



HAL
open science

Les interjections pour détecter les émotions

Amel Fraise, Patrick Paroubek

► **To cite this version:**

Amel Fraise, Patrick Paroubek. Les interjections pour détecter les émotions. 22e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2015), Jun 2015, Caen, France. hal-01617186

HAL Id: hal-01617186

<https://hal.science/hal-01617186>

Submitted on 16 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les interjections pour détecter les émotions

Amel Fraïsse Patrick Paroubek
LIMSI-CNRS, Bât. 508 Université Paris-Sud, 91403 Orsay Cedex, France
fraïsse@limsi.fr, pap@limsi.fr

Résumé. Bien que les interjections soient un phénomène linguistique connu, elles ont été peu étudiées et cela continue d'être le cas pour les travaux sur les microblogs. Des travaux en analyse de sentiments ont montré l'intérêt des émoticônes et récemment des mots-dièses, qui s'avèrent être très utiles pour la classification en polarité. Mais malgré leur statut grammatical et leur richesse sémantique, les interjections sont restées marginalisées par les systèmes d'analyse de sentiments. Nous montrons dans cet article l'apport majeur des interjections pour la détection des émotions. Nous détaillons la production automatique, basée sur les interjections, d'un corpus étiqueté avec les émotions. Nous expliquons ensuite comment nous avons utilisé ce corpus pour en déduire, automatiquement, un lexique affectif pour le français. Ce lexique a été évalué sur une tâche de détection des émotions, qui a montré un gain en mesure F1 allant, selon les émotions, de +0,04 à +0,21.

Abstract.

Interjections for emotion detection

Although interjections have been recognized as linguistic phenomena for a long time, they have somehow been rarely studied and continue to be left aside in works dealing with microblogs. Users of this new kind of communication platforms have popularized widely the use of linguistic constructions, like emoticons or interjections. In spite of their grammatic status and semantic richness for describing emotional states, interjections have been mostly ignored. In this article we show the importance of the role that interjections can play for detecting emotions. We detail how using interjections we have tagged automatically a French microblog corpus with emotion labels. Then we describe how we did deduce automatically from this corpus a fine-grained affective lexicon. The usefulness of the lexicon was evaluated in an emotion recognition task where, depending on the emotion, the F1-measure improvement ranged from +0.04 to +0.21.

Mots-clés : Interjections, détection des émotions, lexique affectif, analyse de sentiments, fouille d'opinions .

Keywords: Interjections, emotion recognition, affective lexicon, sentiment analysis, opinion mining .

1 Introduction

Le microbloggage est devenu aujourd'hui parmi les moyens de communication les plus populaires. En effet, en limitant la longueur de messages à quelques caractères, ce service permet de partager rapidement et facilement une information ou une opinion. Twitter¹ est la plus célèbre plateforme de microbloggage. Elle permet d'envoyer et de recevoir ce que l'on nomme des *tweets*, c'est-à-dire des messages courts dont la longueur est limitée à 140 caractères. D'après les derniers chiffres avancés par la compagnie de Twitter², la plateforme dénombre 284 millions d'utilisateurs actifs mensuels produisant au total environ 500 millions de tweets par jour en moyenne. Le public utilisant ce service est très varié et il croît de jour en jour. Dans la plupart des cas, ce sont des utilisateurs lambda qui souhaitent communiquer une information factuelle, ponctuelle, anecdotique, politique, people, marketing, etc. qui exploitent ce réseau. Mais, ce site communautaire qui a connu une très forte croissance ces dernières années est également utilisé par des professionnels liés pour la plupart à l'industrie des médias (CNN, BBC envoient des flashes d'information sur Twitter), des parties ou des hommes politiques (John Edwards et Barack Obama ont exploité twitter comme outil médiatique lors de la campagne pour l'élection présidentielle américaine de 2008), des personnalités tous domaines confondus à l'intention de leurs fans ou d'un public concerné.

En ce qui concerne, les langues utilisées sur Twitter, même si l'anglais demeure la langue la plus utilisée (34 % des mes-

1. <https://twitter.com/>

2. <https://about.twitter.com/company>

@AnneDuBois hihi ! ça y est ! on est arrivé !
@isa#TeamOlympe Le coup de chaud que j'ai eu. Mon ordi voulais plus démarrer. c'était ma barette de RAM installé récemment qui fonctionnait plus #ouf
@Flavien 2000 : communes, départements, régions. 2014 : intercommunalités en plus ; et bientôt regroupements de communes ! argh .
@Capu ce projet de loi est vraiment inutile, et pas de sens pfff

TABLE 1 – Des exemples de messages de Twitter avec des interjections onomatopéiques exprimant l'état émotionnel des utilisateurs.

sages en septembre 2013 soit 170 millions de Tweets quotidiens), il n'en demeure pas moins qu'aujourd'hui, +60 % des tweets sont rédigés dans d'autres langues. Ce flux incessant de données représente une source d'information multilingue intéressante pour analyser les sentiments et les opinions des utilisateurs à l'égard de différentes questions.

