Alvis-NLPPlatform/	(Hamon <i>et al.</i> , 2008)	Plateforme destinée à traiter de gros volumes documentaires et spécialisés en construisant à partir de ces données des ressources spécialisées (enrichissement des documents avec des annotations linguistiques et sémantiques). Français, anglais (slovène, chinois)
Opinion Finder http://mpqa.cs.pitt.edu/opinionfinder/opinionfinder_2/opinionfinder_2_0/	Pittsburgh university, Cornell university & university of Utah, 2005	Analyse des documents (en anglais) en identifiant les phrases subjectives avec ses différents aspects : agents à l'origine de l'opinion, expressions subjectives elles-mêmes et expressions de sentiment
Opinion Observer	(Liu <i>et al.</i> , 2005)	Visualisation des avantages et inconvénients d'un produit selon les avis d'utilisateurs (selon divers caractéristiques du produit ; Anglais)
S-Sence	(Haruechaiyasak <i>et al.</i> ,)	Analyse des sentiments dans les médias sociaux thaïlandais. Deux modules d'analyse capables de classer un texte comme étant intentionnel ou sentimental
SentiStrength http://sentistrength.wlv.ac.uk/download.html	(Thelwall <i>et al.</i> , 2010)	Estimation de l'intensité polaire (positive, neutre, négative ; de -5 à 5) des textes courts, formels ou informels. Précision proche de l'humain pour l'analyse des textes de réseaux sociaux en anglais. Disponible en 14 langues
Semafor http://www.ark.cs.cmu.edu/SEMAFOR/	(Das <i>et al.</i> , 2010)	Analyse automatique de la structure sémantique de textes, en anglais, via la ressource lexicale FrameNet
Sentiment140 http://help.sentiment140.com/api/registration	(Go <i>et al.</i> , 2009)	Détection du sentiment à l'égard d'une marque, d'un produit ou d'une thématique sur Twitter. Anglais

Sybill	Holmes Semantic Solutions (autrefois, Celi-France)	Analyse lexicale (identification et pondération des mots positifs et négatifs), syntaxique et sémantique de données textuelles diverses. Français
TweetFeel http://web.peanutlabs.com/	Conversation Strategies	Détecte en temps réel le sentiment lié à un mot-clé ou expression clé utilisant les tweets
We Feel Fine http://wefeelfine.org/	(Kamvar & Harris, 2011)	Moteur de recherche émotionnel

TABLE 1.3: Outils de visualisation des émotions, d'analyse de contenus issus de forums, de réseaux sémantiques et de traitements linguistiques

1.3 Objectifs de notre travail

L'objectif de notre travail consiste à modéliser l'évolution des émotions dans le but de faciliter l'accès aux informations contenues dans les fils de discussion des forums de santé pour les professionnels de santé et les patients. Cette modélisation permettra aussi d'observer l'évolution des émotions dans le cadre de deux types d'échanges : patients/patients et patients/professionnels. Afin d'y parvenir nous proposons de mettre en place des mécanismes de visualisation de cette évolution. De plus, les émotions exprimées par les patients dans leurs messages sont souvent contextualisées par la présence d'autres informations sémantiques, comme :

- des marqueurs d'incertitude (*peut-être, possible, si, ?, etc.*),
- des modificateurs d'intensité (*peu, beaucoup, vraiment, à peine, etc.*),
- la négation (*ne...pas, ne...plus, aucun, etc.*),
- des connecteurs syntaxiques (*et, mais, car, etc.*).

Nous proposons également d'étudier la portée de la négation sur les marqueurs émotionnels, qui a pour effet d'inverser la polarité de l'émotion exprimée. Nous nous concentrons aussi sur la détection d'unités linguistiques pertinentes pouvant altérer la polarité et l'intensité d'un marqueur émotionnel. Notre étude s'articule autour de trois aspects de la linguistique (lexical, syntaxique et sémantique) :

- Sur le plan lexical, il s'agit de récupérer les termes correspondant aux différentes classes de marqueurs à partir de lexiques dédiés et à les compléter en fonction des besoins ;
- Sur le plan syntaxique, nous effectuons une étude des contextes associés aux émotions, dans lesquels la position de ces marqueurs peut impacter le degré de subjectivité ainsi que l'orientation sémantique exprimés par les patients ;
- Sur le plan sémantique, nous cherchons à déterminer la nature sémantique des entités grâce à l'étiquetage approprié (*trouble, traitement, négation, l'émotion peur, etc.*) des termes présents dans les fils de discussion.

À travers ces trois aspects, nous proposons d'étudier les traces d'émotions en repérant également dans les messages l'utilisation des smileys (:), :(, :D, les marques de ponctuation !, ?, ? ? ! !, qui sont des substituts de la communication non verbale (Gauducheau, 2008), et des marqueurs linguistiques spécifiques (*avoir peur,*

être rassurée, etc.). Il sera également question de prendre en compte l'expression de l'incertitude dans les textes des patients et des professionnels de santé afin d'en étudier leur utilisation, et d'étudier la relation entre les émotions et l'incertitude. Ces différents marqueurs linguistiques interagissant avec des marqueurs émotionnels seront traités selon un système à base de règles. Nous verrons que l'établissement de priorités dans l'application de ces règles est nécessaire. Ce système essaiera de déduire la polarité des émotions exprimées par les patients et professionnels avec le plus de précision possible grâce à l'application de ces règles. Le niveau de granularité choisi sera l'énoncé : question ou réponse. Nous nous intéresserons à l'énoncé dans son entier (considération de tous les marqueurs émotionnels présents dans une question ou dans une réponse). La dimension temporelle sera marquée par un point de départ : la question posée par un patient, contenant une ou plusieurs émotion(s) initiale(s), puis un point d'arrivée : la réponse formulée par des patients ou par un professionnel de santé, contenant une ou plusieurs émotions nouvelles (transformées). Afin de permettre la visualisation de cette dynamique temporelle, nous générerons différents types de modélisation facilitant la visualisation et l'interprétation de cette évolution :

- des graphiques en barres ou histogrammes ;
- des diagrammes circulaires interactifs ;
- des "constellations" ;
- des nuages de mots interactifs.

Chapitre 2

Méthodes et matériel

L'objectif initial de ce travail était de prendre en considération trois types de textes médicaux : articles scientifiques (rédigés par des chercheurs), rapports cliniques (rédigés par des médecins), messages de forum de santé (rédigés par des patients) (Chauveau-Thoumelin & Grabar, 2014). Cependant, très rapidement, nous nous sommes rendus compte que la quantité d'émotions manifestées par les professionnels de santé et les chercheurs n'était pas significative pour notre étude. Nous avons donc décidé de nous concentrer sur l'étude de corpus de forum médicaux, caractérisés par une manifestation d'émotions importante par les patients. L'évolution des émotions, simulée par les échanges d'informations entre les questions et les réponses, sera caractérisée par l'interaction aussi bien syntaxique que sémantique entre les différents descripteurs lexicaux (*modifieurs, incertitudes, négation et connecteurs*) et les marqueurs émotifs, lexicaux (lexèmes décrivant explicitement une émotion) ou non (association de caractères exprimant une émotion implicitement comme les smileys ou la répétition de ponctuation). L'impact des divers marqueurs sur les marqueurs émotionnels permettra de déterminer l'intensité (atténuation ou renforcement émotionnel) et la polarité (positive, négative ou neutre) des formulations émotionnelles exprimées par les patients.

2.1 Matériel

Nous utilisons deux types de matériel décrits dans la suite de cette section :

- Deux corpus (*Q/A* et *forum*) de nature différente, s'articulant autour de thème assez divers pour le corpus *Q/A* (sexologie, grossesse, dermatologie, etc.) et du thème de la rhumatologie (maladies des os, des articulations ou des muscles) pour le corpus *forum* ;
- Des ressources linguistiques comprenant des terminologies médicales (de trois types différents : *procédures, maladies/troubles et médicaments/traitements*) et des lexiques regroupant les différents descripteurs et les émotions.

2.1.1 Les corpus

2.1.1.1 Description et organisation des corpus

Nos travaux sont axés sur deux types de corpus, dont quelques détails sont fournis dans le tableau 2.1 :

- Le *corpus Q/A* contient un ensemble de questions posées par des patients et un ensemble de réponses fournies par des professionnels de santé. Ce corpus a été construit à partir du site français *MaSante-Net*¹ par (Abdaoui *et al.*, 2014). Le corpus comprend 6,140 fils de discussion formés autour de questions des patients et des réponses correspondantes des médecins. Cela représente un total de 12,280 énoncés. Pour notre analyse, deux jeux de données ont été produites : (1) un sous-corpus d'apprentissage/développement de 1,000 questions et de 1,000 réponses associées puis (2) un sous-corpus de test de 500 questions et les 500 réponses associées, qui nous servira à évaluer notre système. Un échantillon de question et réponse est disponible à l'annexe A.1 ;
- Le *corpus forum* contient un ensemble de fils de discussion où les patients interagissent entre eux. Ces données ont été constituées à partir du site *Doctissimo*² dans le cadre des travaux de (Chauveau-Thoumelin & Grabar, 2014). Dans ce corpus, un fil de discussion est formé autour d'une question posée par un patient et de réponses correspondantes fournies par d'autres patients. Typiquement, les patients parlent de leurs expériences vécues, leurs niveaux de connaissances, donnent des avis sur les traitements, pathologies etc. Ce corpus contient 1,124 fils de discussion composés chacun d'une question et d'une ou plusieurs réponses, pouvant aller jusqu'à une dizaine de réponses. À partir de ces données, deux sous-corpus ont été créés respectant une certaine homogénéité avec le précédent corpus en quantité de données : (1) un sous-corpus pour l'entraînement/développement, composé de 150 fils et (2) un sous-corpus pour les tests afin de pouvoir évaluer notre système sur des types de données différentes, composé de 80 fils de discussion. Un échantillon de question et réponses est disponible à l'annexe A.2.

TABLE 2.1 – Description détaillée des corpus *Q/A* et *forum*

	Corpus Q/A	Corpus forum	Total
Nombre de documents	12,280	1,124	13,404
Nombre de mots	993,383	1,763,022	2,756,405

Ces corpus sont prétraités en suivant les spécifications proposées dans les travaux précédents (Chauveau-Thoumelin & Grabar, 2014). Il s'agit essentiellement de :

- convertir le format des données d'origine au format texte brut ;
- normaliser ces documents en les encodant en UTF-8³ puis en gérant la casse (majuscule et minuscule) et les caractères spéciaux comme les espaces insécables, les apostrophes typographique, les ligatures (*œ*, *æ*, etc.), les retours à la ligne, etc.

2.1.1.2 L'annotation linguistique et sémantique des corpus

Les corpus ont été annotés avec la plateforme *Ogmios* (Hamon *et al.*, 2008). C'est une plateforme adaptée à l'analyse sémantique spécialisée de documents. Elle permet d'enrichir les documents sources avec des

1. www.masantenet.com[collectÃ1' :18/02/2014]

2. <http://forum.doctissimo.fr/sante/douleur-dos>

3. Pour des informations détaillées sur cet encodage : <http://fr.wikipedia.org/wiki/UTF-8>

annotations sémantiques. Ces informations sémantiques sont obtenues à partir de l'articulation entre les différents modules de traitement automatique des langues et l'utilisation de lexiques :

1. *L'annotation linguistique* va permettre de préciser la catégorie morpho-syntaxique des mots (*ADJ* pour adjectif, *ADV* pour adverbe, *NOM* pour nom, etc.) et leur lemme (le lemme *aimer* pour la forme *aime*, le lemme *bon* pour la forme *bonnes*, etc.). C'est l'outil *TreeTagger* (Schmid, 1994), qui assure cette annotation ;
2. *L'annotation sémantique* est effectuée grâce à la reconnaissance de termes provenant de lexiques fournis à la plateforme. De plus, une détection des marques émotives non lexicales, caractéristiques des forums, est aussi réalisée. Il s'agit de :
 - (a) émoticônes ou smileys : (:), :(, ^, etc. ;
 - (b) ponctuations répétées et expressives : ? ? ? ? ?, !!!, ? ? !!, etc. ;
 - (c) communicateurs de rire : *mdr*, *lol*, *haha*, etc. ;
 - (d) répétition de caractères dans un mot : *svppppp*, *merciiii*, etc.

Ces marques émotives non lexicales sont détectées avant la tokenisation afin que l'outil *TreeTagger* les annote comme des unités lexicales à part entière. Cela évite à *TreeTagger* de considérer *!!!* comme trois tokens différents.

2.1.2 Les ressources

Les informations sémantiques sont fournies par l'intermédiaire de diverses ressources :

1. Les *ressources sémantiques* : notions médicales, appartenant à trois types sémantiques (*procédure*, *maladie*, *médicament*) ;
2. Les *ressources linguistiques* : marqueurs de subjectivité (*incertitude*, *négation*, *modificateurs* et marqueurs d'émotions).

Nous décrivons l'ensemble de ces ressources dans la suite de cette section.

2.1.2.1 Les ressources sémantiques

Les *ressources sémantiques*, qui ont permis l'annotation des termes représentant les trois types sémantiques, sont constituées en utilisant différents lexiques spécialisés dans le domaine médical :

- les **procédures** : ces termes proviennent de la terminologie SNOMED international⁴ (Côté, 1996). Ils décrivent des actes médicaux effectués par les médecins ;
- les **maladies** : ces termes proviennent également de la terminologie SNOMED international décrivant des maladies, des problèmes médicaux ou des troubles ;
- les **médicaments** : ces termes proviennent de plusieurs sources :
 1. base Thériaque⁵ créée par le CNHIM⁶ ;

4. Systematized Nomenclature of Human and Veterinary Medicine : AgencedesSystèmesd'InformationPartagésdeSanté : <http://esante.gouv.fr/asip-sante>

5. <http://www.theriaque.org/>

6. Centre National Hospitalier d'Information sur le Médicament : <http://www.cnhim.org/>

2. base de terminologies biomédicales UMLS⁷ ;
3. base UCD⁸.

Ces termes sont étiquetés avec trois types sémantiques : *procedure* pour les procédures, *disorder* pour les maladies et *medication* pour les médicaments. Un échantillon de ces ressources est disponible à l'annexe B.1.

Cependant, il est important de noter que la nature des données issues de forums posent quelques contraintes. En effet, il n'est pas rare que les messages postés sur les forum par les patients contiennent des fautes d'orthographe. Afin de pallier au moins partiellement à cette lacune, l'objectif consiste à calculer la distance de Levenshtein (Levenshtein, 1966). Cette technique permet de mesurer la similarité entre deux chaînes de caractères, en considérant trois opérations :

1. l'ajout d'un caractère ;
2. la substitution d'un caractère par un autre ;
3. la suppression d'un caractère.

Cela permet par exemple, de faire la correspondance entre le mot *convultions* du corpus et le terme *convulsions* du lexique de référence en substituant la lettre *t* par la lettre *s*. Le mot *convultions* peut ainsi être annoté avec un type sémantique. Son étiquette sémantique sera la même que celle du terme présent dans le lexique de référence (*convulsions*) : dans notre cas, il s'agit du type *disorder*. Ces termes mal orthographiés porteront la mention *_otr* à la suite de leur étiquette sémantique (*disorder_otr*) afin de pouvoir les distinguer. De plus, les ressources ont été rendues plus exhaustives par la prise en compte des marques de pluriel (*s* ou *x*) pour chaque entrée simple. Enfin, quelques ajouts manuels ont été effectués en analysant le contenu du corpus.

Un récapitulatif du contenu des différentes terminologies est présenté dans le tableau 2.2, en fonction des trois types sémantiques (Chauveau-Thoumelin & Grabar, 2014).

TABLE 2.2 – Distribution des types sémantiques *procedure*, *disorder* et *medication*

	Procédure	Maladie	Médicament	Total
Nombre d'entrées	24,472	65,787	8,386	98,645
Nombre d'ajouts	1,500	6,593	298	8,391
Total	25,972	72,380	8,684	107,036

2.1.2.2 Les ressources linguistiques

Les *ressources linguistiques* proviennent également des travaux précédents (Chauveau-Thoumelin & Grabar, 2014). Nous utilisons plusieurs lexiques :

- Le *lexique des descripteurs* contient trois types de marqueurs :

7. Unified Medical Language System (développées par la US National Library of Medicine) : <http://www.nlm.nih.gov/research/umls/>

8. Unité commune de dispensation (liste de médicaments mis sur le marché et commercialisés en France) : <http://www.ucdcip.org/>

- Les marqueurs de type *incertitude* sont composés de verbes (*paraître, sembler, risquer, etc.*), de noms (*doute, probabilité, chance, etc.*), d’adjectifs (*plausible, vraisemblable, éventuel, etc.*) et d’adverbes (*sûrement, peut-être, apparemment, etc.*). Ils sont définis par deux niveaux d’intensité :
 - l’incertitude forte (*hypothèse, suspecter, pouvoir, etc.*) atténue fortement la fiabilité d’une information,
 - l’incertitude faible (*certainement, sans doute, supposer, etc.*) atténue faiblement la fiabilité d’une information ;
- Les marqueurs de type *négation* comprennent des adverbes (*ne, pas, jamais, etc.*), des noms (*absence, lacune, etc.*), des adjectifs (*impossible, invraisemblable, etc.*) et des prépositions (*sans*) ou encore avec le préfixe *non-* ;
- Les marqueurs de type *modifieur* sont répartis en deux catégories :
 - les *modifieurs* d’atténuation, annotés *modif-m* (“m” pour “moins”), tels que *peu, à peine, assez peu* ;
 - les *modifieurs* de renforcement, annotés *modif-p* (“p” pour “plus”), tels que *très, beaucoup, vraiment*.

Ces modifieurs auront un impact s’ils sont utilisés conjointement avec des marqueurs de type *incertitude* et *émotion*. En effet, selon la catégorie du *modifieur* employé et le degré d’incertitude ou la polarité émotionnelle d’une unité lexicale, son intensité sera interprétée différemment. Nous développons plus en détail les diverses interprétations possibles dans la section 2.2.5.

Chaque unité lexicale (*incertitude* et *négation*) est étiquetée *pre-wi* (incertitude faible)/*pre-si* (incertitude forte) et *pre-neg*, respectivement, si sa position dans la phrase précède l’entité à laquelle elle se rapporte. De même, l’étiquette *post-neg* indique la post-position de l’unité lexicale par rapport à l’entité ciblée.

- Le lexique de marqueurs d’*émotion* (Augustyn *et al.*, 2008) contient des verbes (*énervé, inquiéter, savourer, etc.*), des noms (*stress, angoisse, optimisme, etc.*) et des adjectifs (*paisible, heureux, triste, etc.*). Ces unités, représentatives d’un état émotionnel, sont associées à une trentaine de catégories d’émotion comme *joie, peur, honte, attirance*. Cette catégorisation est bien plus précise que celles utilisées dans la majorité des travaux sur la classification des émotions (Ekman, 1992) ou (Plutchik, 1980) ;
- Le lexique des *polarités* regroupe les différentes catégories d’émotions associées à une polarité, permettant de généraliser les émotions en trois classes :
 1. positives : *soulagement, attirance, joie* ;
 2. neutres : *étonnement, surprise, ponctuations répétées* ;
 3. négatives : *insatisfaction, souffrance, gêne*.

Les termes exprimant une émotion sont étiquetés selon leur catégorie émotionnelle, définie dans le lexique des *polarités* : *emot*-catégorie émotionnelle.

Un échantillon de ces ressources est disponible à l’annexe B.2. Certains cas d’ambiguïtés ont été traités comme *tendu(e), manque, irritation*, qui peuvent faire référence à l’expression d’émotions : la catégorie *colère* pour *irritation* et *manque*, la catégorie *attirance* pour *tendu*. Dans un contexte médical, ces lexèmes

ont cependant tendance à faire référence à un état du patient (*la jambe tendue d'un patient* ou *une irritation cutanée*) ou d'une situation (*le manque de preuve pour l'établissement d'une procédure*). Afin d'éviter ce bruit, ces termes n'ont pas été pris en compte lors de l'annotation.

Le tableau 2.3 donne quelques détails chiffrés sur le contenu des différents lexiques utilisés pour l'annotation des corpus.

TABLE 2.3 – Distribution des marqueurs de type *incertitude, négation, modifieur* et des émotions

	Incertitude	Négation	Modifieur	Émotion	Total
Nombre d'entrées	101	20	17	1,032	1,170

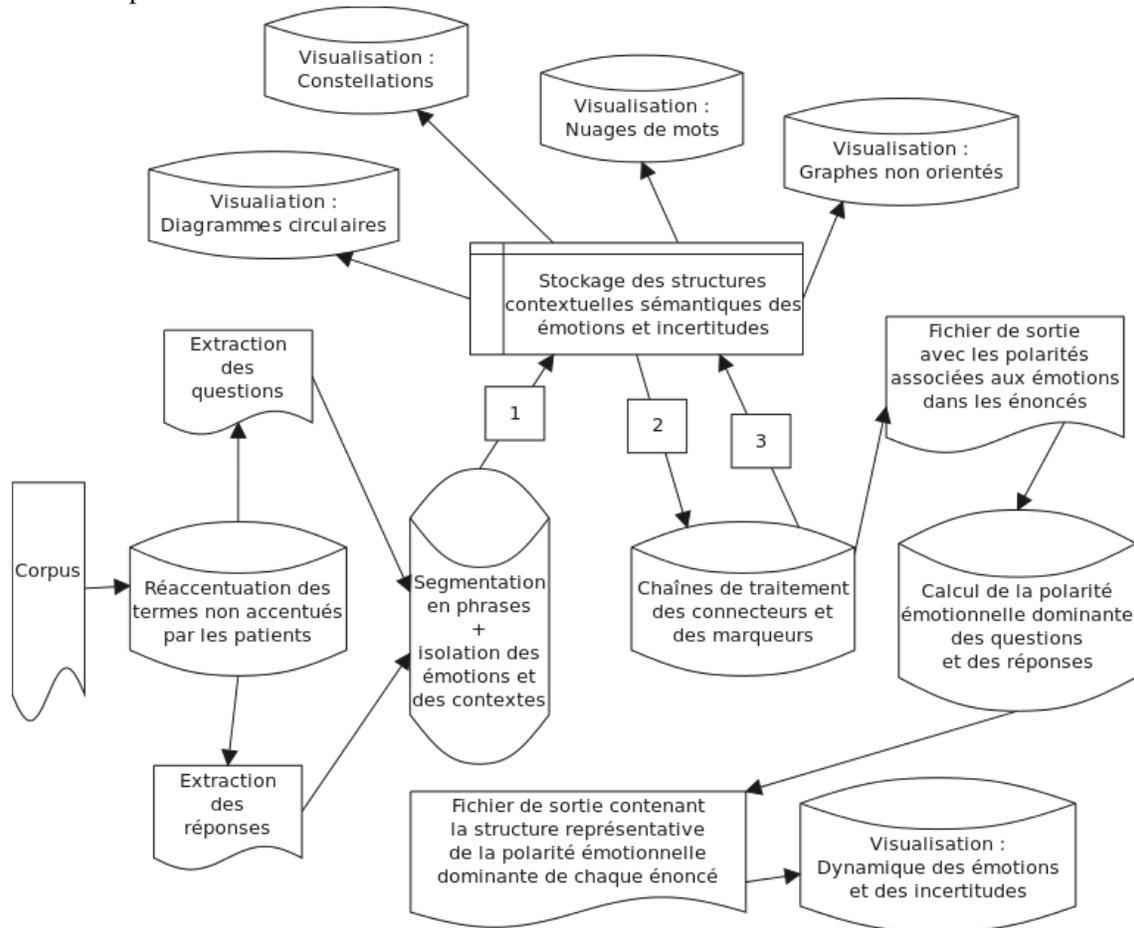
2.2 Méthodes

Notre étude vise l'analyse des émotions et leur visualisation. C'est pourquoi nous décidons d'utiliser les marqueurs *émotionnels* comme forme pôle (unité linguistique sur laquelle seront appliqués les divers traitements) pour nos diverses analyses. Par exemple, nous étudions l'interaction entre les différents *connecteurs* et types de marqueurs avec les émotions exprimées par les patients et les professionnels de santé. C'est à partir de ces émotions et de leurs contextes que nous pouvons faire ressortir leurs relations avec les *types médicaux* dans les différents systèmes de visualisation. Nous procédons également à l'évaluation de notre système en annotant manuellement les corpus de test. Nous procédons d'abord à l'analyse des corpus dans la section 2.2.1 et à l'ajustement des ressources dans la section 2.2.2. Ensuite, nous décrivons comment nous avons sélectionné les connecteurs syntaxiques dans les sections 2.2.4 et , puis notre système à base de règles dans la section 2.2.5. Enfin, nous expliquons la façon dont nous avons évalué les performances du système dans la section 2.2.6 avant de définir les modèles de visualisations dans la section 2.2.7. La figure 2.1 présente le processus des différents traitements des émotions/incertitudes exprimées dans les corpus. Nous décrivons ces étapes dans la suite de la section.

Le système que nous proposons s'organise en plusieurs étapes pour le traitement des émotions et la visualisation de l'évolution des émotions dans les questions et les réponses. Différents processus sont nécessaires depuis l'extraction des informations émotionnelles et sémantiques jusqu'à la visualisation de cette évolution des émotions dans les échanges patients/patients (corpus forum) et patients/professionnels (corpus Questions/Réponses). Les processus s'organisent comme suit :

1. Tout d'abord, nous procédons à la réaccentuation des termes appartenant aux différents types de marqueurs (modifieurs, incertitude, émotions, termes médicaux). En effet, il arrive que les patients omettent volontairement ou pas les caractères accentués : l'écriture des messages est alors plus rapide. Ce type de fautes commises par les patients, et éventuellement par les professionnels de santé, n'est pas excessivement fréquent. Cependant, dans ce type de plateformes sociales, nous nous sommes aperçus à plusieurs reprises que nous perdions de l'information sémantique. En effet, dans la situation où un marqueur comportant des caractères accentués dans les lexiques, n'est pas accentué par les patients, le système n'est pas capable de détecter un tel terme. C'est pourquoi nous décidons de

FIGURE 2.1 – Processus des différents traitements exécutés par le système : application des règles, extraction d'informations pour les modules de visualisation



générer une version supplémentaire des deux corpus dans laquelle nous avons réaccentué les termes non accentués des corpus avec leur version accentuée présente dans les lexiques. De cette façon, nous pouvons évaluer l'influence de la réaccentuation sur les performances du système à détecter la polarité émotionnelle dominante ;

2. Ensuite, nous extrayons de façon distincte les questions et les réponses de chaque corpus que nous stockons dans une structure de données : une par question, une par réponse et ce pour chacun des deux corpus ;
3. Dans cette structure de données nous stockons, pour chaque marqueur émotionnel détecté (portant l'étiquette sémantique du type *emot-[catégorie émotionnelle]* comme *emot-joie*, *emot-peur*,...) les informations sémantiques (modifieurs, incertitude, concepts médicaux – procédure, médicament, maladie –, négation) présentes dans ses contextes (gauche et droit). Nous choisissons de segmenter chaque énoncé en phrases afin de ne récupérer que les informations contextuelles relatives au marqueur émotionnel détecté. Cela permet d'éviter ainsi la sélection de hors-contexte si nous avons décidé de prendre en compte la totalité de l'énoncé sans tenir compte des ponctuations fortes. En effet, avec cette méthode, le risque de hors-contexte aurait augmenté si nous avions conservé des contextes présents dans deux phrases différentes, car les informations dans deux phrases consécutives ne sont pas toujours corrélées ;
4. À partir de ces structures de données, nous récupérons les marqueurs émotionnels et leurs contextes. Ainsi, pour chaque émotion, nous appliquons la chaîne de traitements des différentes règles (pré-traitement, négation, modifieurs, incertitude, connecteurs) sur ses contextes afin de détecter l'éventuelle présence de différents marqueurs (négation, modifieurs, incertitude) et connecteurs (figure 2.2). Cela permet de déterminer la nouvelle polarité de l'émotion que nous ajoutons à la structure de données ;
5. Une fois les règles appliquées, il s'agit ensuite de calculer la polarité émotionnelle dominante pour chaque énoncé (questions et réponses) à partir des résultats fournis par l'étape précédente : pour chaque marqueur émotionnel, les règles déterminent leur nouvelle polarité émotionnelle en fonction des marqueurs présents dans leurs contextes ;
6. Enfin, nous proposons diverses modélisations de visualisations de l'évolution des émotions : (1) la dynamique des émotions et des incertitudes à travers les polarités émotionnelles dominantes, (2) la distribution des polarités et des catégories émotionnelles utilisées dans les énoncés, (3) des polarités émotionnelles et de leurs catégories émotionnelles, (4) des catégories émotionnelles en nuages de mots (5) la relation entre la présence des marqueurs (modifieurs, négation, incertitude) et les différentes polarités émotionnelles.

2.2.1 Analyse des corpus

Nous effectuons une analyse manuelle partielle du corpus Questions/Réponses afin d'étudier la structure des annotations. Chaque mot, que nous appelons *token*, est annoté avec cinq informations. Les termes chargés sémantiquement, que nous appelons *termes*, possèdent une annotation supplémentaire (*types médicaux*, *types*

de marqueur ou catégorie d'émotion). Chaque *token* est annoté avec les informations suivantes (et dans cet ordre) :

1. la forme de surface du token : la forme dans laquelle il est utilisé dans le corpus ;
2. l'étiquette morpho-syntaxique du token : adjectif, nom, adverbe, verbe,...
3. le lemme correspondant au token : unité autonome correspondant au token (pour un verbe conjugué, son lemme est le verbe à l'infinitif ; pour un adjectif au féminin pluriel le lemme est cet adjectif au masculin singulier ;...);
4. deux identifiants internes à la plateforme ;
5. l'information sémantique du token : si le token est un marqueur, une émotion ou un type médical

Le tableau 2.4 donne quelques exemples de ces annotations.

TABLE 2.4 – Description des annotations attachées aux *tokens* et *termes*

Tokens					
commencés	/	VER :pper	/commencer/	16 / 0 /	
belle	/	ADJ	/ beau	/181/ 0 /	
médecin	/	NOM	/ médecin	/141/ 0 /	
Termes					
antalgiques	/	term	/ antalgique	/111/ 0 /	medication
rassurer	/	term	/ rassurer	/ 14 / 0 /	emot-soulagement
jamais	/	term	/ jamais	/274/ 0 /	pre-neg
1	/	2	/	3 / 4 / 5 /	6
forme de surface/étiquette morpho-syntaxique/			lemme	/id1/id2/	Information sémantique

De plus, certains termes peuvent faire référence à deux catégories sémantiques distinctes :

- certains tokens peuvent être annotés avec un type médical *disorder* et une catégorie d'émotion. En effet, des termes comme *peur* ou *anxieu[x,se]* peuvent faire référence à des troubles psychologiques (*disorder*) et exprimer un état émotionnel, ressenti par un patient par rapport à un traitement ou une maladie ;
- d'autres tokens peuvent recevoir deux catégories d'émotion. Par exemple, le terme *terrible* peut faire référence à une émotion *positive* ou *négative*, le terme *frustrée* peut appartenir à deux catégories d'émotion (*déception* et *insatisfaction*).

Lors de l'exécution de la chaîne de traitement (décrite à la section 2.2.5.3), nous traitons la "double émotion" d'un même marqueur *émotionnel* comme deux entités émotionnelles distinctes. Le tableau 2.5 illustre les différentes structures d'annotation générées par la plateforme *Ogmios*.

Après une première analyse de quelques énoncés du corpus Questions/Réponses, nous nous sommes aperçus de quelques lacunes dans les ressources linguistiques et sémantiques. Nous effectuons alors des ajustements supplémentaires de ces ressources afin de les rendre un peu plus exhaustives. La section 2.2.2 présente les différentes observations faites et les modifications effectuées.

TABLE 2.5 – Structures d’annotations des différents marqueurs et types médicaux

Incertitude	Négation
pouvez/term/pouvoir/21/0/pre-si/ ?/term/?/212/0/post-si/ en_cas_de/term/en_cas_de/93/0/pre-wi/ possibilités/term/possibilité/646/0/post-wi/	ne/term/ne/131/0/pre-neg/ pas/term/pas/107/0/post-neg/ rien/term/rien/14/0/pre-neg/ aucune/term/aucun/8/0/pre-neg
Modifieur	Émotion
moins/term/moins/197/0/modif-m/ beaucoup/term/beaucoup/315/0/modif-p/ très/term/très/2/0/modif-p/ peu/term/peu/5/0/modif-m/	souffrir/term/souffrir/62/0/emot-souffrance/ lool/term/lool/662/0/emot-pos/ !!/term/!!/1/0/emot-neu/ plaisir/term/plaisir/7/0/emot-joie/
Combinaison sémantique	
terrible/term/terrible/1/0/emot-peur+emot-attirance/ extraordinaires/term/extraordinaire/606/0/emot-attirance+emot-surprise/ anxieux/term/anxieux/30/0/disorder+emot-peur/ inquiète/term/inquiet/3/0/disorder_orth+emot-inquietude/	
Procédure	Maladie
IRM/term/IRM/281/0/procedure scanner/term/scanner/92/0/procedure_orth/ échographie/term/échographie/12/0/procedure/	scoliose/term/scoliose/28/0/disorder/ allegies/term/allergies/119/0/disorder_orth/ inflammation/term/inflammation/5/0/disorder/
Médicament	
paracétamol/term/paracétamol/453/0/medication/ mopline/term/morphine/406/0/medication_orth/ dafalgan/term/dafalgan/8/0/medication/	

2.2.2 Ajustement des ressources

Nous avons relevé des lacunes d’annotation, inhérentes au caractère informel du contenu des messages de forum. Le domaine de la santé impose également une certaine contrainte sémantique quant aux termes utilisés :

1. “Oublis” d’annotation :

- (a) Ces “oublis” peuvent trouver leur origine dans l’absence d’accentuation des termes. L’impact sur les résultats restera minime. Nous verrons cependant, lors de nos expériences, que cela peut améliorer les résultats (section 2.2.3). En effet, des termes tels que les marqueurs émotionnels ou les modifieurs qui ne sont pas accentués (soit ce sont des fautes d’orthographe, soit par facilité d’écriture) produisent du silence. C’est-à-dire que lors de la phase d’annotation d’un échantillon d’émotions par trois annotateurs, ils vont baser leur jugement sur la présence de marqueurs émotionnels, accentués ou non. Contrairement aux annotateurs humains, le système n’est pas capable de détecter un marqueur émotionnel s’il ne possède pas l’étiquette sémantique correspondante (*emot-peur*, *emot-joie*,...), ce qui est le cas pour les émotions non accentuées présentées dans les exemples ci-dessous.

Exemples :

- i. *Devrais-je m'inquieter ?* ; **inquieter** ne sera pas annoté avec l'étiquette *emot-inquietude* ;
 - ii. *Je prends du zoloft pour anxiete...* ; **anxiete** ne sera pas annoté avec l'étiquette *emot-peur* ;
 - iii. *Ca me degoute.* ; **degoute** ne sera pas annoté avec l'étiquette *emot-degout* ;
 - iv. *Aussi je ne ressens que quelques épisodes de deprime...* ; **deprime** ne sera pas annoté avec l'étiquette *emot-tristesse* ;
 - v. *j ai tres peur que ca se reproduise.* ; **tres** ne sera pas annoté avec l'étiquette *modif-p* ;
- (b) Il s'agit d'une lacune émanant directement de la plateforme d'annotation sémantique *Ogmios*. Malgré la présence de termes dans les corpus qui correspondent aux entrées des lexiques, l'annotation sémantique n'est pas systématique. Ces cas couvrent un panel de plus de 1,500 occurrences et concernent les termes complexes (composés d'au moins deux lexèmes). Ce qui induit une certaine perte d'informations, essentiellement concernant les types médicaux et quelques marqueurs de types *incertitude* et *négation* :
- i. *j'aimerais savoir_si cela est grave ?* ; **savoir_si** ne sera pas annoté avec l'étiquette *pre-wi* ;
 - ii. *La levure_de_bière contient de la vitamine_B.* ; **levure_de_bière** ne sera pas annoté avec l'étiquette *medication* et **vitamine_B** ne sera pas annoté avec l'étiquette *disorder* ;
 - iii. *Mon généraliste m'a prescrit un bilan_sanguin...* ; **bilan_sanguin** ne sera pas annoté avec l'étiquette *procedure* ;
 - iv. *je souffre régulièrement de douleurs_lombaire...* ; **douleurs_lombaire** ne sera pas annoté avec l'étiquette *disorder* ;
 - v. *Cela peut_être une petite cicatrice_fibreuse.* ; **peut_être** ne sera pas annoté avec l'étiquette *pre-si* ;
 - vi. *Aucune cause médicale n a pû être établi...* ; **n** ne sera pas annoté avec l'étiquette *pre-neg* ;
2. Complétion terminologique (un échantillon des ajouts est disponible à l'annexe C.1) :
- (a) Nous nous sommes aperçu que certaines entités médicales sont absentes des listes. Afin de remédier à cette lacune au moins partiellement, nous effectuons une recherche lexicale dans les corpus. Nous prenons en considération des suffixes propres à la terminologie médicale comme *-ome*, *-ose*, *-omie*, *-ite*, *-ase*, *-algie*, etc. Nous avons ainsi pu récupérer et annoter sémantiquement 263 types (dont des variantes d'accentuation et de casse). Cependant, il faut veiller à ne pas sélectionner de lexèmes non spécialisés dans le domaine médical comme *chose*, *ensuite*, *nostalgie*, *astronome*, *ergonomie*, *phrase*, etc., soit 317 formes.⁹
 - (b) D'autre part, en analysant manuellement les données, nous avons découvert de nouveaux lexèmes de type médical *maladies/troubles* à intégrer aux lexiques tels que *tumeur*, *pommade*, *gastro*, *suppo*, *papillomavirus*, *Infection Sexuellement Transmissible*, *virus*, *mythomane*, *lucite*, *mal de tête*, *maux de ventre*, *envie de vomir*.

9. en réalité, nous avons pris l'ensemble des termes du domaine médical, nous avons effectué leur analyse morphologique avec Dérif (Namer, 2009), nous avons récupéré les composants finaux (de tête) de ces termes, entre autres *-ome*, *-ose*, *-omie*, *-ite*, *-ase*, *-algie*. Nous avons gardé les composants qui ne sont pas ambigus quant à leur type sémantique.

3. Complétion linguistique :

- Nous avons remarqué que certains tokens reflètent une information sémantique proche de celle des différents types de marqueurs. Nous les intégrons aux différentes listes (lexiques d'émotions et de marqueurs, *négation*, *incertitude*, *modifieurs*) afin d'éviter une perte d'informations pour les futurs traitements de l'impact des marqueurs sur les émotions. Nous décidons d'ajouter les marqueurs qui nous semblent pertinents pour notre étude.

(a) Les marqueurs de type *incertitude* : *croire*, *penser*, *ressembler*

Exemples :

- je vous écrit car j'ai une infection_urinaire je crois.* ; **crois** sera annoté avec l'étiquette *pre-wi* (bien qu'ici, *post-wi* serait plus appropriée mais la syntaxe correcte consisterait à placer le terme **crois** en début de proposition (juste après *car*) ;
- Je pense qu'il est un peu schizophrène.* ; **pense** sera annoté avec l'étiquette *pre-wi* ;
- Cela ressemble à une crampe...* ; **ressemble** sera annoté avec l'étiquette *pre-wi*.

(b) Les marqueurs de type *négation* : *cesser*, *empêcher*

Exemples :

- Dois-je cesser de m'en inquiéter ?* ; **cesser** sera annoté avec l'étiquette *pre-neg* ;
- La canne est destinée à empêcher les douleurs.* ; **empêcher** sera annoté avec l'étiquette *pre-neg*.

(c) Les marqueurs de type *modifieur* : *à peine*, *moins*, *plein de*, *bcp*, *gros*, *tellement*.

Exemples :

- Celle-ci m'a soulagé deux semaines à peine.* ; **à peine** sera annoté avec l'étiquette *modif-m* ;
- mais bizarrement j'ai moins mal.* ; **moins** sera annoté avec l'étiquette *modif-m* ;
- J'ai plein de point_noir que faire ?* ; **plein de** sera annoté avec l'étiquette *modif-p* ;
- Cela me stresse bcp !!!* ; **bcp** sera annoté avec l'étiquette *modif-p* ;
- j'ai une grosse peur par rapport à l'algodystrophie.* ; **grosse** sera annoté avec l'étiquette *modif-p* ;
- j'ai tellement peur du pire...* ; **tellement** sera annoté avec l'étiquette *modif-p*.

(d) Les marqueurs *émotionnels* : *moquer*, *confortable*

Exemples :

- j'ai peur que l'on se moque de moi...* ; **moque** sera annoté avec l'étiquette *emot-mepris* ;
- je me sens vraiment pas confortable d'aller chez un gynécologue.* ; **confortable** sera annoté avec l'étiquette *emot-joie* ;

4. Il est nécessaire de distinguer, dans la mesure du possible, la cohérence sémantique des termes d'émotion dans le contexte médical :

- (a) Certains termes évoquant l'expression d'une émotion sont retirés du lexique car ils ne sont pas adaptés au contexte médical ou manquent de pertinence : *glacer, brûler, emballer, bouffer, doucher, raser*

Exemples :

- i. *Faut il que je **glace** si c'est une inflammation...* ; **glace** fait référence à un traitement et non à l'émotion *tristesse* ;
 - ii. *Je fais des petits rots et ils **brûle**...* ; **brûle** fait référence à un trouble physique et non à l'émotion *attirance*. Les formes *brûlé* et *brûlure(s)* dénotent la même référence ;
 - iii. *Par contre en sport mon coeur s'**emballe** vite...* ; **emballe** fait référence à un état du coeur à un moment donné et non à l'émotion *joie* ni *attirance* ;
 - iv. *J'ai des spasmes, des étourdissements et ds **bouffées** de chaleurs...* ; **bouffées** fait référence à une sensation et non à l'émotion *fierté* ;
 - v. Quant aux termes *doucher, raser*, ils vont plutôt faire référence au fait de prendre une douche et non à l'émotion *tristesse* ou au fait de réduire sa pillosité et non à l'émotion *lassitude*.
- (b) D'autres termes sont caractérisés par deux traits sémantiques, le type médical *maladie* et un marqueur *émotionnel* : *éclater, enflammer, allergie, malaise, nausées*. Travaillant dans un contexte médical, nous avons finalement décidé de catégoriser ces lexèmes comme étant factuels et non comme des lexèmes définissant un état émotionnel. Nous entendons par "factuel" un état ponctuel caractérisé par des troubles de la santé (ou maladie) sans connotation émotionnelle. En effet, dans un contexte plus général, des termes comme *éclater* ou *nausées* font référence à des états émotionnels bien définis (*je me suis éclaté au foot*, qui évoque une émotion positive ; *ta façon d'agir me donne des nausées*, qui évoque une émotion négative). Mais dans le contexte médical, ils font référence à des états "factuels", définissant des troubles liés à une maladie.

Exemples :

- i. *Le vaisseau **éclaté** sous le sexe n'a pas de conséquence.* ;
- ii. *Votre description correspond vraisemblablement à un ganglion **enflammé**...* ;
- iii. *Le médecin m'a dit que c'était dû à mes **allergies**.* ;
- iv. *le **malaise** vagal se soigne comment ?* ;
- v. *Ce matin encore j'ai la **nausée** et la tête lourde.* ;

5. Locutions :

- Pour éviter certains biais dans les annotations et améliorer les lexiques, nous avons relevé quelques locutions revenant assez régulièrement dans les corpus :
- (a) Nous catégorisons les locutions suivantes comme relevant du type sémantique *maladie* : *envie de vomir/d'uriner, maux de tête/de ventre/etc., mal à respirer/la gorge/la tête/etc.* ;
 - (b) Nous catégorisons les locutions suivantes comme appartenant aux marqueurs de type *modifieur* : *à peine* pour les modifieurs d'atténuation et *pas mal de* pour les modifieurs d'intensification.

2.2.3 Réaccentuation des corpus

Comme nous l'avons signalé, pour faciliter l'écriture des messages, les patients omettent parfois les accents. Or, cela peut causer des pertes d'informations lors de l'extraction des termes (marqueurs médicaux, émotionnels, modificateurs, négation et incertitude) des corpus. Afin de pouvoir vérifier si cela a effectivement un impact sur les performances de notre système à base de règles, nous avons généré une version réaccentuée des corpus. La méthode est effectuée en cinq étapes :

1. récupération des termes accentués dans les différents lexiques (médicaux, marqueurs et émotions) via une expression régulière, contenant les diacritiques les plus utilisés dans la langue française : *éèçâêûîôâëüïöù* ;
2. nous retirons ensuite les accents (grâce au module perl : Unicode : :normalize) des termes récupérés précédemment ;
3. nous vérifions chaque token dans chacun des corpus s'il correspond à un terme dont on a préalablement retiré l'/les accent(s). Cependant, nous avons pris la précaution de ne pas traiter le marqueur d'incertitude *sûr* car la correspondance (après désaccentuation) avec la préposition *sur* aurait été trop fréquente, ce qui aurait eu pour effet d'ajouter de façon excessive ce marqueur d'incertitude ;
4. une fois la correspondance faite, nous procédons à la réaccentuation de ces termes avec le même module perl. Par exemple, si un patient a écrit *desir* ou *inquieter*, le système va alors procéder à leur accentuation (*désir*, *inquiéter*) ;
5. la dernière étape consiste à indiquer l'étiquette sémantique du terme nouvellement accentué. Pour cela, nous avons conservé l'information sémantique correspondante aux termes que nous avons désaccentués dans l'étape 2. De cette manière, il nous est possible d'assigner l'étiquette sémantique correspondante à ce nouveau terme. Par exemple, pour le terme *désir*, nous enregistrons son étiquette sémantique *emot-attirance*, nous retirons l'accent aigu, et nous obtenons le terme *desir*. S'il correspond à un token dans un corpus, il est remplacé par son équivalent accentué *désir* auquel nous associons l'étiquette sémantique *emot-attirance*.

Le tableau 2.6 donne un aperçu quantitatif des termes non accentués pour chacun des marqueurs et des types médicaux. Nous observons dans le tableau 2.6 que, de manière générale, les marqueurs et les concepts médicaux non accentués représentent une très petite portion des données. Nous calculons les proportions de termes nouvellement accentués pour chaque type de marqueur (modificateurs, incertitude, concepts médicaux, émotions). Nous obtenons le pourcentage de termes non accentués pour un type de marqueurs en divisant le nombre de termes nouvellement accentués par le nombre total de termes appartenant à ce type de marqueurs (plus les nouveaux termes accentués). De manière plus précise :

- les *émotions* non accentuées dans les questions du corpus Q/A représente presque 10 % de ce type de marqueur. Comme notre entités linguistiques de travail sont les termes exprimant l'état émotionnel, cela peut représenter un supplément d'informations intéressant ;
- les *modificateurs*, malgré le faible nombre de leurs occurrences non accentués (5,7 % pour le corpus Q/A et 4,5% pour le corpus forum), jouent un rôle important dans la polarisation des énoncés. Ils

TABLE 2.6 – Aperçu quantitatif des termes non accentués dans les deux corpus par rapport aux nombres totaux réels de termes présents (accentués et non accentués)

Corpus		émotions	modifieurs	incertitudes	concepts médicaux (procédures, maladies, médicament)	total
Q/A	questions	305/4,064 7,5%	412/7,195 5,7%	36/13,977 0,2%	1,370/25,687 5,3%	2,123/50,913 4,1%
	réponses	39/1,510 2,5%	4/1,584 0,2%	8/7,925 0,1%	127/22,489 0,5%	178/33,508 0,5%
Forum		284/8,408 3,3%	602/13,246 4,5%	122/19,224 0,6%	1,450/32,282 4,5%	2,458/73,160 3,3%

apparaissent en effet très souvent dans les contextes de marqueurs *émotionnels*. Il nous semble donc intéressant d'étudier l'impact des *modifieurs* non accentués sur les résultats d'évaluation ;

- les marqueurs d'*incertitude* représentent la plus faible portion de termes non accentués. À notre avis, leur impact sur une éventuelle amélioration des résultats fournis par le système ne sera pas significatif.
- les termes *médicaux* non accentués représentent un peu plus de 4 % de ces termes dans chacun des corpus. L'accentuation de ces unités lexicales peut aussi permettre d'obtenir des données plus précises. La plus faible proportion de termes non accentués caractérise les réponses du corpus Q/A, dont les auteurs sont des professionnels de santé, contrairement aux questions, rédigées par les patients.

2.2.4 Sélection des connecteurs syntaxiques

Les fils de discussion dans les forums de santé contiennent de nombreux marqueurs *émotionnels*. En effet, les patients expriment parfois plusieurs émotions au sein d'un même énoncé (question ou réponse). Nous nous proposons comme objectif de dégager la polarité émotive (*positive, négative, neutre*) dominante d'un énoncé. Lorsqu'il n'y a qu'une seule émotion, il est assez simple de détecter la polarité de l'énoncé. Mais lorsque plusieurs émotions sont exprimées, la tâche devient plus complexe, car nous devons décider quelle émotion, et donc quelle polarité, prédomine dans un énoncé. C'est à partir du corpus d'entraînement des Questions/Réponses que nous avons procédé à l'analyse de la structure syntaxique des messages afin de proposer des règles et des unités lexicales discriminantes qui nous permettraient de définir une *relation de dominance* entre plusieurs émotions exprimées dans un énoncé :

- *En tous cas, **tranquillez-vous**, votre problème ne semble pas très **préoccupant**.* ; les termes soulignés sont différents types de marqueurs interagissant directement sur la seconde émotion exprimée **préoccupant** ;
- *L'alimentation **calme** effectivement l'**anxiété** mais vos **angoisses** repartent...* ; le token souligné crée une sorte de cission entre l'état émotionnel le précédant et celui qui lui succède ;
- *Je me demande si ce n'est pas dû au **stress**, car j'ai bcp de **stress**.* ; le token souligné vient appuyer la première émotion exprimée en y ajoutant une nouvelle émotion.

Cette première analyse montre que des connecteurs syntaxiques peuvent nous aider à cerner l'émotion dominante au sein d'un énoncé. Nous attribuons une pondération pour chaque émotion en fonction de la présence

de connecteurs et de marqueurs dans son contexte (section 2.2.5).

Dans le tableau 2.7, nous présentons un certain nombre de connecteurs potentiels pouvant impacter la relation entre les émotions dans un énoncé. Ce tableau indique aussi la distribution des connecteurs dans la totalité des corpus.

TABLE 2.7 – Distribution des *connecteurs* dans les corpus Q/A et forum

Corpus	mais	cependant	car	et	donc	puis	néanmoins	enfin	Total
Q/A	3,677	242	1,296	17,120	879	597	109	82	24,002
forum	6,018	68	2,323	15,892	1,888	550	15	431	27,185
Total	9,695	310	3,619	33,012	2,767	1,147	124	513	51,187

Finalement, nous décidons de conserver les connecteurs suivants :

- *mais* et *cependant* : indiquent une rupture entre le texte qui précède et celui qui succède ;
- *car* : renforce l’argumentation de l’information précédente par l’information qui suit ;
- *et* : marque une énumération d’émotions ;
- *donc* : donne de l’importance à l’information située après.

