

# L'analyse de l'émotion dans les forums de santé

Céline Battaïa (Doctorante en Sciences de l'information et de la communication  
à l'université Stendhal de Grenoble)

Groupe de Recherche sur les Enjeux de la Communication (GRESEC), Université Stendhal, Laboratoire  
Gresec, Grenoble 3 [celine.battaia@u-grenoble3.fr](mailto:celine.battaia@u-grenoble3.fr)

## RÉSUMÉ

---

Les travaux sur l'émotion dans les forums sont nombreux en Linguistique et Psychologie. L'objectif de cette contribution est de proposer une analyse de l'émotion dans les forums de santé selon l'angle des Sciences de l'Information et de la Communication mais également selon une approche interdisciplinaire. Il s'agira ici, d'étudier l'émotion comme un critère de pertinence lorsque des personnes malades effectuent des recherches dans les forums.

Ce papier introduit la méthodologie utilisée en traitement automatique de la langue afin de répondre à cette interrogation. Ainsi, le travail présenté abordera l'exploitation d'un corpus de messages de forums, la catégorisation semi-supervisée et l'utilisation du logiciel NooJ pour traiter de manière automatique les données.

## ABSTRACT

---

### **Analysis of Emotion in Health Fora**

Studies about emotion in fora are numerous in Linguistics and Psychology. This contribution approaches this subject from an Information and Communication Sciences point of view, and studies emotion as a criterion of pertinence for patients in a health forum. This paper introduces the empirical step of automatic language processing in order to answer this question, and uses data processing on the corpus of forum messages, semi-supervised categorisation of messages and use of software NooJ for Natural Language Processing.

**MOTS-CLÉS :** émotion, forum de santé, traitement automatique de la langue, désambiguïsation lexicale.

**KEYWORDS:** emotion, health forum, automatic language processing, lexical disambiguation.

---

# 1 Introduction

La recherche d'information peut être définie comme « l'activité d'un individu qui vise à localiser et traiter une ou plusieurs informations au sein d'un environnement documentaire complexe, afin de répondre à une question ou résoudre un problème » (Dinet et Rouet, 2002). Dans le contexte de la maladie rare, grave, ou chronique, la recherche d'information est une activité subordonnée à d'autres objectifs tels que rechercher des informations pratiques, des renseignements sur les traitements ou encore des contacts avec d'autres malades (Romeyer, 2008). Pour les Sciences de l'Information et de la Communication, la santé est un domaine illustrant les évolutions du processus de recherche d'information (Boubée et Tricot, 2010). En effet, ce champ « est tellement important qu'il influence même le domaine de la recherche d'information en général [...] sur à peu près tous les aspects : besoin d'information, tâche de recherche d'information, démarche de recherche d'information, évaluation des sources, etc. » (Boubée et Tricot, 2010).

Avec le succès d'Internet (Renahy *et al.*, 2007) et plus précisément des forums de santé (Paganelli et Clavier, 2010), se pose la question de l'évaluation de l'information recherchée. En effet, dans ces dispositifs, il n'est pas possible de mobiliser les critères de pertinence précaunisés par les différents modèles représentant le processus de recherche d'information (auteur, date, sources, etc.). Nous nous intéressons, par conséquent, aux critères de pertinence mobilisés par les malades pour valider les informations médicales recherchées dans des forums. En effet, puisqu'il n'est pas possible de mobiliser les critères traditionnels tels que la connaissance de l'auteur, la source, ou encore l'adéquation avec le domaine, nous nous demandons comment les malades valident les différentes informations trouvées.

Lors d'une maladie grave, rare, ou chronique, la motivation des malades à utiliser un forum de santé est informationnelle mais également émotionnelle (Paganelli et Clavier, 2010). Cela nous amène à nous intéresser à l'émotion non pas seulement comme raison pour rechercher de l'information dans les forums de santé mais également comme critère de pertinence pour évaluer les recherches faites dans des dispositifs. Étudier l'émotion comme critère de pertinence est une approche originale car cela permet de mettre en lumière les mutations du processus de recherche d'information et à terme, proposer un nouveaux paradigme de recherche d'information.

Pour répondre à cette question, nous formulons quatre hypothèses : a) les malades utilisent le forum de santé pour des raisons émotionnelles mais également parce qu'il s'apparente à un dispositif de recherche d'information simple à utiliser et non contraignant, b) dans les forums de santé l'évaluation de l'information se fait de manière affective, c) Les malades sont à la recherche d'astuces de vie avec la maladie, d) lorsqu'il y a de l'information médicale, elle est donnée sur un ton émotif.

Afin d'affirmer ou infirmer ces hypothèses, nous avons mobilisé deux méthodes de travail empirique : une méthode d'expérimentation-entretien et une méthode de traitement automatique de la langue. Dans ce travail, nous présenterons d'abord notre posture théorique puis la méthode de traitement automatique de la langue employée. Nous expliciterons, donc, la catégorisation semi-supervisée utilisée afin de classer les différents types de messages et l'utilisation du logiciel NooJ afin d'étudier la proximité entre des termes émotifs et des termes médicaux. Enfin, dans la dernière partie nous présenterons les premiers résultats obtenus.