Du fait, de la taille réduite de messages de Twitter les utilisateurs utilisent très souvent des formes langagières qui, permettent à la fois une réelle concision de l'écriture et d'exprimer de façon très explicite leurs états émotionnels et affectifs tels que les émoticônes ou les interjections.

En effet, ces deux pratiques discursives permettent d'exprimer rapidement des émotions, d'indiquer l'attitude subjective du locuteur, toutes choses qui se font très facilement dans une interaction orale en face à face en recourant au système sémiotique mimo-gestuel, ou encore à l'intonation, mais qui sont moins évidentes à faire passer à l'écrit. Le tableau 1 montre des exemples de tweets contenant des interjections.

Nous nous intéressons dans ce travail, en particulier, aux interjections, utilisées sur le réseau Twitter comme marqueurs de subjectivité et particulièrement d'émotion. En effet, l'interjection est classée par les grammaires parmi les parties du discours, au même titre que le verbe, l'adverbe, l'adjectif, le substantif ou le déterminant. Mais, si elle a bien une étiquette terminologique qui semble lui accorder un statut grammatical, elle est souvent marginalisée dans les analyses malgré sa richesse sémantique.

Il s'agira d'abord d'étudier les caractéristiques des interjections, puis d'évaluer leur apport pour les systèmes d'analyse de sentiments et de fouille d'opinions, en particulier pour la tâche de détection des émotions. Nous montrons, dans la section 4, qu'il est possible d'utiliser les interjections pour produire automatiquement un corpus étiqueté avec des annotations décrivant les émotions exprimés dans les blogs. Ensuite, dans la section 5 nous évaluons la pertinence du corpus collecté pour la détection automatique des émotions. Nous présentons, dans la section 6 notre méthode pour la création automatique de lexique affectif fin à partir du corpus collecté. Dans la section 7 nous évaluons la qualité du lexique produit.

2 Les interjections

2.1 Caractéristiques formelles

Parmi les interjections nous distinguons les deux sous-classes suivantes :

- **Les interjections onomatopéiques** sont celles qui imitent un bruit naturel qui peut être d'origine humaine (*ouille !, ha !, ouf !, hihi !, etc.*) ou non-humaine (*baoum !, tic tac !, flic floc !, etc.*) (Barberis, 1992). En s'appuyant sur les travaux de (Tesnière, 1959), (Kleiber, 2006) a proposé de les nommer "*interjections primaires émotives*". Les grammaires proposent parfois des listes d'interjections les plus fréquentes (Grevisse, 1969; Bonnard, 1971), mais aucune ne peut prétendre en donner une liste exhaustive et stable, du fait qu'il est possible d'introduire sans cesse de nouvelles créations imitatives dans le discours.

@Candy mouah .
@cntdpie héhé c'est vraiment marrant ça :)
@eelv pff ! mais elle sert à quoi cette loi !!
@eelv ouf je l'ai trouvé, je pensais l'avoir perdu dans le bus !
@sam Oh ! c'est quoi ce match !

TABLE 2 – Exemples de tweets contenant des interjections onomatopéiques.

- **Les interjections non onomatopéiques** (Gonçalves, 2008) sont un ensemble de mots ou expressions figées empruntant leur formulation à d'autres classes de mots : noms (*pardon, flûte, sans blague, etc.*), verbes (*allez !, vive !,*

etc.), adjectifs (*hardi!*), adverbes (*là, comment, eh bien*). Dans ce cas, les morphèmes lexicaux perdent leur relation symbolique à l'objet du monde qu'ils représentaient, pour devenir des indices de subjectivité ou de l'émotion du locuteur. Ainsi les expressions "*Mon dieu!*" et "*Diable!*" sont marquées par un déplacement sémantique et pragmatique. Elles ne désignent plus le *le diable et le bon Dieu* mais l'état émotionnel du locuteur. Dans ses travaux de thèse (Halté, 2013), a qualifié cette catégorie d'*interjections secondaires*.

@xsof Purée ça ne marche toujours pas !
@elao non mdr il n'y a rien d'intéressant dans ce programme !
@eelv tiens voilà une bonne mesure
@sam voyons! t'as bien d'autres solutions ?

TABLE 3 – Exemples de tweets contenant des interjections non onomatopéiques.