Plusieurs autres connecteurs ont été écartés :

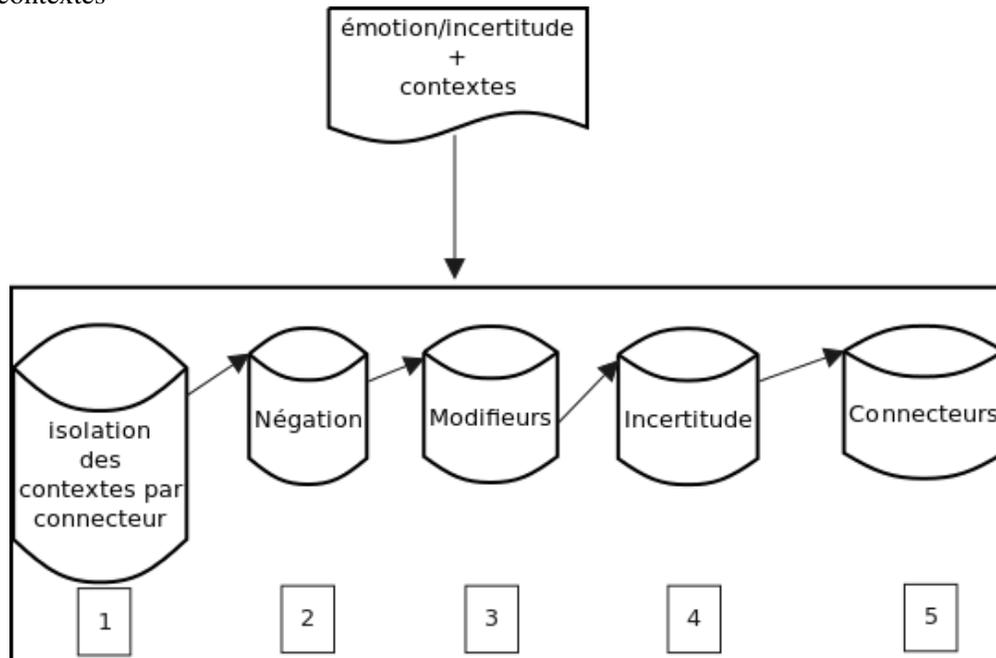
- *néanmoins* montre un faible nombre d’occurrences dans les corpus,
- *puis* est très présent, mais n’a pas été retenu à cause de son ambiguïté avec le verbe *pouvoir* (revenant assez souvent dans les questions),
- *enfin* n’a pas été retenu à cause de son ambiguïté : selon son contexte, il atténue un propos (*bon week end, enfin, autant que possible.*) ou évoque une finalité (*Il vit enfin normalement.*).

Nous recherchons à faire ressortir l’émotion avec la plus forte intensité au sein d’un énoncé. Afin d’y parvenir, nous allons détecter les différentes émotions présentes dans les messages de patients et de professionnels de santé. Une fois les émotions repérées pour chaque question et pour chaque réponse, nous appliquons un système à base de règles (section 2.2.5) sur ces données. Ce système utilise les connecteurs ainsi que les différents types de marqueurs comme paramètres.

2.2.5 Un système à base de règles

Nous proposons un système à base de règles pour détecter la polarité émotionnelle dominante (l’émotion polaire la plus intense) dans les questions et les réponses formulées par les patients et les professionnels de santé. Dans un premier temps, nous allons extraire tous les marqueurs *émotionnels* avec leurs contextes gauche et droit. Si nous ne pouvons pas détecter d’émotions, nous prenons comme forme pôle les marqueurs de type *incertitude*, auxquels les mêmes traitements seront appliqués, mais avec des conséquences différentes dans la modification du degré d’intensité. Le fonctionnement du système à base de règles repose sur trois étapes principales : calcul des contextes (section 2.2.5.1), proposition et application de règles (sections 2.2.5.2 et 2.2.5.3). Nous effectuons également une évaluation du système (section 2.2.6).

FIGURE 2.2 – Chaîne de traitements des connecteurs et des marqueurs appliquée aux émotions/incertitudes et à leurs contextes



2.2.5.1 Les contextes

Nous avons choisi une fenêtre de sept mots pour calculer les contextes (Thoumelin & Grabar, 2014), afin de garantir la cohérence sémantique entre l’émotion (forme pôle) et ses composants contextuels (marqueurs, types médicaux). Le contexte s’arrête aussi lorsque nous rencontrons des marques de *punctuation* forte (?, ., !) et moyenne (;, :, (, []). D’autres études utilisent une fenêtre proche de la nôtre afin de limiter l’analyse des informations hors-contexte (Feng *et al.*, 2013b; Narayanan *et al.*, 2009). Les règles sont définies à l’aide d’expressions régulières pour certaines et de boucles (parcours des contextes) pour d’autres. Pour chacune des émotions ou incertitudes exprimées dans un énoncé, nous extrayons les informations sémantiques (marqueurs de négation, d’incertitude, modifieurs) présentes dans leurs contextes. Les structures de données ainsi générées permettent également de visualiser (section 2.2.7) les informations contextuelles (marqueurs, types médicaux) des émotions (et des incertitudes, le cas échéant).

2.2.5.2 Les règles

Le fonctionnement de notre système est basé sur quatre règles en plus d’un traitement préalable. Cinq étapes sont effectuées. La figure 2.2 décrit la chaîne de traitements par des règles appliquée à chaque émotion (et marqueur d’incertitude le cas échéant).

Nous avons un ordre de priorité dans l’application des règles, que l’on peut voir à la figure 2.2. Les règles vont s’appliquer sur chaque marqueur émotionnel présent dans les énoncés. En fonction des marqueurs (négation, modifieurs, incertitude, connecteurs) présents dans les contextes des marqueurs émotionnels, les règles doivent respecter cet ordre de priorité. Comme le montre la figure 2.2, l’application des règles est

effectuée en cinq étapes :

1. Pré-traitement (code disponible à l'annexe D.1) : cette étape consiste à détecter dans les contextes des marqueurs émotionnels des connecteurs (car, donc, et, mais, cependant). Si c'est le cas, le connecteur, dont la distance avec le marqueur émotionnel ne doit pas dépasser le seuil fixé, est considéré comme la nouvelle frontière contextuelle. Nous décrivons cette étape dans la section 2.2.5.2.1 ;
2. Négation (code disponible à l'annexe D.2) : cette étape consiste à détecter la présence d'un marqueur de négation (ne...pas, rien, aucun, . . .) dans la (nouvelle) fenêtre contextuelle du marqueur émotionnel analysé. Si un marqueur de négation est présent alors la polarité d'origine du marqueur émotionnel est inversée. Nous décrivons cette étape dans la section 2.2.5.2.2 ;
3. Modificateurs (code disponible à l'annexe D.3) : cette étape consiste à vérifier la présence d'un modificateur (amplificateur ou atténuateur) dans la (nouvelle) fenêtre contextuelle du marqueur émotionnel analysé. Si un marqueur de type modificateur est présent alors la polarité du marqueur émotionnel est modifiée (amplifiée ou atténuée). Nous décrivons cette étape dans la section 2.2.5.2.3 ;
4. Incertitude (code disponible à l'annexe D.4) : cette étape consiste à identifier la présence de marqueurs d'incertitude dans la (nouvelle) fenêtre du marqueur émotionnel analysé. Si un ou plusieurs marqueurs d'incertitude ont présents, alors la polarité du marqueur émotionnel est atténuée. Nous décrivons cette étape dans la section 2.2.5.2.4 ;
5. Connecteurs (code disponible à l'annexe D.5) : cette étape consiste à détecter la présence d'un connecteur (car, donc, et, mais, cependant) dans la (nouvelle) fenêtre contextuelle du marqueur émotionnel analysé. Si un connecteur est présent, la polarité du marqueur émotionnel est soit amplifiée soit atténuée en fonction du connecteur. Nous décrivons cette étape dans la section 2.2.5.2.6.

La section 2.2.5.2.5 présente une étape implicite dans l'application des règles. Il s'agit des combinaisons possibles de divers marqueurs. En effet, il est possible que plusieurs types de marqueurs se retrouvent dans les contextes des marqueurs émotionnels. C'est pourquoi l'ordre d'application des règles est importante. Par exemple, si nous observons la présence d'un marqueur de négation et d'un modificateur d'amplification (interagissant ensemble) dans les contextes d'un marqueur émotionnel, le système va appliquer en priorité la règle de la négation et seulement après la règle des modificateurs.

2.2.5.2.1 Pré-traitement. Le pré-traitement consiste à repérer les connecteurs présents dans les contextes des émotions. Ce pré-traitement permet d'éviter les incohérences des autres traitements. Par exemple, dans *je suis très contente ET angoisée à la fois*, le modificateur *très* n'agit pas sur l'émotion *angoissée* car il précède le connecteur *et* auquel succède cette émotion, et dans ce cas, le connecteur *et* devient la nouvelle frontière du contexte gauche de l'émotion *angoissée* :

- *Je suis une personne très anxieuse ET j'ai peur...* ; ici, le modificateur *très* agit sur **anxieuse** mais pas sur **peur** ;
- *J'me force plus CAR j'en pleure.* ; ici, le modificateur *plus*, qui s'avère être une *négation*, n'agit pas sur **pleure** ;
- *qu'il ne veut pas me mentir, MAIS il regrette.* ; ici, la *négation* *ne...pas* n'agit pas sur **regrette**.

Dans les situations où plusieurs connecteurs sont présents dans les contextes d'une émotion comme *j'ai peur CAR j'ai très mal ET rien ne me soulage*, c'est le connecteur le plus proche (*car*) de l'émotion (*peur*) qui est pris en considération :

- *Je suis migraineux, stresser ET fatigué MAIS je ne sais pas...* ; ici, le connecteur MAIS n'agira pas sur **stresser** et le connecteur *ET* constituera la frontière du contexte droit de l'émotion **stresser** ;
- *MAIS aussi ET à ma grande surprise cancer de la gorge...* ; ici, le connecteur MAIS n'agira pas sur **surprise** et le connecteur *ET* constituera la frontière du contexte gauche de l'émotion **surprise**.

2.2.5.2.2 Traitement de la négation. Le traitement de la négation (Paroubek, 2010a; Bakliwal *et al.*, 2013; Volkova *et al.*, 2013) prend en compte la présence des marqueurs de négation dans le contexte d'une émotion. La négation reçoit la priorité sur les incertitudes et modificateurs. La négation inverse la polarité de l'émotion : *positive* -> *négative* ou *négative* -> *positive*. Cette règle exploite la portée de la négation (*pre-neg*, *post-neg*) pour savoir si l'émotion pertinente est attendue avant ou après la négation. De plus, la présence dans le contexte gauche de certains marqueurs de négation appuyée par une marque de négation dans le contexte droit (par exemple, *ne...pas*) implique aussi l'inversion de polarité de l'émotion. La taille de la fenêtre d'application de cette règle va de 0 à 5 tokens. Cette taille varie en fonction du marqueur employé. Dans les travaux de l'état de l'art, différentes tailles de fenêtre ont été testées (pour la langue anglaise) :

- 3 tokens dans (Bakliwal *et al.*, 2013),
- 3 à 6 tokens dans (Narayanan *et al.*, 2009).

Voilà quelques exemples qui illustrent l'application de cette règle :

1. *Je n' éprouve plus de désir sexuel...* ; le marqueur émotionnel **désir** de polarité positive changera pour la polarité négative ;
2. *Je n' ai aucune envie de m'en servir.* ; le marqueur émotionnel **envie** de polarité positive changera pour la polarité négative ;
3. *je suis quelqu'un qui ne stress jamais...* ; le marqueur émotionnel **stress** de polarité négative changera pour la polarité positive ;
4. *mais les prises de sang montre rien alarmant...* ; le marqueur émotionnel **alarmant** de polarité négative changera pour la polarité positive ;
5. *mon sang ne circule plus dans ma jambe droite !!!* ; le marqueur émotionnel **!!** de polarité neutre changera pour la polarité négative (faisant référence ici à l'émotion *inquiétude* ;

Notons que cette règle traite aussi deux cas particuliers :

- *plus* est considéré comme un marqueur de négation et non pas comme un modificateur. Il peut être combiné avec *ne*, *n'* et *n* ;
- emploi conjoint d'une marque de négation avec un marqueur *émotionnel neutre* du type **!!**. Ceci a pour effet de passer d'une polarité *neutre* à une polarité *négative*.

2.2.5.2.3 Traitement des modificateurs. Le rôle des modificateurs est de modifier l'intensité initiale d'une émotion exprimée par un patient ou un professionnel de santé (Maurel *et al.*, 2008). Selon l'étiquette sémantique du modificateur, nous attribuons un degré d'intensité à l'émotion : un "+" pour une intensification

(*modif-p*) et un “-” pour une atténuation (*modif-m*) (section 2.2.5.3). Ce type de marqueur peut se combiner à d’autres marqueurs, comme la négation ou l’incertitude. Cependant, nous devons veiller à ne prendre en compte le modifieur *plus* que s’il fait partie d’une combinaison de marqueurs de négation.

Cette règle est appliquée dans une fenêtre de 0 à 3 tokens à gauche et à droite. La petite taille de la fenêtre permet de respecter une certaine cohérence sémantique entre les modifieurs et l’émotion pôle. De plus, il est possible de trouver une succession de modifieurs, dans cette fenêtre réduite. Cela aura comme conséquence la modification de l’intensité de l’émotion exprimée autant de fois qu’il y a de modifieurs.

Voilà quelques exemples qui illustrent cette règle :

1. *Je suis de plus en plus inquiète...* ; l’intensité polaire (négative) du marqueur émotionnel **inquiète** sera intensifiée (*plus* et *plus*) par deux fois ;
2. *et ne ressens que peu de plaisir physique...* ; l’intensité polaire (positive) du marqueur émotionnel **plaisir** sera atténuée (*peu* une fois) ;
3. *je comence vraiment à avoir peur* ; l’intensité polaire (négative) du marqueur émotionnel **peur** sera intensifiée (*vraiment* une fois) ;

2.2.5.2.4 Traitement de l’incertitude. La présence de marqueurs d’incertitude altère l’intensité des émotions (Abdaoui *et al.*, 2014). En effet, si une émotion est employée avec ce type de marqueurs, alors l’intensité polaire de l’émotion décroît. Comme il existe deux étiquettes sémantiques des incertitudes (*pre-si*, *post-si* indiquent une incertitude forte, *pre-wi*, *post-wi* indiquent une incertitude faible), nous pouvons calculer deux niveaux d’atténuation. Plus l’incertitude est forte et plus l’intensité émotionnelle a tendance à diminuer. Si le degré d’incertitude est faible, alors l’intensité émotionnelle diminue aussi, mais de façon moins importante. Les incertitudes fortes sont représentées par “??” et les incertitudes faibles par “?” (section 2.2.5.3).

L’incertitude est considérée dans la fenêtre de 0 à 7 tokens à gauche et à droite. Nous avons choisi la fenêtre maximale car des marqueurs d’incertitude du type *si*, *?* ont une large portée. Nous avons également essayé de traiter l’utilisation conjointe de l’incertitude et des émotions *neutres* (du type *!!*, *??!!*). Cette situation a également pour effet de modifier la polarité *neutre* en polarité *négative*. En effet, nous remarquons que les patients qui combinent une émotion neutre, marquée par une ponctuation expressive (*!, !!, ...*), avec un marqueur d’incertitude, expriment un état émotionnel négatif.

Voilà quelques exemples qui illustrent cette règle :

1. *pouvez-vous me rassurer ?* ; l’intensité polaire (positive) induite par le marqueur émotionnel **rassurer** sera diminuée fortement par l’utilisation de deux marqueurs d’incertitude forte : *pouvez* et *?* ;
2. *pouvez-vous m’aider ?!* ; la polarité (neutre) induite par le marqueur émotionnel **?!** sera modifiée, pour exprimer un état émotionnel *négatif*, par l’utilisation d’un marqueur d’incertitude fort : *pouvez* ;
3. *Je nai rien vu sur google qui pourrait ressembler à sa!!!* ; la polarité (neutre) induite par le marqueur émotionnel **!!** sera modifiée, pour exprimer un état émotionnel *négatif* (référence à l’émotion *inquiétude*), par l’utilisation d’un marqueur d’incertitude fort : *pourrait* ;

4. Vous avez probablement fait une crise d'**angoisse**. ; l'intensité polaire (négative) induite par le marqueur émotionnel **angoisse** sera diminuée légèrement par l'utilisation d'un marqueur d'*incertitude* faible : probablement ;

2.2.5.2.5 Combinaison de marqueurs. Plusieurs marqueurs peuvent être présents dans les contextes d'une émotion. Nous devons alors tous les prendre en compte, ce qui implique de traiter différentes combinaisons entre les divers types de marqueurs.

Nous proposons plusieurs cas de combinaison de marqueurs.

2.2.5.2.5.1 négation + modifieur de type modif-p : La négation suivie d'un modifieur de type modif-p inverse et atténue la polarité :

1. *ce qui n'est pas* très **encourageant** ; la polarité initiale du marqueur *émotionnel* **encourageant** sera inversée (*positive* -> *négative*) et atténuée, car la *négation* inverse également le degré d'intensification (*modif-p* -> *modif-m*) ;
2. *je ne suis pas* très **rassurée**. ; la polarité initiale du marqueur *émotionnel* **rassurée** sera inversée (*positive* -> *négative*) et atténuée, car la *négation* inverse également le degré d'intensification (*modif-p* -> *modif-m*) ;

2.2.5.2.5.2 négation + modifieur de type modif-m : Néanmoins, lorsque la négation est suivie du modifieur de type modif-m, la règle de l'inversion et de l'amplification de la polarité ne marche pas (à défaut d'exemple concret émanant du corpus, nous proposons cet exemple forgé) :

je ne suis pas peu **inquiet** de la situation. ; d'après le raisonnement précédent : la polarité initiale du marqueur *émotionnel* **stressée** sera inversée (*négative* -> *positive*) et amplifiée, car la *négation* inverse également le degré d'intensification (*modif-m* -> *modif-p*). Or l'interprétation est différente. Dans ce cas précis, la *négation* s'annule avec le modifieur peu et la polarité du marqueur *émotionnel* inquiet reste inchangée.

2.2.5.2.5.3 modifieur de type modif-p + modifieur de type modif-p : Lorsque le modifieur de type modif-p est suivi d'un modifieur de type modif-p, cela conduit à une double amplification de la polarité :

et me sentir de plus en plus **déprimée**. ; la polarité de l'émotion est *négative*, et par la combinaison des modifieurs de type sémantique *modif-p* + *modif-p*, l'intensité polaire sera amplifiée par deux fois.

2.2.5.2.5.4 modifieur de type modif-p + modifieur de type modif-m : Lorsque le modifieur de type modif-p est suivi du modifieur de type modif-m, modif-p amplifie le modif-m, ce qui équivaut à une double atténuation de la polarité :

sans ou avec bcp moins d'**inconfort**. ; la polarité de l'émotion est *négative*, et par la combinaison des modifieurs de type sémantique *modif-p* + *modif-m*, l'intensité polaire sera atténuée par deux fois.

2.2.5.2.6 Traitement des connecteurs. La présence de connecteurs au sein des contextes gauche et/ou droit d'un marqueur émotionnel aura un niveau d'impact différent selon le connecteur (Yu *et al.*, 2013). En effet, la taille de la fenêtre reste propre au connecteur. Pour rappel, un connecteur constituera la nouvelle frontière contextuelle (gauche, s'il apparaît dans le contexte gauche, ou droite s'il apparaît dans le contexte droit) de l'émotion pôle. Cette étape de traitement a pour effet d'augmenter ou de diminuer l'intensité liée à l'émotion exprimée.

2.2.5.2.6.1 mais/cependant :

- Fenêtre maximale (0 à 7) à gauche et à droite ;
- Si le connecteur est présent à gauche alors l'intensité polaire de l'émotion est amplifiée par la rupture qu'il engendre avec le contenu textuel qui précède ;
- Si le connecteur est présent à droite alors l'intensité polaire de l'émotion est atténuée par la rupture qu'il engendre avec le contenu textuel qui succède ;

Quelques exemples avec ce connecteur :

1. *Cependant cela me **stresse**...* ; le connecteur *cependant* est situé dans le contexte gauche du marqueur émotionnel **stresse** donc l'intensité (de polarité *négative*) est amplifiée ;
2. *je ressens beaucoup de **plaisir** pendant l'acte, mais pas d'orgasme.* ; le connecteur *mais* est situé dans le contexte droit du marqueur émotionnel **plaisir** donc l'intensité (de polarité *positive*) est atténuée ;
3. *Je ne doute pas que le magnétisme **apaise** le malade mais n'a-t-il pas un risque de...* ; le connecteur *mais* est situé dans le contexte droit du marqueur émotionnel **plaisir** donc l'intensité (de polarité *positive*) est atténuée ;
4. *tous[tests de grossesse] positifs mais je **crains** la fausse_couche...* ; le connecteur *mais* est situé dans le contexte gauche du marqueur émotionnel **crains** donc l'intensité (de polarité *négative*) est amplifiée.

2.2.5.2.6.2 car :

- Fenêtre de 0 à 4 à droite de l'émotion (afin d'éviter le hors-contexte), de 0 à 7 à gauche (la portée du connecteur *car* est alors assez large) ;
- Que le connecteur soit à gauche ou à droite, l'intensité polaire de l'émotion est amplifiée. La conjonction de coordination vient justifier l'information précédente avec une nouvelle information. Cette relation entre les deux informations permettra d'amplifier l'intensité émotionnelle exprimée, aussi bien dans le contexte gauche que droit. En effet, cette nouvelle information succédant au connecteur *car* fournit une explication sur ce qui précède au connecteur ;

Quelques exemples avec ce connecteur :

1. *J'ai eu très **peur** car sur le mouchoir c'est du sang...* ; le connecteur *car* est situé dans le contexte droit du marqueur émotionnel **peur** donc l'intensité (de polarité *négative*) sera amplifiée par la cause exprimée clairement (*sur le mouchoir c'est du sang*) ;
2. *j me force plus car j en **pleurait** tellement...* ; le connecteur *car* est situé dans le contexte gauche du marqueur émotionnel **pleurait** donc l'intensité (de polarité *négative*) sera amplifiée par la cause (*j en pleurait tellement...*) justifiant la proposition principale (*j me force plus*) ;

3. *Car c'est très **frustrant** Merci beaucoup.* ; le connecteur *Car* est situé dans le contexte gauche du marqueur *émotionnel* **frustrant** donc l'intensité (de polarité *négative*) sera amplifiée car cette justification appuie une précédente information (implicite ici) ;
4. *je **regrette** cet acte car je souhaite un autre enfant.* ; le connecteur *car* est situé dans le contexte droit du marqueur *émotionnel* **regrette** donc l'intensité (de polarité *négative*) sera amplifiée par la précision apportée par la cause (*je souhaite un autre enfant*).

2.2.5.2.6.3 et :

- Fenêtre maximale (0 à 7) à gauche de l'émotion (portée du connecteur *et* assez large) et de 0 à 2 tokens à droite, afin de respecter la cohérence sémantique entre l'émotion et le connecteur ;
- Si le connecteur est présent à gauche, l'intensité polaire de l'émotion sera amplifiée, car cette conjonction de coordination ajoute une information supplémentaire, qui vient amplifier l'intensité polaire de cette émotion ;
- Si le connecteur apparaît dans le contexte droit de l'émotion, il affecte aussi l'intensité de l'émotion. En effet, si une énumération, instanciée par *et*, succède directement (fenêtre de 2 tokens maximum) à un marqueur d'*émotion*, l'intensité polaire de cette émotion est amplifiée.

Quelques exemples avec ce connecteur :

1. *ma mère est âgé de 70 ans et j'ai **peur** que sa vie soit en danger...* ; le connecteur *et* est situé dans le contexte gauche du marqueur *émotionnel* **peur** donc l'intensité (de polarité *négative*) sera amplifiée par le contexte énumératif (*j'ai peur que sa vie soit en danger*) ;
2. *j'ai perdu 34 kg en 4 mois et j'ai fait une **dépression**...* ; le connecteur *et* est situé dans le contexte gauche du marqueur *émotionnel* **peur** donc l'intensité (de polarité *négative*) sera amplifiée par le contexte énumératif (*j'ai fait une dépression*) ;
3. *Je suis très **angoissée** et hypocondriaque et...* ; le connecteur *et* est situé dans le contexte droit du marqueur *émotionnel* **angoissée** donc l'intensité (de polarité *négative*) sera amplifiée par l'amorçage d'une énumération (*hypocondriaque et...*) ;
4. *[avoir] tendance à évacuer mon **chagrin** en pleurant et depuis 4 semaines après avoir **pleuré**...* ; le connecteur *et* est situé dans le contexte droit du marqueur *émotionnel* **chagrin** donc l'intensité (de polarité *négative*) sera amplifiée par l'amorçage d'une énumération (*depuis 4 semaines après avoir pleuré...*) ;

2.2.5.2.6.4 donc :

- Fenêtre de 0 à 5 à gauche comme à droite (portée du connecteur *donc* assez large mais à contrôler pour éviter le hors-contexte) ;
- Que le connecteur soit présent à gauche ou à droite, l'intensité polaire de l'émotion est amplifiée, car cette conjonction de coordination vient conclure l'information précédente. La relation entre les deux informations permet en effet d'amplifier l'intensité émotionnelle exprimée.

Quelques exemples avec ce connecteur :

1. *J'ai un peu **peur** donc voila.* ; le *connecteur donc* est situé dans le contexte droit du marqueur *émotionnel **peur*** donc l'intensité (de polarité *négative*) sera amplifiée par la justification *voila* (très abstraites) qui vient appuyer l'état émotionnel exprimé par le patient ;
2. *Donc pas de **soucis**.* ; le *connecteur donc* est situé dans le contexte gauche du marqueur *émotionnel **soucis*** donc l'intensité (de polarité *positive* – après application de la *négation*) sera amplifiée. En effet, le caractère justificatif de la proposition amorcée par *donc* (*pas de soucis*) appuie l'information principale implicite ;
3. *Donc soyez **sereine**, tout semble normal chez vous.* ; comme dans l'exemple, le *connecteur donc* est situé dans le contexte gauche du marqueur *émotionnel **sereine*** donc l'intensité (de polarité *positive*) sera amplifiée car cette justification appuie une précédente information (implicite ici).

2.2.5.2.7 Incertitude comme forme pôle. Dans les cas où aucune émotion n'est présente dans un énoncé, ce sont les marqueurs d'incertitude qui constituent la forme pôle. Les mêmes traitements que ceux définis pour les émotions, sont alors appliqués, avec quelques variations dans les modifications d'intensité. Nous avons choisi de traiter l'incertitude au même rang que les émotions dans le but de voir le lien qu'il peut y avoir entre l'expression des émotions et de l'incertitude (section 2.2.7). Bien que le processus de traitements est le même que pour les émotions, le résultat est différent :

1. La négation cause l'atténuation (si l'étiquette sémantique est *pre-wi*, *post-wi*) ou l'inhibition (si l'étiquette sémantique est *pre-si*, *post-si*) de l'incertitude :
 - (a) *il n'y a **apparemment** pas de fracture...* ; atténuation de l'*incertitude* exprimée dans **apparemment** par la *négation* ;
 - (b) *il n'est pas **possible** de contracter une maladie...* ; inhibition de l'*incertitude* exprimée dans **possible** par la *négation* ;
2. L'incertitude cause l'amplification de l'incertitude :
 - (a) *Y a t il un **risque** de grossesse ?* ; amplification de l'*incertitude* exprimée dans **risque** par l'*incertitude* exprimée par ? ;
 - (b) *Je **voudrais** savoir que ce que je peux faire pour que sa passe...* ; amplification de l'*incertitude* exprimée dans **voudrais** par l'*incertitude* exprimée par *peux* ;
3. Les modifieurs causent une amplification de l'incertitude si le modifieur est de type *modif-m* sinon une atténuation de l'incertitude avec le modifieur de type *modif-p* :
 - (a) *c'est peu **probable** que ce soit une grossesse.* ; amplification de l'*incertitude* exprimée dans **probable** par le modifieur de type *modif-m* *peu* ;
 - (b) *Votre **ORL** a très **certainement** regardé vos cordes vocales.* ; atténuation de l'*incertitude* exprimée dans **certainement** par le modifieur de type *modif-p* *très* ;
4. Les connecteurs d'amplification (*et*, *car*, *donc* et *mais*), lorsqu'ils se trouvent à gauche ou à droite de l'incertitude, diminuent l'incertitude :

- (a) *il ne s'agit pas de lésions ligamentaires mais d'inflammation **probablement** des os...* ; atténuation de l'incertitude exprimée dans **probablement** par le connecteur mais ;
- (b) *Vous **pouvez donc** être enceinte.* ; atténuation de l'incertitude exprimée dans **pouvez** par le connecteur donc ;
- (c) *est-ce considérer comme prendre un **risque** car les lingettes sont humides...* ; atténuation de l'incertitude exprimée dans **risque** par le connecteur car ;
- (d) *Ce n'est pas douloureux et **semble** en surface.* ; atténuation de l'incertitude exprimée dans **semble** par le connecteur et ;

L'application de ces règles permet de calculer par la suite l'intensité polaire pour chacune des émotions et incertitudes présentes dans les énoncés. Ceci nous permet de définir la polarité émotionnelle dominante par énoncé (section 2.2.5.3).

2.2.5.3 Calcul de la polarité émotionnelle dominante

Le système prend en entrée un corpus de forum, organisés en fils de discussion :

1. Q/A (Questions/Réponses) : contient typiquement une question posée par un patient et sa réponse formulée par un professionnel de santé. Suite aux traitements, nous obtenons une représentation des intensités polaires correspondant aux différents marqueurs émotionnels ou d'incertitude (auquel cas juste l'intensité est calculée). Pour chaque fil de discussion, nous indiquons l'intensité polaire de chaque émotion ou incertitude dans la question et la réponse. Le tableau 2.8 présente un exemple de la structure des données. Un échantillon un peu plus exhaustif de ces données (un avant et un après l'application des règles) est disponible à l'annexe E.1.1.

TABLE 2.8 – Aperçu de la structure des données pour le corpus Q/A

Identifiants du fil	Questions	Réponses
9	pos ??	inc ?? inc ??
10	pos	inc
11	neg ??	inc0 inc
12	inc inc ?? inc ?? inc	inc0- inc
13	inc inc ?? inc ??	pos
64	neg+ ??????	ni émotion ni incertitude
762	inc ?? inc ??	inc- inc inc-
763	neg neg ??	pos
767	neg+ neg+ neg+	ni émotion ni incertitude
768	neg+	neu

Voici les phrases correspondantes pour l'identifiant 763 :

– Question :

- (a) neg : *Le seul soucis, c'est que j'ai peur des conséquences...* ; *soucis* est bien une émotion négative ;

(b) neg ?? : *j'ai peur des conséquences qui peux y avoir sur ma croissance...* ; *peur* est bien une émotion négative avec dans son contexte la présence d'un marqueur d'incertitude peux, d'où les ?? . Cependant, l'incertitude ne porte pas sur l'émotion mais sur les *conséquences*. C'est un exemple des limites du système.

– Réponse :

(a) pos : *Si vous en faites 1 ou 2 fois par semaine, pas de soucis* ; l'émotion *soucis* est négative mais par la présence d'un marqueur de négation, la polarité s'inverse pour devenir positive.

2. Forum : contient typiquement une question posée par un patient et une ou plusieurs réponses formulées par d'autres patients. Nous avons veillé à ne pas traiter les diffusions de publicité, les signatures et les citations, qui altèrent les résultats. Sans ces traitements, le système extrait plusieurs fois les mêmes informations. La structure des données a permis en effet de les repréner assez facilement (ces informations sont amorcées par le symbole typographique “-” en début de phrase). Pour chaque fil de discussion, nous indiquons l'intensité polaire de chaque émotion ou incertitude dans la question et dans chaque réponse. Le tableau 2.9 donne un aperçu de la structure des données. Un échantillon un peu plus exhaustif de ces données (un avant et un après l'application des règles) est disponible à l'annexe E.2.1.

TABLE 2.9 – Aperçu de la structure des données pour le corpus forum

Identifiants	Questions	Réponses
16	neg neu	ni émotion ni incertitude
19	neg	inc inc inc- : inc- inc
20	neu neg ?? neg ?? neg-	ni émotion ni incertitude
21	inc inc ??- inc ??	ni émotion ni incertitude
43	neg neg+ neg	pos ?? : inc : pos : pos : neu neu pos neu : pos ?? : neg+ neg
105	pos pos++ pos ?	neu+ pos+ neu : neg neg pos+ neg neu+ neg pos neu
162	neu neg+	pos pos ?+

Voici les phrases correspondantes pour l'identifiant 162 :

– Question :

1. neu : *Bonsoir à tous !!!;!!!* est bien une émotion neutre ;
2. neg+ : *Tellement j ai peur* ; *peur* est une émotion négative et elle est employée avec un marqueur de négation tellement donc sa polarité négative est amplifiée, d'où le +.

– Réponse :

1. pos : *tu n'as rien à craindre...* ; *craindre* est une émotion négative mais avec la présence du marqueur de négation rien, sa polarité est inversée pour devenir positive ;
2. pos ?+ : *et tu seras mieux soigner et soulagé si le diagnostic est précis!* ; *soulagé* est bien une émotion positive, employée avec le connecteur et qui amplifie l'émotion (d'où le +). De plus, la présence du marqueur d'incertitude si, qui interagit sur l'émotion, va atténuer son intensité (d'où le ?).

Le tableau 2.10 détaille les différents types d'indices d'intensité. Nous choisissons d'attribuer (empiriquement) des valeurs aux différents indices d'intensité. Ces valeurs permettent de calculer la polarité émotionnelle dominante au sein d'un énoncé (question ou réponse(s)). Néanmoins, ces valeurs pondératives caractérisent l'indice d'intensité qu'elles représentent :

– Représentations polaires :

1. *pos* pour les émotions à polarité positive comme *joie*, *soulagement*,... ou celles à polarité d'origine négative, qui ont subi l'inversion de leur polarité due à la présence de marqueurs de négation, comme *inquiétude*, *peur*,... ;
2. *neg* pour les émotions à polarité négative comme *tristesse*, *regret*,... ou celles à polarité d'origine positive, qui ont subi une inversion de polarité ;
3. *neu* pour les émotions à "polarité" neutre ; *inc* pour incertitude, dans les cas où aucune émotion n'est détectée dans un énoncé mais où il y a un ou plusieurs marqueurs d'incertitude ; *inc0* pour l'inhibition d'incertitude (ou certitude), dans les mêmes cas que pour *inc* mais avec la présence de marqueurs de négation qui modifient l'expression d'incertitude en certitude.

– Représentation des intensifieurs :

1. + indique une amplification de la polarité émotionnelle : si nous avons pour un énoncé *pos++ neg+*, la polarité émotionnelle positive sera amplifiée deux fois contre une seule pour la polarité négative. Donc, la polarité émotionnelle dominante de cet énoncé sera positive (car de plus en plus forte intensité) ;
2. - qui indique une atténuation polaire. Si nous avons pour un énoncé *pos- neg*, la polarité émotionnelle positive sera atténuée une fois contre aucune pour la polarité négative. Donc, la polarité émotionnelle dominante de cet énoncé sera négative (car non atténuée) ;
3. ? indique la présence d'un marqueur d'incertitude qui atténue une polarité émotionnelle. Si nous avons pour un énoncé *pos neg ?*, la polarité émotionnelle dominante sera positive car la présence d'un marqueur d'incertitude dans le contexte de l'émotion négative atténue son intensité polaire.

– Description des valeurs pondératives : nous choisissons d'attribuer des valeurs pondératives pour chaque polarité (*pos*, *neg*, *neu*). Ces trois indices servent à initialiser les différentes intensités polaires dans chaque énoncé. La valeur d'intensité de la polarité émotionnelle la plus élevée représente la polarité émotionnelle dominante de l'énoncé. Cette valeur est nuancée par les valeurs pondératives des intensifieurs (+ : +0.1, - : -0.1 et ? : -0.05). En additionnant chacune de ces valeurs pour chacune des polarités, nous sélectionnons la polarité émotionnelle qui obtient le plus haut score et nous la définissons comme polarité émotionnelle dominante de l'énoncé.

À partir de ces données, nous calculons la polarité globale (*positive*, *neutre*, *négative* ou incertitude si aucune émotion n'est détectée) de tous les états émotionnels manifestés dans un énoncé. Pour ceci, nous associons aux différents indices d'intensité des valeurs numériques, qui permettent d'estimer le score de valence (ou pondération polaire) pour chacune des émotions présentes dans les questions et réponses (Moreno-Ortiz *et al.*, 2013; Paroubek, 2010b). Cela permet par la suite de calculer le score de valence pour chaque polarité émotionnelle (*positive*, *neutre*, *négative* et incertitude, si aucune émotion n'a été détectée). C'est seulement

TABLE 2.10 – Description des différents indices d'intensité

Indices	pos	neg
Représentations	polarité positive	polarité négative
Valeur pondérative	+0.1	+0.1
Indices	neu	+
Représentations	“polarité” neutre	amplification
Valeur pondérative	+0.1	+0.1
Indices	-	?
Représentations	atténuation	incertitude
Valeur pondérative	-0.1	-0.05 (forme pôle \in [pos,neu,neg]), sinon +0.01 (forme pôle \in [inc])
Indices	inc	inc0
Représentations	pôle = incertitude (absence de marqueur émotionnel)	inhibition de l'incertitude (= certitude, si forme pôle = incertitude)
Valeur pondérative	+0.1	ignoré

après cette étape que nous sommes en mesure de déterminer la polarité émotionnelle la plus intense. Nous avons déterminé de façon empirique des valeurs d'intensité pour chaque indice présenté dans le tableau 2.10. Nous avons choisi de ne pas prendre en considération l'indice *inc0* puisqu'il ne vient pas altérer le calcul de la polarité globale émotionnelle. Nous estimons le score de valence d'un énoncé suivant ces trois formules :

1. calcul du score de valence par émotion : $S_{emo}^{pol} = \sum_{i \in I} indice(i)$;
2. calcul du score de valence par polarité : $S_{pol} = \sum_{n=1}^N S_{emo}^{pol}(n)$;
3. calcul de la polarité globale émotionnelle d'un énoncé : $PGE_{enonce} = max(S_{pol})$;

où :

- S_{emo} : le score de valence associée à une émotion (ou incertitude) ;
- *pol* : la polarité émotionnelle \in {pos,neu,neg,inc} ;
- *indice(i)* : la valeur pondérative ;
- *I* : les indices \in {pos,neg,neu,+,-,?,inc} ;
- S_{pol} : le score de valence total pour une polarité ;
- *N* : le nombre de S_{emo} pour une même polarité dans un énoncé ;
- PGE_{enonce} : la polarité globale émotionnelle d'un énoncé ;
- $max(S_{pol})$: le S_{pol} qui a la plus grande valeur.

Cependant, lors du calcul de PGE_{enonce} , il arrive que les S_{pol} de différentes polarités soient égales. Dans ces situations, il y a plusieurs cas de figure possibles :

- si $S_{pol=pos} = S_{pol=neg}$ alors $PGE_{enonce} = neutre$;
- si $S_{pol=pos} = S_{pol=neu}$ alors $PGE_{enonce} = positif$;
- si $S_{pol=neg} = S_{pol=neu}$ alors $PGE_{enonce} = négatif$.

Après l'application de ces calculs, chaque fil de discussion est décrit avec les informations suivantes :

- identifiant du fil de discussion,
- la polarité globale émotionnelle de la question du fil,
- la polarité globale émotionnelle de la réponse (corpus Q/A) ou des réponses (corpus forum) du fil.

Les questions et réponses ne contenant ni émotions ni incertitudes portent la mention “nothing”. Cependant, nous effectuons une étape supplémentaire de traitement pour le corpus forum, où il est nécessaire de déterminer une polarité émotionnelle globale pour l'ensemble des réponses. Pour ceci, nous récupérons les polarités globales de chaque réponse et nous sélectionnons la polarité émotionnelle qui obtient le score le plus élevé.

Le tableau 2.11 donne un aperçu de la structure des données pour le corpus forum (un échantillon un peu plus exhaustif de ces données (un avant et un après l'application des règles) est disponible à l'annexe E.2.2). Le tableau 2.12 présente un aperçu de la structure des données que nous exploitons lors de la phase d'évaluation du système. Un échantillon un peu plus exhaustif de ces données (un avant et un après l'application des règles) est disponible à l'annexe E.1.2 pour le corpus Q/A et à l'annexe E.2.3 pour le corpus forum.

TABLE 2.11 – Aperçu de la structure des données affichant les polarités globales émotionnelles par énoncé

Identifiants	Questions	Réponses
9	inc	pos : inc : inc : inc : neu
10	pos	inc : neg : inc : inc
11	inc	inc
12	neg	nothing
13	neg	inc : inc : neg : inc
143	pos	nothing
144	neg	inc : inc
145	inc	inc : pos : neu : neu
146	pos	neu : neg
147	neu	neg : inc
148	neu	inc

TABLE 2.12 – Aperçu de la structure des données affichant les polarités globales émotionnelles par énoncé

Identifiants	Questions	Réponses
17	inc	inc
18	neg	nothing
19	neg	inc
80	nothing	inc
81	neg	neu
82	nothing	pos
83	pos	neg

2.2.6 L'évaluation du système

Pour effectuer une évaluation du système, nous avons demandé à trois annotateurs (un expert AE1, un semi-expert ASE2 et un non expert ANE3) d'annoter un échantillon de fils de discussion de chaque corpus (un échantillon de données annotées par chaque annotateur est disponible à l'annexe F.1 pour le corpus Q/A et à l'annexe F.2 pour le corpus forum). Les annotateurs doivent définir la polarité émotionnelle pour chacune des phrases des énoncés contenant au moins une émotion (détectée par le système). Dans le cas où plusieurs émotions sont présentes, l'annotateur doit spécifier la polarité émotionnelle dominante. Les étapes de la création des données de référence sont les suivantes.

1. le corpus de développement contient 500 fils du corpus *Q/A* et 80 fils du corpus *forum* ;
2. la sélection de ces données a été faite de manière aléatoire, avec la contrainte d'avoir des phrases avec au moins un marqueur émotionnel et provenant des questions et des réponses ;
3. la structure des fichiers d'annotation prévoit quatre parties pour chaque phrase à annoter :
 - (a) l'identifiant de la phrase avec des informations sur le type d'énoncé et le numéro de la réponse pour le corpus *forum* ;
 - (b) la phrase contenant au moins un marqueur émotionnel, qui est alors entouré par des étoiles "***" ;
 - (c) un champs pour ajouter la polarité émotionnelle détectée par l'annotateur : "+" pour *positive*, "-" pour *négative*, "/" pour *neutre* ;
 - (d) un champs pour mettre les indices émotionnels justifiant le choix de l'annotateur. Il s'agit de l'expression textuelle de l'émotion : le marqueur de l'émotion accompagné ou non d'autres marqueurs (incertitude, négation, modificateurs, connecteurs). Cette information peut être facultative si aucune émotion n'est présente ou si l'annotateur n'est pas sûr de ses choix.

Ces données de référence sont utilisées pour effectuer l'évaluation du système à base de règles (section 2.2.5) pour la tâche du calcul de l'émotion dominante. Le tableau 2.13 montre un exemple d'annotations réalisées par les annotateurs. Par exemple, dans la phrase, *J' ai **peur** , mais vraiment une **phobie** de l' hosto .*, nous avons les deux émotions de polarité négative *peur* et *phobie*. Dans ce cas, la polarité émotionnelle dominante est négative, que l'annotateur représentera avec le symbole -. De plus, pour justifier ce choix, l'annotateur a choisi de spécifier l'information émotionnelle *phobie* ainsi que le modifieur d'amplification *vraiment* qui vient intensifier le marqueur émotionnel *phobie*. Dans l'exemple, *je ne sais plus quoi penser , et **hésite** avant cette longue opération risquée .*, l'annotateur décide d'attribuer la polarité neutre, par le symbole / et en justifiant son choix en précisant simplement le marqueur émotionnel *hésite*. Pour la phrase *Bravo pour la réussite de ton opération **!!!**!*, l'annotateur trouve que la polarité émotionnelle qui se dégage de cette phrase est positive. Il justifie son choix en précisant simplement *!!!*. Enfin, dans la phrase *ce traitement n' est pas **agressif** et semble bien vous soulager .*, l'annotateur annote cette phrase avec la polarité émotionnelle positive. En effet, la présence du marqueur de négation *n' . . pas*, la polarité émotionnelle d'origine (négative) de *agressif* est inversée pour devenir positive. L'annotateur justifie son choix en précisant *pas agressif* comme information émotionnelle. Dans notre étude, nous exploitons uniquement l'information sur la polarité émotionnelle. Nous avons donc récupéré les polarités émotionnelles définies par les annotateurs (*positives, négatives, neutres*). Afin de pouvoir évaluer les performances de notre système

automatiquement, nous avons dû aligner les annotations des annotateurs sur celles du système. En effet, après l'application des règles sur chacune des émotions présentes dans un énoncé, le système a pu calculer la polarité émotionnelle dominante de cet énoncé. Donc, pour chacune des annotations des annotateurs, nous regardons, pour chaque énoncé (grâce à son identifiant), la polarité émotionnelle qu'ils ont attribuée à chacune des phrases contenant au moins un marqueur émotionnel. Nous calculons l'intensité pour chaque polarité émotionnelle, en faisant la somme de chaque polarité (nombre de phrases positives, nombre de phrases négatives et nombre de phrases neutres). La polarité émotionnelle qui obtient le plus haut score, est définie comme étant la polarité émotionnelle dominante. Ainsi, nous obtenons des données homogènes que nous pouvons comparer. Nous avons décidé d'évaluer la capacité du système à reconnaître la polarité émotionnelle dominante de la question et de la réponse pour un fil de discussion (patient/professionnel ou patient/patient). C'est pour cette raison que nous choisissons de calculer uniquement la précision.

L'annotation de l'émotion elle-même pourra servir dans de futurs travaux pour évaluer la validité de différents marqueurs détectés automatiquement par notre système.

TABLE 2.13 – Aperçu de la structure d'un fichier de validation proposé aux annotateurs

Corpus	Id.	Phrases avec marqueur(s) émotionnel(s)	Polarité	Émotion
Q/A	Q5657	Cela me **stresse** !	-	stresse
	Q5927	J ' ai **peur** , mais vraiment une **phobie** de l ' hosto .	-	vraiment phobie
	R6116	ce traitement n ' est pas **agressif** et semble bien vous soulager .	+	pas agressif
	R6015	Chère Nadine , Plusieurs travaux , en particulier américains , ont essayé de relier diabète et **dépression** .	/	X
forum	Q2	je ne sais plus quoi penser , et **hésite** avant cette longue opération risquée .	/	hésite
	A2_32	Bravo pour la réussite de ton opération **!!!** !	+	!!!
	A2_34	Essaie de voir avec une Assistante Sociale de la Sécu , mais j ' en **doute** fort **!!!** !	-	doute
	Q4	Malheur **!!!** !	-	malheur !!

L'accord inter-annotateurs est calculé avec le coefficient kappa de Cohen. Le calcul est effectué par paire d'annotateurs : AE1/ASE2, AE1/ANE3, ASE2/ANE3. Nous comparons les annotations sur la polarité des émotions (- + /). Nous calculons d'abord les six combinaisons possibles d'accord inter-annotateurs (trois par corpus) et ensuite les moyennes au niveau des corpus. Le tableau 2.14 montre les résultats de l'accord inter-annotateur. Nous pouvons voir que :

- AE1/ASE2 : avec un coefficient kappa de 0.63 pour le corpus Q/A, l'accord est considéré comme fort d'après la grille d'interprétation du kappa de Cohen (Narayanan *et al.*, 2009). En ce qui concerne le corpus forum, l'accord inter-annotateurs est modéré : 0,58 ;
- AE1/ANE3 : l'accord entre ces deux annotateurs est modéré dans les deux corpus : 0,56 (corpus Q/A) et 0,50 (corpus forum). Comme dans le cas précédent, l'accord inter-annotateurs est moins fort pour le corpus forum ;

TABLE 2.14 – Aperçu de la structure d’un fichier de validation proposé aux annotateurs

Corpus \ Accords	AE1 / ASE2	AE1 / ANE3	ASE2 / ANE3	Moyenne
Q/A	0.63	0.56	0.64	0.61
Forum	0.58	0.50	0.66	0.58

- ASE2/ANE3 : les accords inter-annotateurs sont forts et plus élevés que dans les cas précédents (0,64 pour le corpus Q/A et 0,66 pour le corpus forum), surtout dans le corpus forum, qui montrait jusqu’ici un accord modéré.

De manière générale, les annotateurs ont montré un accord plus faible dans l’annotation de la polarité *neutre*. L’attribution de cette polarité pour les phrases n’évoquant aucune émotion particulière (selon l’interprétation de chaque annotateur), favorise encore plus les désaccords, plus fréquents, sur cette polarité. Les accords les plus élevés sont observés entre les annotateurs ASE2 et ANE3 avec un coefficient kappa de 0,64 pour le corpus Q/A et 0,66 pour le corpus forum. Les accords les plus faibles (0,56 pour le corpus Q/A et 0,50 pour le corpus forum) sont observés entre les annotateurs AE1 et ANE3. Ceci peut être expliqué en partie par la différence du niveau d’expertise du domaine médical et du travail avec les corpus. Par exemple, l’annotateur ANE3 a tendance à catégoriser certains troubles ou maladies comme représentant un état émotionnel alors que l’annotateur expert AE1 peut distinguer l’expression d’une émotion d’un état de santé factuel, qui peut lui aussi être positif (amélioration de l’état de santé) ou négatif (détérioration de l’état de santé).

Notons aussi que les accords plus faibles observés pour le corpus forum, peuvent être dus à la présence fréquente d’ironie et de sarcasmes. En effet, ces deux fonctions rhétoriques peuvent exprimer implicitement un état émotionnel différent de celui exprimé explicitement et restent difficiles à détecter et annoter (Liebrecht *et al.*, 2013).

De manière générale, nous pensons que les différences d’annotations sont dues au caractère très subjectif de l’interprétation des émotions, ce qui est accentué par le contexte médical. Effectivement, il peut être difficile de distinguer l’aspect factuel d’un trouble ou d’une maladie (physique ou psychologique) d’un état émotionnel (peine, peur, soulagement, joie). Par exemple, le terme *dépression*, qui a une connotation négative, porte une signification médicale et factuelle, mais il peut également être assimilé à un état émotionnel, également négatif. Par exemple, dans la phrase *Cher David, Vous décrivez les symptômes typiques d’une **dépression** chronique.*, le terme *dépression*, reconnu comme un marqueur émotionnel par le système, a été perçu comme tel par les annotateurs semi-expert et non-expert alors que l’annotateur expert l’a considéré plutôt comme un état factuel. L’inconsistance interne des annotateurs peut aussi causer une instabilité dans les annotations, y compris pour des émotions apparaissant dans des contextes très proches. Par exemple, un annotateur peut commencer à annoter les phrases avec un certain état d’esprit, puis au fur et à mesure de ses annotations, sa vision de l’expression des états émotionnels et sa compréhension de la tâche peuvent changer.

2.2.7 Les modèles de visualisation

La visualisation des données aide à mieux interpréter les résultats d'une étude. En effet, en plus des données chiffrées, la visualisation offre la possibilité de mettre en lumière des informations supplémentaires. Notre modélisation s'articule autour des échanges entre patients (corpus forum) et des échanges entre patients et professionnels de santé (corpus Q/A). Nous proposons deux types de visualisation :

- visualisation statique, avec une vue cumulative des émotions et de leur distribution ;
- visualisation dynamique, avec une vue représentant la relation entre les questions et les réponses.

Nous réalisons cinq modélisations différentes pour la représentation des états émotionnels et de la relation avec les informations contextuelles.