## 2 État de l'art

### 2.1 Cadre

Plusieurs modèles décrivent les différentes étapes adoptées par un utilisateur lors d'une recherche d'information (Barry et Schamber, 1998), (Guthrie, 1988), (Kuhlthau, 1991), etc. Ces modélisations sont majoritairement issues des Sciences de l'Information et de la Communication et de la Psychologie Cognitive.

Les Sciences de l'Information et de la Communication, mettent l'accent sur le fait que la recherche d'information est un processus dans le sens où il s'agit d'une « suite continue de faits présentant une unité » (Boubée et Tricot, 2010). Il s'agit d'une activité dynamique et évolutive car humaine. Cependant, pour la Psychologie Cognitive, l'activité de recherche d'information n'existe pas intrinsèquement mais dépend de contextes d'utilisation différents. L'individu réalise une tâche et s'aperçoit qu'il manque de connaissances pour la mener à bien. Il décide donc de pallier ce manque et de rechercher les informations qui lui font défauts. Mais pour cela il doit posséder des connaissances sur le contenu qui lui manque car, pour le domaine de la Psychologie Cognitive, il n'est pas possible de rechercher une information sans avoir un minimum de connaissances sur le domaine.

Quelque soit la posture disciplinaire de ces modèles, tous accordent une place importante à l'évaluation de la pertinence de l'information ou du document sélectionné. Ils mettent en évidence des critères de confiance que les individus doivent utiliser pour leur recherche. Ainsi, des critères tels que l'adéquation au thème recherché, la clarté des données, la source d'où provient le document, l'auteur, sont autant d'indices qui peuvent être mobilisés pour évaluer l'information (Barry et Schamber, 1998).

Cependant, force est de constater que les nouvelles technologies ont transformé les pratiques de recherche d'information (Boubée et Tricot, 2010). En effet, différentes enquêtes réalisées en Europe et aux États-Unis, mettent en avant le fait qu'Internet est de plus en plus utilisé pour rechercher de l'information médicale (Aube et Thoër, 2010).

Le champ de la santé est un domaine dans lequel ces évolutions sont le plus visibles (Boubée et Tricot, 2010). Le succès de la recherche santé en ligne est tel que « Google Flu » détecte les épidémies de grippe quelques jours avant leur survenue en analysant les requêtes des malades (Ginsberg *et al.*, 2008). Pour certains chercheurs, Internet conduirait à trois types de recherche de la part des personnes malades (Aube et Thoër, 2010) :

- une « logique professionnelle » : le malade cherche à obtenir des informations pour lui permettre d'appréhender la décision des médecins.
- une « logique consumériste » : le malade utilise Internet pour confronter les informations sur sa maladie, les différents traitements et ainsi faire un choix.
- une « logique communautaire » : le malade participe à des réseaux communautaires, tels que des forums, où s'exercent à la fois échange d'informations de santé et informations personnelles.

Nous nous intéressons à cette dernière logique car, dans les pays occidentaux, les forums de santé rencontrent beaucoup de succès (Paganelli et Clavier, 2010).

Un forum peut se définir comme un dispositif de communication médiaté par ordinateur, asynchrone et anonyme (Blanchard, 2007). C'est un lieux d'échange où l'information est

construite collectivement et le contenu auto-structuré. En effet, les participants sont à la fois producteurs et usagers de l'information (Clavier *et al.*, 2010). Par conséquent, l'inconvénient majeur des informations délivrées dans un forum, que ce soit des informations sur un utilisateur ou dans le contenu du message, est l'impossibilité de mobiliser les critères d'évaluation traditionnellement utilisés, même si le forum est hébergé dans un site fiable. Se pose alors la question de savoir pourquoi les forums de santé ont autant de succès.

Il apparaît que ce dispositif présente, un double objectif : un soutien informationnel et émotionnel (Gaglio, 2010), (Paganelli et Clavier, 2010). Avant de poursuivre, il nous semble important de définir ce que nous entendons par émotion car bien que ce concept ait été très étudié ces dernières années, il reste assez difficile à appréhender de façon claire (Novakova et Tutin, 2009a). Il existe plusieurs systèmes de définitions de l'émotion. Ainsi, les spécialistes ne sont pas toujours d'accord sur les définitions à adopter. Dans notre travail, nous avons choisi de traiter de l'émotion au sens large du terme. Cela signifie que nous englobons l'émotion (joie, tristesse, haine, etc.) et les sentiments (amertume, crainte, honte, etc.) (Blumenthal, 2009).