2.2 Valeurs subjectives et fonctionnement textuel

Les interjections ont ceci de particulier qu'elles sont nécessairement les indices linguistiques d'une émotion. Même si exprimer une émotion n'est pas toujours leur fonction principale (notamment les interjections vocatives comme "*euh!*" ou "*pst!*", ou celles dont la fonction consiste à donner un ordre comme "*Stop!*" ou "*ouste!*"), c'est toujours un mode d'expression linguistique d'une émotion ou plus généralement d'une attitude subjective. Selon (Guillaume, 1973), *l'expressivité* (sens d'intention subjectif et momentané visé par le locuteur dans l'instant de parole) croît aux dépens de *l'expression* (recours au langage institué, avec des sens stabilisés et des propositions grammaticalement formées). L'interjection constitue le cas limite d'expressivité. D'après (Barberis, 1992), si l'on pose que tout message linguistique repose sur l'équation $expressivité + expression = 1$, l'interjection fait tendre l'expressivité vers 1 et l'expression vers 0. Les interjections sont donc des marques de subjectivité. Les valeurs subjectives des interjections sont conjoncturelles, par exemple, l'interjection *Ah* peut exprimer dans certains contextes le *dégoût* et dans d'autres *la peur, l'étonnement, le soulagement, la colère, etc.*

En ce qui concerne son fonctionnement textuel, l'interjection se caractérise par son détachement du reste de l'énoncé. Aucun lien syntaxique ne la relie à son contexte discursif, ce qui lui permet d'y occuper des places variables. Par exemple, à l'intérieur de l'énoncé "*je n'ai pas réussi mon examen*", l'interjection "*Hélas*" peut s'insérer soit au début, après le verbe ou à la fin du texte. Cependant, les interjections qui soulignent l'orientation affective du locuteur (surprise, peur, joie, soulagement, colère, etc.), interviennent souvent au début de l'énoncé : "*Ouf! il est arrivé à l'heure.*" ou "*Ah! il ne tient jamais sa promesse*".

3 Travaux antérieurs

Plusieurs travaux de recherche en traitement automatique de la langue écrite et orale, se sont intéressés à l'étude des interjections dans le discours. (Garcia-Fernandez *et al.*, 2010) ont procédé à une analyse contextuelle de l'interjection onomatopéique d'hésitation "*Euh*" et ont montré qu'elle peut être utilisée comme indicateur de reformulation ou de niveau de confiance d'une réponse utilisateur dans un système de question-réponse.

Concernant, leur utilisation dans les conversations en ligne, (Falaise, 2005) a montré que le recours aux interjections est assez courant. En effet, sur un corpus de 77 messages choisi aléatoirement par l'auteur, les interjections représentent à elles seules 10 % du corpus. Dans un travail plus récent, (Halté, 2013), a démontré que les interjections relèvent des marques modales qui permettent aux locuteurs de faire porter une émotion ou une attitude subjective sur l'énonciation d'un contenu.

Dans le domaine de l'analyse de sentiments et de la fouille d'opinions, plusieurs travaux de recherche utilisent d'autres formes discursives comme marqueur de subjectivité. (Read, 2005; Pak & Paroubek, 2010) ont utilisé les émoticônes comme marqueur de polarité pour distinguer les textes positifs et négatifs. Dans un premier temps, ils ont identifié une liste d'émoticônes positives (:), :-), :-D, etc.) et une liste d'émoticônes négatives (:), :-(, etc.). Ensuite, les deux listes ont été utilisées comme critère de recherche pour récupérer des messages positifs et négatifs depuis Twitter

Dans des travaux plus récents (Mohammad, 2012; Qadir & Riloff, 2013; Fraisse & Paroubek, 2014b) ont utilisé une liste de mot-dièses (*hashtag* en anglais) (*#sad, #happy, #angry, #fear, #anxious, #disappointed, #unhappy, etc.*) pour collecter des corpus émotionnels et construire de façon automatique des lexiques affectifs. Les lexiques ont été ensuite utilisés dans

des tâches de détection automatique des émotions.

4 Création du corpus émotionnel annoté à partir de Twitter

Les utilisateurs de Twitter font souvent recours aux interjections, qui allient la concision d'écriture à la richesse sémantique, pour exprimer leurs états émotionnels et plus généralement leurs attitudes subjectives. Le tableau 4 montre quelques exemples de tweets subjectifs contenant des interjections. Pour les trois premiers messages, il est possible de détecter les émotions exprimées par les trois locuteurs sans lire les interjections employées (*colère* dans le premier message, *Joie* dans le deuxième et *surprise positive* dans le troisième). Cependant, il est quasiment impossible d'identifier avec exactitude les émotions exprimées dans les messages 4 (*Colère*), 5 (*Dégoût*) et 6 (*Peur*) sans prendre en compte l'interjection employée dans chacun de ces messages. En effet, dans ce cas, l'interjection fournit une information importante sur l'état émotionnel du locuteur et qui n'est pas présente (de façon implicite ou explicite) dans le reste du message.

D'autant plus, du fait de la taille réduite du tweet, des fautes d'orthographe, des abréviations et de tous les autres phénomènes linguistiques et extra-linguistiques qu'il peut contenir, il est difficile d'effectuer une analyse correcte de l'intégralité du tweet. Il est donc important de repérer et de concentrer l'analyse sur les parties pertinentes du tweet. Dans notre cas, il s'agit de parties qui contiennent des informations utiles sur l'état émotionnel du locuteur et les interjections en font partie intégrante.