Les différents mécanismes de visualisation que nous proposons offrent chacun un niveau différent de visualisation des émotions :

- Dans la section 2.2.7.1, nous proposons de visualiser trois types d'informations :
 1. Tout d'abord, nous proposons de visualiser la distribution des concepts médicaux (procédure, maladie, médicament) pour chacune des polarités positive, négative, neutre et aussi pour l'incertitude (pour les énoncés où aucune émotion n'est détectée mais avec la présence de marqueurs d'incertitude). Cette représentation offre la possibilité d'observer la proportion des différents types médicaux employés dans les contextes émotionnels en fonction de la polarité des marqueurs émotionnels ;
 2. Ensuite, nous proposons d'observer le dynamisme dans les deux types d'échanges à travers la relation question/réponse : patients/professionnels et patients/patients. Cette représentation nous permet de voir les polarités émotionnelles caractérisant les questions posées par les patients et leur impact sur les polarités émotionnelles caractérisant les réponses des professionnels (corpus Q/A) ou des patients (corpus forum). Cela permet d'observer l'interaction émotionnelle entre les patients et les professionnels ou patients, selon le corpus. Nous obtenons ainsi différentes paires de polarités émotionnelles : une polarité émotionnelle dominante pour la question et une autre pour la réponse. Ceci offre différentes combinaisons possibles (pos/neg, neg/neg, pos/inc,...) ;
 3. Enfin, la dernière visualisation proposée donne un aperçu du concentré émotionnel (positif, négatif, neutre, incertitude) caractérisant l'expression des émotions chez les patients et chez les professionnels. Cette représentation permet de voir les polarités émotionnelles dominantes pour chaque type d'énoncé (les questions posées par les patients et les réponses formulées par les professionnels ou par les patients) et par type d'échanges (corpus Q/A et corpus forum) ;
- Dans la section 2.2.7.2, nous proposons d'observer la significativité de l'emploi des différents types de marqueurs (modificateurs, négation et incertitude) au sein des contextes des marqueurs émotionnels. Cela permet de voir les tendances des patients ou des professionnels à amplifier, atténuer, négativer ou à manifester de l'incertitude quant à leurs émotions en fonction de leur polarité (positive, négative, neutre ou incertitude) ;
- Dans la section 2.2.7.3, nous proposons une troisième modélisation qui consiste à représenter la distribution des émotions parmi les différents types d'énoncés (questions et réponses) pour chacun des

deux corpus (Q/A et forum), en fonction de leur polarité (positive, négative, neutre et incertitude) et de leur catégorie émotionnelle (joie, peur, tristesse, soulagement, . . .). Cela a pour objectif de visualiser l'intensité des polarités émotionnelles et des catégories émotionnelles caractérisant les différents types d'énoncés (questions et réponses). Grâce à cette visualisation, il est ainsi possible de voir comment évoluent les états émotionnels dans les différents types d'échanges (patients → professionnels et patients → patients) ;

- Dans la section 2.2.7.4, nous proposons de visualiser les polarités émotionnelles ainsi que leurs catégories émotionnelles associées à travers une représentation en "constellations". C'est-à-dire que chaque polarité émotionnelle (positive, négative, neutre et incertitude) est représentée par des cercles qui indiquent leur fréquence d'emploi, par questions et par réponses, pour chaque corpus. De plus, pour chacune de ces polarités, nous leur associons d'autres cercles, qui représentent des catégories émotionnelles de cette polarité, reliés par des arcs. La taille des cercles et la longueur des arcs définissent la fréquence d'emploi des marqueurs émotionnels en fonction de leur polarité et de leur catégorie émotionnelle, et ce pour chaque énoncé ;
- Dans la section 2.2.7.5, nous proposons une modélisation basée sur les nuages de mots. Cette modélisation a pour vocation de visualiser les catégories émotionnelles dominantes dans les questions et les réponses, pour chacun des deux corpus. Cette représentation permet de voir les informations sémantiques (marqueurs de négation, modificateurs, types médicaux) présentes dans les contextes de différentes catégories émotionnelles exprimées dans les énoncés. En comparant les nuages de mots correspondant aux questions à celui correspondant aux réponses, il est possible d'observer l'évolution des émotions exprimées dans les différents échanges.

2.2.7.1 Dynamique des émotions et des incertitudes

Une première visualisation permet de représenter la distribution des polarités globales des émotions et des trois types médicaux (troubles, procédures, médicaments) se trouvant dans leur contexte (fenêtres de 7 tokens à gauche et 7 tokens à droite). Nous traitons séparément les questions et les réponses. Nous utilisons deux échelles : en nombre d'occurrences et en pourcentage. Cette visualisation est effectuée avec le programme gnuplot¹⁰. Deux possibilités sont alors offertes :

- la visualisation statique propose une vue cumulative des émotions et de leur contexte pour chaque question et réponse ;
- la visualisation dynamique propose la relation entre les questions et les réponses. Cela indique l'évolution de l'état émotionnel véhiculé par les patients (et professionnels) dans leurs échanges. En effet, lorsqu'un patient pose sa question sur une plateforme du web social, il transmet généralement une émotion en relation avec ses questionnements ou soucis de santé. D'autres personnes (patients ou professionnels), qui se sentent concernées, peuvent répondre à cette question ou exprimer un soutien quelconque. Nous pensons que l'émotion initiale exprimée par l'auteur de la question peut se propager dans le fil de discussion, ou bien elle peut être transformée en d'autres émotions dans les réponses. De ce point de vue, il peut être intéressant d'observer le dynamisme des émotions et de leur évolution dans

10. <http://www.gnuplot.info/>

un fil de discussion. Nous exprimons cette dynamique grâce à la définition de la polarité émotionnelle dominante des questions et des réponses.

Le tableau 2.15 montre les différentes combinaisons possibles représentant la dynamique temporelle des émotions.

TABLE 2.15 – Les différentes combinaisons des couples question/réponse(s) en terme de polarité globale émotionnelle

Question \ Réponse	inc	inc0	neg	neu	no	pos
inc	inc/inc	inc/inc0	inc/neg	inc/neu	inc/no	inc/pos
inc0	inc0/inc	inc0/inc0	inc0/neg	inc0/neu	inc0/no	inc0/pos
neg	neg/inc	neg/inc0	neg/neg	neg/neu	neg/no	neg/pos
neu	neu/inc	neu/inc0	neu/neg	neu/neu	neu/no	neu/pos
no	no/inc	no/inc0	no/neg	no/neu	no/no	no/pos
pos	pos/inc	pos/inc0	pos/neg	pos/neu	pos/no	pos/pos

Les conventions de notation sont les suivantes :

- no : aucun marqueur d’émotion ou d’incertitude détecté,
- inc(0) : aucun marqueur émotionnel détecté, mais un ou plusieurs marqueurs d’incertitude sont présents dans l’énoncé (section 2.2.5),
- 0 : inhibition de l’incertitude, peut être interprété comme une marque de *certitude*,
- pos, neu et neg : représentent les polarités *positive*, *neutre* et *négative*, respectivement.

Ce modèle de visualisation permet aussi d’observer la distribution de ces notations par type d’énoncé et pour chacun des deux corpus. Nous pouvons alors caractériser l’état émotionnel ou d’incertitude dans les différents types échanges : patients/patients et patients/professionnels de santé.

2.2.7.2 Graphes non orientés

L’objectif visé par la deuxième modélisation est de proposer une visualisation de la distribution des différents marqueurs (*incertitude (faible et forte)*, *négation*, *modifieurs (faible et fort)*) et des types médicaux (*procédure*, *maladie/trouble*, *médicament/traitement*) dans les contextes gauche et droit des quatre polarités émotionnelles dominantes : *positive*, *négative*, *neutre* et *incertitude*. Nous calculons un ratio, qui va indiquer l’intensité de la relation entre un marqueur et une polarité émotionnelle (par exemple, est-ce que les modifieurs d’atténuation sont plus employés avec les émotions négatives ? les marqueurs de négation sont-ils plus employés avec des émotions positives ?...). Ce ratio est calculé d’après la formule $R_m = \frac{\text{pourcentage}_m \times N_m}{\text{lissage} \times N_{pol}}$, où :

- R_m est le ratio du marqueur m en fonction du nombre d’émotions d’une polarité donnée pol ;
- pourcentage_m est la proportion du marqueur m présent dans les contextes d’émotions d’une polarité donnée pol ;
- lissage est un indice de lissage pour l’affichage du ratio. Nous le fixons à 100 dans nos expériences ;
- N_m est le nombre total d’occurrences du marqueur m ;

- N_{pol} est le nombre total d’occurrences d’émotions d’une polarité donnée pol .

Cette formule sert à déterminer la significativité d’un marqueur avec une polarité émotionnelle. Par exemple, si nous avons pour un marqueur de type *modifieur* de catégorie *modif-p* 20 % d’occurrences avec des émotions positives, 50 % avec des émotions négatives et 30 % avec des émotions neutres, et si nous obtenons un ratio de 0.05 pour les émotions positives, de 0.02 pour les émotions négatives et de 0.1 pour les émotions neutres, dans ce cas, l’emploi des marqueurs possédant la mention sémantique *modif-p* sera plus significatif avec les émotions neutres. Ceci signifie que nous pouvons déduire que l’intensification des émotions est plus marquée avec les émotions de la catégorie *neutre*. Nous allons prendre un autre exemple avec nos données. Nous prenons l’exemple des marqueurs de négation, répartis comme suit parmi les polarités émotionnelles *positive*, *négative*, *neutre* et *"incertitude"* (corpus forum – questions) :

- positive (808 occurrences de marqueurs émotionnels) : 30,58% avec un ratio de 3,3 ;
- neutre (494 occurrences de marqueurs émotionnels) : 13,90% avec un ratio de 2,45 ;
- négative (1347 occurrences de marqueurs émotionnels) : 43,80% avec un ratio de 2,83 ;
- incertitude (629 occurrences de marqueurs émotionnels) : 11,72% avec un ratio de 1,63.

Nous voyons qu’avec 808 occurrences de marqueurs émotionnels, les émotions à polarité positive contiennent 30,85% des marqueurs de négation contre 43,80% pour les émotions négatives avec 1347 occurrences de marqueurs émotionnels. Mais le ratio des émotions positives est plus élevé (3,3) que celui des émotions négatives (2,83). Cela signifie que les marqueurs de négation sont plus fréquemment employés avec des marqueurs émotionnels positifs que négatifs. Donc, la relation marqueur émotionnel négatif/marqueurs de négation est moins significative que la relation marqueur émotionnel positif/marqueurs de négation. Cette observation permet de déduire que les patients utilisent plus souvent des émotions positives avec des marqueurs de négation, ce qui va inverser la polarité émotionnelle. Cette visualisation est effectuée avec des graphes non orientés générés avec le programme “neato”¹¹.

2.2.7.3 Diagrammes circulaires

La troisième modélisation permet également de visualiser la distribution des émotions par polarité et par marqueur émotionnel. Nous utilisons cette fois des diagrammes circulaires générés avec la technologie SVG¹², qui permet aussi l’affichage d’informations détaillées (section 3.4). Les différentes distributions sont exprimées en pourcentage. Les informations détaillées sont affichées lors du passage de la souris sur une portion du graphique. Elles sont exprimées en nombre d’occurrences. Ces informations détaillées permettent de représenter les différentes catégories émotionnelles (*joie*, *peur*, *tristesse*, *soulagement*, ...) selon la polarité (*positive*, *négative*, *neutre*, *"incertitude"*) employées par les patients et les professionnels. Ces informations détaillées permettent aussi de visualiser les différents types de marqueurs (*incertitude*, *modifieurs*, *négation*, *types médicaux*) présents dans les contextes des différents marqueurs émotionnels.

11. <http://www.graphviz.org/Documentation/neatoguide.pdf>

12. Scalar Vector Graphics : <http://svground.fr/>

2.2.7.4 Constellations

Cette nouvelle modélisation ressemble à celle précédemment décrite (section 2.2.7.3) tout en proposant une vue différente. Elle permet de visualiser les ensembles polaires (*positif, neutre, négatif* et incertitude) associés à leurs catégories émotionnelles. Chaque polarité est représentée par un cercle central, auquel sont liées les catégories émotionnelles, également représentées comme des cercles. La taille de l'arc est indicative de la fréquence des marqueurs émotionnels : plus la catégorie d'émotion est fréquente plus elle est éloignée dans cette représentation. Par exemple, comme pour la polarité *pos* il y a beaucoup de marqueurs exprimant la *joie*, le cercle correspondant à cette catégorie émotionnelle est éloignée du cercle de polarité positive. De plus, la taille des cercles est aussi proportionnelle à la fréquence des marqueurs correspondants. La visualisation est dite *de constellation* et est également effectuée avec la technologie SVG.

2.2.7.5 Nuage de mots

Une dernière visualisation utilise la technologie de nuage de mots¹³. Dans cette représentation, l'importance d'un terme est proportionnelle à sa taille. De cette manière, nous voulons faire ressortir les termes *marqueurs, émotions, médicaux* fréquents. Nous effectuons un contrôle sur les informations à afficher. Par exemple, pour visualiser les types médicaux associés à l'expression d'émotions positives, seuls ces deux types de termes sont sélectionnés. Le graphique est généré à partir de leurs cooccurrences. Le type de terme est représenté par la couleur et leur fréquence par la taille de la police. De plus, une implémentation javascript permet d'obtenir des informations supplémentaires, comme la fréquence des termes ou son type exact. Nous effectuons une optimisation d'affichage (Batra & Rao, 2010) pour les cas où il existe un écart important de fréquences, par exemple lorsqu'un terme apparaît 500 fois et un autre 2 fois.

2.3 Expériences

Nous effectuons différentes expériences pour visualiser les émotions et leur évolution au sein de nos deux corpus de travail. Pour vérifier la qualité des données que nous exploitons pour nos modèles de visualisation, nous mesurons la précision du système dans la détection de la polarité émotionnelle dominante par énoncé, questions et réponses. Nous évaluons plusieurs sorties de notre système :

1. **Baseline** : nous utilisons comme baseline les étiquettes sémantiques brutes, sans aucun traitement sur les marqueurs émotionnels. La polarité émotionnelle dominante d'un énoncé est calculée directement à partir de ces étiquettes sémantiques ;
2. **Processed** : nous appliquons la chaîne de traitement, qui vise à ajuster la polarité émotionnelle d'origine selon les règles définies dans la section 2.2.5.2. La polarité émotionnelle dominante d'un énoncé est alors calculée à partir de ces nouvelles polarités émotionnelles ;
3. **Processed + Accentuation** : nous appliquons les règles de la chaîne de traitements puis nous procédons à l'accentuation des termes (présents dans les lexiques) de type *émotion, modifieur, incertitude* et

13. Sites proposant la génération de nuages de mots : <http://www.netpublic.fr/2013/12/22-outils-gratuits-nuages-de-mots-cles/>

médical afin de d'observer dans quelle mesure l'accentuation peut améliorer les résultats ;

4. **Processed + Dernière émotion** : nous appliquons les règles de la chaîne de traitements puis nous essayons également de ne prendre en compte que la dernière émotion exprimée dans une phrase. En effet, lorsque plusieurs émotions sont présentes dans une même phrase, la dernière émotion exprimée peut avoir une dominance sur les émotions précédentes, comme le montrent les travaux (pour la langue chinoise) de (Yu *et al.*, 2013).
5. **Processed + Accentuation + Dernière émotion** : nous proposons de combiner l'application des règles de la chaîne de traitements, l'accentuation des termes non accentués et la prise en compte de la dernière émotion dans une phrase.

Nous proposons d'évaluer l'impact de ces différents traitements (calcul de nouvelles polarités, réaccentuation, dernière émotion). Nous effectuons l'évaluation de ces traitements en utilisant les deux sous-corpus de test (*Q/A, forum*) et les données de référence, et plus particulièrement, nous prenons en compte les phrases annotées de façon identique par deux annotateurs donnés. Pour chaque énoncé, nous récupérons les phrases associées afin de calculer la polarité émotionnelle dominante des questions et réponses des sous-corpus de test. Dans nos modèles de visualisation, nous utilisons la configuration présentant les meilleurs résultats selon le corpus.

Dans le chapitre suivant, nous présentons et discutons l'évaluation du système à base de règles et les résultats obtenus à partir de différents mécanismes de visualisation.

Chapitre 3

Résultats et Discussion

Nous présentons et discutons les résultats de l'évaluation de notre système dans la section 3.1. Ensuite, nous discutons des résultats d'interprétation des différentes modélisations pour la visualisation de l'évolution des émotions dans les différents types d'échanges (patients/professionnels, dans le corpus Q/A, et patients/professionnels, dans le corpus forum). Dans la section 3.2, nous discutons des résultats obtenus avec la visualisation des concepts médicaux par polarité émotionnelle, des résultats obtenus avec la représentation des polarités émotionnelles dominantes par énoncé (questions et réponses) selon diverses associations possibles et des résultats obtenus en visualisant la distribution des polarités émotionnelles dominantes dans les questions et réponses pour chacun des deux corpus. Dans la section 3.3, nous présentons et discutons les résultats obtenus dans l'expérience consistant à déterminer une possible relation entre les marqueurs (modificateurs, négation et incertitude) et les marqueurs émotionnels en fonction de leur polarité émotionnelle (positive, négative, neutre et incertitude). Dans la section 3.4, nous présentons et discutons des résultats obtenus à partir de la modélisation des diagrammes circulaires. Ces diagrammes présentent la distribution des polarités émotionnelles et leurs catégories émotionnelles pour chaque type d'énoncés et pour chaque forum. Dans la section 3.5, nous présentons et discutons les résultats relatifs à la représentation des émotions (polarités et catégories) par la modélisation en constellation. Enfin, la section 3.6 présente les résultats obtenus en réalisant les nuages de mots.

3.1 Évaluation du système

Lorsque nous traitons les corpus avec notre système, il applique les différentes règles (négation, modificateurs, incertitude, connecteurs) sur les marqueurs émotionnels présents dans les énoncés (que nous avons préalablement extraits et stockés dans une structure de données, par énoncé et par corpus). L'objectif de ces règles est de déterminer la polarité émotionnelle de chacun de ces marqueurs émotionnels en fonction de la présence éventuelle de marqueurs (modificateurs, négation, incertitude) dans leur contexte, pour un énoncé. À partir de ces polarités émotionnelles, le système calcule la polarité émotionnelle dominante de l'énoncé. Ce sont les informations dont nous allons nous servir pour évaluer les performances de notre système dans la détection de la polarité émotionnelle dominante d'un énoncé. Le système proposé est évalué de différentes manières :

- les données de référence correspondent aux annotations identiques produites par deux annotateurs, ce qui fournit trois jeux de référence : AE1 / ASE2, AE1 / ANE3, ASE2 / ANE3 ;
- les données de référence correspondent aux annotations de chaque annotateur, ce qui fournit également trois jeux de référence : AE1, ASE2, ANE3.

Les résultats tiennent compte des deux configurations, *baseline* et *processed*, auxquelles nous avons ajouté la réaccentuation et la considération de la dernière émotion mentionnée dans une phrase. Le tableau 3.1 présente les résultats de l'évaluation par rapport à différents annotateurs. Le tableau 3.2 présente les résultats d'évaluation par rapport aux annotations identiques de deux annotateurs.

TABLE 3.1 – Résultats de l'évaluation par rapport aux annotateurs

Annotateurs	Corpus	baseline	processed	processed + accents	processed + dernière émotion	processed + accent + dernière émotion
AE1	Q/A	39,92%	42,07%	41,64%	41,21%	40,78%
	Forum	41,08%	44,65%	41,08%	40%	40%
ASE2	Q/A	49,79%	55,80%	55,80%	54,94%	54,51%
	Forum	42,86%	42,86%	42,86%	38,19%	41,82%
ANE3	Q/A	47,22%	54%	53,22%	53,22%	51,94%
	Forum	42,86%	48,22%	51,79%	47,28%	50,91%

TABLE 3.2 – Résultats de l'évaluation par rapport aux annotations identiques de deux annotateurs donnés

Annotateurs	Corpus	baseline	processed	processed + accents	processed + dernière émotion	processed + accent + dernière émotion
AE1 / ASE2	Q/A	44,06%	46,54%	46,04%	46,04%	45,55%
	Forum	45,10%	41,18%	43,14%	40%	42%
AE1 / ANE3	Q/A	42,03%	45,75%	45,22%	45,22%	44,69%
	Forum	44%	50%	52%	48,98%	51,03%
ASE2 / ANE3	Q/A	46,54%	53,85%	53,37%	52,89%	51,93%
	Forum	41,82%	47,28%	50,91%	46,3%	50%

Globalement, les résultats montrent que les différents marqueurs (*modifieurs, négation, incertitude*) et connecteurs (*mais, cependant, car, donc, et*) permettent de mieux détecter la polarité émotionnelle dominante d'un énoncé (*processed*). De plus, la réaccentuation des termes améliore les résultats pour le corpus forum. Cependant, la prise en compte de la dernière émotion dans les phrases où plusieurs émotions sont exprimées ne s'avère efficace dans aucun des deux corpus. Nous détaillons maintenant les résultats de l'évaluation.

3.1.1 Évaluation par rapport aux annotateurs

Nous présentons les résultats d'évaluation pour les deux corpus traités :

- *Corpus Q/A*. Les meilleurs résultats sont obtenus en utilisant la configuration *processed*. La précision est alors de 55,88%, avec un gain de 2% à 6% sur la baseline. Les paramètres *accentuation* et *émotion finale* combinés aux *marqueurs* et aux *connecteurs* n'améliorent pas la précision. L'évaluation réalisée

par rapport à l'annotateur expert montre les moins bonnes performances du système. Ceci est dû en partie à l'incapacité du système à distinguer entre l'expression d'un état émotionnel et l'aspect factuel lié à un trouble de la santé :

- *des soucis de cholestérol* ; *soucis* a été annoté avec le symbole “/” (absence d'émotion) par l'annotateur expert. Mais ce terme est interprété par le système comme une émotion négative alors qu'il s'agit d'un état factuel de santé, et non pas d'un état émotionnel.

Nous pensons que les résultats plus élevés des annotateurs semi et non expert résultent d'une objectivité moins prononcée que chez l'annotateur expert. En effet, dans la situation discutée ici, ces deux annotateurs ont tendance à annoter une émotion (positive ou négative) plutôt que de constater un état de santé (absence d'émotion). De cette manière, leur interprétation des états émotionnels est plus proche de celle du système :

- *La pH - métrie n ' est pas un examen agressif ; n'...pas...agressif* a été annoté comme relevant d'une émotion positive par les annotateurs ASE2 et ANE3 alors que l'annotateur AE1 a annoté la phrase avec “/”, précisant l'absence d'émotion. Le système a annoté de la même manière que les annotateurs ASE2 et ANE3.
- *Corpus forum*. Les meilleurs résultats sont obtenus avec la configuration *processed* et la réaccentuation par rapport aux annotateurs ASE2 et ANE3. Le gain atteint jusqu'à 9% pour l'annotateur ANE3. Cependant, la réaccentuation fait baisser les résultats par rapport à l'annotateur AE1, dont le score maximal de 44,65% est obtenu avec la configuration *processed* (+3,5% par rapport à la *baseline*). L'évaluation avec les données de l'annotateur ANE3 fournit le meilleur résultat : 51,79%. Cependant, nous n'observons pas de différence pour l'annotateur ASE2 dans les configurations *baseline*, *processed* et *processed + réaccentuation*.

3.1.2 Évaluation par rapport aux annotations communes

- *Corpus Q/A*. Une fois encore, c'est la configuration qui utilise les marqueurs et les connecteurs qui propose les meilleurs résultats, avec un gain de 2% à 7% par rapport à la *baseline*. L'évaluation avec les annotations identiques ASE2/ANE3 donne le meilleur résultat avec 53,85% contre 46,54% pour la *baseline*. Les scores moins élevés avec l'annotateur expert AE1 peuvent s'expliquer par la différence d'expertise et le niveau d'objectivité moins prononcé chez les annotateurs ASE2 et ANE3. L'accentuation et la prise en compte de la dernière émotion n'améliorent pas la précision. Si la réaccentuation n'améliore pas les résultats, c'est dû au fait que la normalisation des accents est surtout importante dans les énoncés de type question (7,5%), et beaucoup moins dans les énoncés de type réponse (2,5%). En effet, même si le système parvient à mieux détecter la polarité émotionnelle globale d'un énoncé de type question, ce ne sera pas le cas des réponses, qui contiennent moins de fautes à l'origine. Or, un fil est validé lorsque le système reconnaît correctement la polarité émotionnelle globale de la question et de la réponse.
- *Corpus forum*. Les meilleurs résultats sont obtenus en utilisant la configuration *processed* combinée à la réaccentuation avec AE1/ANE3 (52%) et ASE2/ANE3 (50,91%). L'augmentation par rapport à la *baseline* varie alors entre 8% et 9% contre 2% à 3% avec la configuration *processed* seule. La prise en

compte de l'émotion finale ne permet pas d'améliorer les résultats de la configuration *processed*.

Les faibles scores d'évaluation peuvent s'expliquer également par la qualité des données annotées. L'annotation a été effectuée sur une version antérieure des données et des ressources, c'est-à-dire que certains marqueurs émotionnels que nous avons jugés non pertinents ont été retirés du lexique des émotions après la phase d'annotation manuelle. Ceci signifie que le système attribue aux marqueurs émotionnels non pertinents, une polarité émotive. Par exemple, le système considère les termes *rasés*, *emballé* (que nous avons retirés du lexique des émotions) comme s'ils évoquaient une émotion (négative et positive, respectivement). Or, ces deux termes ne doivent pas être considérés comme tels dans un contexte médical :

1. *Chère Elisa , Il est normal qu ' une pilosité se développe lorsque les poils sont rasés . ; le terme rasés s'applique ici à la pilosité et ne reflète pas un état émotionnel.*
2. *sachant que le poulet a été emballé à 13 h30... ; le terme emballé s'applique ici à l'emballage de produits (poulet) et ne reflète pas un état émotionnel.*

La prise en considération de ces termes par le système impacte négativement les résultats d'évaluation. En effet, alors que les annotateurs ont conscience que ces termes n'expriment aucune émotion, ils indiquent que ces phrases (où apparaissent ces termes) ne contiennent pas de marqueur émotionnel. Comme le système considère ces termes comme des émotions à polarité négative (*rasé*) et à polarité positive (*emballé*), lors de la comparaison de ses annotations avec celles des annotateurs, cela génère un désaccord et diminue les résultats. De plus, le corpus d'entraînement est un échantillon du corpus Q/A, ce qui peut expliquer les scores plus élevés pour ce corpus en mode de test. En effet, en utilisant ce corpus pour l'entraînement, le système est logiquement moins adapté au corpus forum. Une autre raison est que le sous-corpus de test forum contient une plus faible quantité de données : une erreur dans un corpus de taille plus petite peut avoir un impact négatif plus important sur les résultats d'évaluation que sur un corpus de grande taille.

Nous allons maintenant présenter et discuter des résultats de la mise en œuvre des différents mécanismes de visualisation des émotions et de leur évolution. Nous sélectionnons la configuration optimale qui fournit les meilleurs résultats pour chacun des deux corpus :

- Corpus Q/A : *processed* ; les questions représentent les patients, les réponses les professionnels ;
- Corpus forum : *processed + accentuation* ; les questions et les réponses représentent les patients (ceux qui recherchent l'information et ceux qui fournissent l'information, respectivement).

3.2 Distributions des polarités émotionnelles

3.2.1 Distribution polaire par couple question/réponse(s)

La dynamique des échanges entre patients/professionnels (corpus Q/A) et patients/patients (corpus forum) de santé évolue différemment en fonction des corpus. La figure 3.1 montre l'interaction des échanges entre patients/professionnels de santé et patients/patients en donnant un aperçu de la répartition des différents couples polaires (une polarité émotionnelle dominante pour la question et une pour la réponse qui lui correspond) :

d'émotions positifs ;

- Une autre particularité est liée à la raison pour laquelle les patients sortent sur les réseaux sociaux :
 - ils cherchent à être rassurés par les professionnels :
 - *merci de me rassurer*
 - *j'ai peur, pouvez-vous m'aider ?*
 - *dois-je m'inquiéter ?*
 - ils recherchent des réponses à leurs questions de santé :
 - *Pourquoi suis-je angoissée ?*
 - *Est-ce la bonne pommade ?*
 - ils décrivent leur état de santé, souvent corrélé à l'expression d'émotions négatives :
 - *j'ai eu des rapports non protégés et j'ai très peur d'être enceinte*
 - *j'ai des douleurs au niveau de la poitrine et je m'inquiète beaucoup*
 - ils veulent confirmer leurs hypothèses quant à leur état de santé :
 - *j'ai eu des rapports donc suis-je enceinte ?*
 - *j'ai une boule au niveau des ganglions, est-ce grave ?*
 - *je victime d'éruption cutanée, est-ce une allergie ?*

Pour les combinaisons polaires des questions et réponses, nous constatons que les couples les plus récurrentes montrent systématiquement des incertitudes dans la réponse des professionnels. Cela peut s'expliquer par :

- le manque de précisions de la part des patients et donc l'impossibilité pour les professionnels de santé de donner une réponse précise,
- la volonté des professionnels de santé d'exprimer une réserve car les diagnostics définitifs ne peuvent être faits que par les médecins en consultation, et non pas par l'intermédiaire des réseaux sociaux. Dans ce dernier cas, cela représenterait un risque trop élevé pour la santé du patient. C'est pourquoi les professionnels de santé prennent des précautions en employant des marqueurs d'incertitude :
 - *il est **possible** que que vous ayez...*
 - *il s'**agirait** de...*
 - *vous avez **probablement** une...*

De plus, l'étiquette "inc0", associée aux réponses des professionnels, décrit la certitude des médecins quant à une incertitude (*y a-t-il un risque ?*) ou à une inquiétude (*j'ai peur de...*) exprimée par un patient. Les professionnels vont alors rassurer ces patients par la négation des marqueurs d'incertitude. En effet, la certitude exprimée par cette combinaison a pour but, une fois de plus, de rassurer les patients.

3.2.1.2 Échanges patients/patients

Dans les échanges patients/patients (figure 3.1(b)), l'incertitude reste dominante. Plusieurs observations peuvent être faites :

- souvent, les réponses ne contiennent pas d'émotions ni d'incertitudes, et ceci indépendamment des polarités exprimées dans les questions ;
- les questions à polarité globale émotionnelle positive représentent la plus petite quantité de fils de discussion. Cependant, elles sont plus fréquentes dans le corpus forum par comparaison avec le cor-

pus Q/A. Nous pouvons supposer que les patients expriment plus facilement leurs états émotionnels avec d'autres patients, en utilisant plus massivement des marqueurs émotionnels du type *émoticônes* et *argot/abréviations*. Cette remarque s'applique également aux émotions neutres exprimées par la ponctuation (!!!, ? ? !);

- les réponses à tendance émotionnelle négative occupent la seconde position, indépendamment de la polarité émotionnelle observée dans les questions. Dans leurs réponses, les patients partagent souvent leurs propres expériences, vécues ou témoignées (*mon père en prend ; ça fait deux ans que j'en prends*) en rapport avec les troubles subis par les patients qui posent la question. Cela introduit une dimension émotive supplémentaire, que l'on ne perçoit pas dans le corpus Q/A : l'*empathie* et la *compassion*. Ici aussi, les patients cherchent à se faire rassurer grâce aux témoignages des autres patients (*si vous connaissez le sujet, merci de me rassurer*). Ce type d'échanges manifeste clairement une solidarité entre les patients (*je me ferais un plaisir de répondre à vos questions sur cette opération*), qui n'est pas observable dans les échanges patients/professionnels.

Le corpus Q/A est caractérisée par l'expression massive de marqueurs d'incertitude chez les professionnels de santé. Pour chacune des polarités émotionnelles (positive, neutre, négative et "incertitude"), les professionnels vont formuler une réponse empreinte d'incertitudes plutôt qu'exprimer des émotions. De cette manière, ils se protègent dans l'établissement de diagnostics ou dans leurs conseils tout en aidant les patients de façon efficace et ce, sans prendre de risque. Quant aux patients, ils favorisent également l'utilisation de marqueurs d'incertitude et montrent leurs inquiétudes et leurs doutes sur les maladies qu'ils affrontent ou sur celles qu'ils pensent avoir. Les patients espèrent alors être rassurés par les professionnels grâce aux connaissances techniques sur les maladies et les symptômes dont les patients sont victimes. Dans le corpus forum, les patients expriment majoritairement des émotions négatives mais aussi plus d'émotions positives que dans le corpus Q/A. Les patients vont tenter de sensibiliser leurs semblables par l'expression plus intensive d'émotions. Cette stratégie vise à toucher d'autres personnes (patients ou ex-patients) ayant déjà eu une expérience avec les troubles décrits par les patients malades. L'emploi plus fréquent d'émotions à polarité positive caractérise l'espoir des patients à obtenir des témoignages positifs qui leur permettraient alors de se soigner ou d'améliorer leur situation de malade (soulagement de douleurs par des médicaments plus efficaces, détecter des spécialistes avec une certaine renommée pour réduire les démarches inutiles, . . .). Les patients qui répondent expriment plus d'émotions à polarité négative dans une optique de partage d'expériences et d'élan de solidarité afin de soutenir les patients malades.

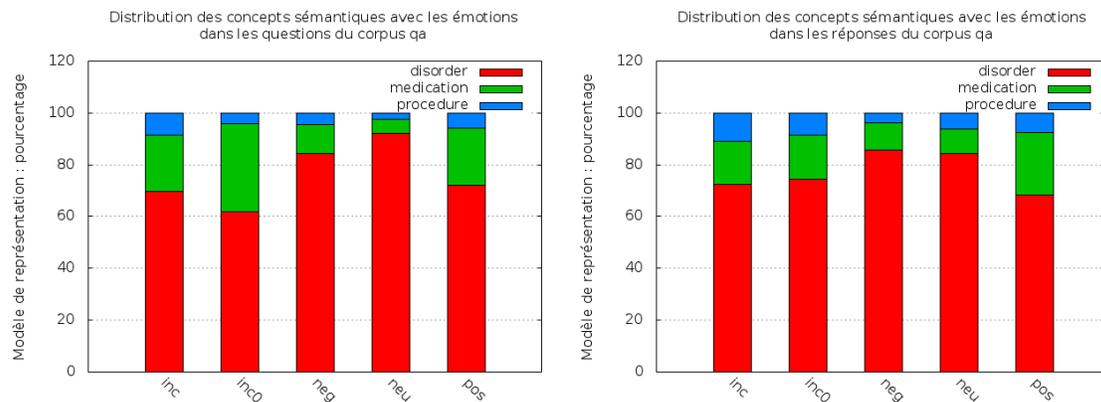
3.2.2 Distribution des types médicaux en fonction des polarités émotionnelles

La distribution des types médicaux en fonction des polarités émotionnelles permet d'observer la proportion de concepts médicaux (*procédures, maladies, médicaments/traitments*) employés dans des contextes émotionnels par polarité (*incertitude/inc, certitude/inc0, positive/pos, négative/neg, neutre/neu*). Nous discutons les résultats en comparant les questions aux réponses pour chaque corpus : QA (section 3.2.2.1) et forum (section 3.2.2.2). Les valeurs chiffrées utilisées dans les résultats présentés ci-après, sont exprimées en pourcentage. Dans l'annexe G.1, les valeurs chiffrées sont exprimées en nombres d'occurrences.

3.2.2.1 Corpus QA

Dans le corpus Q/A, les patients et les professionnels de santé utilisent globalement les mêmes types sémantiques médicaux, quelle que soit la polarité émotionnelle dominante exprimée. Les figures 3.2(a) et 3.2(b) présentent la proportion (en pourcentage sur l'axe des ordonnées) des concepts médicaux (rouge pour les maladies/troubles, vert pour les médicaments/traitements et bleu pour les procédures/diagnostics) employées dans les questions et les réponses du corpus Q/A en fonction de la polarité de l'énoncé (sur l'axe des abscisses avec pos pour positive, neg pour négative, neu pour neutre, inc pour incertitude et inc0 pour inhibition de l'incertitude). Les valeurs chiffrées exploitées dans ces graphiques sont exprimées en pourcentage.

FIGURE 3.2 – Distribution des types sémantiques médicaux dans les questions et dans les réponses pour le corpus Q/A

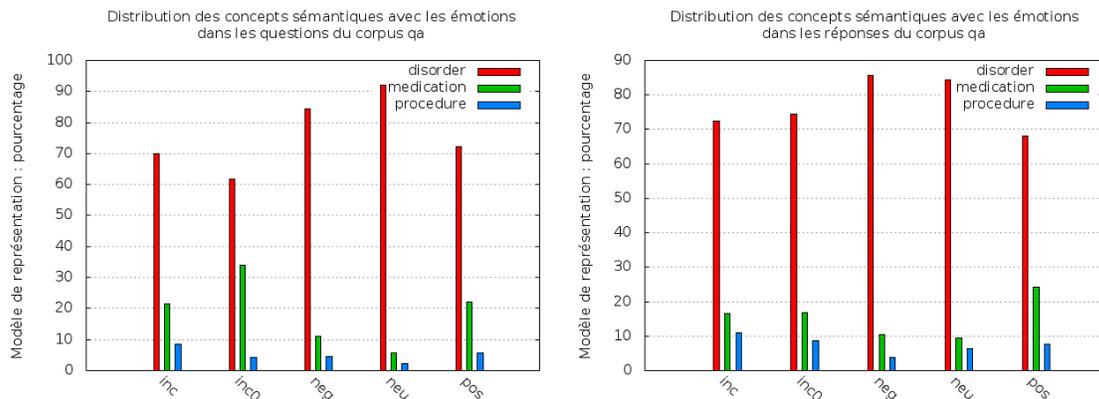


(a) Distribution des types sémantiques médicaux dans les questions (b) Distribution des types sémantiques médicaux dans les réponses

Les figures 3.3(a) et 3.3(b) présentent les mêmes données mais dans un affichage différent :

- le concept sémantique dominant (plus de 60% pour chaque polarité et pour chaque énoncé) concerne les troubles et les maladies dont les patients sont atteints. Quelle que soit la requête des patients (conseils, aide, solution), ils insistent systématiquement sur les maladies qu'ils ont ou qui leur font peur ;
- à l'inverse, les procédures représentent la plus faible proportion (2% à 11% toutes polarités confondues) de concepts sémantiques exprimés par les patients et par les professionnels. Cependant, les professionnels expriment davantage de procédures dans le but de conseiller au mieux les patients et afin de les rassurer :
 - vous devriez **consulter** votre médecin
 - votre cardiologue peut vous rassurer avec des **examens** complémentaires
 - si on vous a fait une **fibroscopie**, vous pouvez être soulagée
- les traitements et les médicaments sont plus nombreux que les procédures mais ne représentent que 5% à 21% des concepts par polarité, à l'exception des 33% pour la catégorie *inc0* dans les énoncés des

FIGURE 3.3 – Distribution des types sémantiques médicaux dans les questions et dans les réponses pour le corpus Q/A



(a) Distribution des types sémantiques médicaux dans les questions (b) Distribution des types sémantiques médicaux dans les réponses

patients. Chez les professionnels de santé, c’est dans les énoncés à polarité positive que la proportion des médicaments est la plus élevée : les professionnels proposent des solutions aux maux des patients afin de les aider et de leur apporter une solution à leurs problèmes ou inquiétudes ;

- les diagnostics représentent la plus forte concentration dans les énoncés à “polarité incertaine” : les professionnels de santé prennent leurs précautions lorsqu’il s’agit d’établir un diagnostic ou de conseiller une procédure à suivre. En effet, ils montrent leur volonté de venir en aide aux patients mais ne peuvent pas prendre le risque sans une consultation médicale.

3.2.2.2 Corpus forum

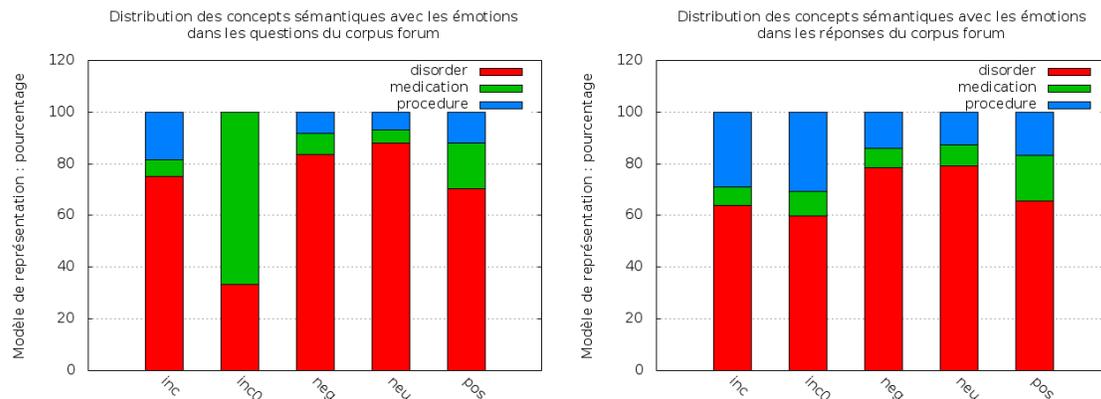
Les figures 3.4(a) et 3.4(b) présentent la proportion (en pourcentage sur l’axe des ordonnées) des concepts médicaux (rouge pour les maladies/troubles, vert pour les médicaments/traitements et bleu pour les procédures/diagnostics) employées dans les questions et les réponses du corpus forum en fonction de la polarité de l’énoncé (sur l’axe des abscisses avec pos pour positive, neg pour négative, neu pour neutre, inc pour incertitude et inc0 pour inhibition de l’incertitude). Les figures 3.5(a) et 3.5(b), pages 78 et 78, respectivement, présentent les mêmes données mais dans un affichage différent

Dans les échanges patients/patients, nous pouvons observer plusieurs tendances :

- les troubles et les maladies représentent toujours la majeure partie (60% à plus de 80%, hors catégories “inc0” et “neu”) de concepts sémantiques exprimés dans les contextes émotionnels ;
- à l’inverse des échanges patients/professionnels, les patients qui répondent utilisent plus de termes liés aux procédures (*scanner, radio, opération*) : entre 12% et 30% contre 21% pour les réponses des professionnels. Cela transparaît souvent lorsque les patients posent des questions au sujet de leurs incertitudes et leurs décisions :

- *j’aimerais avoir votre avis sur votre opération*

FIGURE 3.4 – Distribution des types sémantiques médicaux dans les questions et dans les réponses pour le *corpus forum*



(a) Distribution des types sémantiques médicaux dans les questions (b) Distribution des types sémantiques médicaux dans les réponses

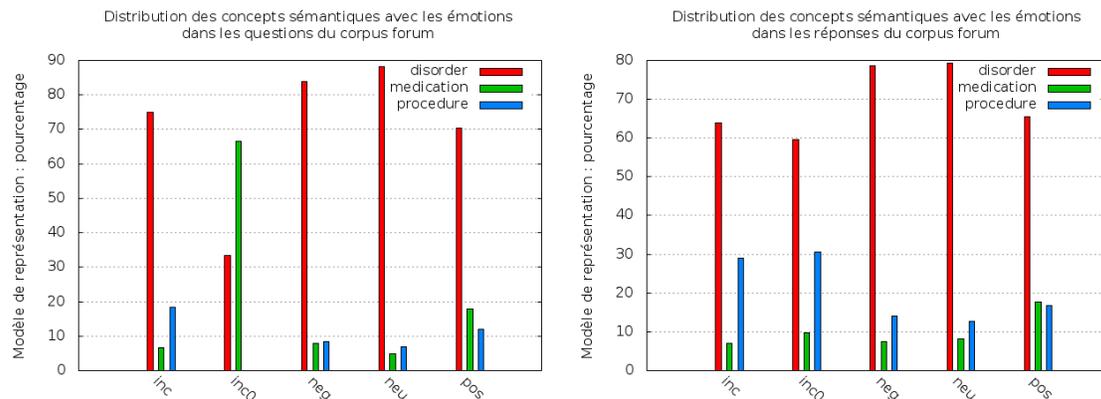
- *j’aimerais avoir quelques témoignages sur l’opération*
- et lorsqu’on les informe et rassure avec des expériences déjà vécues par les patients qui répondent :
 - *ne pas avoir peur si on te propose l’opération*
 - *pour l’opération fonce tu seras drolement soulagé*
- les patients qui répondent emploient moins fréquemment les termes liés aux traitements, contrairement aux professionnels dans le corpus QA. Cependant, dans les énoncés à polarité positive, les termes relatifs aux traitements sont employés plus souvent que les termes de procédures :
 - *l’ixel me fait du bien*
 - *je prends du **IXPRIM** et cela me soulage*
- les procédures occupent une place importante dans les réponses sans tenir compte de la “polarité inc0”. Les patients mentionnent en effet différentes procédures à suivre pour aiguiller les patients indécis ou souffrant de maladies :
 - *tu dois reprendre rendez vous et si tu peux obtenir un contrôle **radio***
 - *tu dois refaire une prise de sang et si c’est positif refais une **IRM***

3.2.2.3 Bilan

Les échanges patients/professionnels divergent des échanges patients/patients de plusieurs points de vue :

- la répartition des informations médicales est spécifique à chaque type d’échanges. Dans les échanges patients/professionnels, les patients comme les professionnels utilisent plus souvent des termes de types *médicaments* que ceux de type *procédures*. Nous pouvons en déduire que les patients se renseignent sur l’usage de médicaments/traitements qui pourraient les soulager. Pour valider la prise d’un médicament par exemple, les patients ont besoin d’être conseillés par des experts qui ont les connaissances techniques nécessaires pour les aider. Or ce n’est pas le cas dans les échanges de types pa-

FIGURE 3.5 – Distribution des types sémantiques médicaux dans les questions et dans les réponses pour le *corpus forum*



(a) Distribution des types sémantiques médicaux dans les questions (b) Distribution des types sémantiques médicaux dans les réponses

tients/patients où ce sont les procédures qui reviennent plus souvent que les médicaments. En effet, dans ce type d'échange, les procédures se partagent et se valident par l'expérience de celles-ci par des patients vivant ou ayant vécu avec telle ou telle maladie. Dans cette optique, ces patients sont en mesure d'indiquer les procédures plus ou moins efficaces en fonction de la maladie d'un patient ;

- la fiabilité des informations fournies est plus élevée avec les professionnels de santé car il s'agit de connaissances et expériences professionnelles, alors que les informations proposées par les patients concernent surtout des expériences vécues ;
- en même temps, les professionnels expriment souvent une réserve plus prononcée que les patients. Ainsi, dans une grande majorité de cas les patients parlent de problèmes médicaux et des soins déjà vécus, et n'hésitent pas à généraliser leur expérience personnelle ; alors que les professionnels de santé sont très prudents et réservés dans leurs réponses ;
- les objectifs de ces échanges sont différents aussi : pour les professionnels de santé il s'agit tout d'abord de conseiller les patients tout en prenant des précautions, alors que pour les patients il s'agit de partager leurs expériences personnelles. Dans les deux cas, l'objectif est de rassurer et d'aider les personnes qui posent des questions.

3.2.3 Distribution polaire en fonction du corpus

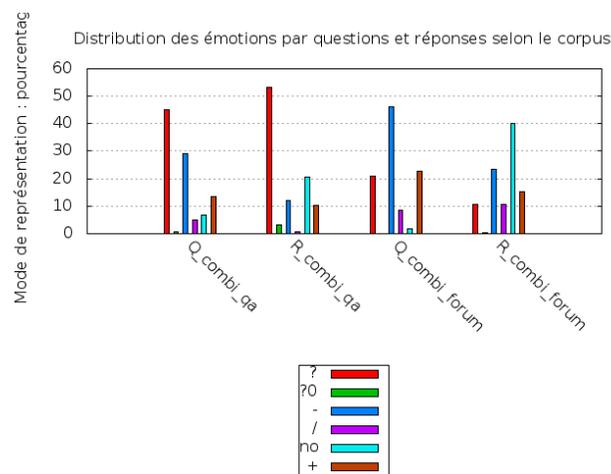
Cette visualisation permet d'observer la distribution polaire parmi les différents énoncés (questions et réponses) selon les corpus. De manière générale, les interactions entre les patients et les professionnels de santé sont très marquées par l'incertitude : environ 40% pour les questions et 60% pour les réponses. Dans les échanges entre patients, les incertitudes sont moins présentes : 20% pour les questions et environ 10% pour les réponses. Nous observons aussi une évolution différente de la manifestation de l'incertitude dans ces deux corpus : QA (section 3.2.3.1) et forum (section 3.2.3.2). Les valeurs chiffrées utilisées dans les résultats

présentés ci-après, sont exprimées en pourcentage. Dans l'annexe G.2, les valeurs chiffrées sont exprimées en nombres d'occurrences.

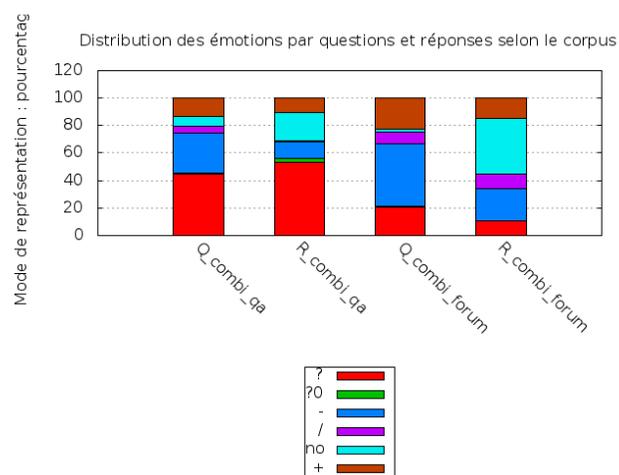
3.2.3.1 Corpus QA

Les figures 3.6(a) page 79 et 3.6(b) page 79 représentent la répartition des polarités émotionnelles dominantes par énoncé et par corpus (les deux figures représentent les mêmes informations mais de façon différente). Cette visualisation permet de voir l'évolution des émotions dans les échanges entre patients/professionnels (corpus Q/A) et patients/patients (corpus forum).

FIGURE 3.6 – Distribution des émotions dans les corpus *Q/A* et *forum* par polarité émotionnelle (? = incertitude, + = positif, - = négatif, / = neutre, ?0 = certitude, no = ni émotion/ni incertitude) et par énoncé (Q_combi_qa = questions du corpus Q/A, R_combi_qa = réponses du corpus Q/A, Q_combi_forum = questions du corpus forum, R_combi_forum = réponses du corpus forum)



(a) Distribution des polarités émotionnelles dans les questions et les réponses pour les corpus Q/A et forum



(b) Distribution des polarités émotionnelles dans les questions et les réponses pour les corpus Q/A et forum

Nous pouvons faire plusieurs observations qui ressortent avec cette visualisation. Globalement, le corpus Q/A laisse peu apparaître les émotions. À notre avis, cela est dû à plusieurs raisons :

- les professionnels privilégient l’emploi de marqueurs d’incertitude (plus de 50% des énoncés). Comme déjà noté auparavant, les médecins essaient au mieux de rassurer les patients mais sans prendre de risque : ils ne peuvent pas établir de diagnostic, en revanche ils peuvent guider les patients le plus précisément possible :

- *en fonction des résultats, on peut aussi proposer une biopsie du foie*
- *compte tenu de votre OPK, je **crois** qu’une échographie sera utile*

- les professionnels évitent de transmettre leurs propres émotions. En effet, environ 20% des réponses ne comportent aucun marqueur sémantique, alors que la majorité des réponses formulées par les professionnels gardent la polarité exprimée par les patients :

- *[c’est] probablement dû à votre **inquiétude***
- *cette **angoisse** peut faire partie d’une dépression*

Il apparaît ainsi que les professionnels cherchent à proposer des solutions adaptées aux maux des patients en les réconfortant (avec 10% des réponses) :

- ***tranquillisez**-vous*
- *soyez **rassuré***

Il nous semble aussi que les émotions à polarité positive exprimées par les patients (plus de 10%) reflètent le désir d’être rassuré :

- *pouvez-vous me **rassurer***
- *besoin de **réconfort***

alors que les émotions négatives (presque 30% des questions) reflètent le mal-être des patients et sont liées à leurs problèmes de santé :

- *je suis **stressé***
- *je suis très **inquiète***

Nous avons également observé un biais potentiel dans l’attribution des polarités émotionnelles dominantes : le lemme *aimer*, exprimant une émotion positive, est annoté correctement, sauf que toutes ses flexions grammaticales y compris au conditionnel (*j’**aimerais** avoir plus d’informations merci ; j’**aimerais** savoir si...*) sont également annotées comme émotion positive. La bonne étiquette est celle de l’incertitude dans ce dernier cas. Le biais est renforcé par la fréquence élevée de cette entité verbale puisque que le système considère alors cette forme conditionnelle comme révélant une émotion positive.