Pour les Sciences de l'Information de la Communication, la prise en compte de l'émotion dans le processus de recherche d'information n'est pas nouveau. Néanmoins, force est de constater que l'émotion n'est étudiée que en tant qu'émotion ressentie par un individu dans les étapes de sa recherche ou dans des lieux documentaires. À titre d'exemple, nous pouvons citer le modèle Information Search Process créé en 1991 (Kuhlthau, 1991). Contrairement aux précédentes modélisations du processus de recherche d'information, celui-ci mettait en évidence les rôle des émotions et de la confiance en soi dans les étapes d'une recherche documentaire. Le modèle ISP se compose de six phases :

- une phase d'initiation qui correspond à l'apparition d'une sensation d'incertitude et d'attente informationnelle,
- une phase de sélection où l'individu définit ce qu'il veut rechercher et a donc un sentiment d'optimisme,
- une phase d'exploration qui correspond à une phase de réflexion et d'acquisition d'information
- une phase de formulation. Cette étape correspond à une phase de sérénité, d'incertitude diminuée et une augmentation de confiance dans sa recherche,
- une phase de collecte des données qui renforce la confiance de l'individu au fur et à mesure qu'il trouve des informations,
- Enfin, une phase de présentation des résultats. L'individu n'a pas d'incertitude mais se trouve dans un état de satisfaction ou d'insatisfaction.

Au cours de ces étapes, l'individu va chercher à réduire son principe d'incertitude. L'incertitude est un état cognitif vague, difficile à formuler pour l'individu et souvent accompagné de sentiments de confusion ou de frustration (Boubée et Tricot, 2010).

Plus récemment, d'autres travaux, se sont interrogés sur le rôle de l'émotion dans la détermination des prises de décision (Julien *et al.*, 2005) ou encore sur le lien entre contrôle de l'émotion et tâche de recherche en ligne (Kim, 2008). Il semblerait qu'Internet favorise le lien entre recherche d'information et émotion (Abbasi *et al.*, 2008). Leur enquête met en avant la richesse émotionnelle des discours sur Internet. Par conséquent, ces dernières années, l'analyse des sentiments a été appliquée à différentes formes du web et dans différents domaines. Par exemple, l'analyse des sentiments dans les forums a déjà été traitée en linguistique mais dans le but d'améliorer la qualité d'un système d'extraction de sentiment grâce au traitement

automatique de la langue naturelle (Maurel *et al.*, 2008). Plus généralement, d'autres travaux en linguistique cherchent à caractériser les émotions, étudient un ou deux champs sémantiques de l'émotion ou encore travaillent sur l'extraction automatique des affects (Novakova et Tutin, 2009a), (Barbé, 2007). Nous n'allons pas étudier l'émotion de cette manière là mais nous aider des précédents travaux pour nous intéresser aux indices émotionnels contenus dans les messages des forums de santé et nous demander si l'émotion peut devenir un critère de pertinence.

## 2.2 Hypothèses

A/ Les malades ou proches de malades cherchent principalement des astuces de vie avec la maladie

Les travaux réalisés précédemment (Romeyer, 2008), (Lemire, 2009), (Clavier *et al.*, 2010) mettent en avant le fait que les internautes, et plus précisément les malades ou proches de malades, recherchent principalement des informations pratiques concernant la vie avec la maladie ; ceux-ci veulent, par exemple, obtenir des astuces sur la gestion des effets secondaires d'un médicament pour la thyroïde. Or, d'autres études ont des résultats contraires et révèlent que : « *Dans son panorama de la « consommation d'informations » sur l'Internet médical, D. Nabarette (Nabarette, 2002) y relie cinq buts : s'informer pour connaître (maladies, traitements...), choisir (praticiens, plans de soins...), superviser l'action des professionnels, produire (un diagnostic...), se coordonner dans le cadre de la relation patient/médecin. Notre enquête révèle une prééminence de la première catégorie (Gaglio, 2010) ».*

Il nous semble important de pouvoir déterminer quel type d'information vont chercher les malades dans les forums de santé. Or, les résultats des travaux réalisés jusqu'à présent ne vont pas toujours dans le même sens.

Après une première observation et un croisement des lectures qui ont pu être faites, nous émettons l'hypothèse que les malades sont plutôt à la recherche d'astuces de vie avec la maladie. Afin de typer les informations contenues dans notre corpus, nous réalisons, à l'aide d'un partenariat avec le Laboratoire d'Informatique d'Orléans, une catégorisation semi-supervisée.

B/ L'information médicale est donnée sur un ton émotif

Par information médicale nous voulons dire information médicale (professionnelle) et information de santé (grand public, non spécialisée) (Romeyer, 2008). Plusieurs enquêtes telles que celle de Joëlle Kivits (Kivits, 2006), d'Hélène Romeyer (Romeyer, 2008), ou encore Gérard Gaglio (Gaglio, 2010) rappellent que la relation avec le praticien ne change pas dans le sens où les professionnels de santé restent les principales sources d'obtention de l'information médicale (Vercher et Touboul, 2011). Les forums de santé ne sont donc pas des lieux uniquement faits pour échanger de l'information médicale. Étant donné que le forum est un lieu d'échange avec d'autres malades, nous pensons que les informations médicales sont délivrées sur un ton émotif car lié à l'expérience et au ressenti que le malade éprouve face à la maladie. Afin d'infirmier ou affirmer cette hypothèse, nous allons nous servir du logiciel NooJ et observer la proximité entre les termes d'émotion et les termes médicaux.