1.	Argh! train en grève ! c'est parti pour des heures d'attentes !
2.	hihi! j'ai gagné mon pari :)
3.	Ah! je ne m'y attendais vraiment pas super !
4.	pff ... je pars seul
5.	Beurk Fast-food. McDonald's dévoile les 19 ingrédients de ses frites
6.	aaah j'ai entendu un grand bruit en bas !

TABLE 4 – Exemples de tweets exprimant des émotions et contenant des interjections.

L'objectif de ce papier est d'étudier l'apport des interjections pour les systèmes d'analyse de sentiments et en particulier pour la tâche de détection des émotions dans les textes courts. Nous nous posons alors les questions suivantes :

- Est-il possible d'utiliser les interjections pour créer et étiqueter un large corpus émotionnel ?
- Est-il possible de considérer les interjections comme une annotation émotionnelle fiable malgré le grand nombre d'annotateurs (les utilisateurs) et leurs différences culturelles et sociales ?

Dans un premier temps, nous nous sommes intéressés au nombre ainsi qu'aux types de classes émotionnelles à considérer. En effet, à ce jour, il est impossible de dénombrer les émotions avec exactitude. Plusieurs travaux (Mohammad, 2012; Qadir & Riloff, 2013) se basent sur les six émotions de base définies par (Ekman, 1970) (*tristesse, colère, peur, surprise, joie, dégoût*). D'autres travaux ont défini leur propres modèles et classes d'émotions (Matsumoto, 2009; Levenson, 2011; Cambria *et al.*, 2012). Cependant ces modèles sont souvent adaptés à une langue, à une culture, à un domaine ou à une application donnée et ils ne couvrent que partiellement le spectre de la subjectivité.

Après une étude complète de l'état de l'art sur ce sujet, nous avons sélectionné le modèle proposé par (Fraisse & Paroubek, 2014a). Ce choix a été motivé par la généricité du modèle proposé (indépendant du domaine) ainsi par sa couverture totale du spectre de la subjectivité (opinion, sentiment, émotion). Les auteurs proposent 18 classes subjectives qui couvrent les différents types d'expressions subjectives (émotion, sentiment, jugement, opinion) (Tableau 5). Les auteurs répartissent les informations subjectives en trois grandes catégories : une première catégorie nommée *Opinion* qui inclut les 4 classes suivantes : *accord, désaccord, valorisation et dévalorisation*, une deuxième catégorie *Sentiment* qui inclut les 2 classes : *satisfaction et insatisfaction* et enfin la catégorie *Émotion* avec les 12 classes affectives : *colère, peur, tristesse, ennui, dérangement, surprise négative, déplaisir, mépris, surprise positive, apaisement, plaisir, amour*. Afin d'identifier les classes d'émotions ayant des interjections connues et utilisées par un grand nombre d'utilisateur. Nous avons, d'abord, extrait manuellement une liste d'interjections depuis différentes sources (Grevisse, 1969; Bonnard, 1971). Ensuite, nous avons utilisé cette liste comme requête de recherche pour collecter les tweets contenant au moins une interjection parmi la liste fournie dans la requête. Enfin, nous avons retenu uniquement les interjections qui sont à la fois fréquentes (ayant un nombre d'occurrence élevé dans le corpus) et souvent utilisées pour exprimer la même émotion par différents utilisateurs. Le tableau 6 décrit l'ensemble des interjections que nous avons retenu ainsi que les classes d'émotions associées.

LES INTERJECTIONS POUR DÉTECTER LES ÉMOTIONS

#	Classe	Catégorie	Spécification
1	SURPRISE NÉGATIVE	e-	surprise négative / étonnement négatif
2	DÉRANGEMENT	e-	dérangement / embarras
3	PEUR	e-	peur / terreur / inquiétude
4	ENNUI	e-	ennui
5	DÉPLAISIR	e-	déplaisir/ déception
6	TRISTESSE	e-	tristesse / chagrin / souffrance / désespoir / résignation
7	COLÈRE	e-	colère / rage / agacement / exaspération / énervement / impatience
8	MÉPRIS	e-	mépris / dédain / dégoût / haine/
9	INSATISFACTION	s-	insatisfaction / mécontentement
10	DÉVALORISATION	o-	dévalorisation / désintérêt / dépréciation
11	DÉSACCORD	o-	désaccord / désapprobation
12	VALORISATION	o+	valorisation / intérêt / appréciation
13	ACCORD	o+	accord / compréhension / approbation
14	SATISFACTION	s+	satisfaction / contentement / fierté
15	SURPRISE POSITIVE	e+	surprise positive / étonnement positif
16	APAISEMENT	e+	apaisement / soulagement / reconnaissance / pardon / sérénité
17	PLAISIR	e+	plaisir / divertissement / joie / euphorie / bonheur / extase
18	AMOUR	e+	amour / tendresse / affection / dévouement / passion / envie

TABLE 5 – Les 18 classes subjectives proposées par (Fraisie & Paroubek, 2014a), e=émotion, s=sentiment, o=opinion, += valence positive, -=valence négative.