3.2.3.2 Corpus forum

Le corpus forum présente quelques particularités :

- les réponses contiennent moins d’incertitude que les questions, 20% contre 10% : les patients qui répondent partagent leurs propres expériences sans les nuancer d’incertitude. Les réponses sont plutôt marquées par des indices émotionnels positifs (plus de 15% d’énoncés soit 5% de plus que les réponses des professionnels) :

- *on peut éventuellement **soulager** une fibrose grâce à certaines tractions vertebrales*

- *pour ma part, je suis totalement **satisfait** de la pose des deux prothèses neutres (10 % des énoncés contre environ 2 % pour les professionnels) :*
 - *fais toi prescrire du tretrazepam !!*
 - *passer une radio et on vous dira !!!!!*
- et négatifs (plus de 20% des réponses contre un peu plus de 10% dans les énoncés des professionnels) :
 - *il a décidé après 5 mois de **souffrance** de me faire passer radio, scanner..*
 - *il avait un patient qui était en **pire** état depuis son opération*
- il y a une forte proportion de réponses sans les émotions et les incertitudes, parfois dû à l'absence de réponses dans les fils de discussion.
- la prédominance d'énoncés à polarité émotionnelle négative (environ 45%) peut s'expliquer par une réticence moins prononcée de la part des patients à se livrer à d'autres patients. Cette proximité sociale permet aux patients malades d'exprimer plus facilement leurs émotions afin de les transmettre à d'autres patients participant à la discussion. Cette "stratégie" infère d'habitude une émotion d'empathie envers le patient malade. Notons aussi que souvent, les émotions d'origine sont transmises, ressenties et transformées par les patients qui répondent, tout en donnant naissance à de nouvelles émotions. Cette évolution émotionnelle permet de détecter le niveau de résolution de problèmes liés à une maladie : si la polarité globale émotionnelle de la question est négative et si celle de la réponse est positive, alors il est probable qu'il existe une solution aux problèmes de santé. De ce point de vue, les échanges entre patients deviennent une mine d'informations pour les professionnels de santé.

3.2.3.3 Bilan

Cette première expérience montre les différentes tendances dans l'évolution des émotions entre les deux types d'échanges :

- dans les échanges patients/professionnels, les patients cherchent à infirmer ou confirmer leurs doutes et incertitudes. Les professionnels de santé établissent alors un diagnostic en émettant une certaine réserve quant à sa validité afin d'éviter tout risque d'erreur et éviter de graves conséquences. Ce type d'échange favorise l'utilisation de marqueurs d'incertitude et modère l'emploi des marqueurs émotionnels ;
- dans les échanges patients/patients, les patients atteints de troubles ont tendance à vouloir trouver des solutions à leurs problèmes médicaux (*quelle procédure entreprendre ? , quels traitements efficaces ?*). Ces problèmes peuvent trouver une solution ou au moins un soutien à travers le vécu d'autres patients ou "ex-"patients. Les échanges de ce type montrent un emploi fréquent des émotions.

3.3 Distribution des marqueurs par polarité émotionnelle (graphes non orientés)

Cette modélisation utilise la configuration *baseline* (sans traitement des marqueurs). L'objectif de cette représentation est de proposer une interprétation de l'utilisation des marqueurs de type *incertitude* (*forte et faible*), *modificateurs* (*atténuation, amplification*) et de la *négation*. Nous avons généré un graphe par polarité

émotionnelle, reliée aux différents types de marqueurs présents dans leurs contextes gauche et droit. Pour chaque type de marqueur et chaque polarité, nous avons sélectionné les plus fortes valeurs de ratio. Elles indiquent la significativité d'un marqueur en fonction du nombre d'occurrences de ce marqueur par rapport au nombre d'occurrences d'émotions appartenant à une polarité. Nous présentons quelques résultats pour chaque corpus : QA (section 3.3.1) et forum (section 3.3.2).

3.3.1 Corpus QA

Les figures 3.7(a) et 3.7(b) page 108 représentent la distribution des marqueurs dans les énoncés de type question, et pour les émotions de polarité positive et neutre.

Nous présentons un nombre restreint de figures afin de ne pas surcharger le contenu et de donner un aperçu des différentes polarités : neutre et positive pour ce corpus Q/A, et négative et incertitude pour le corpus forum. C'est pourquoi les graphes correspondants aux polarités négative et "incertitude" et aux réponses, du corpus Q/A, sont disponibles dans les annexes G.3.1. La première figure représente les marqueurs (négation, modificateurs forts et faibles, incertitude) employés dans le contexte des marqueurs émotionnels neutres. Leur proportion est indiquée dans le nœud central ainsi que le nombre d'occurrences de marqueurs émotionnels correspondant à cette polarité. Les nœuds périphériques donnent une information supplémentaire sur le nombre d'occurrences des différents types de marqueurs avec cette polarité (ici, les marqueurs présents dans les contextes des émotions à polarité neutre) : le nombre entre parenthèses. Ces nœuds précisent également dans quel contexte (gauche ou droit) les marqueurs apparaissent (nous n'exploiterons pas cette information car peu significative pour cette étude). La seconde figure fournit les mêmes informations mais pour les émotions à polarité positives. Nous décrivons plus bas les résultats pertinents que nous avons observés avec l'ensemble des graphes (8 graphes au total : 4 par énoncé pour ce corpus). Deux tendances bien distinctes se profilent : celle typique aux patients (section 3.3.1.1) et celle typique aux professionnels de santé (section 3.3.1.2).

3.3.1.1 Patients

Avec les émotions positives, les patients ont tendance à employer des modificateurs d'*atténuation* (19,72% des modificateurs d'*atténuation* contre 36,62% employés avec des émotions négatives) :

- [traitement] à base de Maléate de trimébutine qui me **soulage un peu**
- je suis **moins optimiste**

Ce type de marqueur montre que les patients connaissent des solutions à leurs troubles ou maladies mais que ces solutions manquent d'efficacité. Les patients vont alors solliciter les professionnels de santé afin de trouver d'autres solutions, potentiellement plus efficaces.

Avec les émotions négatives, les patients utilisent plutôt des modificateurs d'*amplification* (avec 49,38% des modificateurs d'*amplification*) :

- car cela est **très inconfortable**
- sa me **gêne vraiment**

L'emploi de ces modificateurs par les patients montre une certaine insistance sur leur trouble/maladie afin d'inciter les professionnels à leur fournir des informations et des solutions rapidement.

Les marqueurs de *négation* sont caractéristiques des émotions à polarité neutre (9,14% des *négations* contre 46,35% employées avec des émotions d’“incertitude”) :

- *les anti inflammatoires me font plus rien !!!*
- *mon copain n’éprouve plus aucun plaisir*
- *en position allongé je ne ressens plus la douleur*

La négation des émotions neutres par les patients indique qu’ils ne comprennent pas les phénomènes de santé, qu’ils se trouvent dans une situation inconnue et surtout qu’ils ont besoin d’obtenir des explications sur cette situation.

Sans surprise, les incertitudes fortes sont employées de façon plus significative dans les questions. En effet, cela montre que les patients se posent de nombreuses questions, ce qui marque le doute et les incertitudes qu’ils ressentent face à l’absence de connaissances techniques dans le domaine médical.

Nous avons observé un biais dans les données concernant les émotions positives. Le verbe *aimer* est annoté comme une émotion positive, y compris toutes ses formes flexionnelles. Mais la forme la plus fréquente est celle du conditionnel, qui correspond au marqueur de l’incertitude. Il est donc possible que le ratio des incertitudes et des émotions positives soit incorrect.

3.3.1.2 Professionnels de santé

Chez les professionnels de santé, les émotions neutres sont souvent employées avec deux types de modificateurs (2,06% de modificateurs d’*amplification*, 1,78% de modificateurs d’*atténuation* contre 24,63% et 18,35% employés avec les émotions négatives, respectivement) :

- *je suis un peu étonné que votre médecin ait fait le diagnostic*
- *je suis très surpris d’apprendre que vous avez eu...*

À notre avis, cette combinaison de marqueurs avec des émotions neutres, montre l’incompréhension, chez les professionnels, des actions menées par les patients. Le ratio des modificateurs d’*amplification* et d’*atténuation* pour les émotions neutres est de 1,42 et de 0,39 contre 1 et 0,24 pour les émotions négatives. Cela s’explique par le faible nombre de marqueurs émotionnels neutres, dont les contextes sont marqués par une fréquence élevée de modificateurs. En effet, les émotions négatives étant plus nombreuses, elles concentrent plus de modificateurs en terme d’occurrences. En revanche, en terme de fréquence il y a moins de modificateurs employés avec des émotions négatives par rapport à leur quantité.

Selon le même raisonnement, les professionnels utilisent plus fréquemment des *négations* avec des émotions à polarité neutre (1,80% des *négations* contre 17,96% pour les émotions négatives et un ratio respectif de 7,96 et 4,69) :

- *il n’est pas étonnant que la douleur soit retardée*
- *il n’y a rien de surprenant dans ce que vous décrivez*

Cette structure utilisée par les professionnels permet de rassurer les patients quant à leurs inquiétudes sur la maladie. De cette manière, les professionnels signifient aux patients l’absence de gravité de leurs problèmes médicaux.

Les émotions négatives sont liées aux marqueurs d’incertitude faible (29,07% des marqueurs contre 58,04% dans les contextes d’émotions d’“incertitude” et un ratio de 2,15 et de 0,88, respectivement) :

- *si vous êtes inquiète, vous pouvez faire un test de grossesse*
- *si vous continuez à être gêné, vous pouvez revoir votre médecin*
- *votre récit évoque plutôt une dépression*

Comme nous l’avons déjà noté, les professionnels proposent des diagnostics possibles en proposant avec précaution des procédures à suivre ou l’origine du trouble ou de la maladie.

3.3.2 Corpus forum

Les figures 3.8(a) et 3.8(b) page 109 représentent la distribution des marqueurs dans les énoncés de type réponse du corpus forum, pour les émotions de polarité “incertitude” et négative, respectivement. Nous présentons un nombre restreint de figures afin de ne pas surcharger le contenu et de donner un aperçu des différentes polarités (neutre et positive pour le corpus Q/A et négative et incertitude pour ce corpus forum). C’est pourquoi les graphes correspondant aux polarités neutre et positive et aux questions, du corpus forum, sont disponibles dans les annexes G.3.2. Nous décrivons plus bas les résultats pertinents que nous avons observés avec l’ensemble des graphes (8 graphes au total : 4 par énoncé et pour ce corpus). Ici aussi, deux tendances bien distinctes se profilent : celles typiques aux patients qui posent des questions (section 3.3.2.1) et celles typiques aux patients qui répondent aux questions (section 3.3.2.2).

De manière générale, nous avons observé que les marqueurs d’incertitude sont caractéristiques des émotions d’“incertitude”, ce qui s’avère une combinaison assez trivial. Nous préférons discuter les résultats autour des catégories polaires : émotions positives et négatives.

3.3.2.1 Patients qui posent des questions

Les patients ponctuent le contenu des questions avec des marqueurs de *négation* de façon très fréquente (30,58% des *négations* contre 43,8% pour les émotions négatives et un ratio de 3,3 contre 2,83, respectivement) :

- *je n’ ai pas trop envie d’aller voir un médecin*
- *je n’aime pas les médicaments*
- *les antalgiques et myorelaxants ne me soulage pas*
- *les kinés n’ ont rien soulagé*

Les patients décrivent alors leur état de santé à d’autres patients en se confiant émotionnellement et avec plus de détails (et donc plus d’expression d’états émotionnels) que les patients du corpus QA. En effet, dans les échanges patients/patients, les patients malades se livrent plus facilement en donnant un maximum d’informations sur leur mal. Ils indiquent aussi des solutions testées (médicaments, traitements, procédures) qui manquent d’efficacité. À notre avis, avec cette combinaison “marqueurs de *négation* + émotions positives”, les patients cherchent de nouvelles solutions médicales dans l’espoir de trouver le remède qui guérirait leur maladie.

Les patients expriment aussi une quantité importante d’émotions négatives couplées avec des modificateurs d’*amplification* (57,62% de marqueurs dans tous les contextes émotionnels confondus) :

- *j’ai aussi très peur que ce soit un problème très grave*
- *je suis très angoissée, voire dépressive*

- *ce problème de dos me **préoccupe beaucoup***

mais aussi des émotions négatives couplées avec des modificateurs d'*atténuation* (54,84% de marqueur dans tous les contextes émotionnels confondus) :

- *j'ai un **peu peur** car j'ai des spasmes...*
- *je suis un **peu en panique**, j'ai peur de finir paralysé*
- *je suis un **peu en stress***

En ayant recours aux modificateurs d'amplification, les patients essaient de communiquer le plus minutieusement et émotivement des informations sur les difficultés de vivre avec leur maladie. Ceci est fait dans l'espoir de toucher d'autres personnes ayant connaissance des troubles et qui pourraient proposer leur aide (connaissances médicales, expériences de la maladie). Nous pensons que l'utilisation de modificateurs peut ajouter une dimension émotionnelle dont l'objectif est de toucher les lecteurs. Ces vecteurs d'intensité peuvent être indicateurs du niveau de gravité physique ou psychologique du patient malade.

3.3.2.2 Patients qui répondent aux questions

Les patients qui répondent aux questions partagent leur propre expérience personnelle soit par compassion, afin de soutenir le patient malade, soit en s'appropriant eux-mêmes la situation du patient qui pose la question. Ces échanges sont caractérisés par l'expression conjointe d'émotions négatives et de modificateurs d'*amplification* (40,91% de la totalité de ces modificateurs employés dans des contextes émotifs) :

- *J'**appréhende beaucoup** le rhumato qui me suit*
- *courage, tu sais **beaucoup** de gens **souffrent** pour des maux différents*
- *plus tu as mal, **plus** tu **souffres**, plus tu te crispes*

Ces patients ont également recours assez fréquent aux marqueurs de *négation* dans le but d'apporter du réconfort et de rassurer les patients souffrant (39,66% de la totalité des marqueurs de *négation* utilisés dans les contextes émotionnels) :

- *il **ne regrette pas** son opération car il est soulagé*
- *bon courage, il **ne faut pas désespérer***
- *ne **stresse pas** et attends ton rendez-vous*

En ce qui concerne les modificateurs d'*atténuation*, ils sont les plus fréquents dans les contextes d'émotions positives (30,98% des modificateurs de ce type employés dans les contextes émotionnels, contre 36,04% pour la *négation* et un ratio de 0,55 et 0,49, respectivement) :

- *cela me **rassure un peu** de te lire*
- *seuls les antiinflammatoires me **soulage un peu***
- *si ça peut te **réconforter** un petit peu*

Nous pensons que les patients qui s'expriment dans les réponses emploient les marqueurs d'*atténuation* avec des émotions positives afin de partager avec les autres patients leurs solutions provisoires ou leurs frustrations :

- *2 séances par semaine et cela me **soulage un peu** mais... ;*
- *me voilà un **peu moins optimiste** !!! ;*
- *mais cela **calme** juste un **peu** et pas longtemps.*

3.3.2.3 Bilan

Cette visualisation permet de mettre en évidence l'impact des différents marqueurs sur les émotions exprimées par les patients et les professionnels de santé.

Dans le corpus Q/A, les patients ont tendance à modifier l'intensité d'émotions, en employant des modificateurs d'amplification avec des émotions négatives et des modificateurs d'atténuation avec des émotions positives : dans les deux cas, la tendance est d'accentuer l'aspect négatif de leur état. En faisant cela, les patients expriment souvent l'inefficacité des traitements liés à leurs troubles de la santé (atténuation des émotions positives) et attirent l'attention des professionnels sur la gravité de leurs symptômes ou par l'exagération de la description de leurs maux (amplification des émotions négatives) afin d'obtenir des réponses rapides et pertinentes. Les professionnels utilisent une stratégie différente dans la formulation de leurs réponses. Ils associent les modificateurs et la négation avec l'expression d'émotions neutres. En effet, les professionnels ont recours à ce type de combinaisons sémantiques afin d'exprimer avec certaines précautions (caractérisation des émotions neutres contrairement aux émotions positives ou négatives) leur point de vue sur les descriptions que font les patients de leurs troubles, leurs incertitudes ou leurs doutes.

Dans le corpus forum, la caractérisation des émotions positives par les marqueurs de négation et la caractérisation des émotions négatives par les modificateurs (forts et faibles), marquent, à notre avis, la situation difficile des patients atteints de troubles. Les patients cherchent ainsi du réconfort et de la compassion en sollicitant d'autres personnes à partager leur vécu dans le but de trouver des solutions qui pourraient les aider. C'est pourquoi les réponses sont fortement marquées par l'utilisation de différents marqueurs avec les émotions négatives et positives. Ceci montre que les patients qui répondent soutiennent les patients malades en les rassurant ou en leur décrivant leurs propres expériences de la maladie en question.

3.4 Les diagrammes circulaires : distribution des informations sémantiques en fonction des polarités et des catégories d'émotions

Cette modélisation permet de visualiser la proportion des différentes polarités émotionnelles (*positive*, *négative*, *neutre*, "*incertitude*") d'après les émotions exprimées par les patients et les professionnels de santé dans les deux types de corpus. De la même façon, il est possible de visualiser la distribution par catégorie émotionnelle (*joie*, *attirance*, *peur*, *inquiétude*, *étonnement*, *surprise*, ...).

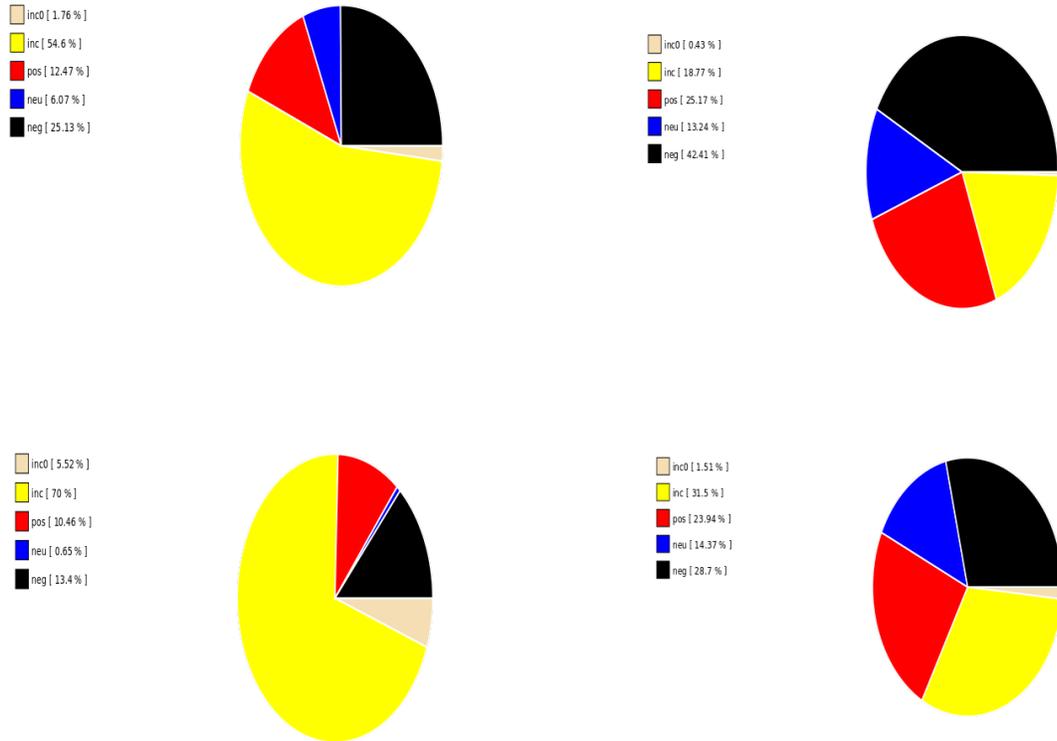
3.4.1 Répartition des polarités émotionnelles

La figure 3.9 montre les quatre diagrammes circulaires représentant la proportion des différentes polarités émotionnelles relatives aux questions et aux réponses des corpus Q/A (figure 3.9(a)) et forum (figure 3.9(b)). La légende correspondante aux polarités :

- *inc0* : inhibition de l'incertitude (ou certitude) – couleur beige
- *inc* : incertitude – couleur jaune
- *pos* : positive – couleur rouge
- *neu* : neutre – couleur bleu

– *neg* : négative – couleur noir

FIGURE 3.9 – Distribution des polarités émotionnelles par corpus et par énoncé



(a) Distribution des polarités émotionnelles dans le corpus Q/A (questions en haut et réponses en bas)

(b) Distribution des polarités émotionnelles dans le corpus forum (questions en haut et réponses en bas)

3.4.1.1 Corpus QA

Dans le corpus Q/A, nous voyons que l’incertitude (en jaune) est dominante pour les questions (54,6%) et pour les réponses (70%). Ces résultats confirment ce que nous avons observé auparavant : les messages des professionnels sont caractérisés par un fort emploi de marqueurs d’incertitude. En ce qui concerne les patients, plus de la moitié des questions sont marquées par l’expression d’incertitudes et de doutes. Il y a très peu d’émotions neutres exprimées par les patients (6,07%) ainsi que par les professionnels (0,65%). Ceci indique que les patients ont tendance à exprimer des émotions plus fortes (positives et négatives) afin d’attirer l’attention des médecins. En effet, si les patients manifestent une émotion positive ou négative, cela peut attirer l’attention des médecins plus facilement sur les troubles décrits par les malades. De manière générale, nous observons qu’il y a environ deux fois plus d’émotions négatives (25,13%) que positives (12,47%) : la description négative de l’état physique ou psychologique des patients est en effet en rapport avec les troubles qu’ils subissent. De plus, dans l’espoir d’attirer l’attention des professionnels et d’obtenir une solution à

leurs problèmes plus rapidement, les patients insistent sur leurs soucis en utilisant des émotions négatives. La proportion de polarité “inc0” est insignifiante (1,76%) chez les patients. Cependant, chez les professionnels de santé, cette inhibition de l’incertitude représente une portion un peu plus importante (5,52%). Dans les messages des professionnels, nous voyons que l’expression d’émotions positives (10,45%) et négatives (13,40%) est moins importantes que dans les questions posées par les patients. Nous retrouvons le schéma utilisé par les médecins, consistant à prendre des précautions et d’utiliser beaucoup d’incertitude dans leurs propos. Quant aux incertitudes inhibées, elles sont aussi plus fréquentes chez les professionnels (5,52%). Cette “polarité” est souvent utilisée pour lever les doutes et inquiétudes que les patients montrent. En effet, en employant des marqueurs d’incertitude combinés à des marqueurs de négation, les médecins manifestent alors leur certitude sur l’absence de troubles et essaient de rassurer les patients :

- *ce que vous décrivez ne m’évoque pas de maladie particulière*
- *il n’y a aucun risque de contracter une maladie*

3.4.1.2 Corpus forum

Nous ne traiterons pas la “polarité inc0”, car son impact est trop insignifiant dans le corpus forum. Dans ce corpus, nous observons que l’expression des émotions est fréquente dans les questions et dans les réponses, et dépasse ce que nous avons pu observer dans le corpus QA.

Les patients qui posent leur question emploient surtout des émotions à polarité négative (42,41 %). Nous pensons que, de cette manière, les patients veulent attirer l’attention et ainsi recevoir des conseils ou témoignages. De plus, lorsque ces patients décrivent des solutions qui les soulagent de façon éphémère (avec 25,17 % d’émotions positives), l’utilisation combinée des émotions à polarité positive permet éventuellement d’obtenir des conseils sur des solutions nouvelles capables de les soulager plus efficacement. Les patients ont aussi recours aux émotions neutres (13,24 %) lorsqu’ils décrivent le ressenti de leur maladie avec plus ou moins d’intensité (ponctuations expressives). Enfin, les incertitudes ne représentent que 18,77 % des “polarités émotionnelles” exprimées dans les questions. Les patients préfèrent en effet ponctuer leurs descriptions par des marqueurs émotionnels plutôt que par des marqueurs d’incertitude.

Les réponses sont plus marquées par l’utilisation de marqueurs d’incertitude que les questions. Dans les réponses, les patients profitent de cette dynamique communicationnelle afin d’obtenir des réponses à leurs interrogations. Les patients qui répondent emploient approximativement la même proportion d’émotions positives et neutres (23,94 % et 14,37 %, respectivement). Ces émotions caractérisent l’échange d’informations entre patients :

- les émotions positives apparaissent avec des solutions qui fonctionnent,
- les émotions neutres ponctuent le partage d’expériences.

Nous voyons que les émotions négatives sont moins nombreuses que dans les questions (28,70 %), ce qui peut s’expliquer par la nature des énoncés. En effet, dans les questions, les patients insistent sur l’aspect négatif de leurs troubles, en exprimant leurs souffrances, leurs peurs, . . . Alors que dans les réponses, il s’agit plus de descriptions d’expériences vécues par d’autres patients ou encore de descriptions explicatives succinctes.

3.4.1.3 Bilan

Nous avons observé quelques différences caractérisant les deux corpus. Dans les échanges patients/professionnels, ce sont les marqueurs d'incertitude qui reviennent le plus fréquemment. Ils définissent les doutes mêlés d'inquiétudes exprimés par les patients, et ils servent aussi aux professionnels à proposer un avis médical ou à renseigner les patients tout en prenant des précautions. Une autre différence marquante est l'utilisation plus importante de marqueurs émotionnels dans les échanges patients/patients. Sachant que leurs questions sont destinées à d'autres patients, les patients malades expriment plus facilement leurs émotions souvent négatives, espérant ainsi solliciter des retours d'expériences de personnes ayant vécu une situation similaire.

3.4.2 Répartition des catégories émotionnelles

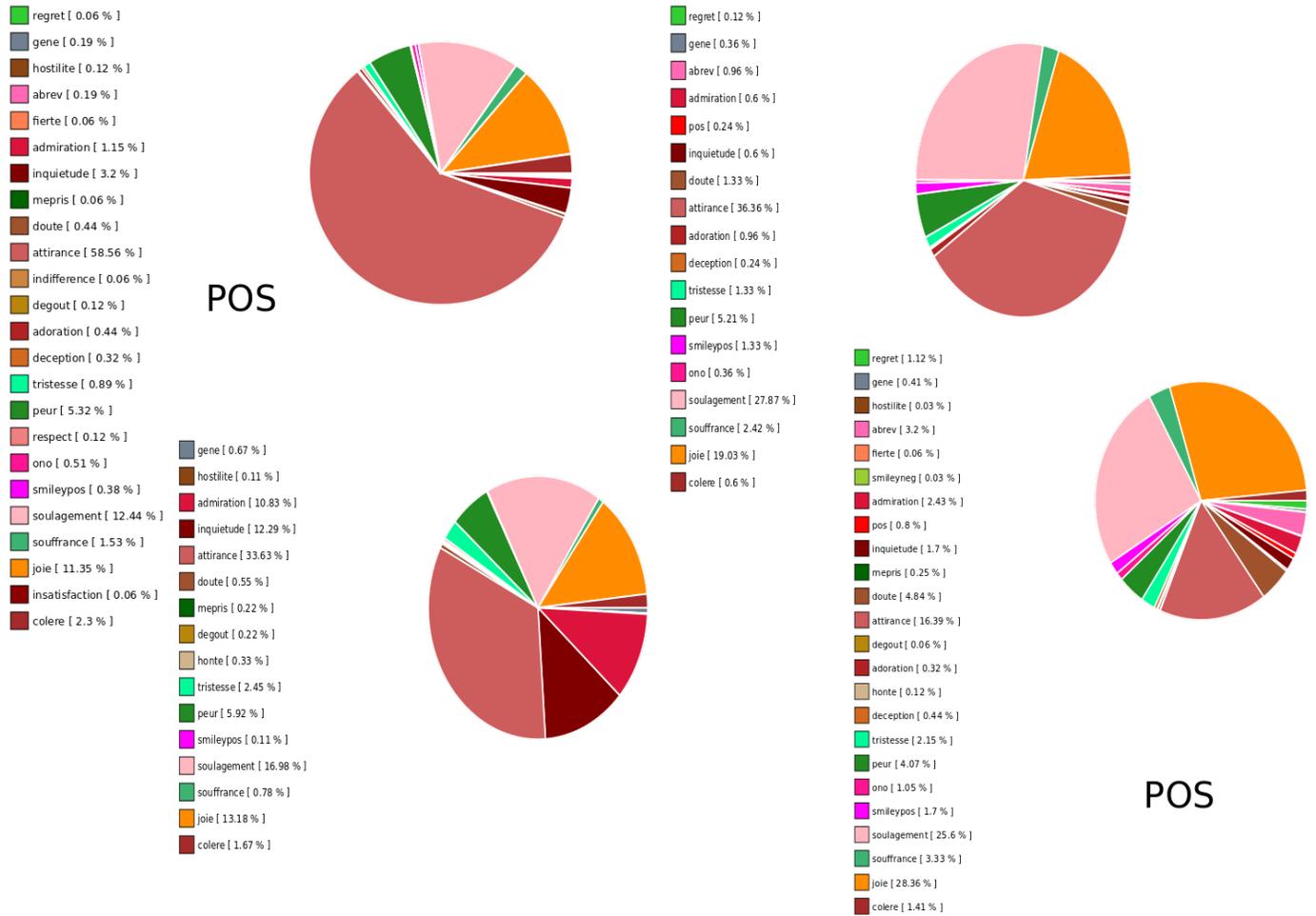
Cette modélisation permet d'avoir un aperçu de la répartition des catégories d'émotions pour chacune des polarités émotionnelles (positive, négative, neutre et incertitude ; nous n'interprétons pas la "polarité inc0" car elle est très peu fréquente). Nous verrons que certaines catégories d'émotions négatives (*peur, inquiétude, gêne,...*) sont catégorisées comme des émotions à polarité positive, et inversement. Ceci est dû à l'application des règles de négation qui inversent la polarité initiale de l'émotion. La légende correspondante aux polarités des catégories d'émotions sur la figure 3.10 sont :

- *incertitude* : incertitude – dégradés de jaune
- *pos* : positive – dégradés de rouge
- *neu* : neutre – dégradés de bleu
- *neg* : négative – dégradés de noir

3.4.2.1 Émotions positives

La figure 3.10 montre les quatre diagrammes circulaires correspondant aux questions et aux réponses des corpus Q/A (figure 3.10(a)) et forum (figure 3.10(b)) pour les émotions à polarité positive.

FIGURE 3.10 – Distribution des catégories d'émotions à polarité positive par corpus et par énoncé



(a) Distribution des catégories d'émotions à polarité positive dans le corpus Q/A (questions en haut et réponses en bas)

(b) Distribution des catégories d'émotions à polarité positive dans le corpus forum (questions en haut et réponses en bas)

3.4.2.1.1 Corpus QA Nous pouvons faire plusieurs observations :

- dans les questions, nous observons que la plupart des émotions positives appartiennent aux trois catégories :
 - *attirance* (50,56%) : *aimer, amour, désir,...*,
 - *soulagement* (12,44%) : *soulager, rassurer, apaiser,...*,
 - *joie* (11,35%) : *plaisir, heureux, tranquille,...*

Nous observons également que les émotions négative ne dépassent pas les 6% :

- *peur* (5,32%) : *souci, peur, angoisse,...*,
- *inquiétude* (3,2%) : *inquiet, inquiétude,...*,
- *colère* (2,30%) : *stresser, ennuyer, énervement,...*
- dans les réponses, les catégories d'émotions sont moins nombreuses mais la répartition est semblable

à celle des questions en ce qui concerne les catégories telles que :

- *attirance* (33,63%) : comme *aimer, apprécier, désirer,...*,
- *soulagement* (16,98%) : *rassurer, soulager, calmer,...*,
- *joie* (13,18%) : *serein, tranquille, satisfaire,...*

La catégorie *admiration* se démarque également, essentiellement marquée par *confiance*.

De manière générale, dans les échanges entre patients et professionnels, les professionnels ont un recours plus fréquent à la négation d'émotions négatives. Cette utilisation permet de rassurer les patients :

- *la fausse couche n'est pas liée au stress*
- *il ne faut pas vous inquiéter*

3.4.2.1.2 Corpus forum Nous pouvons faire plusieurs observations :

- dans les questions, nous observons le même schéma d'expression des émotions positives par les patients que dans le corpus Q/A, où les catégories les plus récurrentes sont :
 - *attirance* (36,36%) : *adorer, envie, formidable,...*,
 - *soulagement* (27,87%) : *soulager, calmer, faire du bien,...*,
 - *joie* (19,03%) : *plaisir, bonheur, confortable,...*

Les catégories d'émotions négatives représentent une proportion très minoritaire. Elles sont les suivantes essentiellement :

- *peur* (5,21%) : *souci, terrible, alarmer,...*,
- *souffrance* (2,42%) : *souffrir*.
- dans les réponses, les patients respectent aussi le même schéma de l'utilisation d'émotions positives, mais dans des proportions un peu différentes :
 - *joie* (28,36%) : *content, heureusement, bonheur,...*,
 - *soulagement* (25,60%) : *soulager, faire du bien, rassurer,...*,
 - *attirance* (16,39%) : *aimer, envie, top,...*

Nous retrouvons également les mêmes émotions à polarité négative telles que :

- *peur* (4,07%) : *peur, stress, affoler,...*,
- *souffrance* (3,33%) : *souffrir, briser, torturer,*
- *tristesse* (2,15%) : *peine, désespérer, pleurer,...*

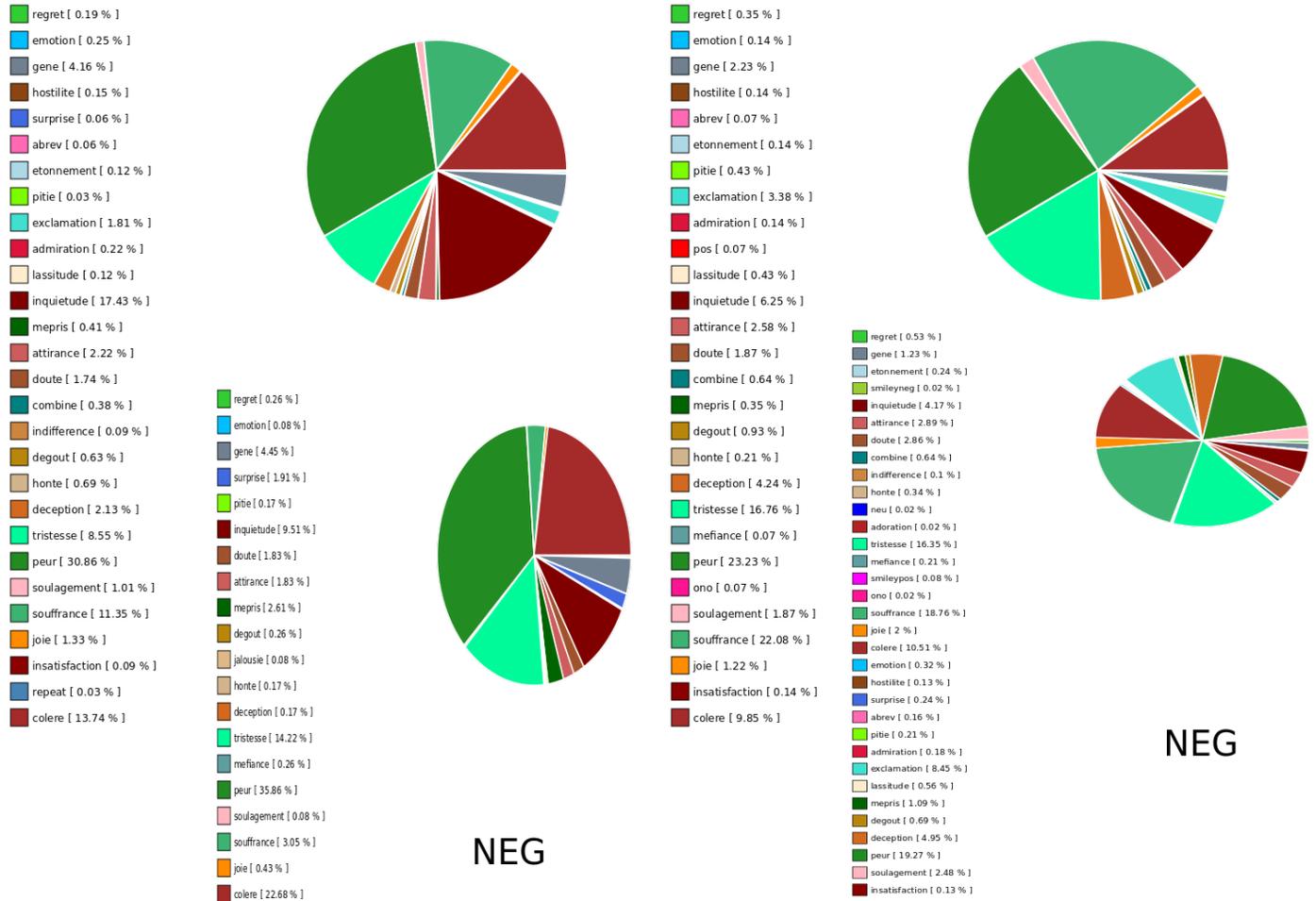
De manière générale, dans les échanges patients/patients, les deux parties ont tendance à employer les mêmes catégories d'émotions, qu'elles soient d'origine positive ou négative. Cela caractérise la volonté des patients qui répondent de partager leurs expériences en décrivant ce qu'ils ont vécu, en rassurant et en aidant les patients qui posent les questions. Ceci permet aux patients malades de s'identifier à une des situations proposées par les patients qui répondent :

- *moi j'ai du lycra qui soulage la douleur de la discopathie*
- *consulter un chiropracteur permet de soulager efficacement les douleurs de dos*

3.4.2.2 Émotions négatives

La figure 3.11 montre les quatre diagrammes circulaires correspondant aux questions et aux réponses des corpus Q/A (figure 3.11(a)) et forum (figure 3.11(b)) pour les émotions à polarité négative.

FIGURE 3.11 – Distribution des catégories d'émotions à polarité négative par corpus et par énoncé (*questions en haut et réponses en bas*)



(a) Distribution des catégories d'émotions à polarité négative dans le corpus Q/A

(b) Distribution des catégories d'émotions à polarité négative dans le corpus forum

3.4.2.2.1 Corpus QA Nous pouvons observer que :

- dans les questions, le nombre de catégories d'émotions négatives est plus importante que d'émotions positives. Parmi celles qui reviennent souvent, nous retrouvons :
 - peur (30,86%) : stress, angoissen anxieux,...
 - inquiétude (17,43%) : inquiétude, inquiet, bizarre,...
 - colère (13,74%) : énerver, nerveux, impatience,...
 - souffrance (11,35%) : souffrir, enfer,...

– *tristesse* (8,55%) : *pleurer, dépression, déprimer,...*

Contrairement aux émotions positives, les émotions à polarité d'origine positive devenues négatives après l'application des règles de négation, sont insignifiantes ;

– dans les réponses, le schéma est assez proche quant à l'utilisation des émotions négatives :

– *peur* (35,86%) : *stress, crainte, phobie,...*,

– *colère* (22,68%) : *nervosité, agressif, stress,...*,

– *tristesse* (14,22%) : *déprimer, triste, peine,...*,

– *inquiétude* (9,51%) : *inquiet, inquiéter, inquiétant,...*

De même que pour les questions, l'inversion de polarité des émotions à l'origine positive est peu fréquente, avec 1,83% pour la catégorie *attirance* et moins de 1% pour les autres catégories, comme *joie* ou *soulagement*.

De manière générale, nous retrouvons dans les messages des professionnels les mêmes émotions dans les questions des patients. En effet, les professionnels confirment souvent l'état de santé dans lequel se trouve le patient :

– *cessez de vous inquiéter*

– *effectivement, vous avez beaucoup souffert dans votre enfance*

Par le même procédé, les professionnels reprennent le ressenti des patients :

– *les symptômes que vous décrivez reflètent des angoisses*

– *il est probable que vous soyez stressée et que vous vous inquiéter pour rien*

Dans les deux cas, il y a une continuité des émotions telles qu'exprimées par les patients et les professionnels.

3.4.2.2.2 Corpus forum Nous pouvons observer que :

– dans les questions, nous retrouvons à nouveau les mêmes catégories d'émotions négatives employées dans les questions du corpus Q/A :

– *peur* (23,23%) : *horrible, terrible, atroce,...*,

– *souffrance* (22,08%) : *souffrir, enfer, meurtrir,...*,

– *tristesse* (16,76%) : *désoler, peine, malheur,...*,

– *colère* (9,85%) : *nerveux, impatience, frustrer,...*,

– *inquiétude* (6,25%) : *inquiéter, bizarre,...*

Nous observons aussi une très faible proportion d'émotions à polarité d'origine positive : *attirance*, *soulagement* et *joie* avec 2,58%, 1,87% et 1,22%, respectivement, ce qui reste assez peu significatif au sein des émotions négatives. Cependant, les émotions à polarité d'origine neutre atteignent maintenant jusqu'à 3,38%. En effet d'après nos règles, une émotion neutre associée avec un marqueur de négation produit souvent une émotion de polarité négative :

– *mais je ne sais pas comment faire !!!*

– *je ne peux rien faire !!!*

– dans les réponses, nous retrouvons à nouveau un schéma similaire dans l'expression des émotions négatives par les patients qui répondent, avec une forte présence de :

– *peur* (19,27%) : *flipper, paniquer, appréhender,...*,

- *souffrance* (18,76%) : *souffrir, enfer, agoniser,...*,
- *tristesse* (16,35%) : *calvaire, pénible, décevoir,...*,
- *colère* (10,51%) : *colère, ennui, embêter,...*

De plus, dans ces énoncés, les émotions à polarité d'origine neutre sont plus fréquentes puisqu'elles représentent 8,45%, soit plus du double par rapport aux questions. Nous retrouvons également les catégories d'émotions telles que *attirance, soulagement* ou encore *joie* mais dans de faibles proportions : 2,89%, 2,48% et 2%, respectivement.

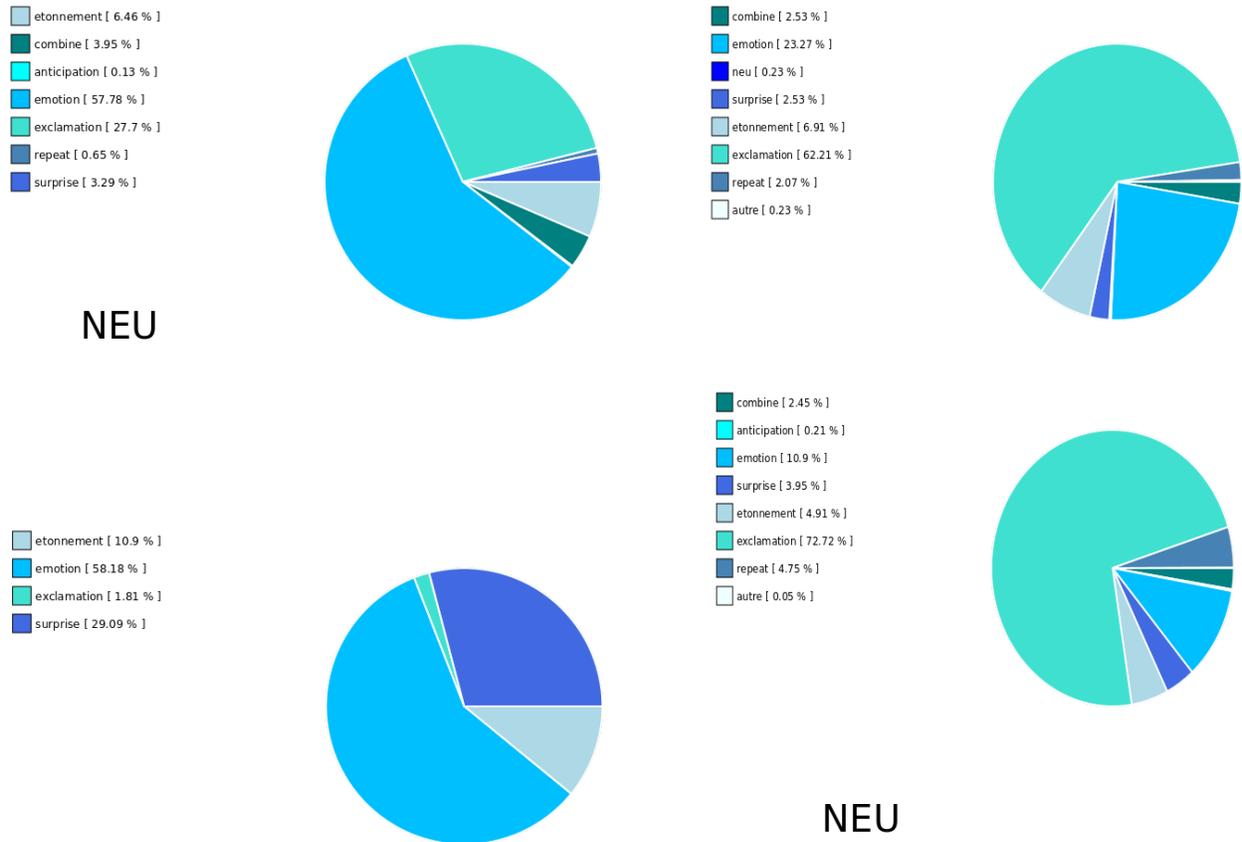
- dans le corpus forum, les patients expriment leurs émotions en utilisant des *modificateurs* d'intensité comme par exemple les émotions neutres véhiculées par des ponctuations expressives :
 - *les médicaments ne font plus aucun effet !!*
 - *je n'ai pas eu plus d'infos !!!!*

De cette façon, les patients communiquent une intensité émotionnelle plus forte, que l'on ne retrouve pas dans le corpus Q/A. En effet, lorsque les patients s'adressent aux professionnels, ils doivent s'adapter à un niveau de langage plus formel, alors qu'en s'adressant à d'autres patients, le niveau de langage peut rester assez familier. Étant dans ce cadre social familier, les patients laissent leurs émotions s'exprimer plus spontanément qu'avec les professionnels.

3.4.2.3 Émotions neutres

La figure 3.12 montre les quatre diagrammes circulaires correspondant aux questions et aux réponses des corpus Q/A (figure 3.12(a)) et forum (figure 3.12(b)) pour les émotions à polarité neutre.

FIGURE 3.12 – Distribution des catégories d'émotions à polarité neutre par corpus et par énoncé



(a) Distribution des catégories d'émotions à polarité neutre dans le corpus QA (*questions* en haut et *réponses* en bas)

(b) Distribution des catégories d'émotions à polarité neutre dans le corpus forum (*questions* en haut et *réponses* en bas)

3.4.2.3.1 Corpus QA Nous pouvons observer que :

- dans les questions, les patients expriment beaucoup d'émotions neutres (57,78%) avec des termes tels que *ressentir, éprouver, ...*. Ces termes permettent en effet de décrire les états liés à leur maladie. Les patients ont également souvent (27,7%) recours aux ponctuations expressives (catégorie *exclamation* sur les diagrammes :!, !!, !!!!!, ...). Comme nous l'avons déjà indiqué, ils utilisent ces marqueurs émotionnels dans le but d'intensifier la description de leur état physique ou psychologique. Les patients emploient aussi des termes liés à d'autres catégories :
 - *étonnement* (6,46%) : *surprendre, étonner, ...*,
 - *surprise* (3,29%) : *curieux, incroyable, ...*,
 - combinaison de ponctuations expressives (catégorie *combine* 3,95%) : *?!, !!!?, ??!!, ...*
- dans les réponses, les professionnels utilisent moins de catégories d'émotions neutres (4 contre 7). Par exemple, les professionnels de santé n'utilisent pas les ponctuations expressives ni leurs combinaisons

(*??!,...*), ni la répétition de caractères (0,65% ; *moiiiiiii, merciiii,...*). Cependant, les professionnels ont recours plus particulièrement aux catégories suivantes :

- *"émotion ressentie"* (58,18%) : *ressentir, éprouver,...*,
- *surprise* (29,09%) : *surprenant, curieux, bizarre,...*,
- *étonnement* (10,90%) : *cho, étonner,...*

– dans les échanges patients/professionnels, nous observons aussi une nette différence dans l'utilisation des émotions neutres. Si les patients y ont recours plus fréquemment que les professionnels, c'est parce que, de cette manière, ils peuvent ajouter une intensité supplémentaire à ce qu'ils écrivent, afin de solliciter une réaction plus rapide :

- *j'ai peur que ce soit un mélanome !!*
- *les anti inflammatoires me font plus rien !!!*

Tandis que dans le cas des professionnels, l'expression d'émotions neutres est plutôt destinée à renseigner les patients sur leurs symptômes ou à les leur confirmer, d'après la description fournie par le patient :

- *la douleur que vous ressentez est probablement due au kyste*
- *il est normal que vous ressentiez de multiples symptômes*

Ces émotions peuvent aussi marquer la difficulté qu'ont les professionnels à comprendre les écrits et les raisonnements des patients :

- *une arythmie à votre âge est surprenant*
- *votre description m'étonne beaucoup*

3.4.2.3.2 Corpus forum Plusieurs observations peuvent être faites :

– dans les questions, contrairement au corpus Q/A, les patients utilisent plus de marqueurs émotionnels de type ponctuations expressives (62,21% des émotions neutres). Ils préfèrent aussi employer des termes de la catégorie *émotion* (23,27% ; *ressentir, choquer, éprouver,...*). Enfin, de façon moins importante, ils ont recours à :

- *étonnement* (6,91%) : *surprise, étonner,...*,
- caractères répétés (2,07%, dénommée *repeat*) : *pffff, grosssssse, trèèèèès,...*,
- combinaisons de ponctuations expressives (2,53%) : *?! , ? ?!, ? ! ?,...*

– dans les réponses, les patients répondant aux messages utilisent aussi très fréquemment les ponctuations expressives (72,72% ; *!, !!!!!, !!!!!!!!!,...*). Ils emploient des marqueurs émotionnels tels que *ressentir, choquer, secouer, éprouver,...* (appartenant à la catégorie *émotion* dans les diagrammes). Nous observons également une proportion légèrement plus importante d'émotions exprimées par la répétition de caractères (4,75%) comme *ouffffff, lolllll, grrrrrr,...*

– dans les échanges patients/patients, les émotions neutres sont caractérisées par l'usage de ponctuations expressives dans les questions :

- *et j'ai toujours des douleurs !!*
- *je nage dans l'inconnu total !!!*

et dans les réponses :

- *la cause de ma douleur était la position assise !!!!*
- *les memes problèmes que toi symptômes identiques !!*

Nous pensons que dans les questions, cela permet de communiquer la difficulté de vivre avec la maladie, tandis que dans les réponses, cela permet de partager l'expérience et de donner des conseils. De plus, les patients ont tendance à exprimer plus librement leurs émotions par l'utilisation de marqueurs émotionnels, comme la ponctuation (!!!!, !! ? ?), les abréviations (*lol, mdr*) et les émoticônes (l :), :(). En ce qui concerne les échanges patients/professionnels, les patients ont recours aux émotions neutres afin de décrire leur état face à la maladie :

- *je ressens de temps à autre des douleurs musculaires*
- *bonjour, je ressens des douleurs aux reins*

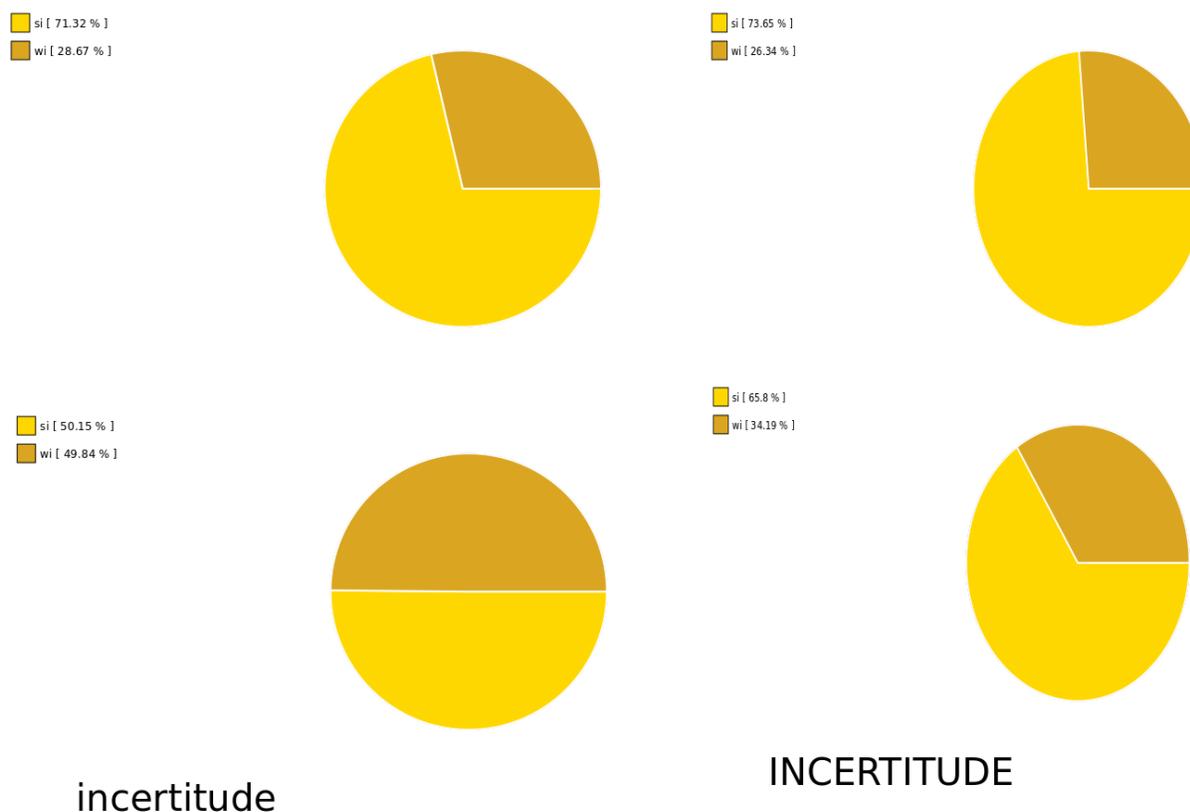
Tandis que les professionnels employent surtout des marqueurs émotionnels neutres dans le but d'exprimer leur incompréhension des écrits ou des conclusions des patients :

- *la durée d'hospitalisation est en effet surprenante*
- *une perte de 10 kg est surprenante*

3.4.2.4 Incertitudes

La figure 3.13 montre les quatre diagrammes circulaires correspondant aux questions et aux réponses des corpus Q/A (figure 3.13(a)) et forum (figure 3.13(b)) pour les émotions à polarité neutre.