Les deux hypothèses suivantes ne font pas l'objet d'un traitement automatique de la langue mais d'une expérimentation complétée par des entretiens. Cela nous permet à la fois de

dépasser le biais méthodologique puisque les entretiens ne sont finalement que des discours sur leurs pratiques faites par les individus interrogés, et non l'observation du processus en train de se faire (Blanchard, 2007). Néanmoins, nous avons choisi de les évoquer brièvement, afin de présenter de manière approfondie le cadre théorique sur lequel la recherche est fondée.

#### C/ Motivation des malades à utiliser le forum de santé

Nous pensons que, outre la dimension émotionnelle et le fait de pouvoir s'exprimer entre personnes malades, que l'on peut retrouver dans d'autres dispositifs, le forum de santé donne l'impression d'être un outil de recherche d'information moins contraignant que les autres pour deux raisons. La première est que l'on peut poser les questions que l'on veut sans avoir besoin de naviguer entre plusieurs sources pour obtenir une réponse. Il est également possible de trouver des informations sans avoir besoin de participer, de s'inscrire. La lecture des messages suffit. La deuxième raison, induite par les observations qui ont pu être menées lors de la composition du corpus de travail, est que lorsque les malades obtiennent une réponse à leur question ou n'en ressentent plus le besoin, il est plus facile de ne plus participer au forum. Nous nous appuyons ici sur les travaux d'H. Romeyer (Romeyer, 2008). En effet, son enquête de 2008 a révélé que les malades ne cherchent pas à établir de lien avec les autres malades. C'est la raison pour laquelle nous pensons que dès qu'ils ont obtenu satisfaction ou non avec leur recherche, les malades ne participent plus ou ne lisent plus les messages des forums de santé .

#### D/ Évaluation affective de l'information faite dans les forums de santé

Nous considérons le forum de santé comme une source d'information à dimension affective dans le sens où les malades partagent aussi bien des informations médicales que personnelles (Clavier *et al.*, 2008), (Clavier *et al.*, 2010), (Paganelli et Clavier, 2010). Les messages sont un mélange d'informations objectives et émotionnelles. Pour nous, l'émotion change la perception qu'un internaute a du message et modifie les critères de sélection habituellement utilisés. Cela signifie que les malades s'attachent aux indices émotionnels contenu dans un message pour l'évaluer. Plus un message contiendra d'indices émotionnels (mots, ponctuation exagérée, smileys), plus les individus auront confiance dans l'information véhiculée. Cela leur donnerait l'impression que le message n'a pas pu être inventé. Par conséquent, nous pensons que plus un message contiendra des termes émotifs, plus les malades le considéreront comme vrai, comme digne de confiance car ils se reconnaissent dans les termes émotifs utilisés.

### 3 Méthodologie du travail empirique

Dans cette partie, nous présentons la méthodologie de traitement automatique de la langue mobilisée pour répondre à nos deux premières hypothèses (A/ et B/).

Le nombre de forums de santé francophones existant sur Internet est trop élevé pour qu'il soit possible de les recenser de manière exhaustive. Notre parti-pris a donc été de les répertorier à l'aide du Catalogue et Index des Sites Médicaux de langue Française (Cismef<sup>1</sup>), du site Health On the Net Foundation (Honcode<sup>2</sup>) et via une requête sans la zone de recherche de deux moteurs de recherche.<sup>3</sup> Le Cismef est un projet initié par le Centre Hospitalier Universitaire de Rouen depuis

---

1. <http://www.chu-rouen.fr/cismef/>

2. <http://www.hon.ch/HONcode/French/>

3. « Forum+santé » dans Google et Yahoo

1995. Des documentalistes et médecins recensent des sites médicaux répondant au référentiel des critères de qualité de l'information de santé sur le net (netscoring<sup>4</sup>). Le netscoring a été développé par un groupement de professionnels de santé, bibliothécaires médicaux et juristes. Le Honcode quant à lui, est financée par le Canton de Genève, des projets Européens, la Haute Autorité de Santé de France (HAS) et la fondation Provisu. Sa mission est de guider les utilisateurs d'internet vers des sources d'informations médicales et de santé fiables.

Les forums accessibles via ces trois types de sources, ont été catégorisés selon leur spécialisation : forums de santé généralistes, spécialisés dans une maladie, modérés, modérés par un médecin. Nous avons sélectionné des fils de discussion traitant de maladies rares, graves, ou chroniques car nous pensons qu'ils sont plus à même de fournir à la fois des données informationnelles et émotionnelles. D'autre part, il est extrêmement difficile de trouver un forum spécialisé sur le rhume par exemple. Or, nous voulons pouvoir effectuer des comparaisons entre forums généralistes et forums spécialisés.