#	Classe	Interjections onomatopéiques	Interjections non onomatopéiques
1	COLÈRE	Argh , pff	Voyons !
2	PLAISIR	hihi , haha	lol , youpi
3	PEUR	aah	
4	APAISEMENT	ouf	
5	TRISTESSE	Aïe, ouille	zut, hélas
6	INSATISFACTION	Bof	
7	MÉPRIS	Beurk	
8	SURPRISE NÉGATIVE	oups	

TABLE 6 – Les classes émotionnelles exprimées ayant des interjections clairement définies sur Twitter.

4.1 Recherche basée sur les interjections

Nous avons utilisé l'API Search³ de Twitter pour collecter et filtrer les messages. L'API permet de spécifier la langue de messages et une requête de recherche par mot clé. Ainsi, pour chaque classe émotionnelle du Tableau 6, nous utilisons la liste d'interjections, qui lui est associée, comme mots clés de la requête. Nous attribuons par la suite l'étiquette de la classe correspondante à tous les messages qui ont été retourné par la requête. Par exemple, tous les messages qui ont été collectés en utilisant la liste d'interjections "*hihi*", "*haha*", "*lol*", "*youpi*" sont étiquetés *PLAISIR*.

Certains utilisateurs n'utilisent pas souvent les formes normalisées des interjections, par exemple l'interjection "*argh*" peut être écrite "*arggh*", "*arrrrgh*", "*arghhhh*", *etc.*. En plus, certaines interjections n'ont pas vraiment de forme d'écriture normalisée. Ainsi, pour collecter aussi les messages contenant des interjections orthographiées de façon non normalisée, nous avons inclus, moyennant des expressions régulières, dans les requêtes de recherche, la plupart des variations orthographiques des interjections par exemple : "*arghh*", "*arrrrgh*", "*arrrrgh*", "*arghhhh*", *etc.* pour l'interjection "*argh*". En total, pour les 8 classes d'émotions nous avons collecté un corpus de 30,123 tweets. Après suppression de *retweets* (messages commençant par l'abréviation "RT" et qui consistent à re-publier tel quel un message par plusieurs personnes.) le corpus final était composé de 19,061 tweets.

4.2 Distribution des interjections dans le corpus

Le tableau 7 décrit la distribution des interjections dans le corpus émotionnel collecté depuis Twitter. Certaines interjections sont plus fréquentes que d'autres tels que "*pff*", "*hihi*", "*haha*". Nous observons aussi que pour certaines émotions comme par exemple la *COLÈRE*, les utilisateurs utilisent plus les interjections onomatopéiques que les non onomatopéiques. Par exemple, parmi les 3896 messages étiquetés *COLÈRE* 3440 contiennent les interjections *argh* et *pff* et seulement 456 messages avec l'interjection non onomatopéique *Voyons !*. Ce phénomène peut être expliqué par le fait que les interjections onomatopéiques sont plus explicites pour exprimer certaines émotions intenses.

Interjection	Nombre de messages	% de messages
Argh	1078	5,65
pff	2362	12,39
Voyons !	456	2,39
hihi	3350	17,57
haha	3987	19,14
lol	1456	7,63
youpi	1211	6,35
aah	844	4,42
ouf	1300	6,82
Aïe	320	1,67
ouille	337	1,76
zut	158	0,82
hélas	987	5,17
Bof	650	3,41
beurk	435	2,82
oups	467	2,45
Total tweets	19061	100
Total utilisateurs	19032	

TABLE 7 – Distribution des différentes interjections dans le corpus émotionnel de Twitter.

5 Pertinence et utilité du corpus collecté depuis Twitter

Un des problèmes majeurs inhérent à la classification supervisée des émotions est d'obtenir un corpus d'apprentissage annoté manuellement. L'annotation est complexe, elle coûte cher et nécessite beaucoup de temps. En outre, du fait de l'ambiguïté des textes émotionnels, il est souvent difficile d'avoir des scores inter-annotateur élevés lors de l'annotation. Dans notre cas, nous considérons que le corpus collecté est un corpus annoté par les utilisateurs eux mêmes, puisque nous

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

attribuons le label émotionnel de l'interjection présente dans le texte à tout le message. Par exemple si le message contient l'une des interjections suivantes : "argh", "pff", "voyons !" alors il est étiqueté *COLÈRE*. Comme le montre le tableau 7, le corpus contient 19061 messages créés par 19032 utilisateurs différents. Ainsi, nous nous posons les deux hypothèses suivantes :

1. Les annotations données par ce grand nombre d'annotateurs sont cohérentes et pertinentes.
2. Ce corpus peut être utilisé par un classifieur pour apprendre à détecter les émotions.