FIGURE 3.13 – Distribution des catégories d’émotions à polarité “d’incertitude” par corpus et par énoncé



(a) Distribution des catégories d’émotions à polarité “d’incertitude” pour le corpus Q/A (*questions* en haut et *réponses* en bas)

(b) Distribution des catégories d’émotions à polarité “d’incertitude” pour le corpus forum(*questions* en haut et *réponses* en bas)

3.4.2.4.1 Corpus QA Nous faisons les observations suivantes :

- dans les messages postés par les patients, nous remarquons des incertitudes fortes (71,32% ; *si* sur le diagramme). En effet, les patients se retrouvent dans une situation où ils sont impuissants : ils n’ont pas le savoir nécessaire pour comprendre la maladie et encore moins pour se soigner. Ils sollicitent donc les professionnels. Pour cela, les patients ont recours à des marqueurs d’incertitude forte (?, *risquer, possible, souhaiterais,...*) afin de former une interrogation. Les patients utilisent aussi des marqueurs d’incertitude faible (*wi* sur le diagramme) mais de façon moins importante (28,68% ; *si, peut, savoir si, en cas de,...*). En effet, ils ont recours à ce type de marqueur dans le but de lever leurs doutes et inquiétudes liées à leur maladie.
- dans les réponses, les professionnels utilisent les deux catégories de marqueurs de façon équitable : 50,15% pour les incertitudes fortes et 49,84% pour les incertitudes faibles. Les médecins renseignent

les patients sur leurs inquiétudes, mais, grâce à ces marqueurs, ils ne s'engagent pas dans la confirmation des solutions proposées. En effet, il serait dangereux pour un professionnel de santé d'établir un diagnostic uniquement sur une description textuelle des symptômes et basé seulement sur la perception des patients (manque de précisions et de détails) ;

- les patients cherchent à récolter des informations pour améliorer la qualité de vie avec la maladie, et pour ceci ils posent des questions :
 - *Y'a t il des risques de grossesse ?*
 - *Est-ce dû au kyste ?*
 - *est-ce possible que se soit dû à une mauvaise digestion ?*

Tandis que les professionnels tentent de satisfaire ces requêtes en prenant certaines précautions quant aux conseils qu'ils prodiguent :

- *Steve, il peut s'agir d'hémorroïdes*
- *Paul, votre grain de beauté semble normal*
- *ce que vous décrivez évoque plutôt une infection virale*

3.4.2.4.2 Corpus forum Le corpus forum offre aussi quelques particularités :

- dans les questions, les patients utilisent un schéma presque identique à celui observé dans le corpus Q/A, dans l'emploi des marqueurs d'incertitude : 63,65% pour les incertitudes fortes et 26,35% pour les faibles. L'objectif des patients est alors d'obtenir des informations de la part d'autres personnes. Dans tous les cas, les patients cherchent à être rassurés en espérant de trouver des solutions (démarches, médicaments, spécialiste à consulter, ...) afin d'améliorer la qualité de vie avec la maladie ;
- dans les réponses, les patients qui répondent font globalement la même utilisation des incertitudes (34,19% faibles contre 65,81% fortes). Ces personnes ont tendance à privilégier les incertitudes fortes afin d'obtenir plus de précisions sur la situation des patients qui posent les questions, ceci dans le but de donner une réponse la plus précise et efficace possible ;
- les patients qui posent les questions cherchent des informations capables de les éclairer :
 - *qui a eu des neuralgies intercostales, pouvez vous me décrire les symptômes*
 - *je voulais savoir si quelques personnes ont ces malformations*
 - *pour quel type de hernie peut on avoir recour à la chirurgie au laser ?*

D'autre part, les personnes qui partagent leurs témoignages ont parfois besoin d'avoir plus de précisions sur les troubles subis :

- *ton poignet gonfle t'il ?*
- *as tu des douleurs importantes ?*
- *tu ne prends pas d'antidouleur ?*

Ces personnes donnent également des conseils pour améliorer la vie avec la maladie :

- *si la scoliose est confirmée à ce moment la il faudrait consulter un spécialiste*
- *je peux vous conseiller un chirurgien très compétent*

Comme dans le cas de réponse des professionnels, les patients qui répondent emploient aussi les marqueurs d'incertitude :

- *d’après ce que vous me dites, cela ne serait donc pas une hernie*
- *dans ce que tu décris, on pourrait aussi penser à un syndrome du canal carpien*

3.4.2.5 Bilan

Les diagrammes circulaires permettent de visualiser les émotions et leur évolution au sein des questions et des réponses dans chaque corpus. Chaque type d’échanges (patients/professionnels et patients/patients) présente des caractéristiques particulières. Les échanges entre les patients et les professionnels sont clairement marqués par une forte présence de marqueurs d’incertitude et une faible présence de marqueurs émotionnels (beaucoup moins nombreux chez les professionnels). Alors que dans les échanges entre patients, ces derniers expriment des marqueurs émotionnels de manière plus fréquente et plus de marqueurs d’incertitude. Dans cette représentation des polarités et catégories émotionnelles, nous pouvons dire que les patients ciblent le type d’échanges suivant les informations qu’ils cherchent. S’ils préfèrent avoir accès à un savoir technique de qualité, ils vont favoriser une communication avec des professionnels de santé (comme le corpus Q/A). Ils vont donc devoir observer une certaine retenue quant à l’expression de leurs émotions dans leur communication avec les professionnels. En revanche, si les patients préfèrent une communication plus souple et moins codée (communication entre personnes de "même niveau social" contrairement aux échanges précédents où les patients doivent respecter certaines normes dans leur manière d’écrire leur message), ils vont opter pour des échanges avec d’autres patients. Ces derniers vont partager leurs expériences et leurs témoignages sur les maladies vécues par les patients posant les questions. Cette modélisation permet également d’obtenir des informations sémantiques en relation avec les polarités émotionnelles et les catégories émotionnelles. En effet, il est possible d’obtenir des informations comme les marqueurs émotionnels utilisés, les types médicaux présents dans les contextes des marqueurs émotionnels, . . .

Notons que cette visualisation peut être générée pour différentes polarités (positive, négative, neutre, “incertitude”) et différentes catégories d’émotions. L’ensemble des informations associées (lemmes, marqueurs, types médicaux. . .) sont accessibles dans un navigateur après y avoir chargé le fichier XML choisi selon trois critères :

- le type de corpus : Q/A ou forum,
- le type de l’énoncé : question ou réponse,
- le type de diagramme circulaire : par polarité, par catégorie émotionnelle.

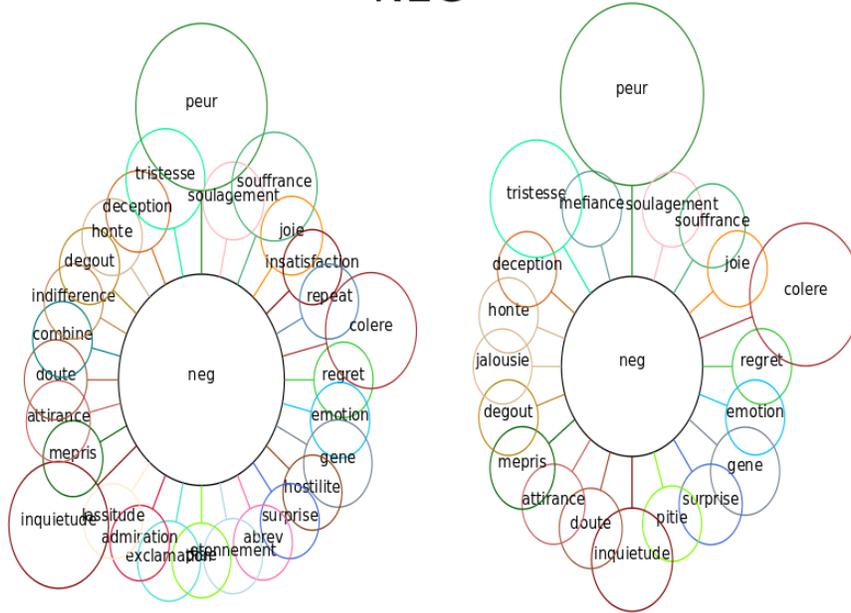
3.5 Les “constellations“ : représentation quantitative de l’expression des émotions par polarité

Le mécanisme utilisé consiste à générer des cercles pour les polarités (marquées par *pos – polarité positive, neg – polarité négative, neu– polarité neutre, inc– polarité "incertitude"*) et des catégories d’émotions correspondantes. Par exemple, sur la figure 3.14, la polarité du cercle du milieu (polarité *neg*) est instanciée par toutes les catégories d’émotions liées gravitant autour : *peur, inquiétude, colère, souffrance*, etc. La taille du diamètre des cercles varie proportionnellement en fonction du nombre d’occurrences de ses termes et marqueurs : plus il y a d’occurrences, plus la surface du cercle correspondant est importante.

La figure 3.14 représente les proportions d'émotions négatives employées dans les corpus Q/A (figure 3.14(a)) et forum (figure 3.14(b)).

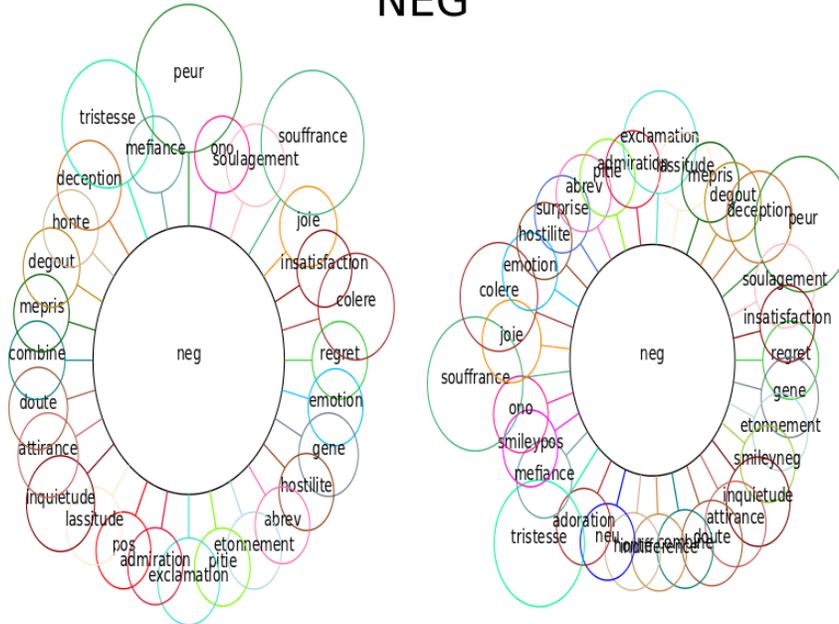
FIGURE 3.14 – Distribution des catégories d'émotions à polarité négative par corpus et par énoncé

NEG



(a) Distribution des catégories d'émotions à polarité négative pour le corpus Q/A (*questions* à gauche et *réponses* à droite)

NEG



(b) Distribution des catégories d'émotions à polarité négative pour le corpus forum (*questions* à gauche et *réponses* à droite)

Cette modélisation permet également de visualiser la proportion des différentes polarités des émotions (*positive, négative, neutre, "incertitude"*). Comme auparavant, il est possible de visualiser cette distribution par catégorie émotionnelle (*joie, attirance, peur, inquiétude, étonnement, surprise, ...*). Pour ce mode de visualisation, nous utilisons les structures de données identiques à celles utilisées précédemment (section 3.4) : une par type d'énoncé et pour chacun des deux corpus. Comme seul le mode d'affichage est différent, nous détaillons le mécanisme de visualisation seulement. Pour l'interprétation des résultats, il faut se reporter à la section 3.4 page 86.

Nous pensons que ce mode de représentations permet d'observer l'évolution des émotions dans les deux corpus. Par exemple, dans le corpus Q/A (en haut), nous pouvons remarquer que l'emploi d'émotions négatives est plus important chez les patients que chez les professionnels : le diamètre du cercle *neg* des questions est plus grand que celui des réponses. De plus, si l'on compare les deux corpus, cette visualisation indique que la proportion d'émotions négatives dans le corpus forum est plus importante que dans le corpus Q/A.

Grâce à cette visualisation, nous pouvons observer, dans la figure 3.15, les particularités de répartition des émotions positives dans les corpus Q/A (figure 3.15(a)) et forum (figure 3.15(b)) :

- dans le corpus Q/A, une plus grande diversité de marqueurs émotionnels positifs apparaît dans les questions des patients, et non pas dans les réponses des professionnels,
- dans le corpus forum, la situation est inverse car nous observons une plus grande variété de catégories d'émotions dans les réponses.

Enfin, cette visualisation permet de voir que les catégories émotionnelles telles que *attirance, soulagement, joie* dominent dans les questions et les réponses dans les deux corpus. En effet, l'expression d'émotions positives dans les questions des patients des corpus Q/A et forum se caractérise par la recherche de solutions médicales (plus) efficaces et de réconfort :

- *avez-vous une parole **réconfortante**, j'ai tellement peur ;*
- *je désirerais donc être **rassure** ;*
- *je bouge ma jambe ou l'étire pour **soulager** la douleur pendant quelque seconde ;*
- *qu'est-ce que je peux mettre ou prendre pour **calmer** vraiment la douleur ;*
- *vous avez des solutions à mon problème pour **soulager** ma douleur merci de me répondre*

Quant aux patients qui répondent dans le corpus forum, ils essaient plutôt de proposer des solutions aux patients qui souffrent ou expriment leur empathie :

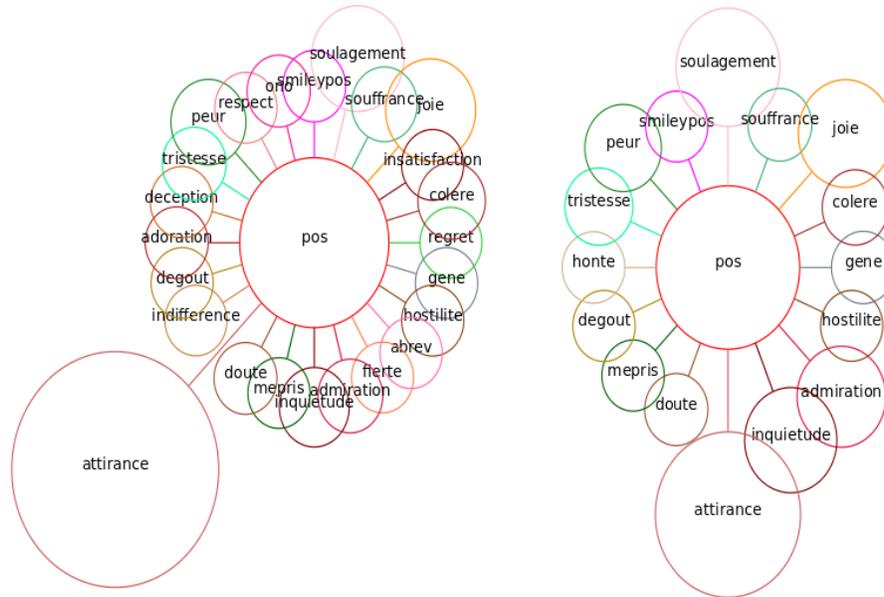
- *un kiné peut être utile pour **soulager** les douleurs ;*
- *Iplikator peut beaucoup t'aider à **soulager** la douleur ;*
- *j'**espère** que la prochaine infiltration pourra te **soulager** ;*
- *bravo pour ton **optimisme** et ta volonté.*

Dans les réponses formulées par les professionnels de santé dans le corpus Q/A, ils tentent de rassurer les patients et proposent des solutions à leurs troubles :

- ***heureusement**, les séquelles sont très rares ;*
- *soyez **tranquille** et en cas d'inquiétude, voyez votre médecin ;*
- *les boissons chaudes comme le thé peut aussi **soulager** ;*
- *le rayon peut également **soulager** les douleurs osseuses.*

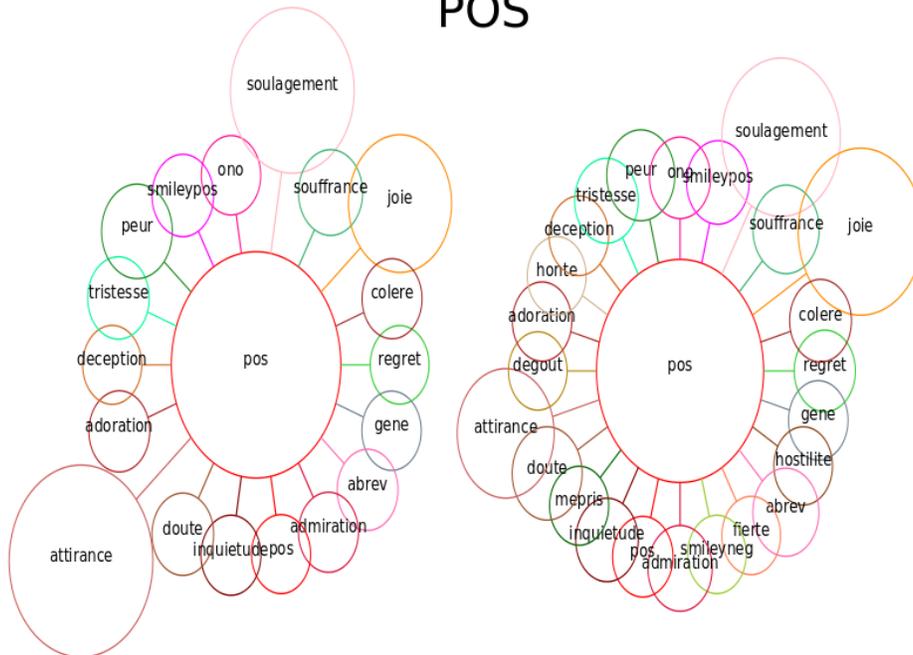
FIGURE 3.15 – Distribution des catégories d'émotions à polarité positive par corpus et par énoncé

POS



(a) Distribution des catégories d'émotions à polarité positive pour le corpus Q/A (*questions* à gauche et *réponses* à droite)

POS



(b) Distribution des catégories d'émotions à polarité positive pour le corpus forum (*questions* à gauche et *réponses* à droite)

Les visualisations correspondantes aux émotions à polarité neutre et "incertitude" sont disponibles à l'annexe G.4.

L'avantage de cette visualisation par rapport aux diagrammes circulaires est qu'elle permet d'observer les proportions des émotions en fonction de leur polarité, en même temps que les proportions des catégories d'émotions.

3.6 Les nuages de mots : modélisation du contenu des contextes émotionnels

Le principe de la visualisation avec des nuages de mots consiste à afficher l'ensemble de mots-clés d'un document ou corpus, où la taille de ces mots-clés varie proportionnellement au nombre de leurs occurrences. Les couleurs des mots-clés sont évocatrices des polarités des émotions : *positive* en rouge, *négative* en noir, *neutre* en bleu et *incertitude* en jaune. Nous avons généré des nuages de mots pour chaque corpus : un pour les questions et un pour les réponses. En passant le curseur de la souris sur une catégorie d'émotions, une fenêtre d'affiche en bas de la page et présente différentes informations : la polarité de l'émotion sélectionnée, le nombre d'occurrences de marqueurs émotionnels de cette catégorie émotionnelle, le nombre d'occurrences des différents concepts médicaux (*procédures, maladies, médicaments*),... La figure 3.16 présente les nuages de mots correspondant aux questions (figure 3.16(a)) et aux réponses (figure 3.16(b)) du corpus Q/A.

FIGURE 3.16 – Distribution des marqueurs émotionnels en fonction de leur catégorie émotionnelle par énoncé pour le corpus Q/A



(a) Distribution des marqueurs émotionnels en fonction de leur catégorie émotionnelle pour les questions



(b) Distribution des marqueurs émotionnels en fonction de leur catégorie émotionnelle pour les réponses

Nous pouvons faire les mêmes observations que pour les visualisations précédentes : les patients emploient plus de marqueurs émotionnels que les professionnels. Les catégories émotionnelles qui reviennent le plus fréquemment pour les émotions à polarité positive sont *attirance*, *soulagement*, *joie*, tandis que pour les émotions à polarité négative nous avons *peur*, *inquiétude*, *colère*, *souffrance*. Dans le discours des professionnels, nous retrouvons également les mêmes catégories émotionnelles que celles détectées grâce aux diagrammes circulaires : *peur*, *inquiétude*, *tristesse*, *joie*, *attirance*, *soulagement*. La figure 3.17 présente les nuages de mots correspondant aux questions (figure 3.17(a)) et aux réponses (figure 3.17(b)) du corpus forum.

FIGURE 3.17 – Distribution des marqueurs émotionnels en fonction de leur catégorie par énoncé pour le corpus *forum*



(a) Distribution des marqueurs émotionnels en fonction de leur catégorie émotionnelle pour les questions



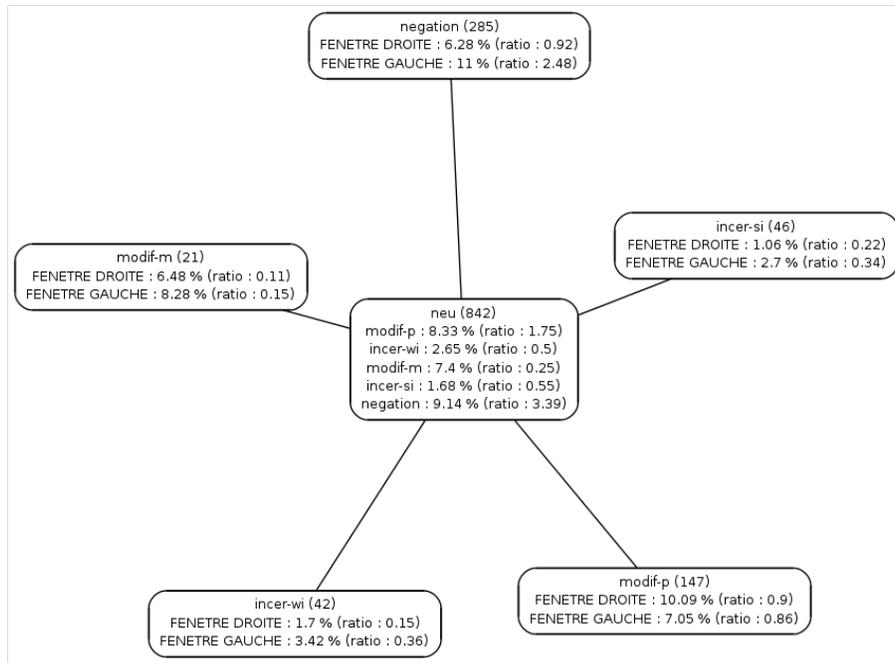
(b) Distribution des marqueurs émotionnels en fonction de leur catégorie émotionnelle pour les réponses

Ici aussi, nous pouvons faire les mêmes observations sur la distribution des catégories d'émotions qu'avec les diagrammes circulaires : dans les questions et les réponses, les patients expriment leurs émotions beaucoup plus que dans le corpus Q/A. De plus, nous remarquons que l'emploi de ponctuations expressives (catégorie *exclamation*) est plus fréquente dans ce corpus.

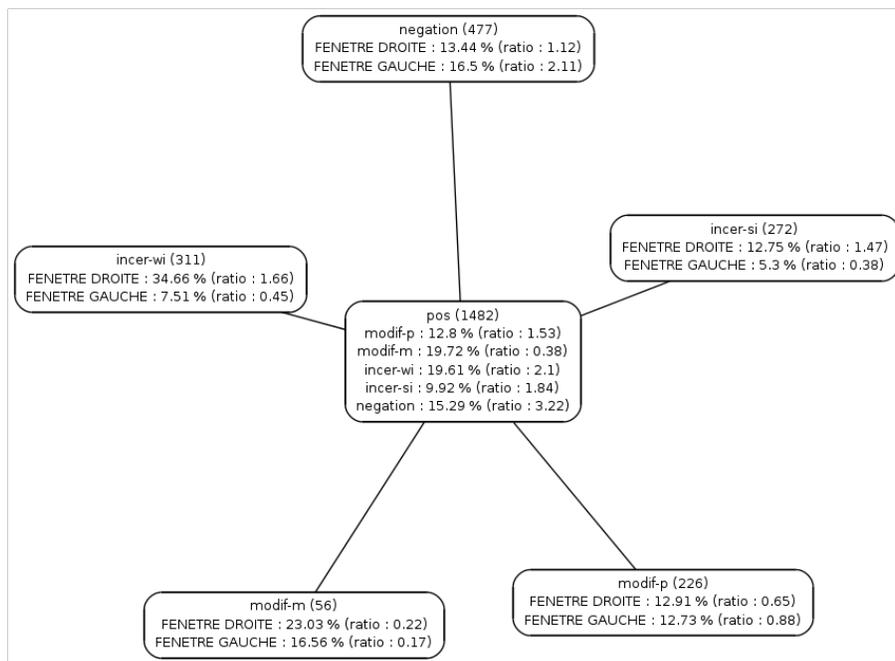
Nous pouvons donc voir que cette modélisation propose un autre type de visualisation de l'évolution des émotions dans les échanges patients/patients et patients/professionnels. En utilisant les nuages de mots, nous pouvons voir directement quelles émotions et quelles polarités émotionnelles sont les plus caractéristiques d'un type d'énoncé. En effet, la taille d'une catégorie d'émotions renseigne sur sa fréquence d'utilisation, alors que la couleur renseigne sur sa polarité. Notons qu'une même émotion peut être définie par deux polarités différentes si elle contient dans ses contextes un ou des marqueurs de négation, qui inverse alors la polarité d'origine de cette émotion. Par exemple, dans les réponses formulées par les patients dans le corpus forum, les termes émotionnels définissant la catégorie émotionnelle *soulagement* sont aussi bien employés

dans un contexte positif (rouge) que négatif (noir). Pour cet exemple, la taille de la police indique que cette catégorie d'émotions est surtout employée dans des descriptions positives.

FIGURE 3.7 – Association de différents marqueurs *incertitude*, *modifieurs*, *négation* avec les polarités *neutre* et *positive* dans le corpus *Q/A*

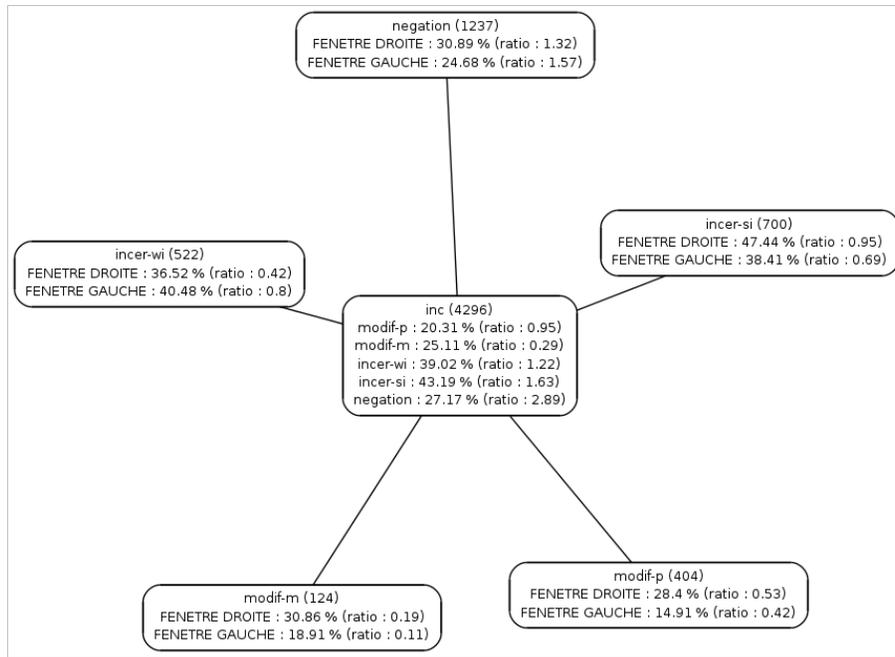


(a) Association de différents marqueurs *incertitude*, *modifieurs*, *négation* avec la polarité *neutre* dans le corpus *Q/A*

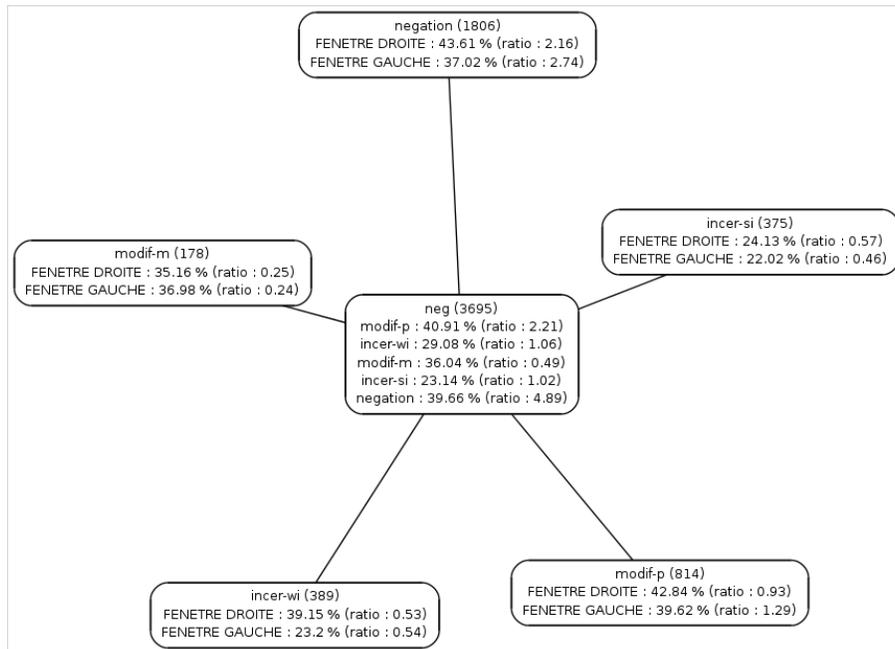


(b) Association de différents marqueurs *incertitude*, *modifieurs*, *négation* avec la polarité *positive* dans le corpus *Q/A*

FIGURE 3.8 – Association des marqueurs *incertitude*, *modifieurs*, *négation* avec les polarités émotionnelles *incertitude* et *négative* dans le *corpus forum*



(a) Association des marqueurs *incertitude*, *modifieurs*, *négation* avec la polarité émotionnelle *incertitude* dans le *corpus forum*



(b) Association des marqueurs *incertitude*, *modifieurs*, *négation* avec la polarité émotionnelle *negative* dans le *corpus forum*

Chapitre 4

Conclusion et perspectives

4.1 Conclusion

4.1.1 Particularités des discours et des corpus étudiés

Dans le cadre de ce travail, nous avons proposé un système permettant de visualiser les émotions exprimées par les patients entre eux et celles exprimées lors de l'interaction entre patients et professionnels de santé. Dans les deux corpus étudiés, les interactions sont organisées selon le principe de questions-réponses. Ainsi, cette structure en questions-réponses nous a permis de simuler la représentation d'une évolution des émotions dans ces fils de discussion. De plus, les différents types de marqueurs viennent interagir avec les marqueurs émotionnels dans cette évolution. Différents types de visualisations ont été mis en place afin de faire ressortir différentes interprétations liées à l'expression et l'évolution des émotions dans les forums de santé analysés. De cette manière, il est possible de voir quels genres d'interactions existent dans les situations de communication patients/patients et patients/professionnels de santé.

4.1.1.1 Interactions patients/professionnels de santé dans le corpus Q/A

Dans les échanges patients/professionnels, nous avons souvent vu que les patients cherchent à être rassurés, informés sur la prise de risque dans l'attribution d'un diagnostic, avoir quelqu'un pour se plaindre sur la qualité insuffisante de soins médicaux ou sur un professionnel de santé, être conseillés sur des démarches à suivre pour améliorer la qualité de vie avec la maladie, . . . Cependant, malgré une certaine réserve quant à l'expression des émotions, les professionnels de santé essaient de venir en aide aux patients en les conseillant, en les informant, en les rassurant et en tentant de lever leurs doutes. Le trait typique est que les professionnels prennent souvent des précautions dans la formulation de leur réponse. Ceci se caractérise par l'emploi fréquent de marqueurs d'incertitude. En effet, il serait trop risqué pour un médecin d'établir et de confirmer un diagnostic sans avoir examiné au préalable le patient. L'emploi des incertitudes par les professionnels est également lié aux lacunes de description des symptômes de la part des patients.

4.1.1.2 Interactions patients/patients dans le corpus forum

Dans les échanges entre patients, les personnes expriment plus facilement et librement leurs émotions. Ceux qui posent des questions essaient de solliciter de la compassion et de l'empathie chez d'autres patients ayant déjà vécu avec la maladie en question. Ils essaient d'obtenir des témoignages, des partages d'expériences et des conseils (sur les procédures les plus pertinentes à entreprendre, sur les spécialistes plus ou moins compétents et efficaces, . . .). Enfin, dans ce type d'échanges, l'emploi de marqueurs d'incertitude est peu fréquent.

Comme le décrivent les auteurs dans (Chauveau-Thoumelin & Grabar, 2014), la manifestation du degré de subjectivité diffère selon la catégorie socio-professionnelle (patient ou professionnel de santé) de l'auteur d'un message. De plus, l'emploi des différents marqueurs (émotions, incertitude) est spécifique à chaque catégorie, comme le montre les travaux de (Abdaoui *et al.*, 2014). De manière générale, le discours des patients est caractérisé par la subjectivité, exprimée essentiellement à travers les marqueurs émotionnels qui permettent d'obtenir des réponses pour améliorer la qualité de vie avec la maladie. Nous pensons qu'à travers l'emploi de différents marqueurs (modificateurs, négation, incertitude et émotions) les patients tentent d'attirer l'attention sur leurs problèmes. Même si les professionnels de santé ont également recours à ces marqueurs, ils restent plus objectifs dans leurs réponses et les utilisent moins fréquemment, mise à part l'utilisation fréquente de marqueurs d'incertitude. En fonction de ce que recherchent les patients malades, ils préféreront soit des réponses émanant de professionnels, pouvant leur apporter des connaissances techniques sur la maladie et des solutions dans lesquelles ils peuvent avoir confiance, soit ils opteront pour la recherche de témoignages d'autres patients ayant vécu la maladie afin de se faire une idée sur les divers traitements possibles et plus ou moins efficaces.

4.1.2 Les concepts médicaux

Concernant l'emploi des concepts médicaux, les différents modes de visualisations ont montré que les concepts médicaux de type *maladies* sont les plus fréquents dans les deux corpus. D'autre part, les échanges patients/professionnels sont plus marqués par les concepts *médicaments* que par *procédures*, plus caractéristiques du corpus forum (partage d'expériences par des patients qui ont déjà vécu avec la maladie). Dans les deux types de communication, les émotions à polarité négative sont les plus utilisées. En effet, elles marquent les doutes, les peurs, les inquiétudes liées à la maladie ou à l'absence de connaissances techniques des patients sur leurs troubles physiques ou psychologiques.

4.1.3 Efficacité du système à base de règles

Nous avons proposé un système à base de règles pour détecter la polarité émotionnelle dominante d'un énoncé afin de pouvoir visualiser l'évolution des émotions dans deux types d'échanges : lorsque les patients posent leur question et des professionnels de santé y répondent, et lorsque les patients posent leurs questions et d'autres patients y répondent. Dans ces deux types d'échanges, nous avons appliqué un certain nombre de règles (dans un ordre bien défini) sur les marqueurs émotionnels présents dans les énoncés (questions et réponses). Ces règles sont définies suivant les types de marqueurs : négation, modificateurs (plus et moins),

incertitude (forte et faible), connecteurs. Pour chacun de ces marqueurs, une règle y est associée, permettant alors de déterminer la polarité de chaque émotion exprimée dans les énoncés, selon leur présence dans le contexte de l'émotion (sept tokens à gauche et sept tokens à droite). Nous avons constaté que l'application de ces règles sur les marqueurs émotionnels permet d'améliorer la détection, par le système, de la polarité associée à cette émotion. En effet, si une émotion positive comme *soulager* contient un marqueur de négation dans son contexte et si la règle s'applique, alors la polarité du marqueur émotionnel *soulager* deviendra négative (par inversion de polarité). De plus, la détection de *connecteurs* pertinents (*car, donc, et, mais, cependant*) a permis de construire une règle permettant d'améliorer la détection de la polarité des marqueurs émotionnels tout en définissant de nouvelles frontières contextuelles pour les marqueurs émotionnels. D'autre part, à partir d'un lexique polaire (une polarité associée à chaque catégorie émotionnelle), nous avons essayé de détecter la polarité des marqueurs émotionnels dans les énoncés, en fonction de la présence de marqueurs (négation, incertitude, . . .) dans leur contexte. Pour cela, nous avons calculé l'intensité polaire de chaque émotion dans les questions et les réponses afin de définir la polarité émotionnelle dominante d'un énoncé. Cela permet d'avoir une idée du type d'émotion dominante d'après le contenu plutôt positif, négatif, neutre ou marqué d'incertitude d'un énoncé.

4.1.4 Difficultés

Nous avons rencontré quelques difficultés inhérentes aux données.

4.1.4.1 Création des données de référence

Lors de la création des données de référence, la plus grande difficulté d'annotation s'est présentée pour les annotateurs semi-expert et non-expert. En effet, il leur a été assez difficile de distinguer deux situations assez récurrentes dans les corpus traités, à savoir pouvoir distinguer entre :

- une émotion, qui peut être positive, négative ou neutre,
- un état factuel ponctuel lié à une maladie, mais qui peut aussi être positif, négatif ou neutre.

Ces annotations se sont fait ressentir sur les données de référence produites. Notons aussi que le système actuel a plus de facilité à reproduire les annotations effectuées par les annotateurs semi-expert et non-expert.

4.1.4.2 Détection de l'ironie et du sarcasme

Les messages dans les forums peuvent contenir de l'ironie et du sarcasme, alors que le système automatique est incapable de les détecter. Selon nos observations, ces situations sont parfois marquées par l'utilisation de smileys (*l :-D, . . .*) ou d'abréviations spécifiques (*mdr, lol*). Dans l'état actuel du système, ces marqueurs émotionnels sont identifiés comme positifs alors que les annotateurs peuvent être capables de les identifier comme exprimant une émotion négative.

4.1.4.3 Orthographe et syntaxe particulières

Nous avons repérés deux grands problèmes inhérents aux forums qui réduisent les performances du système :

- *Oubli de ponctuation*. Sans marque de ponctuation forte, comme le *point*, le système récupère les contextes gauche et droit d'un mot pivot même si ces mots n'ont aucune relation syntaxique ou sémantique avec l'émotion exprimée ;
- *Fautes d'orthographe/de frappe*. Si un patient saisit le mot *gène* au lieu de *gêne*, le système ne sera pas capable de détecter qu'il s'agit de l'émotion *gêne*. De la même manière, si le patient écrit *peur* au lieu de *peut*, le système détectera alors la présence d'un marqueur émotionnel négatif. De telles situations peuvent induire du bruit et des silences lors de l'application des règles.

4.1.4.4 Couverture et précision des ressources linguistiques et sémantiques

Nous nous sommes rendus compte de plusieurs difficultés liées à la couverture et à la précision des ressources linguistiques et sémantiques :

- la détection de locutions (ou expressions composées) comme *laisser à désirer* est très faible actuellement faute de ressources nécessaires, alors que cela peut introduire du bruit dans les annotations. Par exemple, *laisser à désirer* qui dénote un état émotionnel négatif, n'est pas annoté, tandis que *désirer*, que notre système sait détecter et annoter, est dénoté comme un état émotionnel positif ;
- malgré les adaptations de ressources linguistiques et sémantiques effectuées, ces ressources contiennent encore des termes émotionnels non pertinents dans un contexte médical. Par exemple, nous avons remarqué que certaines flexions verbales sont à tort annotées comme émotions. Cela concerne essentiellement du verbe *aimer*, qui est un terme émotionnellement positif, tandis que sa forme conditionnelle *aimerais* dénote une incertitude. D'autre part, nous avons repéré d'autres termes comme *trouble* ou *terrible* sont annotés comme étant des manifestations d'émotions positives. Ceci est vrai dans un contexte de la langue générale et figurée mais, dans le contexte médical, *trouble* doit être annoté comme *disorder* :
 - *trouble de la santé*
 - *troubles gastriques*
 tandis que *terrible* doit être annoté comme un marqueur émotionnel négatif :
 - *terribles douleurs*
 - *terrible mal de tête*

4.2 Perspectives

Une des perspectives de notre travail est d'intégrer le système proposé dans une plateforme de traitement et visualisation des forums, comme celle du doctorant en informatique à l'IRIT¹, Ghazar Chahbandarian, qui consiste en l'analyse contextuelle des données de santé à partir des informations présentes dans les forums médicaux. De cette manière, notre contribution peut trouver sa place dans le cadre du projet *Parlons de nous !* à côté de la contribution de (Capeyron, 2014), dont l'objectif a consisté à analyser les citations et leur rôle dans les forums de santé. Dans l'optique d'améliorer la recherche d'informations disponibles dans les forums de santé, la combinaison de ces différents travaux (citations et émotions) offrirait davantage

1. Institut de Recherche et d'Informatique de Toulouse

d'outils pour visualiser les informations pertinentes pour les médecins, spécialistes, . . . De plus, notre travail pourrait être adapté afin d'intégrer d'autres types de plateformes de visualisation de contenus de forum comme *Calico* (voir section 1.2.5.3). Notre travail permettrait d'ajouter une dimension émotionnelle à ces différentes plateformes.

Pour améliorer la qualité de l'analyse fournie par notre système, il serait intéressant d'ajouter des modules supplémentaires ou bien d'ajuster le contenu des ressources afin d'améliorer la détection des émotions et de leurs polarités. Plusieurs pistes existent :

- L'exploitation des étiquettes morpho-syntaxiques peut servir de marqueur d'intensité émotionnelle. Par exemple, si dans un syntagme verbal, nous avons un verbe conjugué à l'imparfait et si son complément d'objet représente une émotion, alors l'émotion exprimée devrait recevoir une intensité moindre : *J'étais comme vous, j'étais **angoissée**, j'avais **peur** de ne plus pouvoir marcher. . .* Dans ces situations, l'auteur (patient) du message partage une expérience passée avec des émotions qui appartiennent au passé. En effet, ces émotions (*peur,angoissée*) ne reflètent pas les émotions exprimées dans le présent donc elles perdent en intensité. Ces émotions font plus l'objet d'une description que d'une expression émotionnelle forte ;
- Il serait également intéressant de prendre en considération d'autres types de marqueurs comme la double négation, comme dans *non sans peine*, qui a pour objet d'annuler l'effet de la négation. La polarité doit donc correspondre à la polarité initiale (négative) et non pas à l'inversion de la polarité (positive) ;
- De manière générale, il faudrait enrichir les ressources linguistiques afin de détecter plus efficacement les émotions et leur polarité. Notons que les travaux de ce type sont en cours de réalisation dans le projet (Abdaoui *et al.*, 2014) grâce à la traduction, avec l'aide d'une traductrice professionnelle, du lexique de d'émotions (Mohammad & Turney, 2010) contenant plus de 20,000 entrées, et à l'ajout de synonymes à ce lexique.

Une autre perspective consiste à prendre en compte des entités nommées représentant les professionnels de la santé. Ceci permettrait d'ajouter un autre type d'information sémantique et de détecter la perception des médecins ou spécialistes par les patients. Cette focalisation peut donner des informations sur les pratiques des professionnels, ce qui permettrait d'obtenir un aperçu des raisons pour lesquelles les patients sont favorables ou non à la consultation de tel ou tel professionnel de santé. Pour effectuer la détection d'entités nommées, nous pouvons utiliser un lexique spécialisé ou bien utiliser les suffixes employés pour la formation de noms d'agents ou de métiers, comme *-ogue, -iste, -ien*

L'identification du genre des patients peut aussi se révéler utile pour les professionnels à des fins d'observations des populations les plus touchés par une maladie, par exemple. De plus, le lien avec des données de géolocalisation peut permettre d'identifier des zones à risques : point de départ d'une pandémie, les populations les plus touchées par un symptôme, . . . (Bertrand *et al.*, 2013).

D'autres perspectives sont liées à la détection de la portée des marqueurs. Par exemple, il pourrait être intéressant de faire varier la fenêtre contextuelle des émotions (actuellement de 7 tokens à gauche et à droite) afin de vérifier s'il est possible d'optimiser le système. Nous pouvons peut-être trouver un meilleur compromis entre la précision du système et son rappel. De plus, il serait intéressant de comparer notre approche avec

l'exploitation des arbres syntaxiques. Pour ceci, l'utilisation d'un analyseurs syntaxique fonctionnant sur le français et traitant les données des médias sociaux est nécessaire.

Enfin, il serait intéressant d'observer la portabilité du système à d'autres domaines comme celui des romans : la dynamique évolutive pourrait alors être basée sur la progression de l'histoire sur la ligne temporelle.

Bibliographie

- ABDAOUI, A., AZÉ, J., BRINGAY, S., PONCELET, P. & GRABAR, N. (2014). Analyse des messages des patients et des médecins dans les fora de santé.
- ATHAR, A. & TEUFEL, S. (2012). Context-Enhanced Citation Sentiment Detection. In *HLT-NAACL*, pp. 597–601 : The Association for Computational Linguistics.
- AUGUSTYN, M., HAMOU, S. B., BLOQUET, G., GOOSSENS, V., LOISEAU, M., RINCK, F. *et al.* (2008). Lexique des affects : constitution de ressources pédagogiques numériques. In *Autour du langage et des langues : perspective pluridisciplinaire, Sélection d'articles du Colloque International des étudiants-chercheurs en didactique des langues et linguistique*.
- BACCIANELLA, S., ESULI, A. & SEBASTIANI, F. (2010). Sentiwordnet 3.0 : An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pp. 2200–2204.
- BAKLIWAL, A., FOSTER, J., VAN DER PUIL, J., O'BRIEN, R., TOUNSI, L. & HUGHES, M. (2013). Sentiment analysis of political tweets : towards an accurate classifier.
- BALAHUR, A. & TANEV, H. (2013). Detecting event-related links and sentiments from social media texts. In *ACL (Conference System Demonstrations)*, pp. 25–30 : Citeseer.
- BASILE, V. & NISSIM, M. (2013). Sentiment analysis on italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 100–107.
- BATRA, S. & RAO, D. (2010). Entity based sentiment analysis on twitter. *Science*, **9**(4), 1–12.
- BATTAÏA, C. (2012). L'analyse de l'émotion dans les forums de santé. *Actes de la conférence conjointe JEP-TALN-RECITAL, RECITAL*, pp. 267–280.
- BERMINGHAM, A. & SMEATON, A. F. (2011). On using twitter to monitor political sentiment and predict election results.
- BERTRAND, K. Z., BIALIK, M., VIRDEE, K., GROS, A. & BAR-YAM, Y. (2013). Sentiment in new york city : A high resolution spatial and temporal view. *arXiv preprint arXiv :1308.5010*.
- BISTA, S. K., NEPAL, S. & PARIS, C. (2014). Multifaceted visualisation of annotated social media data. In *Big Data (BigData Congress), 2014 IEEE International Congress on*, pp. 699–706 : IEEE.
- BLUMENTHAL, P. (2009). Les noms d'émotion : trois systèmes d'ordre. In *NOVAKOVA, I. et TUTIN, A., éditeurs : Le lexique des émotions*, pp. 41–79, Grenoble, France.
- BRINGAY, S., KERGOISIEN, E., POMPIDOR, P. & PONCELET, P. (2014). Identifier la cible des émotions dans les forums de santé.

- BRODY, S. & DIAKOPOULOS, N. (2011). Cooooooooooooooooo!!!!!!!!!!!!!!! : using word lengthening to detect sentiment in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 562–570 : Association for Computational Linguistics.
- BRODY, S. & ELHADAD, N. (2010). An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies : The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 804–812 : Association for Computational Linguistics.
- CAMBRIA, E., HUSSAIN, A., HAVASI, C. & ECKL, C. (2010). SenticSpace : visualizing opinions and sentiments in a multi-dimensional vector space. In *Knowledge-Based and Intelligent Information and Engineering Systems*, pp. 385–393. Springer.
- CAPEYRON, M. (2014). Analyse des citations dans les forums de sante.
- CAPLAN, S. E. & TURNER, J. S. (2007). Bringing theory to research on computer-mediated comforting communication. *Computers in human behavior*, **23**(2), 985–998.
- CERINI, S., COMPAGNONI, V., DEMONTIS, A., FORMENTELLI, M. & GANDINI, G. (2007). Micrownop : A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. *Language resources and linguistic theory : Typology, second language acquisition, English linguistics*, pp. 200–210.
- CHAUVEAU-THOUMELIN, P. & GRABAR, N. (2014). La subjectivité dans le discours médical : sur les traces de l’incertitude et des émotions. In *EGC 2014*, Rennes, France.
- CHETVIORKIN, I. & LOUKACHEVITCH, N. (2013). Evaluating sentiment analysis systems in russian. *ACL 2013*, pp. 12.
- COSNIER, J. (1994). *Psychologie des émotions et des sentiments*. Retz.
- CÔTÉ, R. A. (1996). Répertoire d’anatomopathologie de la snomed internationale, v3. 4. *Université de Sherbrooke, Sherbrooke, Québec*.
- DAS, D., SCHNEIDER, N., CHEN, D. & SMITH, N. A. (2010). Semafor 1.0 : A probabilistic frame-semantic parser. *Language Technologies Institute, School of Computer Science, Carnegie Mellon University*.
- EKMAN, P. (1992). An argument for basic emotions. *Cognition & Emotion*, **6**(3-4), 169–200.
- ELHADAD, N., ZHANG, S., DRISCOLL, P. & BRODY, S. (2014). Characterizing the sublanguage of online breast cancer forums for medications, symptoms, and emotions. In *Proc AMIA Annual Fall Symposium*.
- ESULI, A. & SEBASTIANI, F. (2006). Sentiwordnet : A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pp. 417–422.
- ETTER, D., FERRARO, F., COTTERELL, R., BUZEK, O. & VAN DURME, B. (2013). *Nerit : Named entity recognition for informal text*. Technical report, Technical Report 11, Human Language Technology Center of Excellence, Johns Hopkins University, July.
- FENG, S., KANG, J. S., KUZNETSOVA, P. & CHOI, Y. (2013a). Connotation Lexicon : A Dash of Sentiment Beneath the Surface Meaning. In *ACL (1)*, pp. 1774–1784 : The Association for Computer Linguistics.