Nous travaillons sur un corpus de 2481 messages répertoriés comme suit :

- Forums généralistes modérés par un médecin= catégorie GD  
Forum Atoute (8 fils de discussion, 631 messages)
- Forums généralistes modérés= catégorie GM  
Forums Au féminin (1 fil de discussion, 101 messages), Doctissimo (1 fil de discussion, 159 messages), E-santé (1 fil de discussion, 172 messages) et Santé médecine (2 fils de discussion, 150 messages).
- Forums spécialisés modérés par un médecin= catégorie SD  
Forums Ligue Cancer (6 fils de discussion et 329 messages) et Traitement du pied bot à la naissance (14 fils de discussions, 292 messages).
- Forums spécialisés modérés= catégorie SM  
Forums maladies Lysosomales (2 fils de discussion, 153 messages), Renaloo (2 fils de discussion, 157 messages), Solhand (3 fils de discussion, 173 messages) Vivre sans thyroïde (1 fil de discussion, 164 messages).

Il aurait été intéressant pour notre enquête de pouvoir également analyser des forums de santé non modérés mais nous n'en avons pas trouvé.

Un traitement semi-automatique (avec le logiciel Python<sup>5</sup>) a été réalisé sur notre corpus. Nous avons en effet, répertorié et classé les informations externes au contenu des messages tel que le titre des fils de discussion, la source, l'url ou encore le nombre de messages. Nous avons souhaité traiter ces informations afin de pouvoir effectuer des comparaisons entre les différents fils de discussion mais également afin d'avoir une trace d'archivage. Nous avons ensuite créé un filtre Python afin de numéroter automatiquement les messages et de ne conserver que le texte. L'inconvénient majeur est la suppression des smileys car le filtre ne conserve que le texte et non les images. Les smileys ont ensuite été intégrés au corpus en tant que signification (par exemple : SMILEY=SOURIRE).

---

4. <http://www.chu-rouen.fr/netscoring/>

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

### 3.1 Catégorisation semi-supervisée

La catégorisation semi-supervisée se fait en partenariat avec le Laboratoire d'Informatique Fondamentale d'Orléans<sup>6</sup>. La classification semi-supervisée est une hybridation entre la classification supervisée (ou classification par apprentissage) et la classification par extraction (non-supervisée). La classification par apprentissage permet d'établir des classifications à partir de documents pré-classés tandis que, la classification par extraction consiste en une fouille de données. Cela permet de pallier le nombre trop élevé de documents en les classant de manière automatique. Faire du semi-supervisé dans le cadre de ce travail, permet d'orienter la classification et de faire ainsi du clustering semi-supervisé avec plusieurs vues : lexicale, smileys, ponctuation exagérée. L'objectif est donc de répondre à l'hypothèse concernant le type d'information recherchée par les malades (hypothèse A/) et de pouvoir ainsi catégoriser de manière semi-automatique les messages. L'originalité de cette méthode est également de permettre de travailler en parallèle sur les différentes manières de transmettre l'émotion dans un forum de santé, à savoir les mots, la ponctuation et les smileys. En effet, quelques fois l'émotion n'est pas transmise par les termes émotifs.

Je conduis pour aller au boulot , mais c'est mais arriver de m'endormir en voiture après le travail SMILEY=PEUR fatigue .  
Je vais a Créteil environ tout les 18 mois comme la maladie et stable ( pour le moment ,rien ne dit que sa peut évaluer !!!!!

Forum spécialisé modère Solidarité Handicap, Fil de discussion Syndrome de Poems (numéro 1), Message SMSH01-0014

Nous avons, dans un premier temps, créé une liste de catégories d'après le compte rendu du colloque : web participatif et santé (Paganelli et Clavier, 2010), d'une exploration de notre corpus, et de forums de santé sur Internet. La raison pour laquelle nous avons décidé de procéder ainsi est que nous ne voulons pas avoir de catégories correspondant uniquement à notre corpus car cela risquerait de fausser les résultats.

Une pré-catégorisation manuelle de 586 messages du corpus a ainsi été faite et envoyée au laboratoire d'Informatique Fondamentale d'Orléans. 9 « items » ont ainsi pu être mis en évidence :

- Astuces face à la maladie
- Autre  
Demande de présentation de soi ou d'un proche sans évoquer son récit de vie, présentation de soi sans évoquer son récit de vie avec la maladie, encouragements sans apports d'informations, demande de contact physique ou virtuel, réponse à une demande de contact, proposition de contact sans demande préalable dans les autres messages.
- Feed Backs

---

6. le LIFO



- Hors sujet
- Informations ponctuelles
- Informations sur le traitement suivi
- Information sur la maladie en général
- Ressenti vis-à-vis de la maladie
- Témoignage

Ces classes ont ensuite été divisées en sous-catégories. Le typage des messages et le corpus de travail ont ensuite été envoyés au LIFO pour traitement informatique.

### 3.2 Traitement automatique de la langue avec NooJ

Une phase de lemmatisation précède le traitement du corpus avec NooJ<sup>7</sup>. Elle permet de ramener les termes à leur forme de lemme, mais également de corriger les fautes (nombreuses dans les forums). Lemmatiser le corpus permet, en effet, de mettre en évidence les fautes puisque le logiciel ne reconnaîtra pas les mots mal orthographiés et les catégorisera comme « unknowns ». La lemmatisation a été faite à l'aide du logiciel Treetagger<sup>8</sup>. Ne pas corriger les erreurs auraient compliqué l'analyse du corpus puisque les termes mal orthographiés n'auraient pas pu être analysés par NooJ.