Pour vérifier ces deux hypothèses, nous avons mené une expérience de classification automatique des émotions en utilisant ce corpus comme corpus d'apprentissage. Nous avons entraîné un système de classement à vecteurs supports en utilisant des traits à base de n-grammes extraits du corpus d'apprentissage. Nous avons utilisé la bibliothèque LIBLINEAR (Fan *et al.*, 2008) avec un noyau linéaire et un paramétrage par défaut. Pour pondérer chaque terme nous avons eu recours aux deux fonctions de poids suivantes :

- **Binaire** : 1 si le terme apparaît dans le message et 0 sinon.
- **TF-IDF normalisée** : nous avons utilisé une forme normalisée de la fonction *TF-IDF* (Pak *et al.*, 2014). Cette normalisation consiste à diviser la fonction de poids *TF-IDF* par la fréquence moyenne d'un terme (*avg.tf*) . La normalisation à base de fréquence moyenne d'un terme est basée sur l'observation que les utilisateurs ont tendance à utiliser un vocabulaire riche quand ils expriment leur attitude subjective. Ainsi, les termes subjectifs et émotionnels (par exemple *adorable*, *satisfaisant*) ont une fréquence moyenne proche de 1, tandis que les termes non subjectifs ont une fréquence moyenne plus élevée. Ainsi, afin de normaliser le poids de chaque n-gramme (unigramme dans notre cas) du document, nous le divisons par sa fréquence moyenne (Équation 1).

$$w(t_i) = \frac{tfidf(t_i)}{avg.tf(t_i)} \quad (1)$$

$$tfidf(t_i) = tf(t_i) \cdot \log \frac{|D|}{df(t_i)} \quad (2)$$

$$avg.tf(t_i) = \frac{\sum_{\{d|t_i \in d\}} tf(t_i)}{|\{d|t_i \in d\}|} \quad (3)$$

où $\{d|t_i \in d\}$ est l'ensemble de messages qui contient le terme t_i .

Pour la classification multi-étiquettes, nous avons entraîné indépendamment un système de classement pour chaque émotion. Le tableau 8 montre la répartition des données en apprentissage et test par classe d'émotion. Afin de diminuer le bruit dans les tweets, nous avons pré-traiter le corpus en supprimant les mentions utilisateurs (les mots qui commencent par "@" et qui font référence aux utilisateurs) et les URL. Enfin, pour ne pas biaiser le classement, nous avons supprimé des tweets toutes les interjections qui ont servi pour collecter et étiqueter le corpus.

Classes	Apprentissage	test
COLÈRE	2598	1298
PLAISIR	6445	3222
PEUR	563	281
APAISEMENT	867	433
TRISTESSE	1202	600
INSATISFACTION	434	216
MÉPRIS	290	145
SURPRISE NÉGATIVE	312	155
Total	12711	6350

TABLE 8 – Caractéristiques du corpus Twitter utilisé pour l'apprentissage.

Les résultats de classification illustrés dans le tableau 9, montrent que certains classifieurs atteignent des scores acceptables en mesure F1 : 0,66 pour le classifieur *COLÈRE* et 0,73 pour le classifieur *PLAISIR*. Ces deux bons scores s'expliquent par le fait que les deux classifieurs disposent de plus de données d'entraînement que les autres classifieurs. Nous notons aussi que les systèmes utilisant les vecteurs de traits *unigrammes+TF-IDF/avg.tf* performant mieux la classification. En effet, la fonction *TF-IDF/avg.tf* permet de favoriser les termes qui caractérisent au mieux une émotion et donc de mieux apprendre les traits représentatifs d'une classe donnée. Nos résultats sont comparables à d'autres travaux similaires comme ceux de (Roberts *et al.*, 2012), qui annoncent un score de 0.74 en mesure F1 atteint par leur classifieur supervisé et entraîné sur un corpus de tweets annoté manuellement par des experts.

Classes	unigr.+binaire			unigr.+TF-IDF/avg.tf		
	Précision	Rappel	F1	Précision	Rappel	F1
COLÈRE	0,72	0,61	0,66	0,43	0,37	0,39
PLAISIR	0,42	0,51	0,46	0,71	0,76	0,73
PEUR	0,21	0,13	0,16	0,11	0,09	0,09
APAISEMENT	0,13	0,23	0,16	0,25	0,31	0,27
TRISTESSE	0,31	0,42	0,35	0,34	0,37	0,35
INSATISFACTION	0,12	0,22	0,15	0,27	0,12	0,16
MÉPRIS	0,09	0,12	0,10	0,19	0,12	0,14
SURPRISE NÉGATIVE	0,07	0,11	0,08	0,04	0,09	0,05

TABLE 9 – Résultats de la classification des émotions, les meilleurs score en mesure F1 sont montrés en gras.