- FENG, S., ZHANG, L., LI, B., WANG, D., YU, G. & WONG, K.-F. (2013b). Is twitter a better corpus for measuring sentiment similarity ? In *EMNLP*, pp. 897–902.
- FILLMORE, C. J., JOHNSON, C. R. & PETRUCK, M. R. (2003). Background to framenet. *International journal of lexicography*, **16**(3), 235–250.
- GANU, G., ELHADAD, N. & MARIAN, A. (2009). Beyond the stars : Improving rating predictions using review text content. In *WebDB*.
- GAUDUCHEAU, N. (2008). La communication des émotions dans les échanges médiatisés par ordinateur : bilan et perspectives. *Bulletin de Psychologie*, **61**, **4**, 389–404.
- GLASER, A. & SCHÜTZE, H. (2012). Automatic generation of short informative sentiment summaries. In W. DAELEMANS, M. LAPATA & L. MÀRQUEZ, Eds., *EACL*, pp. 276–285 : The Association for Computer Linguistics.
- GO, A., HUANG, L. & BHAYANI, R. (2009). Twitter sentiment analysis. *Entropy*, **17**.
- GONZÁLEZ-IBÁÑEZ, R., MURESAN, S. & WACHOLDER, N. (2011). Identifying sarcasm in twitter : a closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies : short papers-Volume 2*, pp. 581–586 : Association for Computational Linguistics.
- GRABAR, N. & HAMON, T. (2009). Exploitation of speculation markers to identify the structure of biomedical scientific writing. In *AMIA 2009*, pp. 203–7.
- GREENE, S. & RESNIK, P. (2009). More than words : Syntactic packaging and implicit sentiment. In *Proceedings of human language technologies : The 2009 annual conference of the north american chapter of the association for computational linguistics*, pp. 503–511 : Association for Computational Linguistics.
- HABERNAL, I., PTÁCEK, T. & STEINBERGER, J. (2013). Sentiment analysis in czech social media using supervised machine learning. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 65–74.
- HAMON, T., NAZARENKO, A. *et al.* (2008). Le développement d’une plate-forme pour l’annotation spécialisée de documents web : retour d’expérience. *Traitement Automatique des Langues*, **49**(2), 127–154.
- HANCOCK, J. T., TOMA, C. & ELLISON, N. (2007). The truth about lying in online dating profiles. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 449–452 : ACM.
- HARUECHAIYASAK, C., KONGTHON, A., PALINGOON, P. & TRAKULTAWEEKOON, K. S-sense : A sentiment analysis framework for social media sensing. In *Sixth International Joint Conference on Natural Language Processing*, pp. 6.
- HASEGAWA, T., KAJI, N., YOSHINAGA, N. & TOYODA, M. (2013). Predicting and eliciting addressee’s emotion in online dialogue. In *ACL (1)*, pp. 964–972.
- 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*, pp. 174–181 : Association for Computational Linguistics.

- HUH, J., YETISGEN-YILDIZ, M. & PRATT, W. (2013). Text classification for assisting moderators in online health communities. *Journal of biomedical informatics*, **46**(6), 998–1005.
- IKEDA, D., TAKAMURA, H., RATINOV, L.-A. & OKUMURA, M. (2008). Learning to shift the polarity of words for sentiment classification. In *IJCNLP*, pp. 296–303 : Citeseer.
- JIANG, L., YU, M., ZHOU, M., LIU, X. & ZHAO, T. (2011). Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies-Volume 1*, pp. 151–160 : Association for Computational Linguistics.
- KAMVAR, S. D. & HARRIS, J. (2011). We feel fine and searching the emotional web. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 117–126 : ACM.
- KEMPTER, R., SINTSOVA, V., MUSAT, C. & PU, P. (2014). Emotionwatch : Visualizing fine-grained emotions in event-related tweets.
- KIKUCHI, Y., TAKAMURA, H., OKUMURA, M. & NAKAZAWA, S. (2014). Identifying a demand towards a company in consumer-generated media. In *Computational Linguistics and Intelligent Text Processing*, pp. 153–163. Springer.
- KOULOUMPIS, E., WILSON, T. & MOORE, J. (2011). Twitter sentiment analysis : The good the bad and the omg ! *ICWSM*, **11**, 538–541.
- KUHN, M., CAMPILLOS, M., LETUNIC, I., JENSEN, L. J. & BORK, P. (2010). A side effect resource to capture phenotypic effects of drugs. *Molecular systems biology*, **6**(1).
- KUNNEMAN, F., LIEBRECHT, C. & VAN DEN BOSCH, A. (2014). The (un) predictability of emotional hashtags in twitter.
- LEI, J., RAO, Y., LI, Q., QUAN, X. & WENYIN, L. (2014). Towards building a social emotion detection system for online news. *Future Generation Computer Systems*, **37**, 438–448.
- LEVENSHTAIN, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. In *Soviet physics doklady*, volume 10, pp. 707.
- LI, R., SHI, S., HUANG, H., SU, C. & WANG, T. (2014). A method of polarity computation of chinese sentiment words based on gaussian distribution. In *Computational Linguistics and Intelligent Text Processing*, pp. 53–61. Springer.
- LI, S., LEE, S. Y. M., CHEN, Y., HUANG, C.-R. & ZHOU, G. (2010). Sentiment classification and polarity shifting. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pp. 635–643 : Association for Computational Linguistics.
- LIEBRECHT, C., KUNNEMAN, F. & VAN DEN BOSCH, A. (2013). The perfect solution for detecting sarcasm in tweets# not.
- LIU, B., HU, M. & CHENG, J. (2005). Opinion observer : analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web*, pp. 342–351 : ACM.
- LIU, Y., YU, X., LIU, B. & CHEN, Z. (2014). Sentence-level sentiment analysis in the presence of modalities. *Computational Linguistics and Intelligent Text Processing*, pp. 1–16.

- MATHIEU, Y. Y. (2006). A computational semantic lexicon of french verbs of emotion. In *Computing attitude and affect in text : Theory and applications*, pp. 109–124. Springer.
- MAUREL, S., CURTONI, P. & DINI, L. (2008). L'analyse des sentiments dans les forums. *Atelier Fouille des Données d'Opinion*.
- MELZI, S., ABDAOUI, A., AZÉ, J., BRINGAY, S., PONCELET, P. & GALTIER, F. (2014). Que ressentent les patients ? In *EGC*, pp. 449–454.
- MIHALCEA, R. & STRAPPARAVA, C. (2012). Lyrics, music, and emotions. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 590–599 : Association for Computational Linguistics.
- MITCHELL, M., AGUILAR, J., WILSON, T. & VAN DURME, B. (2013). Open domain targeted sentiment.
- MITTAL, N., AGARWAL, B., CHOUHAN, G., BANIA, N. & PAREEK, P. (2013). Sentiment analysis of hindi review based on negation and discourse relation. In *11th Workshop on Asian Language Resources (ALR), In Conjunction with IJCNLP* : Citeseer.
- MOHAMMAD, S. M., KIRITCHENKO, S. & ZHU, X. (2013). Nrc-canada : Building the state-of-the-art in sentiment analysis of tweets. *arXiv preprint arXiv :1308.6242*.
- MOHAMMAD, S. M. & TURNEY, P. D. (2010). Emotions evoked by common words and phrases : Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pp. 26–34 : Association for Computational Linguistics.
- MOHAMMAD, S. M. & YANG, T. W. (2011). Tracking sentiment in mail : how genders differ on emotional axes. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (ACL-HLT 2011)*, pp. 70–79.
- MORENO-ORTIZ, A., PÉREZ-HERNÁNDEZ, C., DEL-OLMO, M. Á. *et al.* (2013). Managing multiword expressions in a lexicon-based sentiment analysis system for spanish. *NAACL HLT 2013*, **13**, 1.
- NAKAGAWA, T., INUI, K. & KUROHASHI, S. (2010). Dependency tree-based sentiment classification using crfs with hidden variables. In *Human Language Technologies : The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 786–794 : Association for Computational Linguistics.
- NALISNICK, E. T. & BAIRD, H. S. (2013). Character-to-character sentiment analysis in shakespeare's plays.
- NAMER, F. (2009). *Morphologie, Lexique et TAL : l'analyseur DériF. TIC et Sciences cognitives*. London : Hermes Sciences Publishing.
- NARAYANAN, R., LIU, B. & CHOUDHARY, A. (2009). Sentiment analysis of conditional sentences. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing : Volume 1-Volume 1*, pp. 180–189 : Association for Computational Linguistics.
- PAROUBEK, A. P. (2010a). Le microblogage pour la microanalyse des sentiments et des opinions.

- PAROUBEK, A. P. P. (2010b). Construction d'un lexique affectif pour le français à partir de twitter.
- PÉRINET, A., GRABAR, N. & HAMON, T. (2011). Identification des assertions dans les textes médicaux : application à la relation {patient, problème médical}. *TAL*, **52**(1), 97–132.
- PETROVIC, S., OSBORNE, M. & LAVRENKO, V. (2010). The edinburgh twitter corpus. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*, pp. 25–26.
- PIANTA, E., BENTIVOGLI, L. & GIRARDI, C. (2002). Developing an aligned multilingual database. In *Proc. 1st Int'l Conference on Global WordNet*.
- PLUTCHIK, R. (1980). A general psychoevolutionary theory of emotion. *Emotion : Theory, research, and experience*, **1**(3), 3–33.
- PTASZYNSKI, M., MACIEJEWSKI, J., DYBALA, P., RZEPKA, R. & ARAKI, K. (2010). Cao : A fully automatic emoticon analysis system based on theory of kinesics. *Affective Computing, IEEE Transactions on*, **1**(1), 46–59.
- PURVER, M. & BATTERSBY, S. (2012). Experimenting with Distant Supervision for Emotion Classification. In W. DAELEMANS, M. LAPATA & L. MÀRQUEZ, Eds., *EACL*, pp. 482–491 : The Association for Computer Linguistics.
- QADIR, A. & RILOFF, E. (2013). Bootstrapped learning of emotion hashtags# hashtags4you. *WASSA 2013*, pp.2.
- QUAN, C. & REN, F. (2014). Visualizing emotions from chinese blogs by textual emotion analysis and recognition techniques. *International Journal of Information Technology & Decision Making*, pp. 1–20.
- RAMTEKE, A., MALU, A., BHATTACHARYYA, P. & NATH, J. S. (2013). Detecting turnarounds in sentiment analysis : Thwarting. In *ACL (2)*, pp. 860–865 : Citeseer.
- SAYEED, A. B., BOYD-GRABER, J. L., RUSK, B. & WEINBERG, A. (2012). Grammatical structures for word-level sentiment detection. In *HLT-NAACL*, pp. 667–676 : The Association for Computational Linguistics.
- SCHMID, H. (1994). Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the international conference on new methods in language processing*, volume 12, pp. 44–49.
- SCHULER, K. K. (2005). Verbnets : A broad-coverage, comprehensive verb lexicon.
- SMITH, P. & LEE, M. (2012). Cross-discourse development of supervised sentiment analysis in the clinical domain. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pp. 79–83 : Association for Computational Linguistics.
- SOUZA, M. & VIEIRA, R. (2013). Entity-centric sentiment analysis on twitter data for the portuguese language.
- SOUZA, M., VIEIRA, R., Busetti, D., Chishman, R., Alves, I. M. *et al.* (2011). Construction of a portuguese opinion lexicon from multiple resources. *STIL*.
- STRAPPARAVA, C. & MIHALCEA, R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, pp. 1556–1560 : ACM.

- STRAPPARAVA, C. & VALITUTTI, A. (2004). Wordnet affect : an affective extension of wordnet. In *LREC*, volume 4, pp. 1083–1086.
- SYSSAU, A. & FONT, N. (2005). Evaluations des caractéristiques émotionnelles d'un corpus de 604 mots. *Bulletin de psychologie*, (3), 361–367.
- TABOADA, M. & GRIEVE, J. (2004). Analyzing appraisal automatically. In *Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text (AAAI Technical Report SS# 04# 07)*, Stanford University, CA, pp. 158q161. AAAI Press.
- TAKAMURA, H., INUI, T. & OKUMURA, M. (2005). Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 133–140 : Association for Computational Linguistics.
- TAKASE, S., MURAKAMI, A., ENOKI, M., OKAZAKI, N. & INUI, K. (2013). Detecting Chronic Critics Based on Sentiment Polarity and User's Behavior in Social Media. In *ACL (Student Research Workshop)*, pp. 110–116 : The Association for Computer Linguistics.
- THELWALL, M., BUCKLEY, K., PALTOGLOU, G., CAI, D. & KAPPAS, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, **61**(12), 2544–2558.
- THOUMELIN, P. C. & GRABAR, N. (2014). La subjectivité dans le discours médical : sur les traces de l'incertitude et des émotions. In *EGC*, pp. 455–466.
- TOKUHISA, R., INUI, K. & MATSUMOTO, Y. (2008). Emotion classification using massive examples extracted from the web. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pp. 881–888 : Association for Computational Linguistics.
- VELIKOVICH, L., BLAIR-GOLDENSOHN, S., HANNAN, K. & McDONALD, R. (2010). The viability of web-derived polarity lexicons. In *Human Language Technologies : The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 777–785 : Association for Computational Linguistics.
- VINCENT, M. & WINTERSTEIN, G. (2013). Construction et exploitation d'un corpus français pour l'analyse de sentiment.
- VO, B. & COLLIER, N. (2013). Twitter emotion analysis in earthquake situations. *International Journal of Computational Linguistics and Applications*, **4**(1), 159–173.
- VOLKOVA, S., DOLAN, W. B. & WILSON, T. (2012). Clex : a lexicon for exploring color, concept and emotion associations in language. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 306–314 : Association for Computational Linguistics.
- VOLKOVA, S., WILSON, T. & YAROWSKY, D. (2013). Exploring sentiment in social media : Bootstrapping subjectivity clues from multilingual twitter streams. In *ACL (2)*, pp. 505–510.
- VU, H. T., NEUBIG, G., SAKTI, S., TODA, T. & NAKAMURA, S. (2014). Acquiring a dictionary of emotion-provoking events. *EACL 2014*, pp. 128–132.

- WECKER, A. J., LANIR, J., MOKRYN, O., MINKOV, E. & KUFLIK, T. (2014). Semantize : visualizing the sentiment of individual document. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*, pp. 385–386 : ACM.
- WILSON, T., HOFFMANN, P., SOMASUNDARAN, S., KESSLER, J., WIEBE, J., CHOI, Y., CARDIE, C., RILOFF, E. & PATWARDHAN, S. (2005). Opinionfinder : A system for subjectivity analysis. In *Proceedings of hlt/emnlp on interactive demonstrations*, pp. 34–35 : Association for Computational Linguistics.
- WILSON, T., WIEBE, J. & HOFFMANN, P. (2009). Recognizing contextual polarity : An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, **35**(3), 399–433.
- WU, Y., ZHANG, Q., HUANG, X. & WU, L. (2009). Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing : Volume 3-Volume 3*, pp. 1533–1541 : Association for Computational Linguistics.
- YU, H.-H. H. C.-H., LIN, T.-W. C. C.-K. & CHEN, H.-H. (2013). Analyses of the association between discourse relation and sentiment polarity with a chinese human-annotated corpus. *LAW VII & ID*, pp.70.

Annexe A

Les corpus

A.1 Corpus Q/A

A.2 Corpus forum

FIGURE A.1 – Échantillon des énoncés du corpus Q/A

Cher/NAM/Cher/1/0// (/(/(/4/0// e/NAM/e/2/0//))//)/8/0// Patrick/NAM/Patrick/3/0// ,././,12/0//
 Votre/DET:POS/votre/4/0// partenaire/NOM/partenaire/5/0// n/term/ne/14/0/pre-neg/ '/'/'/20/0//
 est/VER:pres/être/7/0// probablement/term/probablement/5/0/pre-wi/ pas/term/pas/7/0/post-neg/
 enceinte/term/enceindre/3/0/disorder/ car/KON/car/11/0// elle/PRO:PER/elle/12/0//
 a/VER:pres/avoir/13/0// pris/VER:pper/prendre/14/0// la/DET:ART/le/15/0//
 pilule/term/pilule/13/0/medication/ du/PRP:det/du/17/0// lendemain/NOM/lendemain/18/0//
 ./SENT./45/0//
 D/PRP/de/19/0// '/'/'/49/0// ailleurs/ADV/ailleurs/20/0// ,././,53/0// ce/PRO:DEM/ce/21/0//
 que/PRO:REL/que/22/0// vous/PRO:PER/vous/23/0// racontez/VER:pres/raconter/24/0// n/term/ne/18/0/pre-
 neg/ '/'/'/65/0// est/VER:pres/être/26/0// pas/term/pas/11/0/post-neg/ en/PRP/en/28/0//
 faveur/NOM/faveur/29/0// d/PRP/de/30/0// '/'/'/77/0// une/DET:ART/un/31/0//
 grossesse/term/grossesse/10/0/disorder/ ./SENT./82/0//
 Vous/PRO:PER/vous/33/0// êtes/VER:pres/être/34/0// d/PRP/de/35/0// '/'/'/90/0// un/DET:ART/un/36/0//
 tempérament/NOM/tempérament/37/0// très/term/très/4/0/modif-p/
 anxieux/term/anxieux/8/0/disorder+emot-peur/ et/KON/et/40/0// cela/PRO:DEM/cela/41/0//
 est/VER:pres/être/42/0// le/DET:ART/le/43/0// plus/term/plus/12/0/modif-p/
 inquiétant/term/inquiétant/6/0/emot-inquietude/ ./SENT./111/0//
 Si/term/si/17/0/pre-wi/ cela/PRO:DEM/cela/47/0// dure/VER:subp/durer/48/0// ,././,119/0//
 vous/PRO:PER/vous/49/0// devrez/VER:futu/devoir/50/0// consulter/VER:infi/consulter/51/0//
 ./SENT./126/0//
 Bien/ADV/bien/52/0// cordialement/ADV/cordialement/53/0// ./SENT./131/0//
 Ceci/PRO:DEM/ceci/54/0// n/term/ne/15/0/pre-neg/ '/'/'/137/0// est/VER:pres/être/56/0//
 pas/term/pas/1/0/post-neg/ une/DET:ART/un/58/0// consultation/term/consultation/9/0/procedure/
 médicale/ADJ/médical/60/0// et/KON/et/61/0// n/term/ne/16/0/pre-neg/ '/'/'/153/0//
 a/VER:pres/avoir/63/0// pas/term/pas/2/0/post-neg/ pour/PRP/pour/65/0// objet/NOM/objet/66/0//
 de/PRP/de/67/0// la/PRO:PER/la/68/0// remplacer/VER:infi/remplacer/69/0// ./SENT./168/0// |

(a) Échantillon de questions

Bonjour/NOM/bonjour/1/0// ,././,4/0// J/PRO:PER/je/2/0// '/'/'/8/0// ai/VER:pres/avoir/3/0//
 eu/VER:pper/avoir/4/0// un/DET:ART/un/5/0// début/NOM/début/6/0// de/PRP/de/7/0// rapport/NOM/rapport/8/0//
 non/term/non/12/0/pre-neg/ protégé/VER:pper/protéger/10/0// il/PRO:PER/il/11/0// y/PRO:PER/y/12/0//
 a/VER:pres/avoir/13/0// une/DET:ART/un/14/0// semaine/NOM/semaine/15/0// pendant/PRP/pendant/16/0//
 plusieurs/PRO:IND/plusieurs/17/0// secondes/NOM/second|seconde/18/0// ,././,42/0// avant/PRP/avant/19/0//
 de/PRP/de/20/0// me/PRO:PER/me/21/0// protéger/VER:infi/protéger/22/0// ./SENT./51/0//
 La/DET:ART/le/23/0// fille/NOM/fille/24/0// m/PRO:PER/me/25/0// '/'/'/59/0// a/VER:pres/avoir/26/0//
 dit/VER:pres/dire/27/0// avoir/VER:infi/avoir/28/0// pris/VER:pper/prendre/29/0// la/DET:ART/le/30/0//
 pilule/term/pilule/15/0/medication/ du/PRP:det/du/32/0// lendemain/NOM/lendemain/33/0// mais/KON/mais/34/0//
 je/PRO:PER/je/35/0// n/term/ne/25/0/pre-neg/ '/'/'/83/0// étais/VER:impf/être/37/0// pas/term/pas/9/0/post-
 neg/ présent/ADJ/présent/39/0// ./SENT./90/0//
 Je/PRO:PER/je/40/0// suis/VER:pres/suivre|être/41/0// une/DET:ART/un/42/0// personne/NOM/personne/43/0//
 très/term/très/1/0/modif-p/ anxieuse/term/anxieux/24/0/disorder+emot-peur/ et/KON/et/46/0//
 j/PRO:PER/je/47/0// '/'/'/108/0// ai/VER:pres/avoir/48/0// peur/term/peur/10/0/disorder+emot-peur/
 depuis/PRP/depuis/50/0// de/PRP/de/51/0// l/PRO:PER/la/le/52/0// '/'/'/120/0// avoir/VER:infi/avoir/53/0//
 mise/VER:pper/mettre/54/0// enceinte/term/enceinte/3/0/disorder/ ./SENT./127/0//
 Je/PRO:PER/je/56/0// lui/PRO:PER/lui/57/0// ai/VER:pres/avoir/58/0// demandé/VER:pper/demander/59/0//
 de/PRP/de/60/0// faire/VER:infi/faire/61/0// un/DET:ART/un/62/0// test/NOM/test/63/0// de/PRP/de/64/0//
 grossesse/term/grossesse/18/0/disorder/ pour/PRP/pour/66/0// être/NOM/être/67/0// sûr/term/sûr/7/0/pre-wi/
 mais/KON/mais/69/0// ça/PRO:DEM/cela/70/0// l/PRO:PER/la/le/71/0// '/'/'/161/0// a/VER:pres/avoir/72/0//
 juste/ADJ/juste/73/0// énervé/term/énervé/4/0/disorder+emot-colere/ et/KON/et/75/0//
 depuis/PRP/depuis/76/0// elle/PRO:PER/elle/77/0// a/VER:pres/avoir/78/0// coupé/VER:pper/couper/79/0//
 tout/PRO:IND/tout/80/0// contact/NOM/contact/81/0// en/PRP/en/82/0// disant/VER:pper/dire/83/0//
 qu/KON/que/84/0// '/'/'/189/0// elle/PRO:PER/elle/85/0// n/term/ne/20/0/pre-neg/ '/'/'/195/0//
 était/VER:impf/être/87/0// pas/term/pas/17/0/post-neg/ enceinte/term/enceinte/2/0/disorder/ et/KON/et/90/0//
 que/KON/que/91/0// même/ADV/même/92/0// si/term/si/19/0/pre-wi/ elle/PRO:PER/elle/94/0// l/PRO:PER/la/
 le/95/0// '/'/'/215/0// était/VER:impf/être/96/0// elle/PRO:PER/elle/97/0// n/term/ne/21/0/pre-neg/
 '/'/'/223/0// en/PRO:PER/en/99/0// voudrait/term/vouloir/100/0/pre-si/ jamais/term/jamais/11/0/pre-neg/
 ./SENT./230/0//
 Je/PRO:PER/je/102/0// n/term/ne/22/0/pre-neg/ '/'/'/236/0// arrive/VER:pres/arriver/104/0//
 plus/term/plus/13/0/modif-p/ à/PRP/à/106/0// dormir/VER:infi/dormir/107/0// ,././,246/0//
 j/PRO:PER/je/108/0// '/'/'/250/0// ai/VER:pres/avoir/109/0// repris/VER:pper/reprendre/110/0//
 une/DET:ART/un/111/0// activité/NOM/activité/112/0// professionnelle/ADJ/professionnel/113/0//
 et/KON/et/114/0// je/PRO:PER/je/115/0// ressens/term/ressentir/23/0/emot-emotion/ que/KON/que/117/0//
 je/PRO:PER/je/118/0// suis/VER:pres/suivre|être/119/0// en/PRP/en/120/0// train/NOM/train/121/0//
 de/PRP/de/122/0// partir/VER:infi/partir/123/0// en/PRP/en/124/0//
 dépression/term/dépression/6/0/disorder+emot-tristesse/ ,././,286/0// j/PRO:PER/je/126/0// '/'/'/290/0//
 ai/VER:pres/avoir/127/0// besoin/NOM/besoin/128/0// d/PRP/de/129/0// '/'/'/298/0// être/VER:infi/être/130/0//
 rassuré/term/rassurer/16/0/emot-joie/ sur/PRP/sur/132/0// le/DET:ART/le/133/0// fait/NOM/fait/134/0//
 qu/PRO:REL/que/135/0// '/'/'/312/0// elle/PRO:PER/elle/136/0// ne/term/ne/14/0/pre-neg/
 soit/VER:subp/être/138/0// pas/term/pas/8/0/post-neg/ enceinte/term/enceinter/5/0/disorder/ ./SENT./323/0//

(b) Échantillon de réponses

FIGURE A.2 – Échantillon des énoncés du corpus forum

Athroplastie/term/arthroplastie/7/0/procedure_orth/ lombaire/ADJ/lombaire/2/0// ./SENT./3/0//
 -/-/5/0// -/-/7/0// -/-/8/0// -/-/9/0// -/-/10/0// -/-/11/0// -/-/12/0// -/-/13/0// -/-/14/0// -/-/15/0//
 -/-/16/0// -/-/17/0// -/-/18/0// -/-/19/0// -/-/20/0// -/-/21/0// -/-/22/0// -/-/23/0// -/-/24/0// -/-/25/0//
 -/-/26/0// -/-/27/0// -/-/28/0// -/-/29/0// -/-/30/0// -/-/31/0// -/-/32/0// -/-/33/0// -/-/34/0// -/-/35/0//
 -/-/36/0// -/-/37/0// -/-/38/0// -/-/39/0// -/-/40/0// -/-/41/0// -/-/42/0// -/-/43/0// -/-/44/0// -/-/45/0//
 -/-/46/0// -/-/47/0// -/-/48/0// -/-/49/0// -/-/50/0// -/-/51/0// -/-/52/0// -/-/53/0// -/-/54/0// -/-/55/0//
 -/-/56/0// -/-/57/0// -/-/58/0// -/-/59/0// -/-/60/0// -/-/61/0// -/-/62/0// -/-/63/0// -/-/64/0// -/-/65/0//
 -/-/66/0// -/-/67/0// -/-/68/0// -/-/69/0// -/-/70/0// -/-/71/0// -/-/72/0// -/-/73/0// -/-/74/0// -/-/75/0//
 ./SENT./76/0//
 Bonjour/NOM/bonjour/3/0// je/PRO:PER/je/4/0// recherche/VER:pres/rechercher/5/0// un/DET:ART/un/6/0//
 chirurgien/NOM/chirurgien/7/0// ortho/ADJ/ortho/8/0// pour/PRP/pour/9/0// pratiquer/VER:infi/pratiquer/10/0// une/DET:ART/un/11/0//
 arthroplastie/term/arthroplastie/2/0/procedure_orth/ lombaire/ADJ/lombaire/13/0// sur/PRP/sur/14/0// Paris/NOM/Paris/15/0//
 ou/KON/ou/16/0// région/NOM/région/17/0// parisienne/ADJ/parisien/18/0// de/PRP/de/19/0// préférence/NOM/préférence/20/0//
 a/VER:pres/avoir/21/0// l/NUM/L/22/0// hôpital/NOM/hôpital/23/0// ou/KON/ou/24/0// en/PRP/en/25/0// clinique/NOM/clinique/26/0//
 en/PRP/en/27/0// dernier/ADJ/dernier/28/0// recours/VER:impe/recourir/29/0// car/KON/car/30/0// payant/VER:ppre/payer|payer/31/0//
 il/PRO:PER/il/32/0// me/PRO:PER/me/33/0// semble/term/sembler/5/0/pre-wi// ./SENT./141/0//
 Merci/NOM/merci/35/0// a/VER:pres/avoir/36/0// tous/PRO:IND/tout/37/0// ./SENT./147/0//
 =/=/149/0// =/=/150/0// =/=/151/0// =/=/152/0// =/=/153/0// =/=/154/0// =/=/155/0// =/=/156/0// =/=/157/0//
 =/=/158/0// =/=/159/0// =/=/160/0// =/=/161/0// =/=/162/0// =/=/163/0// =/=/164/0// =/=/165/0// =/=/166/0//
 =/=/167/0// =/=/168/0// =/=/169/0// =/=/170/0// =/=/171/0// =/=/172/0// =/=/173/0// =/=/174/0// =/=/175/0//
 =/=/176/0// =/=/177/0// =/=/178/0// =/=/179/0// =/=/180/0// =/=/181/0// =/=/182/0// =/=/183/0// =/=/184/0//
 =/=/185/0// =/=/186/0// =/=/187/0// =/=/188/0// =/=/189/0// =/=/190/0// =/=/191/0// =/=/192/0// =/=/193/0//
 =/=/194/0// =/=/195/0// =/=/196/0// =/=/197/0// =/=/198/0// =/=/199/0// =/=/200/0// =/=/201/0// =/=/202/0//
 =/=/203/0// =/=/204/0// =/=/205/0// =/=/206/0// =/=/207/0// =/=/208/0// =/=/209/0// =/=/210/0// =/=/211/0//
 =/=/212/0// =/=/213/0// =/=/214/0// =/=/215/0// =/=/216/0// =/=/217/0// =/=/218/0// =/=/219/0// =/SENT/=220/0//

(a) Échantillon de questions

Bonjour/NOM/bonjour/38/0// ./././224/0// je/PRO:PER/je/39/0// me/PRO:PER/me/40/0// suis/VER:pres/suivre|être/41/0//
 fait/VER:pres/faire/42/0// opérer/VER:infi/opérer/43/0// à/PRP/à/44/0// l/PRO:PER/la|le/45/0// './././240/0//
 Hopital/PRO:POS/Hopital/46/0// Atoine/PRO:POS/Atoine/47/0// Bèclère/PRO:POS/Bèclère/48/0// de/PRP/de/49/0//
 Clamart/PRO:POS/Clamart/50/0// d/PRP/de/51/0// './././254/0// une/DET:ART/un/52/0// arthroplastie/term/arthroplastie/1/0/procedure/orth/
 discale/ADJ/discal/54/0// par/PRP/par/55/0// un/DET:ART/un/56/0// ortho/ADJ/ortho/57/0// qui/PRO:REL/qui/58/0//
 me/PRO:PER/me/59/0// semble/term/sembler/9/0/pre-wi// est/VER:pres/être/61/0// super/term/super/6/0/modif-p-ps/ ./././278/0//
 opérée/PRO:POS/opérée/63/0// le/DET:ART/le/64/0// mois/NOM/mois/mois/65/0// dernier/ADJ/dernier/66/0// et/KON/et/67/0//
 pour/PRP/pour/68/0// l/DET:ART/le/69/0// './././294/0// instant/NOM/instant/70/0// tout/ADV/tout/71/0// va/VER:pres/aller/72/0//
 pour/KON/pour/73/0// le/PRO:PER/le/74/0// mieux/ADV/mieux/75/0// './././308/0// aucun/term/aucun/8/0/pre-neg/ DP/PRO:POS/DP/77/0//
 ne/term/ne/4/0/pre-neg/ ma/DET:POS/mon/79/0// été/NOM/été/80/0// demandé/VER:ppre/demander/81/0// ./SENT./321/0//
 Bon/ADJ/bon/82/0// courage/NOM/courage/83/0// ./SENT./326/0//
 =/=/328/0// =/=/329/0// =/=/330/0// =/=/331/0// =/=/332/0// =/=/333/0// =/=/334/0// =/=/335/0// =/=/336/0//
 =/=/337/0// =/=/338/0// =/=/339/0// =/=/340/0// =/=/341/0// =/=/342/0// =/=/343/0// =/=/344/0// =/=/345/0//
 =/=/346/0// =/=/347/0// =/=/348/0// =/=/349/0// =/=/350/0// =/=/351/0// =/=/352/0// =/=/353/0// =/=/354/0//
 =/=/355/0// =/=/356/0// =/=/357/0// =/=/358/0// =/=/359/0// =/=/360/0// =/=/361/0// =/=/362/0// =/=/363/0//
 =/=/364/0// =/=/365/0// =/=/366/0// =/=/367/0// =/=/368/0// =/=/369/0// =/=/370/0// =/=/371/0// =/=/372/0//
 =/=/373/0// =/=/374/0// =/=/375/0// =/=/376/0// =/=/377/0// =/=/378/0// =/=/379/0// =/=/380/0// =/=/381/0//
 =/=/382/0// =/=/383/0// =/=/384/0// =/=/385/0// =/=/386/0// =/=/387/0// =/=/388/0// =/=/389/0// =/=/390/0//
 =/=/391/0// =/=/392/0// =/=/393/0// =/=/394/0// =/=/395/0// =/=/396/0// =/=/397/0// =/=/398/0// =/SENT/=399/0//
 bonjour/NOM/bonjour/84/0// ./SENT./402/0//
 c/PRO:DEM/ce/85/0// './././406/0// est/VER:pres/être/86/0// quoi/PRO/quoi/87/0// exactement/ADV/exactement/88/0// qu/KON/que/89/0//
 './././416/0// il/PRO:PER/il/90/0// t/PRO:PER/te/91/0// './././422/0// a/VER:pres/avoir/92/0// fait/VER:ppre/faire/93/0//
 ?/term/?/427/0/post-si/
 une/DET:ART/un/94/0// prothèse/NOM/prothèse/95/0// discale/ADJ/discal/96/0// ?/term/?/434/0/post-si/
 sur/PRP/sur/97/0// quelle/PRO:REL/quel/98/0// vertèbre/PRO:POS/vertèbre/99/0// ?/term/?/441/0/post-si/
 comment/ADV/comment/100/0// est/VER:pres/être/101/0// ta/DET:POS/ton/102/0// convalescence/NOM/convalescence/103/0//
 tu/PRO:PER/tu/104/0// as/VER:pres/avoir/105/0// mal/ADV/mal/106/0// ?/term/?/456/0/post-si/
 tu/PRO:PER/tu/107/0// peux/term/pouvoir/10/0/pre-si/ me/PRO:PER/me/109/0// raconter/VER:infi/raconter/110/0// un/DET:ART/un/111/0//
 peu/term/peu/3/0/modif-m/ ton/DET:POS/ton/113/0// parcours/NOM/parcours/114/0// ?/term/?/473/0/post-si/
 merci/NOM/merci/115/0// de/PRP/de/116/0// flolero/PRO:POS/flolero/117/0// ./SENT./479/0//
 =/=/481/0// =/=/482/0// =/=/483/0// =/=/484/0// =/=/485/0// =/=/486/0// =/=/487/0// =/=/488/0// =/=/489/0//
 =/=/490/0// =/=/491/0// =/=/492/0// =/=/493/0// =/=/494/0// =/=/495/0// =/=/496/0// =/=/497/0// =/=/498/0//
 =/=/499/0// =/=/500/0// =/=/501/0// =/=/502/0// =/=/503/0// =/=/504/0// =/=/505/0// =/=/506/0// =/=/507/0//
 =/=/508/0// =/=/509/0// =/=/510/0// =/=/511/0// =/=/512/0// =/=/513/0// =/=/514/0// =/=/515/0// =/=/516/0//
 =/=/517/0// =/=/518/0// =/=/519/0// =/=/520/0// =/=/521/0// =/=/522/0// =/=/523/0// =/=/524/0// =/=/525/0//
 =/=/526/0// =/=/527/0// =/=/528/0// =/=/529/0// =/=/530/0// =/=/531/0// =/=/532/0// =/=/533/0// =/=/534/0//
 =/=/535/0// =/=/536/0// =/=/537/0// =/=/538/0// =/=/539/0// =/=/540/0// =/=/541/0// =/=/542/0// =/=/543/0//
 =/=/544/0// =/=/545/0// =/=/546/0// =/=/547/0// =/=/548/0// =/=/549/0// =/=/550/0// =/=/551/0// =/SENT/=552/0//

(b) Échantillon de réponses

Annexe B

Les ressources

B.1 Ressources sémantiques

FIGURE B.1 – Échantillons des ressources sémantiques (marqueurs de type *médical*)

```
# AJOUTS
envie de vomir : envie de vomir : disorder
envie d'uriner : envie d'uriner : disorder
examen : examen : procedure
examen anatomopathologique : examen anatomopathologique : procedure
examen endoscopique : examen endoscopique : procedure
# SNOMED
symptôme ménopausique : : disorder
symptôme musculosquelettique : : disorder
symptôme nasal : : disorder
symptôme pancréatique : : disorder
antihypertenseur : : medication
antihypertenseur central : : medication
anti-infectieux local : : medication
anti-infectieux pour la peau : : medication
amputation par la main : : procedure
amputation partielle du pénis : : procedure
amputation radicale : : procedure
chirurgie pour acné : : procedure
chirurgie pour strabisme : : procedure
chirurgie reconstructive sur la cornée : : procedure
```

B.2 Ressources linguistiques

B.2.1 Les descripteurs

FIGURE B.2 – Échantillons des ressources linguistiques : échantillons des marqueurs de type *négation*, *incertitude* et *modifieur*

```
# marqueurs de négation
lacune||pre-neg
n' ||pre-neg
ne||pre-neg
ni||pre-neg
# marqueurs d'incertitude forte
eventuellement||pre-si
eventuel||post-si
possible||post-si
pourrait||pre-si
# marqueurs d'incertitude faible
probablement||pre-wi
probable||post-wi
en cas de||pre-wi
# marqueurs de type modifieur fort
énormément||modif-p
pas mal de||modif-p
tellement||modif-p
# marqueurs de type modifieur faible
moins||modif-m
à peine||modif-m
```

B.3 Les émotions

B.3.1 Les marqueurs émotionnels

FIGURE B.3 – Échantillons des ressources linguistiques : échantillons des marqueurs de type *émotion*

```
effrayer : : emot-peur
effrayé : : emot-peur
#emballer : : emot-attirance+emot-joie
#emballé : : emot-joie
embarrasser : : emot-inquietude+emot-gene
embarrassé : : emot-peur
ras le bol : : emot-colere
#rasant : : emot-tristesse
#raser : : emot-lassitude
#rasoir : : emot-tristesse
rassurant : : emot-joie
rassurer : : emot-soulagement
rassuré : : emot-joie
```

B.3.2 L'association polarité/catégorie émotionnelle

FIGURE B.4 – Échantillons des ressources linguistiques : associations des catégories émotionnelles avec une polarité (*pos* = *positive*, *neg* = *négative* et *neu* = *neutre*)

pos		emot-admiration
pos		emot-adoration
neu		emot-anticipation
pos		emot-attirance
neu		emot-autre
pos		emot-bienveillance
neg		emot-colere
pos		emot-convoitise
neg		emot-deception
neg		emot-degout
neg		emot-doute
neu		emot-emoi
neu		emot-emotion
neu		emot-etonnement
pos		emot-fierte
neg		emot-gene
neg		emot-honte
neg		emot-hostilite
neg		emot-indifference
neg		emot-inquietude
neg		emot-insatisfaction
neg		emot-jalousie
pos		emot-joie
neg		emot-lassitude
neg		emot-mefiance
neg		emot-mepris
neg		emot-neg
neu		emot-neu
pos		emot-pos
neg		emot-peur
neg		emot-pitie
neg		emot-regret
pos		emot-respect
neg		emot-souffrance
pos		emot-soulagement
neu		emot-surprise
neg		emot-tristesse
pos		emot-abrev
pos		emot-ono
neu		emot-repeat
pos		emot-smileypos
neg		emot-smileyneg
neu		emot-combine
neu		emot-exclamation

Annexe C

Ajustements des ressources

C.1 Ressources sémantiques

FIGURE C.1 – Ajouts de marqueurs de type *médicaux*

```
#####TERMES#####  
tousser : : disorder  
fou : : disorder  
cancer de la gorge : : disorder  
sentir mal : : disorder  
remontées gastriques : : disorder  
cachats épais : : disorder  
gastro : gastro-entérite : disorder  
hypocondriaque : : disorder  
mythomane : : disorder
```

FIGURE C.2 – Ajouts de marqueurs de type *médicaux* à partir des terminologies médicales : *-ose, -ite, -algie,...*

abdoplastie, acromioplastie, adénocarcinome, adenofibrome, adénofibrome, adénolymphite
 adenome, adénome, Adénome, adénomes, adenopathie, adenopathie
 alexandrite, alitose, aminiocentèse, aminocentèse, aménorrhée, amniocentèse
 analyse, Analyse, analyses, Analyses, ANALYSES, angioliopome
 angiome, angiomes, anite, annalyse, annexectomie, antitransglutaminase
 apindicite, arteriographie, arthrite, arthroscopie, atticotomie, bcgite
 biométrie, blenorragie, brochite, bulbite, byopsie, carcinome
 cardite, carsinome, célioscopie, celulite, cervicarthrose, cholecystectomie
 cholesteatome, chondropathie, cimentoplastie, cystite, coagulase, coeloscopie
 colioscopie, colopathie, condropathie, condymome, coxopathie, depilepsie
 discarthose, discarthrose, dishidrose, drépanocytose, échoendoscopie, échographie
 Echographie, Échographie, ecographie, écographie, électrolyse, ematome
 endometriose, endrométrie, enterite, epathite, épatite, épendinome
 Epidermose, epididymite, épilepsie, episotomie, epycondilite, érythromelalgie
 ésteroscopie, ethmoidite, ethmoïdite, fascite, fibrome, fibromes
 fistulographie, flébite, fybromalgie, gastroentérite, gastro-entérite, gastroentérite
 hapathite, hartrite, hémangiome, hémangiomes, hematocrite, hématocrite
 Hématocrite, hematome, Hematrocrite, hématurie, hémocromatose, hépaite
 hépatite, Hepatite, hépate, Histérectomie, homéopathe, homeopathie
 homéopathie, hypersensibilité, hysterectomie, Hysterectomie, hystérogographie, hystérogographie
 hysteroscopie, Idiopathe, impindicite, integrite, invalidite, istérosocopie
 keratose, kérostose, kystectomie, labirintite, labyrintite, laparotomie
 lechographie, léchographie, lhepatite, ligamentoplastie, ligementoplastie, lipome
 lipomes, listeriose, litériose, lombalgies, lymphangiome, lymphogranulomatose
 lymphome, lymphoréticulose, macroadénome, maelyse, manométrie, mastose
 maxiliase, melanome, mélanome, mélanomes, mellanome, ménigite
 méningiome, meningite, menopause, mésothéliome, métastase, métastase
 métastases, micose, microbiopsie, miopathie, mononucleose, mononucléose
 mononucléose, mucose, mychose, myélome, nephrectomie, nephrite
 neuralgie, nevrite, nevrose, obesite, oesogastroduodenoscopie, orchi-épididymite
 ostéochodrite, osteochondrite, ostéome, osteopathe, ostéopathe, ostéopathes
 osteopathie, osteophytose, ostéophytose, osteoporose, ostéosarcome, Osteopathe
 osthéopathe, osthéopathie, ostrodose, ottite, pancreatite, papillome
 Papillome, paradontite, paraphlébite, peniscopie, peritonite, pharingite
 pheochromocytome, phéochromocytome, phibriscopie, phlebite, phmétrie, pHmétrie
 phosphatase, phybrosopie, pièlonéphrite, pilosite, poliomyélite, polygraphie
 polynec=vrite, pyelonéphrite, pyèlonéphrite, pyelonéphrite, rachianesthésie, rhizarthrose
 rhynopharigite, rigidite, sarcoidose, sarcome, schwannome, schwanome
 sclerose, segmentectomie, sendrome, septoplastie, Serrapeptase, sigmoidite
 simptome, sinistrose, slérose, spondylophytose, steatose, stenose
 stétopose, surdite, symphome, symptome, syptome, synchondrose
 syntrome, synusite, tandinite, ténocynovite, tetralyse, thyroidectomie
 thyroidite, tracheite, trapézectomie, tréponematose, tymanoplastie, tympanoplastie
 uncarthrose, vaginose, virose, vitréorétinopathie, ytomie

FIGURE C.3 – Filtrage des termes non *médicaux* composés des terminologies médicales : *-ose, -ite, -algie...*

chose, rose, prope, pose, close, parasite, suite
 Suite, droite, maudite, ermite, visite, vite, ensuite
 bénite, marmite, carmélite, théocrite, truite, barnabite, benedicite
 ite, traite, retraite, israélite, israélite, fuite, guérite
 copie, écurie, nostalgie, antipathie, photographie, phrase, Viédase
 vase, phase, base, extase, "droit|droite", majordome, Théocrite
 Benedicite, grase, paraphrase, emphase, antonomase, please, cease
 case, pégase, périphrase, calligraphie, tome, Marguerite, marguerite
 Plaudite, Rome, poursuite, glose, marguerite, limite, hétéroclite
 physionomie, élite, moscovite, misanthropie, idiome, économie, conduite
 satellite, mérite, réussite, Tacite, tacite, favorite, rite
 atome, Sinite, sunite, astronome, Favourite, apothéose, quirite
 jésuite, faillite, pénurie, choose, purpose, site, seigneurie
 entrefaite, hypocrite, termite, moite=> "0",, Allouaite, Despite, Some
 come, géométrie, loose, biographie, composite, hippodrome, cénobite
 défaite, grandiose, Larose, jacobite, illicite, utopie, axiome
 Roumanie, économe, sybarite, emphytéose, métalloïde, rogome, métamorphose
 cosmopolite, géographie, stalactite, acite, redite, Sodome, sodome
 Germanie, virtuose, Percose, amphitrite, Amphitrite, orbite, dose
 préadamite, Disome, licite, hermite, explicite, Chantefleurie, Éthiopie
 démérite, gnome, lèchefrite, Ecnome, activite, necessite, "disque.Petite"
 hésite, monographie, tite, nessecite, autonome, evite, hesite
 Ensuite, reduite, "20llose", ésite, boite, quelquechose, diplôme
 website, qqchose, autonomie, bicose, ergonomie, assome, white
 increase, disease, souhaite, faite, Securite, securite, habite
 repose, ptite, reussite, inegalite, efficacase, Jérôme, Jerome
 jérôme, jerome, quelquechose, intensite, ose, shouaite, some
 consomme, pase, doite, drome, Tite, arobase, verite
 realite, home, soite, Maite, petite, globalite, revisite
 mobilite, évite, pornographie, nécessite, anfaite, cuase, nhesite
 serenite, précite, spécialite, chjose, ontraite, traite, ensuite
 zerolimitite, jose, Evite, croite, fantome, nome, peite
 prescrite, motricite, discraite, desuite, consite, unepetite, snowwhite
 pitite, retaite, esite, desuite, essite, Desease, Petite
 gymnase, profite, solidarite, suppose, merrite, welcome, prescrite
 ecrase, écrase, peite, possibilite, motricite, autrechose, brownbase
 propose, sexualite, insolite, précose, precose, jhabite, possibilite
 exhaustivite, efficacite, "09ensuite", nose, inquite, persite, immunité
 quantite, acidite, peetite, sensibilite, cavite, dépose, flexibilite
 sollicite, fertilite, cose, bite, frite, enceite, Laurie
 Exsite, specificite, nececite, Sexualite, sexualite, grose, éficase
 elasticite, minquite, chromosome, acudite, curiosite, possiblite, encenite
 ensuite, nécessite, "cuit|cuite", kome, capacite, exite, dépromé
 Rose, laurie, lucite, méérite, Lederhose, lederhose, monsité
 Pamellarose, quelchose, peopose, interdite, déboite, anatomie, Maëlyse
 refuge, iconoclaste

C.2 Ressources linguistiques

FIGURE C.4 – Ajouts de marqueurs de type *émotion*, *incertitude*, *modifieur*

#####EMOTIONS#####	
pire : : emot-deception	
#bizarre : : emot-inquietude	##### NEGATION#####
dispute : : emot-colere	cesser pre-neg
disputer : : emot-colere	cesse pre-neg
espérer : : emot-joie	empêcher pre-neg
espère : : emot-joie	empêche pre-neg
espéré : : emot-joie	##### INCERTITUDE#####
espoir : : emot-joie	croire pre-wi
valoir la peine : : emot-joie	impression post-wi
valoir le coup : : emot-joie	penser pre-wi
mal à l'aise : : emot-gene	ressemble pre-wi
bravo : : emot-joie	ressembler pre-wi
clocher : : emot-inquietude	##### MODIFIEURS#####
fabuleux : : emot-admiration	partiellement modif-m
enfer : : emot-souffrance	atrocement modif-p
malheur : : emot-tristesse	bcp modif-p
confiance : : emot-admiration	gros modif-p
faire du bien : : emot-soulagement	grosse modif-p
faire du mal : : emot-tristesse	grand modif-p
dommage : : emot-deception	grande modif-p
aberrant : : emot-deception	hyper modif-p
juger : : emot-mepris	plein de modif-p
avoir marre : : emot-lassitude	pleine de modif-p
	souvent modif-p
	super modif-p
	totaletement modif-p

(a) Marqueurs émotionnels

(b) Marqueurs de négation, d'incertitude et de type modifieur

Annexe D

Le système à base de règles

D.1 Pré-traitement des contextes émotionnels/incertitude : mise à jour des frontières contextuelles des marqueurs émotionnels et d'incertitude en fonction de la présence de connecteurs

FIGURE D.1 – Restriction des contextes émotionnels et d'incertitude droit et gauche en fonction de la présence de connecteurs dans ces contextes (1)

```
# Rupture du contexte si présence de connecteurs
sub get_connector_bunch {
  » my ($contexts, $lemmas) = @; # récupération des informations sémantiques et des lemmes dans les contextes
  » gauche et droit du marqueur émotionnel analysé
  » my @emotions = split(/\^\^\^\^/, $lemmas); # stocke dans une liste les lemmes des contextes gauche, droit et
  » du marqueur analysé
  » my $emotion = $emotions[1];
  » my @el = split(/\*/, $emotions[0]); # stocke dans une liste les lemmes du contexte gauche
  » @el = reverse(@el); # inversion de l'ordre des lemmes dans la liste pour parcourir le
  » contexte gauche à partir du lemme le plus proche du marqueur analysé
  » my @er = split(/\*/, $emotions[2]) unless (scalar(@emotions) < 3); # stocke dans une liste les lemmes du
  » contexte droit, s'il existe
  » my @windows = split(/\^\^\^\^/, $contexts); # stocke dans une liste les informations sémantiques des contextes
  » gauche, droit et du marqueur analysé
  » my @wl = split(/\*/, $windows[0]); # stocke dans une liste les informations sémantiques du contexte gauche
  » @wl = reverse(@wl); # inversion de l'ordre des informations sémantiques dans la liste pour
  » parcourir le contexte gauche à partir du lemme le plus proche du marqueur analysé
  » my @wr = split(/\*/, $windows[2]) unless (scalar(@windows) < 3); # stocke dans une liste les informations
  » sémantiques du contexte droit, s'il existe
  » my (@new_contexts_l, @new_contexts_r); # construction des nouveaux contextes gauche et droit pour les
  » informations sémantiques
  » my (@new_lemmas_l, @new_lemmas_r); # pareil pour les lemmes
  » my %connectors = ('car' => 1, 'donc' => 1, 'et' => 1, 'mais' => 1, 'ou' => 1, 'cependant' => 1);
  » my $break_l = 0; # vaut 1 en cas de présence d'un connecteur dans le contexte gauche, sinon 0
  » my $break_r = 0; # vaut 1 en cas de présence d'un connecteur dans le contexte droit, sinon 0
  » for(my $i = 0; $i <= 6; $i++) {
  »     if (exists $el[$i]) {
  »         if (!exists $connectors{lc($el[$i])} && !$break_l) {
  »             push(@new_contexts_l, $wl[$i]);
  »             push(@new_lemmas_l, $el[$i]);
  »         } # (1) on conserve les contextes tant qu'on ne rencontre pas de connecteur
  »         else {
  »             if (!$break_l) {
  »                 $break_l = 1;
  »                 push(@new_contexts_l, $wl[$i]);
  »                 push(@new_lemmas_l, $el[$i]);
  »             }
  »         } # (2) on définit la nouvelle fenêtre contextuelle en cas de présence d'un connecteur
  »     } # on parcourt le contexte gauche
  »     if (exists $er[$i + 1]) {
  »         if (!exists $connectors{lc($er[$i + 1])} && !$break_r) {
  »             push(@new_contexts_r, $wr[$i + 1]);
  »             push(@new_lemmas_r, $er[$i + 1]);
  »         } # même étape qu'en (1)
  »     } else {
  »         }
  »     }
  » }
}
```