NooJ est un environnement linguistique créé en 1993, d'abord sous le nom d'INTEX, par M. Silberztein (Svetla *et al.*, 2007). Il permet de formaliser les langues naturelles, de développer des applications de traitement automatique du langage mais également de travailler sur de gros corpus (Yamouni-Aoughlis, 2010).

Afin de pouvoir affirmer ou infirmer notre hypothèse sur la proximité de termes émotifs et de termes médicaux (hypothèse B/), nous avons créé des grammaires afin de localiser les adjectifs, noms et verbes émotifs. La figure 1 présente un extrait de la grammaire faite pour les adjectifs.

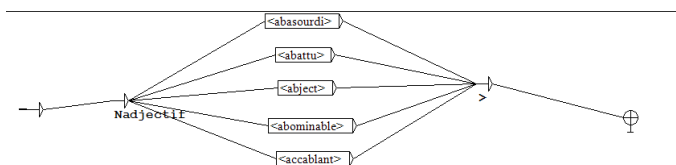


FIGURE 1 – Extrait de la grammaire NAdjectif

7. <http://www.nooj4nlp.net/pages/nooj.html>

8. <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

Les affects ne constituent pas une classe homogène (Novakova et Tutin, 2009b). Certains se rapprochent des sentiments (par exemple honte), d'autres des émotions (par exemple peur). Néanmoins, force est de constater que nous ne travaillons pas sur la catégorisation de l'émotion en tant que tel. Pour le moment, nous n'étudions pas la valeur des termes utilisés par les internautes mais nous nous servons de listes de termes émotifs pour analyser la manière dont est transmise l'information médicale. Le but est d'analyser la proximité entre les termes émotifs et les termes médicaux. Les trois grammaires (Nadjectif, Nnom et Nverbe) permettent de mettre en lumière les phrases contenant des affects. Une comparaison est ensuite faite afin de vérifier la présence ou non de termes médicaux transmis aussi bien sous la forme spécialisée (Carcinome à cellules claires) que sur le mode grand public (Tumeur rénale).

L'inconvénient de cette méthode est de ne pas pouvoir mettre en lumière des expressions ne mobilisant pas a priori d'affects mais transmettant néanmoins une émotion : en avoir gros sur la patate. Pour pallier ce désagrément, nous établirons une liste d'expressions reflétant des émotions et nous créerons une quatrième grammaire.

Enfin, il nous semble important de préciser, que les paramètres standards de Treetagger et NooJ ont été utilisés pour traiter le corpus.

## 4 Premiers résultats et discussion

### 4.1 Les malades recherchent des témoignages

Nous sommes en attente des résultats de la catégorisation semi-supervisée. Néanmoins, le typage manuel transmis au LIFO, nous permet d'affirmer, à priori, que les malades ne sont pas à la recherche d'astuce de vie (un saignement du nez provoqué par un médicament pour le foie). Les messages analysés contiennent essentiellement des témoignages, des récits de vie liés à la maladie.

En effet, sur 586 messages traités, 150 proviennent d'échanges de témoignages (17 demandes, 24 apports de témoignages suite à une demande, 109 apports de témoignages sans demande particulière au préalable).

116 messages sont liés au ressenti émotionnel vis-à-vis de la maladie, 95 correspondent à des demandes de contact et/ou encouragements, 93 concernent des échanges d'information sur les traitements, 66 sont hors sujet. 47 messages sont des retours sur les réponses (feedback), 46 concernent des échanges d'informations ponctuelles, d'astuce de vie et 19 échanges traitent de la maladie en général.

Il s'agit majoritairement d'échanges de témoignages et de ressentis avec la maladie. Cela confirme l'affirmation de Céline Paganelli et Viviane Clavier (Paganelli et Clavier, 2010), à savoir que les motivations des malades sont d'ordre émotionnel. Par contre cela remet en cause les travaux faits par (Clavier *et al.*, 2010) puisque les résultats de leur étude montrent que les malades sont plutôt à la recherche d'astuce de vie avec la maladie.

Le traitement entier du corpus doit être attendu pour généraliser les résultats. Néanmoins les premières données montreraient que les malades n'utilisent pas les forums dans un but informationnel. Ces résultats doivent être mis en parallèle avec la thèse réalisée par (Sénis, 2003), pour qui les malades restent attachés au monde médical pour valider les informations recherchées sur Internet. Cela expliquerait également un autre de ses résultats, à savoir que les

patients sont hostiles à la présence de médecins comme intervenants dans les forums. Ce ne sont pas, pour eux des lieux de recherche mais d'échange de leur vie.