6 Création du lexique d'émotions

Un lexique d'émotions est une liste de mots associés à une ou plusieurs émotions. Par exemple, l'adjectif "épanoui" peut être associé aux deux émotions *joie* ou *amour* en fonction du contexte. Ce type de lexique est très utile et peut être utilisé dans plusieurs applications et tâches tels que la détection des émotions dans le texte, l'identification de passages émotionnels et affectifs dans les documents ou encore l'analyse de personnalité. Nous présentons dans cette section notre approche pour la construction automatique d'un lexique affectif fin à partir du corpus émotionnel de Twitter.

6.1 Méthode

En se basant sur une approche statistique, notre méthode consiste à extraire, à partir du corpus émotionnel collecté (décrit dans la section 4), et pour chaque classe d'émotion (du Tableau 6), l'ensemble de mots qui lui est associé. Afin de mesurer l'association entre un mot m du corpus et une émotion e , nous nous sommes basés sur l'information mutuelle introduite par (Fano, 1961; Church & Hanks, 1990), qui pour chaque couple de variables aléatoires (X, Y) mesure leur degré de dépendance au sens probabiliste. L'information mutuelle est donnée par la formule suivante :

$$\text{IM}(X, Y) = \log_2 \left(\frac{P(X, Y)}{P(X) \cdot P(Y)} \right) \quad (4)$$

Ainsi, dans notre cas, il s'agit de mesurer le degré de dépendance entre un mot m et une émotion e .

$$\text{IM}(m, e) = \log_2 \left(\frac{\text{freq}(m, e)}{\text{freq}(m) \cdot \text{freq}(e)} \right) \quad (5)$$

Avec $\text{freq}(m, e)$ est le rapport entre le nombre de messages contenant le mot m et étiqueté e ($|T_{m,e}|$) et le nombre total de messages ($|T|$).

$$\text{freq}(m, e) = \frac{|T_{m,e}|}{|T|} \quad (6)$$

$\text{freq}(m)$ est le rapport entre le nombre total de messages contenant le mot m ($|T_m|$) et le nombre total de messages.

$$\text{freq}(m) = \frac{|T_m|}{|T|} \quad (7)$$

et $\text{freq}(e)$ est le rapport entre le nombre total de messages étiquetés e ($|T_e|$) et le nombre total de messages.

$$\text{freq}(e) = \frac{|T_e|}{|T|} \quad (8)$$

Certains mots peuvent être associés à plusieurs émotions. Nous calculons donc comme dans (Mohammad, 2012), le degré de dépendance entre les deux variables m et $\neg e$ ($\neg e$ représente toutes classes d'émotions sauf e).

$$\text{IM}(m, \neg e) = \log_2 \left(\frac{\text{freq}(m, \neg e)}{\text{freq}(m) \cdot \text{freq}(\neg e)} \right) \quad (9)$$

Ainsi, le degré d'association entre un m et une e est donné par l'équation 10. Nous considérons qu'un mot m est fortement associé à une émotion e , si son degré d'association $\text{Asso}(m, e)$ est supérieur à 0.

$$\text{Asso}(m, e) = \text{IM}(m, e) - \text{IM}(m, \neg e) \quad (10)$$

6.2 Expérimentation et résultats

Les messages Twitter peuvent contenir : des URLs, des mentions utilisateur (par exemple @oXc11), des retweets commençant par la mention "RT", des caractères spéciaux, etc. Ainsi, avant de procéder à l'extraction des mots affectifs, depuis le corpus collecté, nous avons tout d'abord effectué certaines opérations de pré-traitements : (1) suppression des liens URL, retweets ainsi que les mentions utilisateurs (2) segmentation et (3) suppressions des mots outils. Ensuite, nous avons calculé pour chaque m du corpus son degré d'association $\text{Asso}(m, e)$ selon la méthode présentée dans la section précédente. En fonction de son degré d'association, un mot peut être attribué à une ou plusieurs classes d'émotions.

La taille du lexique affectif produit est de 1530 mots répartis sur les 8 classes d'émotions (Tableau 10).

#	Classe émotionnelle	Nombre de termes	Exemples
1	COLÈRE	234	colère, furieux, fâché
2	PLAISIR	423	super, enjoy, cool
3	PEUR	184	peur, tremble, inquiet
4	APAISEMENT	178	ouf, enfin, fini
5	TRISTESSE	168	triste, seul, mal
6	INSATISFACTION	58	bof, insatisfait, moyen
7	MÉPRIS	198	déteste, dégoûtant, dégueulasse
8	SURPRISE NÉGATIVE	87	non, stupéfait, impossible

TABLE 10 – Nombre de termes par classe d'émotion dans le lexique.

7 Évaluation du lexique

7.1 Expérimentations

Afin, d'évaluer le lexique produit, nous l'avons utilisé dans une tâche de détection des émotions à partir de textes dont le corpus de référence est issu du projet DOXA (Paroubek *et al.*, 2010). À notre connaissance, c'est le seul corpus en français qui est annoté avec des classes d'émotions fines et auquel nous avons accès. Il s'agit d'un corpus de critiques de films annoté manuellement avec les classes d'émotion correspondantes (18 classes en total). Les annotations étaient effectuées sur deux niveaux :

- le niveau *macro* : annotations au niveau du document ;
- et le niveau *meso* : annotation au niveau du paragraphes.