FIGURE D.2 – Restriction des contextes émotionnels et d’incertitude droit et gauche en fonction de la présence de connecteurs dans ces contextes (2)

```

»     »     »     »     if (!$break_r) {
»     »     »     »     $break_r = 1;
»     »     »     »     push(@new_contexts_r, $wr[$i + 1]);
»     »     »     »     push(@new_lemmas_r, $er[$i + 1]);
»     »     »     »     }
»     »     »     } # même étape qu'en (2)
»     } # on parcourt le contexte droit
» } # on parcourt les fenêtres gauche et droite du marqueur analysé pour vérifier la présence d'un connecteur
» # reconstruction des contextes (lemmes et informations sémantiques) selon la présence d'un connecteur
» my $return_contexts = join(" ", reverse(@new_contexts_l))."^^^^$windows[1]^^^".join(" ", @new_contexts_r);
» my $return_lemmas = join(" ", reverse(@new_lemmas_l))."^^^^$emotions[1]^^^".join(" ", @new_lemmas_r);
» $return_contexts =~ s/\^\^\^\^\*$/;
» $return_lemmas =~ s/\^\^\^\^\*$/;
» return $return_contexts, $return_lemmas; # renvoie les nouveaux contextes en fonction de la présence de
» connecteur
}

```

D.2 Traitement de la négation

FIGURE D.3 – Traitement de la présence de marqueurs de négation dans les contextes des marqueurs émotionnels ou d’incertitude (1)

```

# Application des règles pour la négation
sub apply_negation_rules {
  >> my $contexts = shift; # récupération des contextes gauche et droit du marqueur émotionnel ou d'incertitude
  >>   analysé, représenté par les étiquettes sémantiques (pre-neg, modif-m, ...)
  >> my $lemmas = shift; # récupération des contextes avec les lemmes
  >> my $polarity = shift; # récupération de la polarité initiale du marqueur émotionnel ou d'incertitude
  >> my $return; # vaut 0 si pas de prise en compte de la négation, 1 si on la prend en compte, 2 pour
  >>   négativer une émotion neutre
  >> my $inc; # stocke le marqueur émotionnel ou d'incertitude
  >> my @windows = split(/\\^\\^/, $contexts); # permet de récupérer le marqueur dans $window[1] pour un post-
  >>   traitement
  >> if ($contexts =~ m/(?post-neg\\*)(?:[\\^\\^*]+\\*){0,3}?(pre-neg\\*)(?:[\\^\\^*]+\\*){0,4}\\^\\^\\^([\\^\\^*]+)/i) {
  >>   >> $return = 0;
  >>   >> $inc = $1;
  >> } # si un "pas" est antérieur à une négation de type "pre-neg" dans une fenêtre de 4 tokens dans le contexte
  >>   gauche, la polarité n'est pas inversée
  >> elsif ($lemmas =~ m/(?\\*|^)(?:ne|n'|n)\\*(?:[\\^\\^*]+\\*){0,4}\\^\\^\\^([\\^\\^*]+\\*){0,4}\\^\\^\\^\\*(?:\\.+?\\*){0,4}?(jamais|
  >>   plus|aucun(?:e)?|ni|rien)\\*/i) {
  >>   >> $return = 1;
  >>   >> $inc = $1;
  >> } # inversion de la polarité si on a : ne/n/n'...[emotion]...jamais/ni/rien/aucun/plus
  >> elsif ($lemmas =~ m/(?\\*|^)(?:ne|n'|n)\\*(?:[\\^\\^*]+\\*){0,4}?(jamais|plus|aucun(?:e)?|ni|rien)\\*(?:
  >>   [\\^\\^*]+\\*){0,4}\\^\\^\\^([\\^\\^*]+)/i) {
  >>   >> $return = 2;
  >>   >> $inc = $windows[1];
  >> } # inversion de la polarité si on a : ne/n/n'...jamais/ni/rien/aucun/plus...[emotion]
  >> elsif ($contexts =~ m/(?pre-neg\\*)(?:[\\^\\^*]+\\*)*(?:post-neg\\*)(?:[\\^\\^*]+\\*){0,4}\\^\\^\\^([\\^\\^*]+)/i) {
  >>   >> $return = 2;
  >>   >> $inc = $windows[1];
  >> } # inversion de la polarité si on a : ne/n/n'...pas...[emotion]
  >> elsif ($contexts =~ m/(?pre-neg\\*)(?:[\\^\\^*]+\\*)*(?:[\\^\\^*]+\\*){0,4}\\^\\^\\^([\\^\\^*]+)\\^\\^\\^\\*(?:\\.+?\\*){0,4}
  >>   (?:post-neg)/i) {
  >>   >> $return = 1;
  >>   >> $inc = $1;
  >> } # inversion de la polarité si on a : ne/n/n'...[emotion]...pas
  >> elsif ($contexts =~ m/(?post-neg\\*)(?:[\\^\\^*]+\\*){0,4}\\^\\^\\^([\\^\\^*]+)/i) {
  >>   >> $return = 2;
  >>   >> $inc = $windows[1];
  >> } # inversion de la polarité si on a : pas...[emotion] (concerne essentiellement les émotions neutres)
  >> elsif ($lemmas =~ m/(?\\*|^)(?:jamais|aucun(?:e)?|ni|sans|non|rien)\\*(?:[\\^\\^*]+\\*){0,4}\\^\\^\\^([\\^\\^*]+)/i) {
  >>   >> $return = 1;
  >>   >> $inc = $windows[1];
  >> } # inversion de la polarité si on a : jamais/ni/rien/aucun/plus...[emotion]
  >> else {
  >>   >> return $polarity, 0;
  >> }
}

```

FIGURE D.4 – Traitement de la présence de marqueurs de négation dans les contextes des marqueurs émotionnels ou d'incertitude (2)

```

» } # pas d'inversion de polarité
» if ($return) {
»     if ($polarity eq "pos") {
»         return "neg", 1; # on renvoie la polarité neg et l'indice 1 précisant que la règle de négation a été
»             appliquée (indice utilisé pour les traitements suivant : modifieurs)
»     } # si polarité = pos, inversion => neg
»     elsif ($polarity eq "neg") {
»         return "pos", 1;
»     } # si polarité = neg, inversion => pos
»     elsif ($polarity eq "neu" && $return == 2) {
»         return "neg", 1;
»     } # si polarité = neu et que la règle appliquée renvoie 2 dans "$return", modification polaire => neg
»     elsif ($polarity eq "inc") {
»         if ($lemmas =~ m/\^\^\^(si|\?)/) {
»             my $strong_inc = $1;
»             if ($contexts =~ m/\^\^\^\*(?:[^\*]+\)*pre-neg\*(?:[^\*]+\)*post-neg/ && $strong_inc ne "?" ) {
»                 return "inc+", 1;
»             } # atténuation de l'incertitude si on a : si...ne...pas
»             elsif ($contexts =~ m/pre-neg\*(?:[^\*]+\)*post-neg(?:[^\*]+\)*{0,2}\^\^\^\$strong_inc/ &&
»                 $strong_inc ne "si") {
»                 return "inc-", 1;
»             } # amplification de l'incertitude si on a : ne...pas...?
»             else {
»                 return "inc", 1;
»             } # sinon aucune modification de la polarité
»         } # si le marqueur d'incertitude est "si" ou "?"
»         elsif ($inc =~ m/(si|wi)/ && $lemmas !~ m/\^\^\^\^apparaître/) {
»             return "inc--", 1 if ($1 eq "wi"); # atténuation forte si c'est un marqueur d'incertitude faible
»             return "inc0", 1 if ($1 eq "si"); # inhibition ou atténuation forte si c'est un marqueur
»                 d'incertitude fort
»         } # si la polarité est incertitude
»     } # si polarité = inc (incertitude) et qu'une règle est appliquée, on altère l'intensité du marqueur
»         d'incertitude
»     else {
»         return $polarity, 0;
»     } # sinon aucune modification de l'intensité de l'incertitude
» } # si une des règles est appliquée, on altère la polarité
» else {
»     return $polarity, 0;
» } # sinon aucune modification de la polarité
}

```

D.3 Traitement des modifieurs

FIGURE D.5 – Traitement de la présence de marqueurs de type modifieur dans les contextes des marqueurs émotionnels ou d’incertitude (1)

```
# Application des règles pour les modifieurs
sub apply_modifieurs_rules {
  >> my $contexts = shift; # récupération des contextes gauche et droit du marqueur émotionnel ou d'incertitude
  >> analysé, représenté par les étiquettes sémantiques (pre-neg, modif-m, ...)
  >> my $lemmas = shift; # récupération des contextes avec les lemmes
  >> my $polarity = shift; # récupération de la polarité initiale du marqueur émotionnel ou d'incertitude
  >> my $is_neg = shift; # vaut 1 si une règle de négation à été appliquée sur le marqueur analysés, sinon vaut
  >> 0
  >> my $return; # contiendra la polarité du marqueur et des indices d'amplification (+) ou
  >> d'atténuation (-)
  >> my $intensififier = ""; # stocke la chaîne d'indices (par exemple : ++, ---)
  >> my $modif; # stocke l'indice
  >> my @sides = split(/\^\^\^\^/, $lemmas); # (1) stocke le contexte gauche, droit et le marqueur dans une liste,
  >> avec les lemmes
  >> my @sl = split(/\*/, $sides[0]); # (2) stocke chaque lemme du contexte gauche dans une liste
  >> @sl = reverse(@sl); # (3) inversion de l'ordre du tableau pour pouvoir boucler sur le
  >> contexte gauche en partant du marqueur (de la droite vers la gauche)
  >> my @sr = split(/\*/, $sides[2]) unless (scalar(@sides) < 3); # même étape que (1) pour le contexte droit en
  >> vérifiant s'il y a bien un contexte droit : si on a le marqueur "?" pour l'incertitude, il n'y aura pas de
  >> contexte droit car "?" est une ponctuation forte (frontière phrastique)
  >> my @windows = split(/\^\^\^\^/, $contexts); # même étape que (1) mais avec les étiquettes sémantiques
  >> my $emotion = $windows[1];
  >> my @wl = split(/\*/, $windows[0]); # stocke les informations sémantiques du contexte gauche dans une
  >> liste
  >> @wl = reverse(@wl); # même étape que (3) mais avec les étiquettes sémantiques
  >> my @wr = split(/\*/, $windows[2]) unless (scalar(@windows) < 3); # stocke les informations sémantiques du
  >> contexte droit dans une liste en s'assurant qu'il existe
  >> for(my $i = 0; $i <= 2; $i++) {
  >> >> return $polarity, 0 if (exists $wl[0] && $sl[0] eq "comme" && $sides[1] eq "si" || $lemmas =~
  >> m/\^\^\^\^apparaître^\^\^\^/); # pas d'altération du marqueur "si" s'il est précédé directement par "comme"
  >> >> if (exists $wl[$i] && $wl[$i] =~ m/modif-(p|m)/) {
  >> >> >> $modif = $1;
  >> >> >> if ($sl[$i] eq "plus" && $windows[0] !~ m/(? :ne|n'|n)\*(?:.+?)*{0,4}plus\*/i) {
  >> >> >> >> $return = 1;
  >> >> >> >> $intensififier .= $modif;
  >> >> >> } # récupération du type de modifieur (p = fort, m = faible) en vérifiant que "plus" n'est pas
  >> >> >> employé dans un contexte négatif
  >> >> >> else {
  >> >> >> >> next if ($sl[$i] eq "plus");
  >> >> >> >> $return = 1;
  >> >> >> >> $intensififier .= $modif;
  >> >> >> } # récupération du type de modifieur
  >> >> } # si dans la fenêtre gauche, il y a un modifieur, on altère l'intensité du marqueur
  >> >> if (exists $wr[$i + 1] && $wr[$i + 1] =~ m/modif-(p|m)/ && $lemmas !~ m/\^\^\^\^?{1}/) {
  >> >> >> $modif = $1;
  >> >> >>
```

FIGURE D.6 – Traitement de la présence de marqueurs de type modifieur dans les contextes des marqueurs émotionnels ou d’incertitude (2)

```

» |         if ($sr[$i + 1] eq "plus" && $sides[0] !~ m/(?ne|n'|n)\*/i) {
» |             $return = 1;
» |             $intensifier .= $modif;
» |         } # récupération du type de modifieur
» |         else {
» |             last if ($sr[$i + 1] eq "plus");
» |             $return = 1;
» |             $intensifier .= $modif;
» |         } # récupération du type de modifieur
» |     } # si dans la fenêtre droite, il y a un modifieur, on altère l'intensité du marqueur
» | } # on boucle sur les contextes gauche et droit dans une fenêtre de 3 tokens
» | if ($return) {
» |     if (substr($intensifier, -1) eq "p") {
» |         if (length($intensifier) > 1 && substr($intensifier, -2, 1) eq "p") {
» |             if (length($intensifier) > 2 && substr($intensifier, -3, 1) eq "p") {
» |                 $return = "$polarity+++, 1";
» |             } # présence de trois modifieurs forts = 3 amplifications
» |             elsif (length($intensifier) > 2 && substr($intensifier, -3, 1) eq "m") {
» |                 $return = "$polarity---, 1";
» |             } # présence de deux modifieurs forts et d'un faible = 3 atténuations
» |             else {
» |                 $return = "$polarity++, 1";
» |             } # présence de deux modifieurs forts = 2 amplifications
» |         } # un second modifieur de type fort est détecté
» |         elsif (length($intensifier) > 1 && substr($intensifier, -2, 1) eq "m") {
» |             $return = "$polarity--, 1";
» |         } # présence d'un modifieur fort et d'un faible = 2 atténuations
» |         else {
» |             $return = "$polarity+, 1";
» |         } # présence d'un seul modifieur fort = 1 amplification
» |     } # le modifieur est fort
» |     else {
» |         $return = "$polarity-, 1";
» |     } # présence d'un seul modifieur faible = 1 atténuation
» | } # un modifieur est détecté
» | else {
» |     $return = "$polarity, 0";
» | } # sinon aucune altération
» | if ($is_neg && substr($return, -1) eq "1" && $emotion =~ m/emot-/ || !$is_neg && substr($return, -1) eq "1"
» | && $emotion !~ m/emot-/ ) {
» |     if ($is_neg && $return =~ m/\-+/) {
» |         $return =~ s/pos/neg/ if ($return =~ m/pos/);
» |         $return =~ s/neg/pos/ if ($return =~ m/neg/);
» |     } # combinaison NEGATION + MODIF-M
» |     else {

```

(a)

```

» |         $return =~ m/\+/? $return =~ s/\+/-/g : $return =~ s/-/\+/?g;
» |     } # combinaison NEGATION + MODIF-P
» |     return split(/, /, $return);
» | } # si une règle de négation a été appliquée précédemment
» | elsif ($is_neg && substr($return, -1) eq "1" && $emotion !~ m/emot-/ ) {
» |     $return =~ m/\+/? $return =~ s/\-/\+/?g : $return =~ s/\+/-/g;
» |     return split(/, /, $return);
» | } # même étape mais pour un marqueur d'incertitude
» | else {
» |     return split(/, /, $return);
» | } # on renvoie la polarité avec des indices d'altération éventuels
» | }

```

(b)

D.4 Traitement l'incertitude

FIGURE D.7 – Traitement de la présence de marqueurs d'incertitude dans les contextes des marqueurs émotionnels ou d'incertitude

```
# Application des règles pour les incertitudes
sub apply_uncertainty_rules {
  > my $contexts = shift;# récupération des contextes gauche et droit du marqueur émotionnel ou d'incertitude analysé,
  > représenté par les étiquettes sémantiques (pre-neg, modif-m, ...)
  > my $lemmas = shift; # récupération des contextes avec les lemmes
  > my $polarity = shift; # récupération de la polarité initiale du marqueur émotionnel ou d'incertitude
  > my $is_uncertain = 0; # vaut 1 s'il y a un marqueur d'incertitude dans les contextes du marqueur analysé, sinon vaut 0
  > my $uncertain = ""; # stocke les indices d'incertitude (par exemple : "?", "???", "??")
  > my @emotions = split(/\^\^\^\^/, $lemmas); # stocke les lemmes des contextes gauche, droit et le marqueur analysé dans une
  > liste
  > my $emotion = $emotions[1];
  > my @er = split(/\^*/, $emotions[2]) unless (scalar(@emotions) < 3); # stocke les lemmes dans le contexte droit dans une
  > liste, s'il existe
  > my %unc = ("si" => "??", "wi" => "?");
  > my @windows = split(/\^\^\^\^/, $contexts); # stocke les informations sémantiques des contextes gauche et droit du marqueur
  > analysé dans une liste
  > my @wl = split(/\^*/, $windows[0]); # stocke les informations sémantiques du contexte gauche
  > @wl = reverse(@wl); # inverse l'ordre de la liste pour parcourir les tokens en commençant à partir
  > du marqueur
  > my @wr = split(/\^*/, $windows[2]) unless (scalar(@windows) < 3); # stocke les informations sémantiques du contexte droit,
  > s'il existe
  > for(my $i = 0; $i <= 6; $i++) {
  > # contexte gauche
  > next if (exists $wl[$i] && $wl[$i] eq "apparaître" || exists $wr[$i] && $wr[$i] eq "apparaître");
  > if (exists $wl[$i] && $wl[$i] =~ m/-(si|wi)/ && $polarity =~ m/neu/ && $emotion =~ m/!+/) {
  > | $polarity = "neg-";
  > } # modifie la polarité neutre en négative en présence de marqueur(s) d'incertitude si l'émotion est marquée par un
  > "!"
  > elsif (exists $wl[$i] && $wl[$i] =~ m/-(si|wi)/) {
  > | $uncertain .= $unc{$i};
  > } # sinon on ajoute un indice d'incertitude ("??" si incertitude forte sinon "?" pour incertitude faible)
  > else {
  > | $uncertain .= "";
  > } # sinon pas d'altération du marqueur
  > # contexte droit
  > if (exists $wr[$i + 1] && ($wr[$i + 1] =~ m/-(si|wi)/)) {
  > | $uncertain .= $unc{$i} unless ($emotion =~ m/\?+/);
  > } # on ajoute un indice d'incertitude en présence d'un marqueur d'incertitude dans le contexte droit
  > elsif (exists $er[$i + 1] && $er[$i + 1] =~ m/\?+/) {
  > | $uncertain .= "?";
  > | last;
  > } # on ajoute un indice d'incertitude et on sort de la boucle si on rencontre une fin de phrase avec "?"
  > $is_uncertain++; # indique la présence de marqueur(s) d'incertitude
  > } # on parcourt les fenêtres gauche et droite du marqueur analysé (7 tokens)
  > return $polarity.$uncertain, $is_uncertain; # renvoie la polarité suivi du/des indices d'incertitude s'il y en a
  > ("is_uncertain" = 1 sinon 0 s'il n'y en a pas
  > }
}
```

(a)

D.5 Traitement des connecteurs

FIGURE D.8 – Traitement de la présence de connecteurs dans les contextes des marqueurs émotionnels ou d’incertitude

```
# Traitement des connecteurs : ET, DONC, CAR, CEPENDANT, MAIS
sub apply_connectors_rules {
  > my $contexts = shift; # récupération des contextes gauche et droit du marqueur émotionnel ou d'incertitude
  > analysé, représenté par les étiquettes sémantiques (pre-neg, modif-m, ...)
  > my $lemmas = shift; # récupération des contextes avec les lemmes
  > my $polarity = shift; # récupération de la polarité initiale du marqueur émotionnel ou d'incertitude
  > my @emotions = split(/\^\^\^/, $lemmas); # stocke les lemmes des contextes gauche, droit et le marqueur
  > analysé
  > my $emotion = $emotions[1];
  > my @el = split(/\*/, $emotions[0]); # stocke dans une liste les lemmes du contexte gauche
  > @el = reverse(@el); # inversion de l'ordre des lemmes dans la liste pour parcourir le
  > contexte gauche à partir du lemme le plus proche du marqueur analysé
  > my @er = split(/\*/, $emotions[2]) unless (scalar(@emotions) < 3); # stocke dans une liste les lemmes du
  > contexte droit, s'il existe
  > my @windows = split(/\^\^\^/, $contexts); # stocke dans une liste les informations sémantiques des contextes
  > gauche, droit et le marqueur analysé
  > if ($lemmas =~ m/(?:(?:car|donc|et)\*(?:[^\*\^]+\*)|(?:(?:donc|et)\*(?:[^\*\^]+\*)|(?:(?:car|donc|et)\*(?:[^\*\^]+\*))\^\^\^/i) {
  > {0,4}|(?:[^\*\^]+\*)|(?:(?:car|donc|et)\*(?:[^\*\^]+\*))\^\^\^/i) {
  >     if ($windows[1] =~ m/emot-/) {
  >         return "$polarity+", 1;
  >     } # on ajoute un indice d'amplification si le marqueur est une émotion
  >     else {
  >         return "$polarity-", 1;
  >     } # on ajoute un indice d'amplification si c'est un marqueur d'incertitude
  > } # traite les connecteurs "car", "donc" et "et" dans le contexte gauche
  > elsif ($lemmas =~ m/\^\^\^*(?:[^\*\^]+\*){0,3}(?:car|donc|et)\*(?:[^\*\^]+\*)|(?:(?:car|donc|et)\*(?:[^\*\^]+\*))\^\^\^/i) {
  >     if ($windows[1] =~ m/emot-/) {
  >         return "$polarity+", 1;
  >     } # on ajoute un indice d'amplification si le marqueur est une émotion
  >     else {
  >         return "$polarity-", 1;
  >     } # on ajoute un indice d'amplification si c'est un marqueur d'incertitude
  > } # traite les connecteurs "car", "donc" et "et" dans le contexte droit
  > for(my $i = 0; $i <= 6; $i++) {
  >     if (exists $el[$i] && $el[$i] =~ m/(mais|cependant)/i) {
  >         if ($windows[1] =~ m/emot-/) {
  >             return "$polarity+", 2;
  >         } # on ajoute un indice d'amplification si le marqueur est une émotion si le connecteur apparaît à
  >         gauche du marqueur émotionnel
  >         else {
  >             return "$polarity-", 2;
  >         } # on ajoute un indice d'amplification si c'est un marqueur d'incertitude
  >     } # traite les connecteurs "mais" et "cependant" dans le contexte gauche
  >     if (exists $er[$i + 1] && $er[$i + 1] =~ m/(mais|cependant)/i) {
  >         return "$polarity+", 2;
  >     } # on ajoute un indice d'amplification si le marqueur est une émotion si le connecteur apparaît à
  >     droite du marqueur émotionnel
  >     else {
  >         return "$polarity-", 0;
  >     } # aucune altération de polarité si c'est un marqueur d'incertitude
  > } # traite les connecteurs "mais" et "cependant" dans le contexte droit
  > } # on parcourt les contextes gauche et droit dans leur intégralité pour détecter la présence d'un
  > connecteur
  > return "$polarity", 0; # aucune altération de polarité
  > }
}
```

(a)

```
>     if ($windows[1] =~ m/emot-/) {
>         return "$polarity-", 2;
>     } # on ajoute un indice d'atténuation si le marqueur est une émotion si le connecteur apparaît à
>     droite du marqueur émotionnel
>     else {
>         return "$polarity", 0;
>     } # aucune altération de polarité si c'est un marqueur d'incertitude
> } # traite les connecteurs "mais" et "cependant" dans le contexte droit
> } # on parcourt les contextes gauche et droit dans leur intégralité pour détecter la présence d'un
> connecteur
> return "$polarity", 0; # aucune altération de polarité
> }
```

(b)

Annexe E

Calcul de la polarité émotionnelle dominante

E.1 Corpus Q/A

E.1.1 Détection de la polarité pour chaque marqueur émotionnel présent dans les questions et les réponses

E.1.2 Détection de la polarité émotionnelle dominante pour chaque question et réponse

E.2 Corpus forum

E.2.1 Détection de la polarité pour chaque marqueur émotionnel présent dans les questions et les réponses

E.2.2 Détection de la polarité émotionnelle dominante pour chaque question et réponse

E.2.3 Détection de la polarité émotionnelle dominante pour l'ensemble des réponses

FIGURE E.1 – Détection de la polarité des marqueurs émotionnels suivant les deux types de configuration : *baseline* et *processed* (corpus Q/A)

1 : neg neg neg neu neg pos : neg neg	1 : neg++ neg+ neg+ neu+ neg pos : neg++ neg++
2 : pos neg neg pos pos : pos neg pos neg	2 : pos neg neg pos?+ neg : pos+ neg??+ pos+ neg
3 : inc inc inc : inc inc	3 : inc inc inc : inc inc
4 : neu neg neg neg	4 : neu neg+ neg+ neg+
5 : neu : neg	5 : neu : neg+
6 : inc inc inc : inc	6 : inc inc inc : inc
7 : neg neg pos neg neg : neg neg neg neg neg	7 : neg+ neg+ pos+ neg neg : neg? neg? neg? neg?? neg??
8 : inc : inc inc	8 : inc : inc- inc
9 : pos : inc inc	9 : pos : inc? inc??
10 : neg pos : inc	10 : pos+ pos+ : inc
11 : neg : inc inc	11 : neg?? : inc0- inc
12 : inc inc inc inc : inc inc	12 : inc inc?? inc?? inc : inc0- inc
13 : inc inc : inc	13 : inc?? inc? : inc-
14 : pos : inc inc inc inc inc	14 : pos++ : inc- inc inc inc?? inc
15 : : inc	15 : : inc
16 : neg : inc inc inc inc	16 : neg???? : inc0 inc- inc? inc??
17 : inc inc inc inc inc inc inc : inc inc inc inc	17 : inc inc inc inc?? inc?? inc inc : inc+ inc inc inc0
18 : neg	18 : neg++
19 : neg : inc inc inc inc inc	19 : neg : inc0 inc inc inc?? inc-
20 : neu neg : inc	20 : neu+ neg+ : inc
21 : : inc	21 : : inc
22 : inc inc inc : pos	22 : inc inc inc : pos
23 : inc : pos	23 : inc : pos
24 : inc inc inc inc inc : inc inc inc inc inc inc inc	24 : inc inc inc inc?? inc? : inc0 inc inc- inc- inc- inc- inc
25 : inc : neg	25 : inc : neg
26 : neg : inc inc inc	26 : neg+ : inc0 inc0- inc
27 : inc	27 : inc
28 : inc inc inc inc inc inc : inc inc inc	28 : inc- inc- inc inc- inc- inc- : inc inc- inc
29 : neg neg : inc inc	29 : neg+ neg+ : inc0 inc
30 : neg : pos	30 : neg : pos
31 : inc inc : inc inc inc inc	31 : inc inc : inc0 inc inc?? inc?
32 : neg neg neg neg : pos neg neg neg neg neg neg neg	32 : neg neg+ pos+ neg+ : pos neg neg neg neg- neg? neg? neg-
33 : neg : pos neg	33 : neg+ : pos+ neg+
34 : inc inc : inc inc	34 : inc inc : inc inc
35 : inc : neg	35 : inc- : neg?
36 : neg : pos neg neg pos pos neu	36 : neg+ : pos neg neg? pos neg neu
37 : inc inc : inc inc inc	37 : inc??- inc? : inc inc inc
38 : pos neg neg pos neg : neg neg	38 : pos neg+ neg pos- neg? : neg neg
39 : pos : neg pos	39 : pos : neg pos
40 : inc : inc	40 : inc : inc-
41 : pos neg neg neg pos : pos	41 : pos? neg+?? neg+ neg+ pos+ : pos+
42 : inc inc inc inc inc	42 : inc?? inc? inc inc?? inc
43 : neg neg neu : inc inc inc	43 : neg? neg+ neu : inc?? inc? inc0
44 : neg : neu neg pos	44 : neg+ : neg pos pos
45 : : pos	45 : : pos

(a) Polarité de chaque marqueur émotionnel pour la configuration *baseline*

(b) Polarité de chaque marqueur émotionnel pour la configuration *processed*

FIGURE E.2 – Détection de la polarité émotionnelle dominante suivant les deux types de configuration : *baseline* et *processed* (corpus Q/A)

1 : neg : neg
 2 : pos : neg
 3 : inc : inc
 4 : neg : nothing
 5 : neu : neg
 6 : inc : inc
 7 : neg : neg
 8 : inc : inc
 9 : pos : inc
 10 : neg : inc
 11 : neg : inc
 12 : inc : inc
 13 : inc : inc
 14 : pos : inc
 15 : nothing : inc
 16 : neg : inc
 17 : inc : inc
 18 : neg : nothing
 19 : neg : inc
 20 : neg : inc
 21 : nothing : inc
 22 : inc : pos
 23 : inc : pos
 24 : inc : inc
 25 : inc : neg
 26 : neg : inc
 27 : inc : nothing
 28 : inc : inc
 29 : neg : inc
 30 : neg : pos
 31 : inc : inc
 32 : neg : neg
 33 : neg : neg
 34 : inc : inc
 35 : inc : neg
 36 : neg : pos
 37 : inc : inc
 38 : neg : neg
 39 : pos : neg
 40 : inc : inc
 41 : neg : pos
 42 : inc : nothing
 43 : neg : inc
 44 : neg : neg
 45 : nothing : pos

(a) Polarité émotionnelle dominante dans les questions et les réponses pour la configuration *baseline*

1 : neg : neg
 2 : neg : pos
 3 : inc : inc
 4 : neg : nothing
 5 : neu : neg
 6 : inc : inc
 7 : neg : neg
 8 : inc : inc
 9 : pos : inc
 10 : pos : inc
 11 : neg : inc
 12 : inc : inc
 13 : inc : inc
 14 : pos : inc
 15 : nothing : inc
 16 : neg : inc
 17 : inc : inc
 18 : neg : nothing
 19 : neg : inc
 20 : neg : inc
 21 : nothing : inc
 22 : inc : pos
 23 : inc : pos
 24 : inc : inc
 25 : inc : neg
 26 : neg : inc
 27 : inc : nothing
 28 : inc : inc
 29 : neg : inc
 30 : neg : pos
 31 : inc : inc
 32 : neg : neg
 33 : neg : neg
 34 : inc : inc
 35 : inc : neg
 36 : neg : neg
 37 : inc : inc
 38 : neg : neg
 39 : pos : neg
 40 : inc : inc
 41 : neg : pos
 42 : inc : nothing
 43 : neg : inc
 44 : neg : pos
 45 : nothing : pos

(b) Polarité émotionnelle dominante dans les questions et les réponses pour la configuration *processed*

FIGURE E.3 – Détection de la polarité des marqueurs émotionnels suivant les deux types de configuration : *baseline* et *processed + accentuation* (corpus forum)

```

12 : neg
13 : neg neg : pos : pos neg : neu neg : pos
14 : neg neg
15 : pos pos : neg : neg neg pos neg : neg pos neg : neg pos neg : inc inc inc inc inc : neu neg neg : inc :
inc inc inc : pos pos pos : pos pos : neg : inc : pos pos : neu
neu neu neu neu neu : pos : inc inc inc : inc inc inc : neg : pos : inc : inc : pos pos pos : pos pos pos :
inc : neg pos : neg neu neu neg : pos pos pos : pos : pos : pos : inc inc inc : pos : neg pos : inc : neg pos
: inc : neg pos neg : inc inc inc inc inc inc : inc inc inc inc : neu neg pos neu pos neg : inc : inc inc :
neu pos : inc : inc : neg : neg neg pos neg pos : inc : neg
16 : neu neg neu neg neg
17 : neg neg neg
18 : neg neu neg pos pos pos : neu
19 : neg pos neg : pos : inc inc
20 : neu neg neg neu
21 : pos
22 : neu neg pos neg neg neg neg : neg : neu neu pos neu : pos : neg : inc inc : neu : neg neu pos neu neu :
inc : neu : inc : neu : inc : pos neu pos neu : neg neu neg neu : pos : pos pos : inc inc : neg : pos pos :
pos neu neu pos : inc inc : neu : neu : neu : pos pos neu pos pos neg pos neg pos neu pos : neu pos pos :
neu pos neu neu : neu neu neg neg neg neg pos neu neu : pos : pos neg pos pos : inc inc : neg : pos : inc
: pos neg : neg neg neu neu : neg : inc : inc : neg pos : neg neu : inc inc : neg neg : neg neg : neu : neg
23 : neg : inc inc : neg : neg neu : inc inc : neg neu neu neu neu neu : pos : pos neg pos : neg neg neu
neg neg : inc
24 : neg neg : inc inc inc : inc inc : inc inc : pos
25 : neg pos neu pos : inc inc
26 : inc inc : neg
27 : neg pos pos : inc : inc inc : pos : pos : neu : pos neg pos : inc inc inc inc : neu neg neu neu : inc
inc : pos pos : neu
28 : inc inc inc inc : inc inc
29 : neg : pos neg pos : neg pos pos pos : neu neu neg neu : neu : neu pos pos neu pos neu neu : pos pos neg
: pos neg : neg neu neg neg : neg neg neg neg pos neu pos : neu pos neg neg neu
30 : pos neg pos neg neu neg neg : pos : neg neg neu : neg : neg pos : neg neu neg neg neg : inc : neg
neg pos neu neu : neg : neg pos : inc
31 : neg pos : neu : pos neg pos neg : neg pos neu : neg neu
32 : neg neg : pos neg

```

(a) Polarité de chaque marqueur émotionnel pour la configuration *baseline*

```

12 : neg+
13 : neg+ neg+ : pos? : pos- pos+ : neg+ neg?? : pos??
14 : neg neg+
15 : pos pos : neg+ : neg++ neg+ pos+ neg??+ : neg pos? neg : pos+ pos++ neg? : inc0 inc- inc- inc?? inc?? :
neu pos? pos+ : inc : inc inc inc : pos+ pos++ pos??+ : pos++ pos+?? : neg+ : inc- inc inc inc?? inc??
inc?? inc inc inc inc inc inc : pos pos??+ : neu neu+ neg- neu neg neu : pos???? : inc inc inc : inc inc
inc : neg+ : pos?+ : inc : inc : neg?? pos pos : pos? pos+ pos : inc : neg+ pos : neg neu neu neg+ : pos+ pos
pos+ : pos? : pos : neg : inc inc inc+ : pos?? : pos neg- neu : inc : neg- pos : inc- : neg?? neg neg+ : inc- inc
inc0- inc inc inc : inc0 inc inc inc : neu+ neg- pos+ neu pos+ neg- : inc : inc??- inc0?- : neu pos?+ : inc-
inc : neg- : neg neg- pos? neg+ pos : inc : neg
16 : neg neg neu neg neg
17 : neg? neg neg
18 : neg neu- neg pos+ pos+ pos : neg+
19 : neg pos neg : pos : inc inc
20 : neu neg?? neg?? neg-
21 : pos????
22 : neu+ neg pos neg neg neg neg : neg+ : neu+ neu pos neu : pos++ : neg : inc inc : neg : pos neu pos+
neu+ neu : inc : neu : inc : neu : inc : pos?? neu pos+ neu : neg neu neg+ neg neu : pos : pos pos+ : inc
inc0- : neg : pos+ pos+ pos+ neg neu pos : inc inc : neu++ : neu : neu+ neu?? : neu : pos pos? neu? pos pos
neg++ pos+ neg+ neg+ neg+++ pos neu+ pos??+ : neu pos neu neu : neu neu+ neg+ neg neg+ neg+ neg+?? pos? neu
neu : pos+ : pos+ neg+ pos+ pos+ : inc??- inc?- : neg : pos? : inc- : neg neg : neg+ pos+ neu neg+ : neg :
inc : inc : neg++ pos+ : neg neu : inc- inc : neg neg+ : neg neg : neg- : neg+
23 : neg : inc?? inc0? : neg+ : neg neu : inc+- inc : neg??+ neg- neu neu neu neu : pos : pos? neg pos :
neg neg+ neu++ neg+ pos : inc-
24 : neg neg : inc inc- inc- : inc inc : inc inc : pos??
25 : neg++ pos neu pos+ : inc?? inc??
26 : inc- inc : neg+
27 : neg pos pos+ : inc : inc- inc- : pos+ : pos++ : neu : pos??? neg pos? : inc inc inc?? inc?? : neu pos
neu neu+ : inc inc : pos+ pos : neu++
28 : inc inc inc?? inc?? : inc inc
29 : neg : pos??? neg pos+ : neg pos- pos+ pos??? : neu neg-?? neg+ neg- : neu : neg- pos pos neu pos neu neu
: pos- pos- neg : pos+ pos : pos neu+ neg+ neg : neg?? neg- neg??+ neg+ pos+ neu+ pos : neu+ pos+ neg neg neu
30 : pos neg+ pos neg+ neu+ neg+ neg+ : pos? : neg neg- neu : pos- : pos+ pos- : neg+ neu neg+ neg neg??+
neg : inc- : neg neg+ pos neu neu : neg : neg? pos+ : inc
31 : neg+ pos+++ : neu+ : pos????+ neg+ neg neg+ : neg+ pos neu+ : neg+ neu+
32 : neg+ neg+ : pos neg+

```

(b) Polarité de chaque marqueur émotionnel pour la configuration *processed + accentuation*

FIGURE E.4 – Détection de la polarité émotionnelle dominante (une par question et une pour chaque réponse) suivant les deux types de configuration : *baseline* et *processed + accentuation* (corpus forum)

12 : neg : nothing
 13 : neg : pos : neg : neg : pos
 14 : neg : nothing
 15 : pos : neg : neg : neg : neg : neg : inc : neg : inc : inc : pos : pos : neg : inc : pos : neu : pos : inc :
 inc : neg : pos : inc : inc : pos : pos : inc : neg : neg : pos : pos : pos : pos : inc : pos : neg : inc :
 neg : inc : neg : inc : inc : neg : inc : inc : pos : inc : inc : neg : neg : inc : neg
 16 : neg : nothing
 17 : neg : nothing
 18 : pos : neu
 19 : neg : pos : inc
 20 : neg : nothing
 21 : pos : nothing
 22 : neg : neg : neu : pos : neg : inc : neu : neu : inc : neu : inc : neu : inc : pos : neg : pos : pos :
 inc : neg : pos : inc : neu : neu : neu : neu : pos : neu : neg : pos : pos : inc : neg : pos : inc : neg :
 neg : neg : inc : inc : neg : neg : inc : neg : neg : neu : neg
 23 : neg : inc : neg : neg : inc : neu : pos : pos : neg : inc
 24 : neg : inc : inc : inc : pos
 25 : pos : inc
 26 : inc : neg
 27 : pos : inc : inc : pos : pos : neu : pos : inc : neu : inc : pos : neu
 28 : inc : inc
 29 : neg : pos : pos : neu : neu : neu : pos : neg : neg : neg : neg
 30 : neg : pos : neg : neg : neg : neg : inc : neg : neg : neg : inc
 31 : neg : neu : neg : neg : neg
 32 : neg : neg

(a) Polarité émotionnelle dominante dans les questions et les réponses pour la configuration *baseline*

12 : neg : nothing
 13 : neg : pos : pos : neg : pos
 14 : neg : nothing
 15 : pos : neg : neg : neg : pos : inc : pos : inc : inc : pos : pos : neg : inc : pos : neu : pos : inc :
 inc : neg : pos : inc : inc : pos : pos : inc : neg : neg : pos : pos : pos : neg : inc : pos : neg : inc :
 pos : inc : neg : inc : inc : pos : inc : inc : pos : inc : inc : neg : neg : inc : neg
 16 : neg : nothing
 17 : neg : nothing
 18 : pos : neg
 19 : neg : pos : inc
 20 : neu : nothing
 21 : pos : nothing
 22 : neg : neg : neu : pos : neg : inc : neg : neu : inc : neu : inc : neu : inc : pos : neg : pos : pos :
 inc : neg : pos : inc : neu : neu : neu : neu : neg : neu : neg : pos : pos : inc : neg : pos : inc : neg :
 neg : neg : inc : inc : neg : neg : inc : neg : neg : neg : neg
 23 : neg : inc : neg : neg : inc : neu : pos : pos : neg : inc
 24 : neg : inc : inc : inc : pos
 25 : neg : inc
 26 : inc : neg
 27 : pos : inc : inc : pos : pos : neu : neg : inc : neu : inc : pos : neu
 28 : inc : inc
 29 : neg : pos : pos : neg : neu : pos : neg : pos : neg : neg : neu
 30 : neg : pos : neg : pos : pos : neg : inc : neg : neg : pos : inc
 31 : pos : neu : neg : neg : neg
 32 : neg : neg

(b) Polarité émotionnelle dominante dans les questions et les réponses pour la configuration *processed + accentuation*

FIGURE E.5 – Détection de la polarité émotionnelle dominante (une par question et une pour l'ensemble des réponses) suivant les deux types de configuration : *baseline* et *processed + accentuation* (corpus forum)

```

12 : neg : nothing
13 : neg : neu
14 : neg : nothing
15 : pos : neg
16 : neg : nothing
17 : neg : nothing
18 : pos : neu
19 : neg : pos
20 : neg : nothing
21 : pos : nothing
22 : neg : neg
23 : neg : neg
24 : neg : pos
25 : pos : inc
26 : inc : neg
27 : pos : pos
28 : inc : inc
29 : neg : neg
30 : neg : neg
31 : neg : neg
32 : neg : neg

```

(a) Polarité émotionnelle dominante dans les questions et les réponses pour la configuration *baseline*

```

12 : neg : nothing
13 : neg : pos
14 : neg : nothing
15 : pos : pos
16 : neg : nothing
17 : neg : nothing
18 : pos : neg
19 : neg : pos
20 : neu : nothing
21 : pos : nothing
22 : neg : neg
23 : neg : neg
24 : neg : pos
25 : neg : inc
26 : inc : neg
27 : pos : pos
28 : inc : inc
29 : neg : neu
30 : neg : neu
31 : pos : neg
32 : neg : neg

```

(b) Polarité émotionnelle dominante dans les questions et les réponses pour la configuration *processed + accentuation*

Annexe F

Création des données de référence

F.1 Corpus Q/A

FIGURE F.1 – Annotation des phrases (+,-,/ et indications sur l'émotion exprimée) contenant au moins un marqueur émotionnel pour le corpus Q/A (1) : Échantillon des données annotées par l'annotateur expert AE1

Q5842|je suis enceinte c ' est pour ca que je suis plus ****angoissé**** . | - |angoissé
Q5845|Mon seul problème c ' est que j ' ai des accès de ****nervosité**** et de ****stress**** , qui ne durent pas longtemps (5 minutes maxi) mais je suis très ****énervée**** , et pour rien du tout . | - |énervée
Q5845|J ' ai beaucoup de transports , beaucoup de marche , et un métier à dimension commerciale qui est on ne peut plus ****stressant**** . | - |stressant
Q5845|J ' ai besoin d ' être ****rassurée**** , mais dommage , en plein mois d ' aout , tout le monde est en congés . | / |rassurée
Q5845|Afin de savoir_si mon ****stress**** a eu une conséquence et pr pouvoir repartir du bon pied après , si il devait se passer quoi que ce soit (fausse_couche) combien de temps cela interviendrait - il après le ****stress**** présumé ? | - |stress
Q5846|on ma dit aussi qu ' il ny avait pas de ****souci**** , deplus j ' ai deja eu un enfant . | + |pas de souci
Q5847|bonsoir , j ' ai 2 chiens et je voulais savoir_si jamais j ' attrape des vers intestinaux (ma chienne ma lécher la bouche pendant ma sieste sur le canapé ****!!!**** ! | - |!!!
Q5848|Et j ' ai déjà fais un analyse de sang pour savoir et il n ' y a rien d ' anormale mais c ' est comme si j ' étais une enfant encore sur les analyses au niveau hormonale alors que cela fais quand même 5ans que j ' ai les premiers signes pubertaires donc je m ' ****inquiète**** . | - |inquiète
Q5851|****Désolée**** mais je ne comprends pas à la lecture de votre réponse si l ' opération est envisageable d ' ici le 7 Septembre ou non . | / |Désolée
Q5851|Le chirurgien est pour mais moi , j ' ai une grosse ****peur**** par rapport à l ' algodystrophie . | - |peur|
R5842|Vous êtes ****angoissée**** parce que vous êtes enceinte , prenez quelques précautions : lavez - vous les mains fréquemment , séparer les aliments crus des aliments cuits , ne mangez pas de charcuterie ni de fruits_de_mer . | - |angoissée
R5844|Le plus souvent , on peut ****brûler**** une verrue avec du froid (azote) ou bien mettre dessus une pommade pour la détruire (à mettre tous les jours pendant plusieurs jours) en protégeant la peau autour | / |
R5845|Cher (e) Audrey , Soyez ****sereine**** . | + |sereine
R5845|Les modifications hormonales au cours de la grossesse peuvent modifier votre caractère et vous ****stresser**** . | - |peuvent stresser
R5845|Les fausses couches ne sont pas liées au ****stress**** . | / |
R5847|Soyez ****rassurée**** . | + |rassurée
R5848|En attendant soyez ****rassurée**** , votre poids et votre taille sont normaux pour votre âge . | + |rassurée
R5849|Il s ' agit plutôt d ' une ****allergie**** , le spray est peut-être en cause . | / |
R5854|Il faut se ****méfier**** des produits non vendus en pharmacie . | / |méfier
R5856|Ne soyez pas ****angoissée**** . | + |pas angoissée
R5860|Il n ' est pas ****étonnant**** qu ' avec votre métier , votre pied enfle et devienne douloureux . | / |pas étonnant
R5864|Il n ' y a pas de quoi s ' ****angoisser**** , votre neurologue vous dira ce qu ' il en est. . | + |pas de quoi s ' angoisser
R5865|Les allergies_alimentaires les plus fréquentes sont les poissons , les fraises , oeufs , arachide , etc. les tests d ' ****allergie**** sont faits par un un spécialiste et sont compliqués . | / |
R5865|Le vaccin_ROR peut_être allergisant (surtout dans l ' ****allergie**** à l ' oeuf) mais il n ' y a pas de moyen facile de le savoir . | / |
R5874|Il n ' y a pas de quoi s ' ****angoisser**** mais vous avez raison de consulter un médecin . | + |pas de quoi s ' angoisser

FIGURE F.2 – Annotation des phrases (+,-,/ et indications sur l'émotion exprimée) contenant au moins un marqueur émotionnel pour le corpus Q/A (2) : Échantillon des données annotées par l'annotateur semi-expert

ASE2

Q5842|je suis enceinte c ' est pour ca que je suis plus ****angoissé**** . |-|plus angoissé
Q5845|Mon seul problème c ' est que j ' ai des accès de ****nervosité**** et de ****stress**** , qui ne durent pas longtemps (5 minutes maxi) mais je suis très ****énervée**** , et pour rien du tout . |-|très énervée
Q5845|J ' ai beaucoup de transports , beaucoup de marche , et un métier à dimension commerciale qui est on ne peut plus ****stressant**** . |-|on ne peut plus stressant
Q5845|J ' ai besoin d ' être ****rassurée**** , mais dommage , en plein mois d ' aout , tout le monde est en congés . |-|dommage
Q5845|Afin de savoir_si mon ****stress**** a eu une conséquence et pr pouvoir repartir du bon pied après , si il devait se passer quoi que ce soit (fausse_couche) combien de temps cela interviendrait - il après le ****stress**** présumé ? |//savoir_si stress
Q5846|on ma dit aussi qu ' il ny avait pas de ****souci**** , deplus j ' ai deja eu un enfant . |+|pas souci
Q5847|bonsoir , j ' ai 2 chiens et je voulais savoir_si jamais j ' attrape des vers intestinaux (ma chienne ma lécher la bouche pendant ma sieste sur le canapé ****!!**** ! |//!!
Q5848|Et j ' ai déjà fais un analyse de sang pour savoir et il n ' y a rien d ' anormale mais c ' est comme si j ' étais une enfant encore sur les analyses au niveau hormonale alors que cela fais quand même 5ans que j ' ai les premiers signes pubertaires donc je m ' ****inquiète**** . |-|inquiète
Q5851|****Désolée**** mais je ne comprends pas à la lecture de votre réponse si l ' opération est envisageable d ' ici le 7 Septembre ou non . |-|désolée
Q5851|Le chirurgien est pour mais moi , j ' ai une grosse ****peur**** par rapport à l ' algodystrophie . |-|grosse peur
R5842|Vous êtes ****angoissée**** parce que vous êtes enceinte , prenez quelques précautions : lavez - vous les mains fréquemment , séparer les aliments crus des aliments cuits , ne mangez pas de charcuterie ni de fruits_de_mer . |-|angoissée
R5844|Le plus souvent , on peut ****brûler**** une verrue avec du froid (azote) ou bien mettre dessus une pommade pour la détruire (à mettre tous les jours pendant plusieurs jours) en protégeant la peau autour ||/
R5845|Cher (e) Audrey , Soyez ****sereine**** . |+|sereine
R5845|Les modifications hormonales au cours de la grossesse peuvent modifier votre caractère et vous ****stresser**** . |-|stresser
R5845|Les fausses couches ne sont pas liées au ****stress**** . |+|pas liées au stress
R5847|Soyez ****rassurée**** . |+|rassurée
R5848|En attendant soyez ****rassurée**** , votre poids et votre taille sont normaux pour votre âge . |+|rassurée
R5849|Il s ' agit plutôt d ' une ****allergie**** , le spray est peut-être en cause . |//
R5854|Il faut se ****méfier**** des produits non vendus en pharmacie . |-|méfier
R5856|Ne soyez pas ****angoissée**** . |+|pas angoissée
R5860|Il n ' est pas ****étonnant**** qu ' avec votre métier , votre pied enfle et devienne douloureux . |//pas étonnant
R5864|Il n ' y a pas de quoi s ' ****angoisser**** , votre neurologue vous dira ce qu ' il en est. . |+|pas angoisser
R5865|Les allergies_alimentaires les plus fréquentes sont les poissons , les fraises , oeufs , arachide , etc. les tests d ' ****allergie**** sont faits par un un spécialiste et sont compliqués . |//
R5865|Le vaccin_ROR peut être allergisant (surtout dans l ' ****allergie**** à l ' oeuf) mais il n ' y a pas de moyen facile de le savoir . |//
R5874|Il n ' y a pas de quoi s ' ****angoisser**** mais vous avez raison de consulter un médecin . |+|pas angoisser

FIGURE F.3 – Annotation des phrases (+,-,/ et indications sur l'émotion exprimée) contenant au moins un marqueur émotionnel pour le corpus Q/A (3) : Échantillon des données annotées par l'annotateur non-expert