## 4.2 L'information médicale n'est pas donnée sur un ton émotif

La question de l'émotion dans les forums est majoritairement traitée selon une approche communicationnelle dichotomisante. Ce type de communication favoriserait l'émergence d'une communication émotive entre écrit et oral (ponctuation exagérée, smileys) tandis que, à l'inverse, d'autres études considèrent que la communication dans les forums et plus généralement la communication médiatisée par ordinateur, entraverait la communication émotionnelle (Atifi *et al.*, 2010). Les premiers résultats obtenus avec l'analyse du corpus par NooJ, confirment la catégorisation semi-supervisée du contenu et laissent à penser que la communication émotionnelle existe bien mais elle est réservée aux témoignages, aux expériences de vie avec la maladie et non avec la diffusion d'informations médicales. Ainsi, sur 200 messages analysés, 153 contiennent de l'émotion que cela soit par les smileys, les mots ou la ponctuation exagérée :

```
Ma mère se fait opérer mercredi et depuis 2 jours,  
comme vous, son moral est au plus bas.  
Elle angoisse et est très négative.  
Heureusement elle n'a ni vertige ni maux  
de tête mais elle est très fatiguée
```

Forum Généraliste Modéré par un médecin Atoute,  
Fil de discussion Opération d'un méningiome, Message GDATA02-0002

Sur les 57 messages restants, 22 messages contiennent des informations objectives comme des informations pratiques par exemple et seulement 35 messages contiennent des informations médicales. D'après nos premiers résultats, les informations médicales ne semblent pas être transmises sur un ton émotionnel :

```
mon médecin me prescrit un scanner et le 5 Juillet : méningiome  
4 x 3,5 x 2,5 avec œdème. Premier contact avec un  
chirurgien de la Pitié-Salpêtrière (en vacance) RDV pour  
consultation le 11/8, diagnostique : opération pour le 30  
Août. Je suis entrée à l'hôpital le 29 car j'avais passé les  
examens demandés chez moi. Le 29 au soir double douche avec  
shampooing à la Bétadine, léger somnifère et le lendemain  
matin 6H.
```

Forum Généraliste Modéré par un médecin Atoute,  
Fil de discussion Opération d'un méningiome, Message GDATA02-0006

Les données recueillies, donnent l'impression que l'émotion est vraiment réservée à l'apport d'informations personnelles. On peut alors s'interroger sur l'émotion comme critère de pertinence. En effet, cela laisse à penser que les individus ont peur de ne plus être crédibles s'ils délivrent une information médicale avec beaucoup d'émotion. Le forum de santé serait, avant d'être un lieu de recherche d'information, un dispositif de maintien de lien entre des personnes vivant la même expérience de maladie.

## 5 Conclusion

Ce travail s'intéresse à l'analyse de l'émotion dans l'étude des pratiques de recherche d'information, que les spécificités propres aux forums font évoluer (Paganelli et Clavier, 2010) et tout particulièrement dans le domaine de la santé où la fiabilité des informations médicales recherchée est difficilement vérifiable pour des malades. C'est la raison pour laquelle il est intéressant d'étudier comment l'émotion peut devenir un critère de pertinence.

La méthodologie présentée dans l'article a mis en évidence que les malades semblent plutôt être à la recherche de récit de vie avec la maladie. Les conversations, lorsqu'elles concernent la maladie en elle-même sont plus objectives. L'émotion intervient lorsqu'il s'agit d'échange plus personnels sur la vie avec la maladie. Les premiers résultats nous laissent à penser que l'émotion n'est pas un critère de pertinence en tant que tel car elle n'intervient pas dans les échanges d'informations médicales mais dans l'échange de témoignages, de ressenti vis-à-vis de la maladie. Cela confirme les résultats de l'enquête menée par (Clavier *et al.*, 2008), à savoir que les malades ne viennent pas pour rechercher des informations médicales *stricto sensu* mais pour partager leur expérience avec la maladie. La mutation du processus de recherche d'information interviendrait dans le sens où des personnes recherchent des informations filtrées par l'expérience de l'individu et non plus des informations objectives. Néanmoins, les résultats sont à compléter par la continuation du travail de terrain (catégorisation semi-supervisée et traitement NooJ de tout le corpus), mais également par la phase d'expérimentation-entretien afin d'appréhender le discours des malades mais aussi leur manière de faire.

## Remerciements

Un grand merci à Julien Corman qui a lemmatisé le corpus de travail avec Treetagger et à Agnès Tutin pour nous avoir transmis les listes de termes émotifs.

## Références

ABBASI, A., CHEN, H. et SALEM, A. (2008). Sentiment analysis in multiple languages : Feature selection for opinion classification in web forums.