ANE3

Q5842|je suis enceinte c ' est pour ca que je suis plus ****angoissé**** . |-|angoissé
Q5845|Mon seul problème c ' est que j ' ai des accès de ****nervosité**** et de ****stress**** , qui ne durent pas longtemps (5 minutes maxi) mais je suis très ****énervée**** , et pour rien du tout . |-|problème, énervée
Q5845|J ' ai beaucoup de transports , beaucoup de marche , et un métier à dimension commerciale qui est on ne peut plus ****stressant**** . |-|stressant
Q5845|J ' ai besoin d ' être ****rassurée**** , mais dommage , en plein mois d ' aout , tout le monde est en congés . |-|rassurée
Q5845|Afin de savoir_si mon ****stress**** a eu une conséquence et pr pouvoir repartir du bon pied après , si il devait se passer quoi que ce soit (fausse_couche) combien de temps cela interviendrait - il après le ****stress**** présumé ? |-|stress
Q5846|on ma dit aussi qu ' il ny avait pas de ****souci**** , deplus j ' ai deja eu un enfant . |+|pas de souci
Q5847|bonsoir , j ' ai 2 chiens et je voulais savoir_si jamais j ' attrape des vers intestinaux (ma chienne ma lécher la bouche pendant ma sieste sur le canapé ****!!!**** ! |-|!!
Q5848|Et j ' ai déjà fais un analyse de sang pour savoir et il n ' y a rien d ' anormale mais c ' est comme si j ' étais une enfant encore sur les analyses au niveau hormonale alors que cela fais quand même 5ans que j ' ai les premiers signes pubertaires donc je m ' ****inquiète**** . |-|m'inquiète
Q5851|****Désolée**** mais je ne comprends pas à la lecture de votre réponse si l ' opération est envisageable d ' ici le 7 Septembre ou non . |-|désoléé
Q5851|Le chirurgien est pour mais moi , j ' ai une grosse ****peur**** par rapport à l ' algodystrophie . |-|grosse peur
R5842|Vous êtes ****angoissée**** parce que vous êtes enceinte , prenez quelques précautions : lavez - vous les mains fréquemment , séparer les aliments crus des aliments cuits , ne mangez pas de charcuterie ni de fruits_de_mer . |/
R5844|Le plus souvent , on peut ****brûler**** une verrue avec du froid (azote) ou bien mettre dessus une pommade pour la détruire (à mettre tous les jours pendant plusieurs jours) en protégeant la peau autour| |/
R5845|Cher (e) Audrey , Soyez ****sereine**** . |+|sereine
R5845|Les modifications hormonales au cours de la grossesse peuvent modifier votre caractère et vous ****stresser**** . |-|stresser
R5845|Les fausses couches ne sont pas liées au ****stress**** . |/
R5847|Soyez ****rassurée**** . |+|rassurée
R5848|En attendant soyez ****rassurée**** , votre poids et votre taille sont normaux pour votre âge . |+|rassurée
R5849|Il s ' agit plutôt d ' une ****allergie**** , le spray est peut-être en cause . |/>
R5854|Il faut se ****méfier**** des produits non vendus en pharmacie . |-|méfier
R5856|Ne soyez pas ****angoissée**** . |+|pas angoissée
R5860|Il n ' est pas ****étonnant**** qu ' avec votre métier , votre pied enfle et devienne douloureux . |-|pas étonnant
R5864|Il n ' y a pas de quoi s ' ****angoisser**** , votre neurologue vous dira ce qu ' il en est. . |+|pas de quoi s'angoisser
R5865|Les allergies_alimentaires les plus fréquentes sont les poissons , les fraises , oeufs , arachide , etc. les tests d ' ****allergie**** sont faits par un un spécialiste et sont compliqués . |/
R5865|Le vaccin_ROR peut_être allergisant (surtout dans l ' ****allergie**** à l ' oeuf) mais il n ' y a pas de moyen facile de le savoir . |/
R5874|Il n ' y a pas de quoi s ' ****angoisser**** mais vous avez raison de consulter un médecin . |+|pas de quoi s'angoisser

F.2 Corpus forum

FIGURE F.4 – Annotation des phrases (+,-,/ et indications sur l'émotion exprimée) contenant au moins un marqueur émotionnel pour le corpus forum (1) : Échantillon des données annotées par l'annotateur expert

AE1

A1_1|Vous pouvez monter les escaliers sans problème du moment que vous courez pas ******** ! |+|!!
A1_1|Surtout ne vous ****inquiétez**** pas lors de votre hospitalisation il vous expliquerons tout cela , mais surtout respectez a la lettre les consignes une fois a la maison car interdit également de s ' asseoir sur une chaise et interdit de monter en voiture que ce soit conducteur ou passager . |-/ne vous inquiétez pas
A1_1|N ' ****hésitez**** pas , sopjer1974 . |-/hésitez
A1_2|Encore une source d ' ****inquiétude**** pour " l ' après " . |-/inquiétude
A1_3|Non mon lit et canapé sont a la meme hauteur que toi , il faut donc comme tu le dis ****fléchir**** légèrement les genoux afin de ne pas plier le dos ******** ! |-/!!!
A1_3|Bon courage et n ' ****hésitez**** pas . |-/n ' hésite pas
A1_4|Merci , ça me ****rassure**** un peu . |+|rassure
Q2|solution envisagée par lui , remplacement des deux disques par des prothèses lombaires ******** ! |-/!!!!
Q2|je ne sais plus quoi penser , et ****hésitez**** avant cette longue opération risquée . |-/hésitez
Q2|qui peut me renseigner ****las**** dessus ????? ? |-/|
A2_2|Je suis en arrêt depuis le 12 09 , le neuro m ' a dit qu ' il fallait 6 mois après l ' intervention pour être au ****top**** . |-/top
A2_2|Et aujourd ' hui je ne ****regrette**** pas d ' avoir effectué l ' intervention qui a durée 4 heures et 5 jours d ' hospitalisation . |+|ne regrette pas
A2_3|slt j ' ai ete opere en 2003 pour une prothese de disque et je ne remercierei jamais assez mon chirurgien qui ma opere sa a ete une reussite j ' ai ete opere a lyon et je ne ****regrette**** pas . |+|ne regrette pas
A2_4|Je m ' appelle stef et je viens de ma faire opérée d ' une hernie discale bilatérale avec débord important en L5 - S1 ; tt s ' est bien passé mais le neuro - chir m ' a di qu ' on se reverrai dans quelques années pour mettre une prothse de disque car mon disque est complètement vide ******** ! |-/|!!
A2_11|mais ca vaut la ****peine**** peine. . |+|
A2_11|il est ****content**** content. . |+|content
A2_19|et j ' ai pas ****envie**** d ' avoir tous mes disques d ' âbimés . |-/pas envie
A2_22|Suite a une hernie discale paralysante mon disque l5s1 a été très abimé ce qui , depuis un an me donne de ****terribles**** douleurs: douleurs:après plusieurs infiltration sous radioscopie , et traitements x , y rien n ' a pu me soulager . |-/rien n ' a pu me soulager
A2_22|fr merci a toutes et tous pour vos renseignements car je suis très ****angoissé**** . |+|merci
A2_24|est ce que d ' ici une quinzaine de jours je pourrais rester plus longtemps debout et pourrais je enfin m ' assoir (ça par contre je ne m ' y attendais pas et c ' est assez ****pénible**** de ne pas pouvoir " se_poser " enfin bref ça me remonterais le moral voilà ****désolé**** pour le " coup de blues " c ' est peut_être le contre coup de l ' opération . |-/désolé
A2_25|Merci d ' avance et ****désolé**** de vous prendre de votre temps . |-/désolé
A2_26|Je vous remercie d ' avance de votre réponse qui me sera d ' un grand ****réconfort**** . |+|réconfort
A2_27|D ' apres mes deux rendez vous avec lui il n ' est pas normal d ' avoir mal comme vous le décrivez ******** ! |-/avoir mal
A2_29|Le seul conseil que je vous donne est d ' en discuter avec votre chirurgien afin d ' éviter toute ****surprise**** ******** ! |-/éviter toute surprise
A2_29|N ' ****hésitez**** pas en_cas_de besoina me poser vos questions . |-/hésitez|

FIGURE F.5 – Annotation des phrases (+,-,/ et indications sur l'émotion exprimée) contenant au moins un marqueur émotionnel pour le corpus forum (2) : Échantillon des données annotées par l'annotateur semi-expert ASE2

A1_1|Vous pouvez monter les escaliers sans problème du moment que vous courez pas **** ! |/>!!
A1_1|Surtout ne vous **inquiétez** pas lors de votre hospitalisation il vous expliquerons tout cela , mais surtout respectez a la lettre les consignes une fois a la maison car interdit également de s ' asseoir sur une chaise et interdit de monter en voiture que ce soit conducteur ou passager . |+|ne vous inquiétez pas
A1_1|N ' **hésitez** pas , soper1974 . |+|n'hésitez pas
A1_2|Encore une source d ' **inquiétude** pour " l ' après " . |-|inquiétude
A1_3|Non mon lit et canapé sont a la meme hauteur que toi , il faut donc comme tu le dis **fléchir** légèrement les genoux afin de ne pas plier le dos **** ! |/>!!!
A1_3|Bon courage et n ' **hésite** pas . |+|n'hésite pas
A1_4|Merci , ça me **rassure** un peu . |+|rassure un peu
Q2|solution envisagée par lui , remplacement des deux disques par des prothèses lombaires **** ! |-|!!!!
Q2|je ne sais plus quoi penser , et **hésite** avant cette longue opération risquée . |-|hésite
Q2|qui peut me renseigner **las** dessus ????? ? |/>|
A2_2|Je suis en arrêt depuis le 12 09 , le neuro m ' a dit qu ' il fallait 6 mois après l ' intervention pour être au **top** . |+|top
A2_2|Et aujourd ' hui je ne **regrette** pas d ' avoir effectué l ' intervention qui a durée 4 heures et 5 jours d ' hospitalisation . |+|ne regrette pas
A2_3|slt j ' ai ete opere en 2003 pour une prothese de disque et je ne remercierei jamais assez mon chirurgien qui ma opere sa a ete une reussite j ' ai ete opere a lyon et je ne **regrette** pas . |+|ne regrette pas
A2_4|Je m ' appelle stef et je viens de ma faire opérée d ' une hernie discale bilatérale avec débord important en L5 - S1 ; tt s ' est bien passé mais le neuro - chir m ' a di qu ' on se reverrai dans quelques années pour mettre une prothse de disque car mon disque est complètement vide **** ! |-|!!
A2_11|mais ca vaut la **peine** peine. . |+|
A2_11|il est **content** content. . |+|content
A2_19|et j ' ai pas **envie** d ' avoir tous mes disques d ' âbimés . |-|pas envie
A2_22|Suite a une hernie discale paralysante mon disque l5s1 a été très abimé ce qui , depuis un an me donne de **terribles** douleurs: douleurs:après plusieurs infiltration sous radioscopie , et traitements x , y rien n ' a pu me soulager . |-|rien n'a pu me soulager
A2_22|fr merci a toutes et tous pour vos renseignements car je suis très **angoissé** . |+|merci
A2_24|est ce que d ' ici une quinzaine de jours je pourrais rester plus longtemps debout et pourrais je enfin m ' assoir (ça par contre je ne m ' y attendais pas et c ' est assez **pénible** de ne pas pouvoir " se_poser " enfin bref ça me remonterais le moral voilà **désolé** pour le " coup de blues " c ' est peut_être le contre coup de l ' opération . |-|assez pénible
A2_25|Merci d ' avance et **désolé** de vous prendre de votre temps . |-|désolé
A2_26|Je vous remercie d ' avance de votre réponse qui me sera d ' un grand **réconfort** . |+|réconfort
A2_27|D ' apres mes deux rendez vous avec lui il n ' est pas normal d ' avoir mal comme vous le décrivez **** ! |-|!!!!
A2_29|Le seul conseil que je vous donne est d ' en discuter avec votre chirurgien afin d ' éviter toute **surprise** **** ! |/>|surprise
A2_29|N ' **hésitez** pas en_cas_de besoina me poser vos questions . |/>|n'hésitez pas

FIGURE F.6 – Annotation des phrases (+,-,/ et indications sur l'émotion exprimée) contenant au moins un marqueur émotionnel pour le corpus forum (3) : Échantillon des données annotées par l'annotateur non-expert ANE3

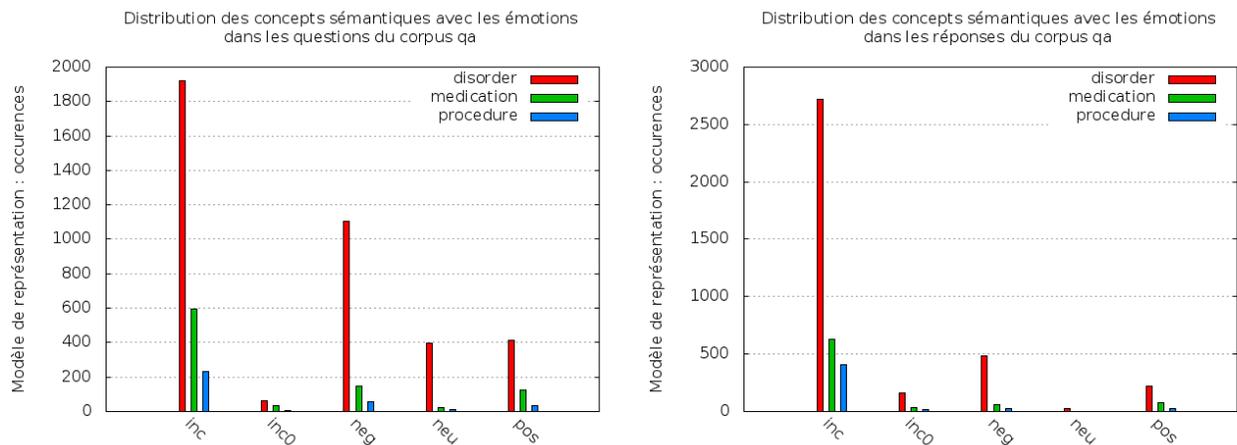
A1_1|Vous pouvez monter les escaliers sans problème du moment que vous courez pas ******** ! |+|sans probleme
A1_1|Surtout ne vous ****inquiétez**** pas lors de votre hospitalisation il vous expliquerons tout cela , mais surtout respectez a la lettre les consignes une fois a la maison car interdit également de s ' asseoir sur une chaise et interdit de monter en voiture que ce soit conducteur ou passager . |+|ne vous inquiétez pas
A1_1|N ' ****hésitez**** pas , s'opjler1974 . |/\n'hésitez pas
A1_2|Encore une source d ' ****inquiétude**** pour " l ' après " . |-|inquiétude
A1_3|Non mon lit et canapé sont a la meme hauteur que toi , il faut donc comme tu le dis ****fléchir**** légèrement les genoux afin de ne pas plier le dos ******** ! |/
A1_3|Bon courage et n ' ****hésite**** pas . |+|n'hésite pas
A1_4|Merci , ça me ****rassure**** un peu . |+|rassure
Q2|solution envisagée par lui , remplacement des deux disques par des prothèses lombaires ********* !
|-|*****
Q2|je ne sais plus quoi penser , et ****hésite**** avant cette longue opération risquée . |-|je ne sais plus quoi penser
Q2|qui peut me renseigner ****las**** dessus ????? ? |/
A2_2|Je suis en arrêt depuis le 12 09 , le neuro m ' a dit qu ' il fallait 6 mois après l ' intervention pour être au ****top**** . |/
A2_2|Et aujourd ' hui je ne ****regrette**** pas d ' avoir effectué l ' intervention qui a durée 4 heures et 5 jours d ' hospitalisation . |+|je ne regrette pas
A2_3|slt j ' ai ete opere en 2003 pour une prothese de disque et je ne remercierei jamais assez mon chirurgien qui ma opere sa a ete une reussite j ' ai ete opere a lyon et je ne ****regrette**** pas . |+|je ne regrette pas
A2_4|Je m ' appelle stef et je viens de ma faire opérée d ' une hernie discale bilatérale avec débord important en L5 - S1 ; tt s ' est bien passé mais le neuro - chir m ' a di qu ' on se reverrai dans quelques années pour mettre une prothse de disque car mon disque est complètement vide ******** ! |-|mon disque est complètement vide !!
A2_11|mais ca vaut la ****peine**** peine . |+|vaut la peine
A2_11|il est ****content**** content . |+|content
A2_19|et j ' ai pas ****envie**** d ' avoir tous mes disques d ' abimés . |-|pas envie
A2_22|Suite a une hernie discale paralysante mon disque l5s1 a été très abimé ce qui , depuis un an me donne de ****terribles**** douleurs: douleurs:après plusieurs infiltration sous radioscopie , et traitements x , y rien n ' a pu me soulager . |-|terribles douleur, y rien n'a pu me soulager
A2_22|fr merci a toutes et tous pour vos renseignement car je suis très ****angoissé**** . |-|angoissé
A2_24|est ce que d ' ici une quinzaine de jours je pourrais rester plus longtemps debout et pourrais je enfin m ' assoir (ça par contre je ne m ' y attendais pas et c ' est assez ****pénible**** de ne pas pouvoir " se poser " enfin bref ça me remonterais le moral voilà ****désolé**** pour le " coup de blues " c ' est peut_être le contre coup de l ' opération . |-|pénible, coup de blues
A2_25|Merci d ' avance et ****désolé**** de vous prendre de votre temps . |+|merci
A2_26|Je vous remercie d ' avance de votre réponse qui me sera d ' un grand ****réconfort**** . |+|réconfort
A2_27|D ' apres mes deux rendez vous avec lui il n ' est pas normal d ' avoir mal comme vous le décrivez ******** ! |-|pas normal
A2_29|Le seul conseil que je vous donne est d ' en discuter avec votre chirurgien afin d ' éviter toute ****surprise**** ******** ! |/
A2_29|N ' ****hésitez**** pas en_cas_de besoina me poser vos questions . |+|n'hésitez pas

Annexe G

Modélisations de l'évolution des émotions

G.1 Distribution des types médicaux dans les énoncés en fonction des polarités émotionnelles dominantes (échelle : nombre d'occurrences)

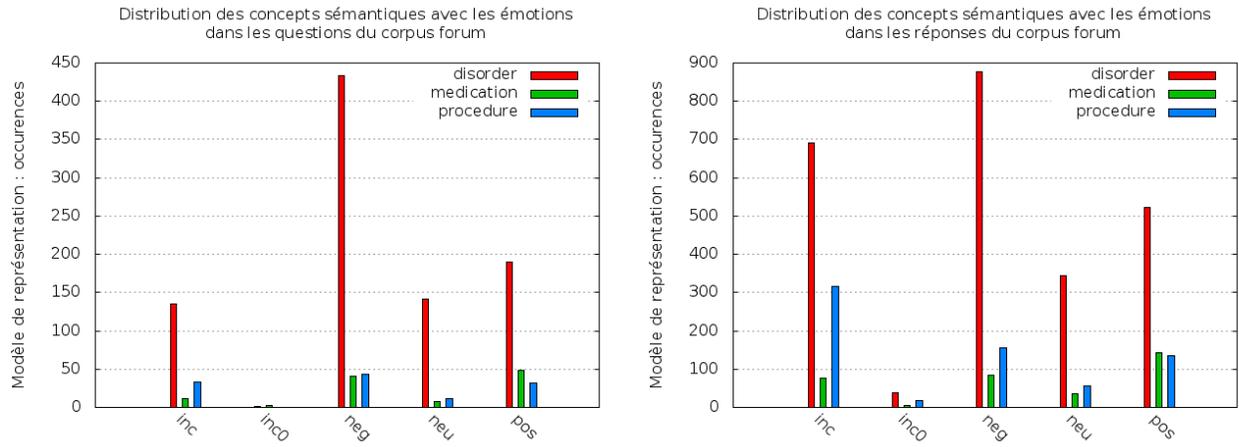
FIGURE G.1 – Distribution des types médicaux présents dans les contextes des marqueurs émotionnels dans le corpus Q/A, en fonction des polarités émotionnelles



(a) Distribution des types médicaux dans les questions, en fonction des polarités émotionnelles

(b) Distribution des types médicaux dans les réponses, en fonction des polarités émotionnelles

FIGURE G.2 – Distribution des types médicaux présents dans les contextes des marqueurs émotionnels dans le corpus forum, en fonction des polarités émotionnelles

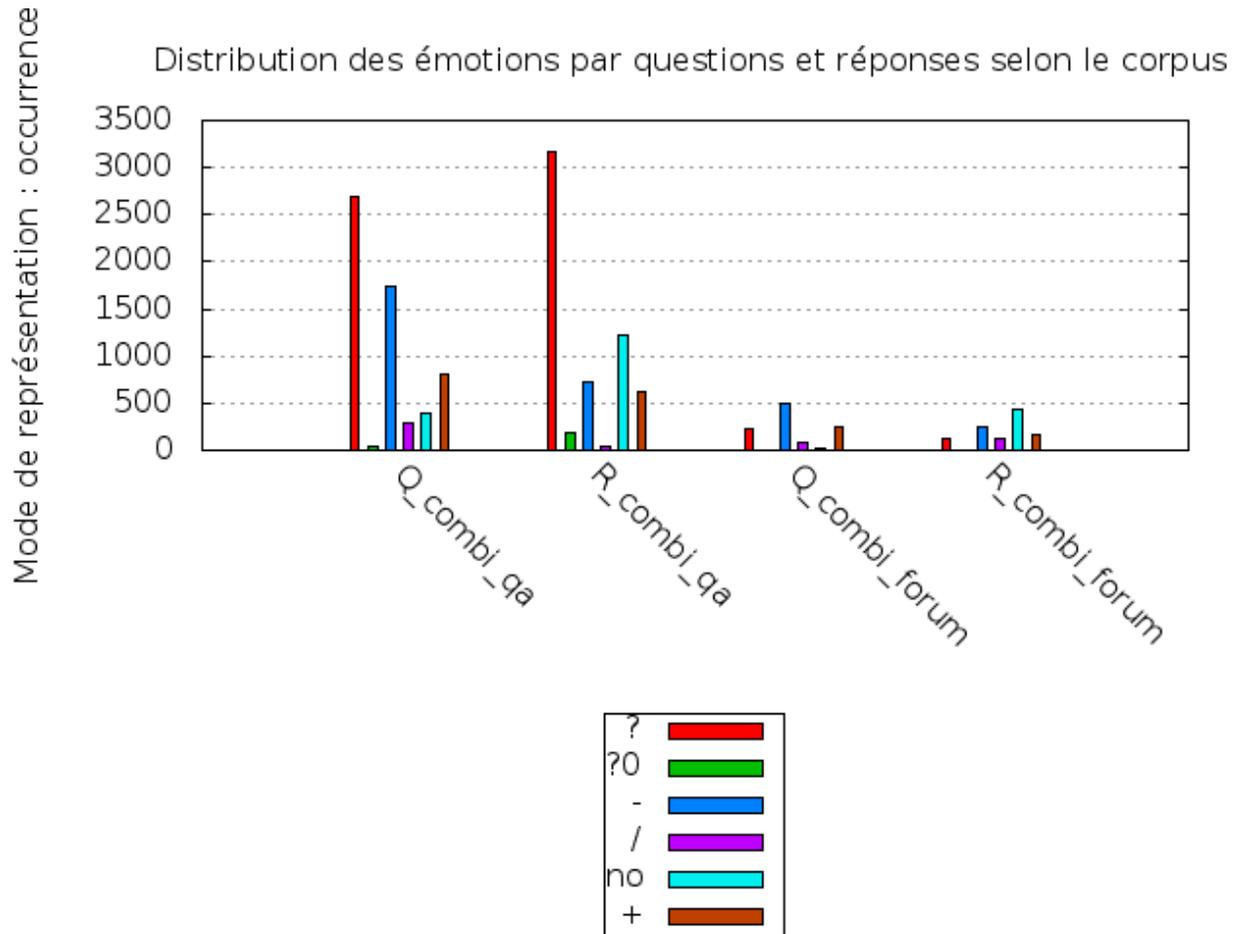


(a) Distribution des types médicaux dans les questions, en fonction des polarités émotionnelles

(b) Distribution des types médicaux dans les réponses, en fonction des polarités émotionnelles

G.2 Distribution des polarités émotionnelles dominantes dans les énoncés en fonction du corpus (échelle : nombre d'occurrences)

FIGURE G.3 – Distribution polaire en fonction du type d'énoncé (questions : *Q_combi_qa* et *Q_combi_forum* ; réponses : *R_combi_qa* et *R_combi_forum*) et du corpus : Q/A et forum

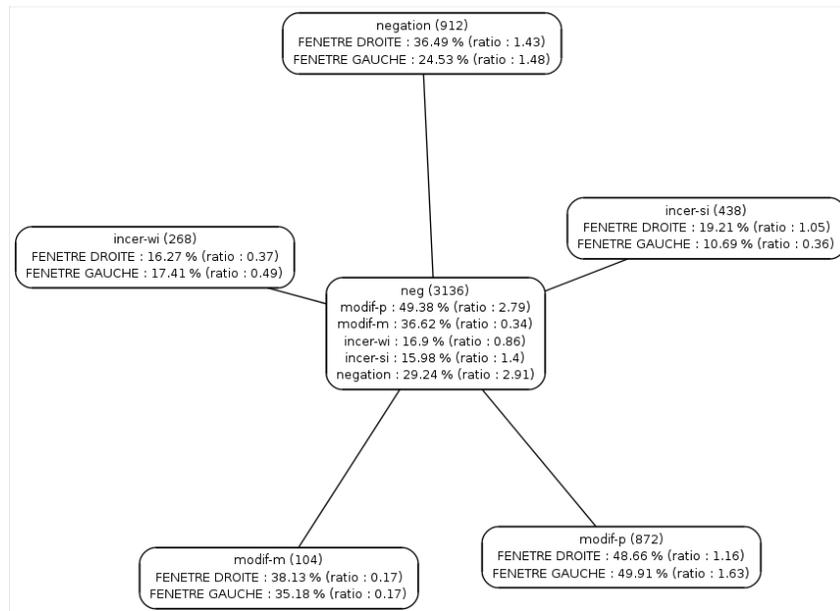


(a) Distribution polaire dans les questions et les réponses

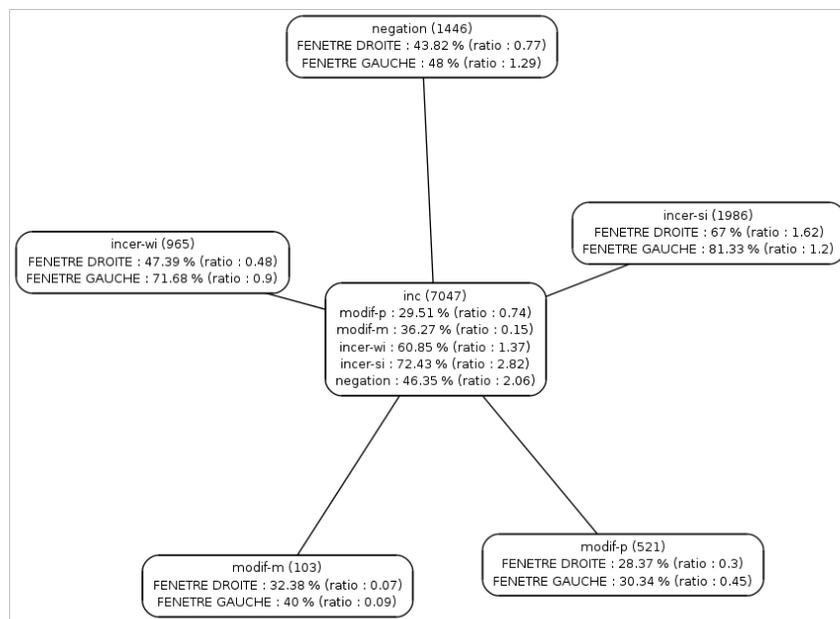
G.3 Distribution des marqueurs par polarité émotionnelle

G.3.1 Corpus Q/A

FIGURE G.4 – Distribution des marqueurs par polarité émotionnelle dans les questions du corpus Q/A

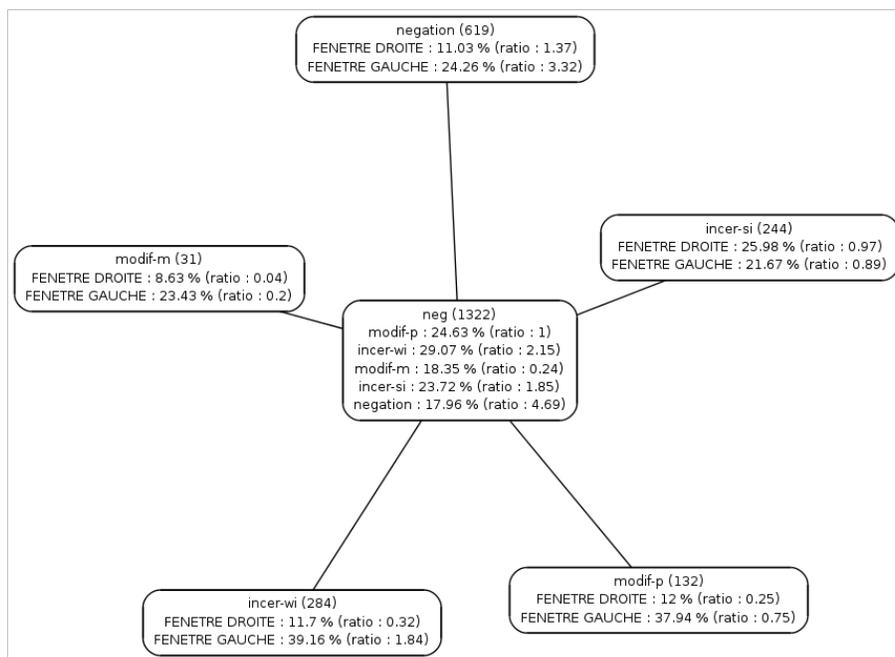


(a) Distribution des marqueurs pour la polarité négative

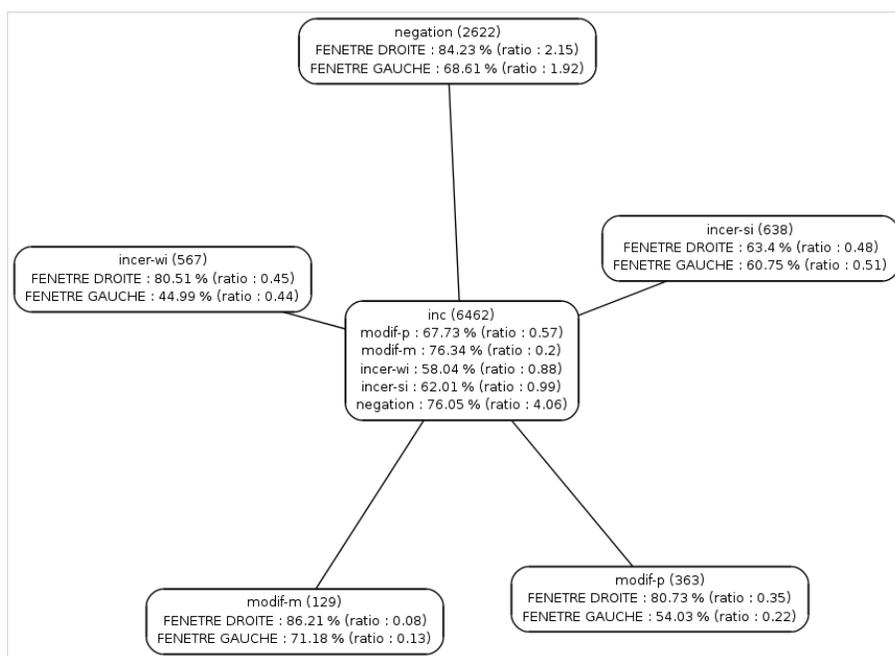


(b) Distribution des marqueurs pour la polarité incertitude

FIGURE G.5 – Distribution des marqueurs par polarité émotionnelle dans les réponses du corpus Q/A

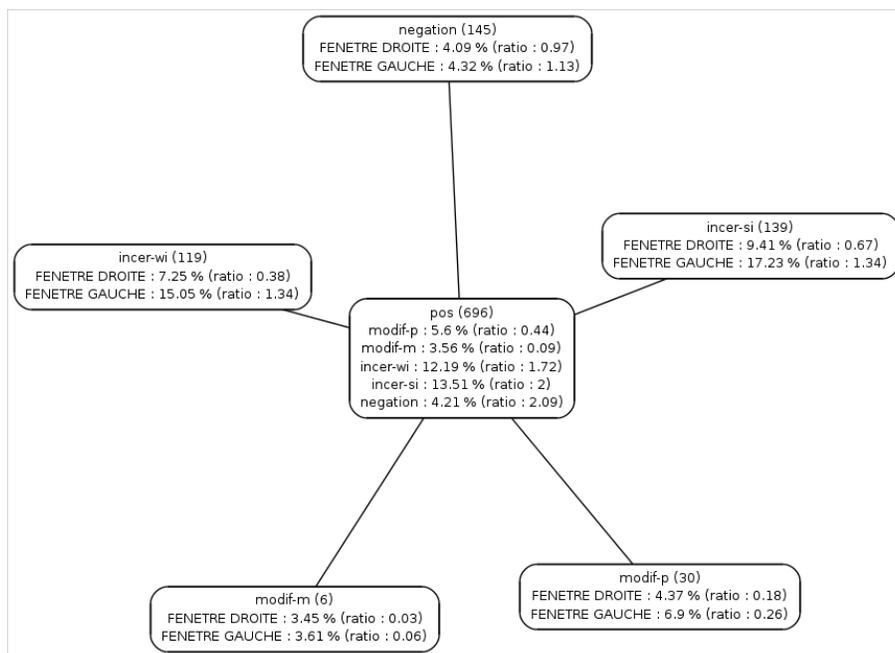


(a) Distribution des marqueurs pour la polarité négative

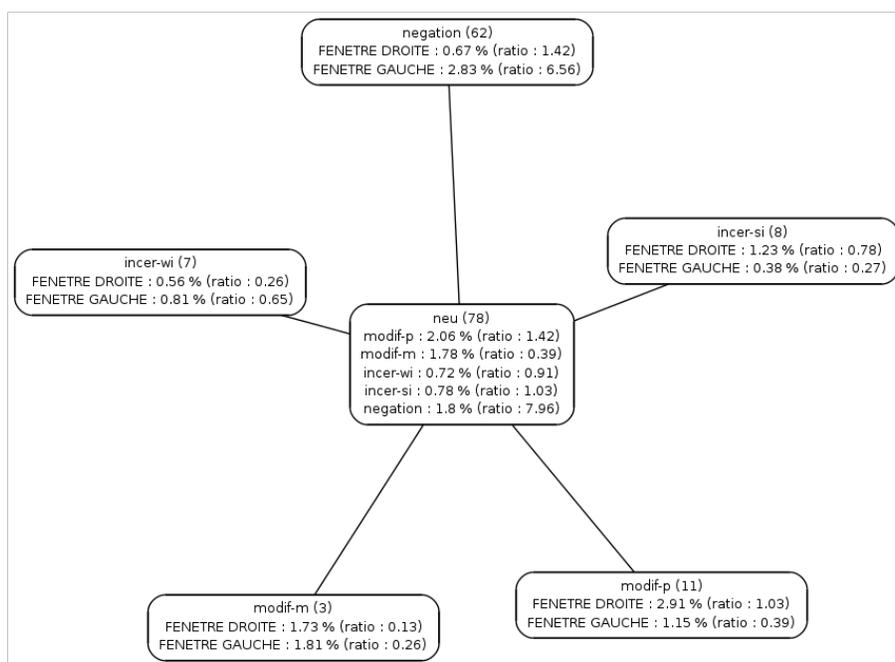


(b) Distribution des marqueurs pour la polarité incertitude

FIGURE G.6 – Distribution des marqueurs par polarité émotionnelle dans les réponses du corpus Q/A



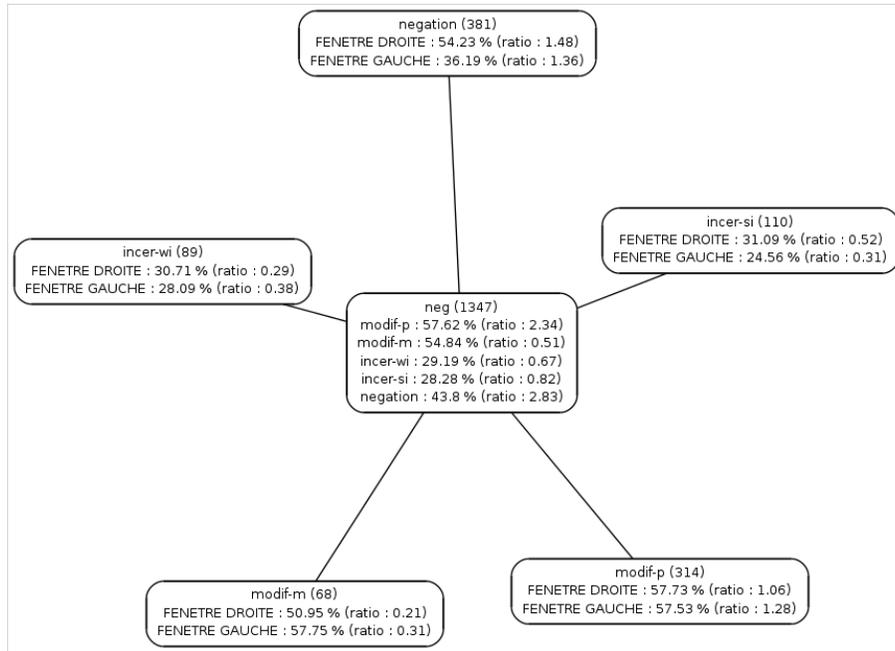
(a) Distribution des marqueurs pour la polarité positive



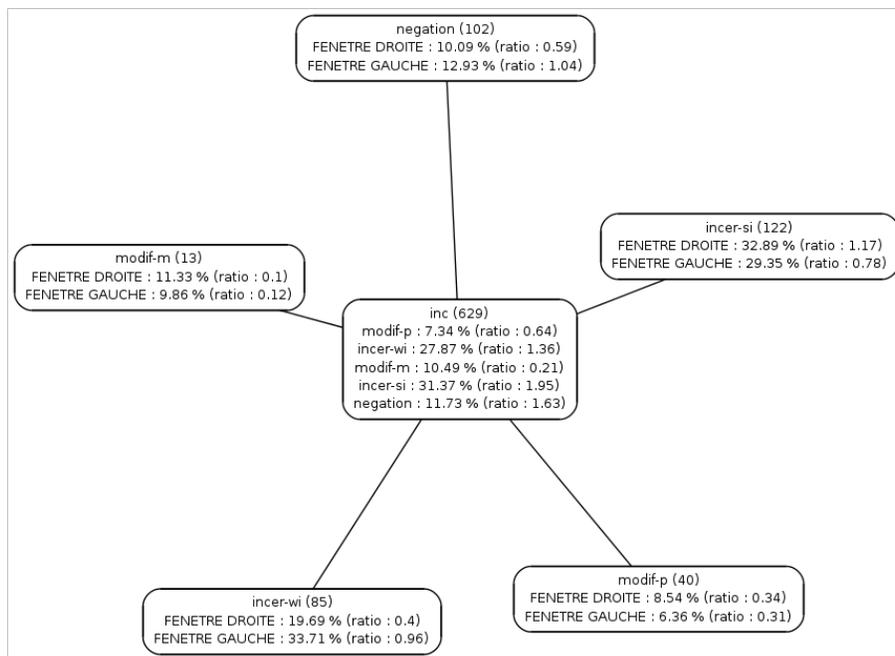
(b) Distribution des marqueurs pour la polarité neutre

G.3.2 Corpus forum

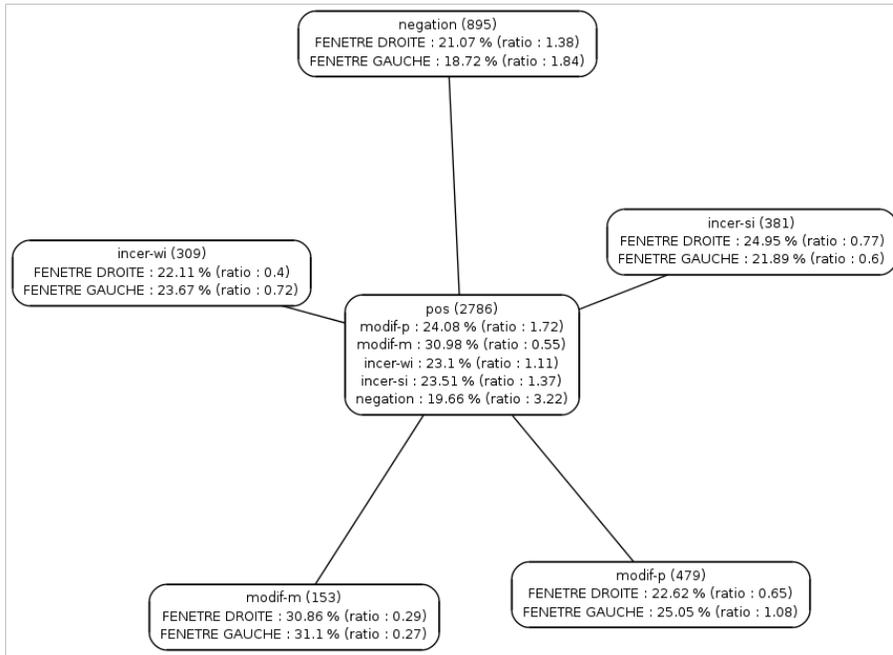
FIGURE G.7 – Distribution des marqueurs par polarité émotionnelle dans les questions du corpus forum



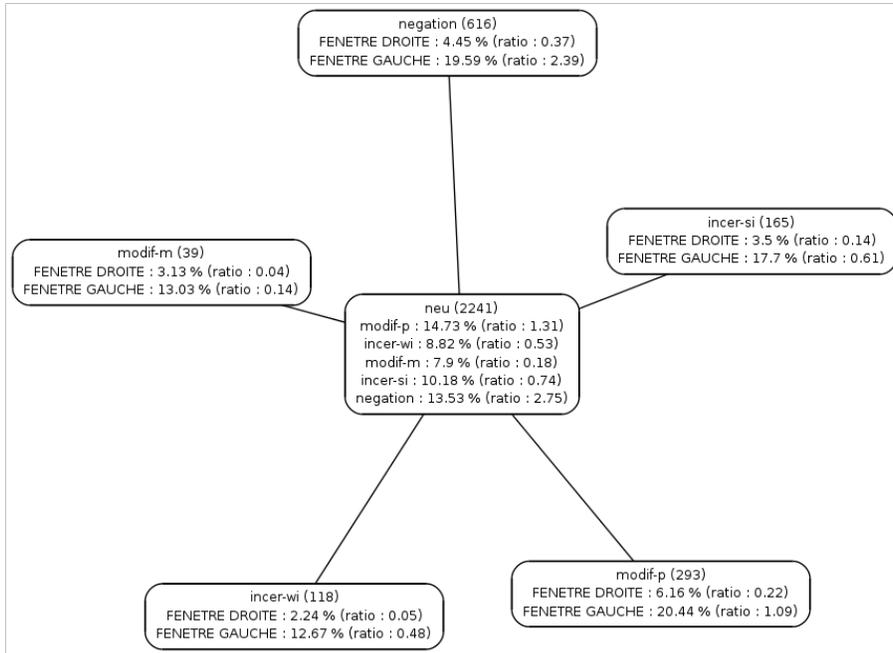
(a) Distribution des marqueurs pour la polarité négative



(b) Distribution des marqueurs pour la polarité incertitude

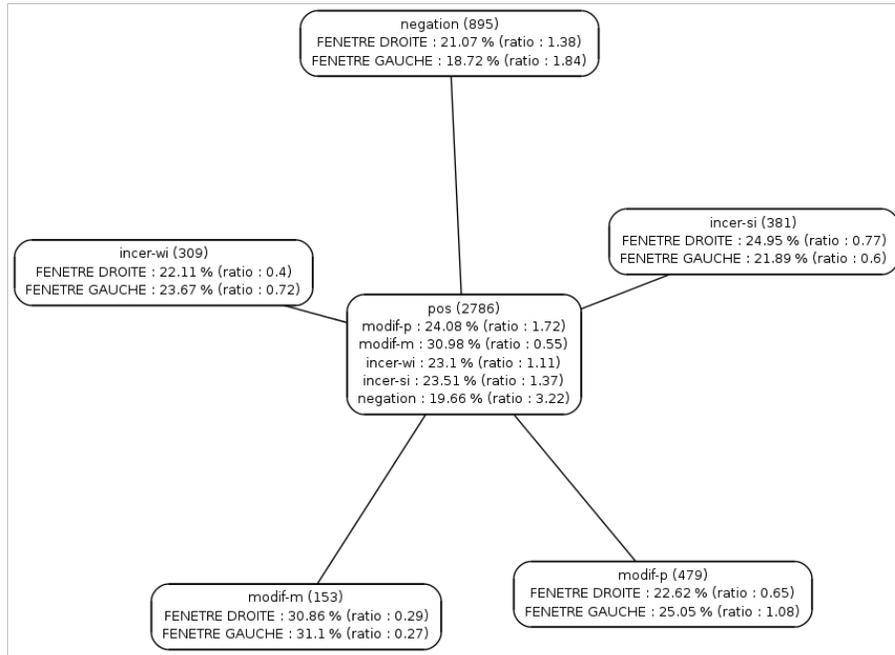


(c) Distribution des marqueurs pour la polarité positive

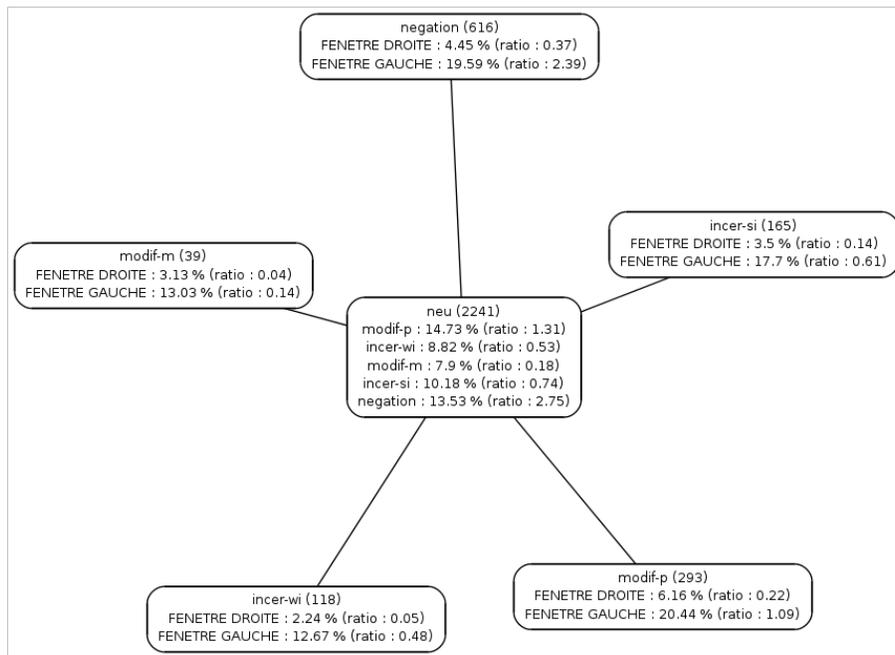


(d) Distribution des marqueurs pour la polarité neutre

FIGURE G.8 – Distribution des marqueurs par polarité émotionnelle dans les réponses du corpus forum



(a) Distribution des marqueurs pour la polarité positive

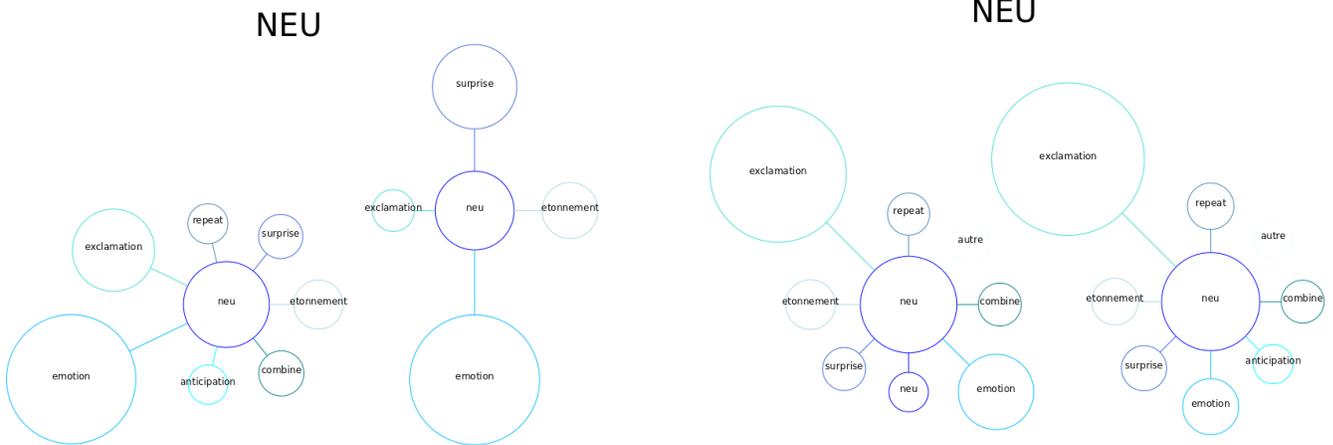


(b) Distribution des marqueurs pour la polarité neutre

G.4 Les "constellations" : représentation quantitative de l'expression des émotions par polarité et par catégorie émotionnelle

G.4.1 Les émotions à polarité neutre

FIGURE G.9 – Distribution des catégories d'émotions à polarité neutre pour les corpus Q/A et forum

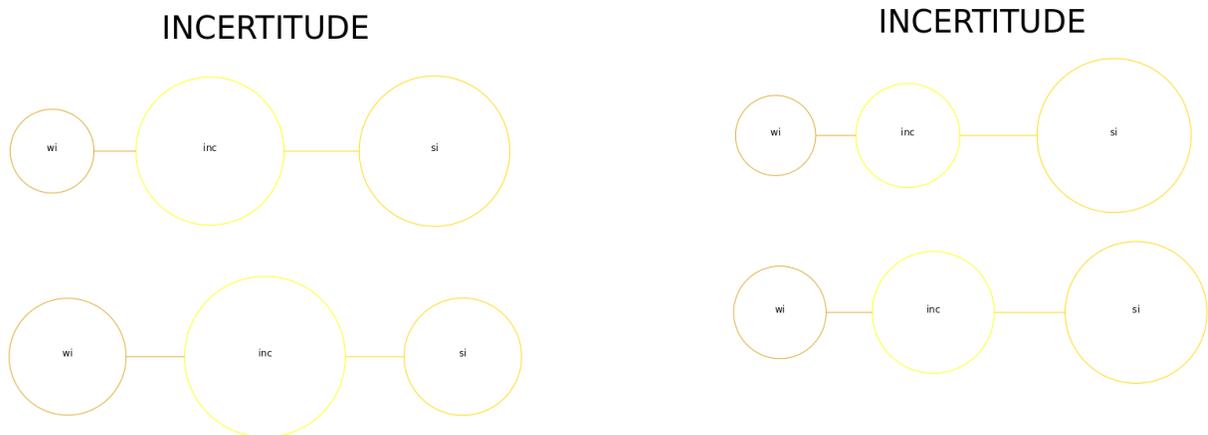


(a) Corpus Q/A : les questions sont représentées à gauche et les réponses à droite

(b) Corpus forum : les questions sont représentées à gauche et les réponses à droite

G.4.2 Les émotions à polarité incertitude

FIGURE G.10 – Distribution des catégories d'émotions à polarité incertitude pour les corpus Q/A et forum



(a) Corpus Q/A : les questions sont représentées en haut et les réponses en bas

(b) Corpus forum : les questions sont représentées en haut et les réponses en bas