- ATIFI, H., GAUDUCHEAU, N. et MARCOCCIA, M. (2010). L'expression et le rôle des émotions dans les forums de discussion. In YASRI-LABRIQUE, E., éditeur : *Les forums de discussion : agora du XXIème siècle ? Théorie, enjeux et pratiques discursives*, pages 71–87. Paris.
- AUBE, S. et THOËR, C. (2010). La construction des savoirs relatifs aux médicaments sur internet : étude exploratoire d'un forum sur les produits amaigrissants utilisés sans supervision médicale. In RENAUD, L., éditeur : *Les médias et la santé : de l'émergence à l'appropriation des normes sociales*, pages 239–266. Québec.
- BARBÉ, M. (2007). *Analyses linguistiques et modélisations des connaissances en vue d'un traitement automatique des e-mails entrants : vers un système de veille dans le tourisme aérien*.
- BARRY, L. C. et SCHAMBER, L. (1998). User's criteria for relevance evaluation : a cross-situational comparison. *Information Processing and Management*, 34(2-3):219–236.
- BLANCHARD, G. (2007). *La communication politique partisane sur Internet : des pratiques et des stratégies nouvelles ?*
- BLUMENTHAL, P. (2009). Les noms d'émotion : trois systèmes d'ordre. In NOVAKOVA, I. et TUTIN, A., éditeurs : *Le lexique des émotions*, pages 41–79. Grenoble.
- BOUBÉE, N. et TRICOT, A. (2010). *Qu'est-ce-que rechercher de l'information ? : état de l'art*, page 204. Villeurbanne.
- CLAVIER, V., PAGANELLI, C., MANES-GALLO, M.-C., MOUNIER, E., ROMEYER, H. et STAIH, A. (2010). Dynamiques interactionnelles et rapport à l'information dans les forums de discussion médicale. In MILLERAND, F. e. a., éditeur : *Web social : mutation de la communication*, pages 297–314. Québec.
- CLAVIER, V., PAGANELLI, C., MANES-GALLO, M.-C., MOUNIER, E., ROMEYER, H. et STAIH, A. (Québec 6-7 mai 2008). Web participatif et santé : de nouveaux rapports à l'information ? In *Web participatif : mutation de la communication ?*
- DINET, J. et ROUET, J.-F. (2002). La recherche d'information : processus cognitifs, facteurs de difficultés et dimension de l'expertise. In PAGANELLI, C., éditeur : *Interaction Homme-Machine et recherche d'information*, pages 113–161. Paris.
- GAGLIO, G. (2010). Consommation d'informations sur internet et modulation de la relation aux médecins. le cas d'aïdantes de malades atteints d'une pathologie lourde. *Sociologies Pratiques*, 1(20):63–74.
- GINSBERG, J., MOHEBBI, M.-H., P. R.-S. B. L., SMOLINSKI, M.-S. et BRILLIANT, L. (2008). Detecting influenza epidemics using search engine query data.
- GUTHRIE, J. T. (1988). Locating information in documents : Examination of a cognitive model. *Reading Research Quarterly*, 23:178–199.
- JULIEN, H., MCKECHNIE, L. E. F. et HART, S. (2005). Affective issues in library and information science systems work : A content analysis. *Library and Information Science Research*, 27(4):453–466.
- KIM, K. (2008). Effects of emotion control and task on web searching behavior. *Information Processing and Management*, 44(1):373–385.
- KIVITS, J. (2006). Informed patients and the internet : a mediated context for consultations with health professionals. *Journal of Health Psychology*, 11(2):269–282.
- KUHLTHAU, C. C. (1991). Inside the search process : Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5):361–371.

- LEMIRE, M. (2009). Internet et responsabilisation : perspective de l'utilisateur au quotidien. *Santé Publique*, 21:13–25.
- MAUREL, S., CURTONI, P. et DINI, L. (Fontainebleau - 2008). L'analyse des sentiments dans les forums. In *Atelier Fouille des Données d'Opinion*.
- NABARETTE, H. (2002). L'internet médical et la consommation d'information par les patients. *Réseaux*, 4:249–286.
- NOVAKOVA, I. et TUTIN, A. (2009a). *Le lexique des émotions*. Grenoble.
- NOVAKOVA, I. et TUTIN, A. (2009b). Les émotions sont-elles comptables. In NOVAKOVA, I. et TUTIN, A., éditeurs : *Le lexique des émotions*, pages 65–79. Grenoble.
- PAGANELLI, C. et CLAVIER, V. (2010). Le forum de discussion : une ressource informationnelle hybride entre information grand public et information spécialisée. In YASRI-LABRIQUE, E., éditeur : *Les forums de discussion : agora du XXIème siècle ? Théorie, enjeux et pratiques discursives*, pages 39–54. Paris.
- RENAHY, E., PARIZOT, I. et CHAUVIN, P. (2007). Whist : a web-based survey on health information seeking on internet in france. Rapport technique, Paris, INSERM.
- ROMEYER, H. (2008). Tic et santé : entre information médicale et information de santé.
- SVETLA, K., MAUREL, D. et SILBERZTEIN, M. (2007). *Formaliser les langues avec l'ordinateur : de INTEX à NooJ*.
- SÉNIS, F. (2003). *Pourquoi accéder à l'information médicale sur Internet par le biais des groupes de discussion ? Qualité, centres d'intérêt et motivations des participants aux forums médicaux. À propos du forum Usenet Fr.bio.medecine*.
- VERCHER, E. et TOUBOUL, A.-L. (2011). L'information santé : entre vulgarisation scientifique et circulation communautaire, le cas des sites santé 2.0. In *Internet et santé : regards croisés France/Québec*. Lyon : Médiathèque Université Lyon 1, 21 février.
- YAMOUNI-AOUGHILIS, Y. (2010). *Construction d'un dictionnaire électronique de terminologie informatique et analyse automatique de textes par grammaires locales*.