Les paragraphes peuvent avoir de 1 à 5 étiquettes d'émotions. Le corpus total compte 7126 paragraphes dont 612 annotées avec plus d'une émotion (609 annotées avec 2 émotions et 3 paragraphes avec 3 émotions). Pour nos expérimentations, nous considérons les paragraphes comme des documents. Ainsi, dans un premier temps, nous avons constitué un corpus avec l'ensemble de paragraphes regroupés par classe d'émotions. Ensuite, nous avons écarté les paragraphes qui sont étiquetées avec une classe d'émotion non présente dans notre lexique. Au total, nous avons formé un corpus de 1,000 documents repartis sur les 8 classes d'émotions de notre lexique. Le Tableau 11 décrit la répartition des données, en apprentissage et en test, à travers les différentes classes d'émotions du corpus.

Nous avons entraîné un système de classement à vecteurs supports utilisant différents traits extraits du corpus d'apprentissage. Nous avons utilisé la bibliothèque LIBLINEAR (Fan *et al.*, 2008) avec un noyau linéaire et un paramétrage par défaut. Pour la classification multi-étiquettes, nous avons utilisé une stratégie en parallèle, c'est-à-dire que nous avons

#	Classe émotionnelle	Apprentissage	test	Total
1	COLÈRE	86	42	128
2	PLAISIR	288	144	432
3	PEUR	16	7	23
4	APAISEMENT	82	40	122
5	TRISTESSE	51	25	76
6	INSATISFACTION	9	4	13
7	MÉPRIS	117	58	175
8	SURPRISE NÉGATIVE	21	10	31
Total		670	330	1000

TABLE 11 – Caractéristiques du corpus d’apprentissage DOXA.

entraîné indépendamment un système de classement pour chaque émotion. Chaque système de classement fournit pour chaque message une indication de présence ou d’absence de l’émotion qu’il a été entraîné à détecter. Ainsi nous pouvons obtenir pour chaque document, de 0 à 8 étiquettes d’émotions. La liste des traits utilisés pour l’apprentissage comprenait :

- **n-grammes** : nous avons utilisé des unigrammes et des bigrammes avec la fonction de poids TF-IDF normalisée (décrite dans la section 5).
- **Lexique** : nous avons créé un vecteur de traits avec les noms de classes d’émotions présents dans le lexique (8 classes en total) et dont la valeur est initialisée à 0. Ensuite pour chaque occurrence de mot présent dans le tweet, nous vérifions s’il est présent dans le lexique. Si c’est le cas alors nous incrémentons de 1 la valeur numérique de classes d’émotions correspondantes dans le vecteur de traits. Par exemple si le tweet contient 2 occurrences de mots qui sont présents dans le lexique et étiquetées *PLAISIR* alors l’élément *PLAISIR* du vecteur de traits aura comme valeur 2.

Afin, d’évaluer l’apport de notre lexique pour la tâche de détection des émotions, nous avons utilisé l’algorithme de classification suivant :

1. D’abord nous avons entraîné le système uniquement avec les deux vecteurs *unigr.TF-IDF/avg.tf* et *bigr.TF-IDF/avg.tf*.
2. Ensuite, nous avons rajouté les traits *Lexique* aux deux vecteurs *unigr.TF-IDF/avg.tf* et *bigr.TF-IDF/avg.tf* et évalué la différence.

7.1.1 Discussion de résultats

Afin d’affiner le paramétrage de notre système de classement, nous avons effectué une validation croisée à 10 replis sur le corpus d’apprentissage. Nous avons utilisé *la précision, le rappel et la mesure F1* pour évaluer la performance du système.

D’abord, nous avons analysé la performance de différents vecteurs de traits utilisés pour la détection des émotions : *unigr.TF-IDF/avg.tf*, *unigr.TF-IDF/avg.tf+lexique*, *bigr.TF-IDF/avg.tf*, *bigr.TF-IDF/avg.tf+lexique*. La figure 1 montre les performances du système en mesure F1 pour chaque émotion détectée et pour chaque vecteur de traits utilisé par le classifieur. La performance de classification de classes fréquentes comme *PLAISIR*, *COLÈRE*, *MÉPRIS*, *TRISTESSE* et *DÉGOÛT* est meilleure que celles qui sont moins fréquentes dans les données d’apprentissage comme *PEUR*, *SURPRISE NÉGATIVE* et *INSATISFACTION*. Cependant dans les deux cas, les vecteurs de traits ayant obtenus les meilleurs scores en mesure F1 sont les *unigr.TF-IDF/avg.tf+lexique* et *bigr.TF-IDF/avg.tf+lexique*. D’après les résultats illustrés dans le tableau 12, le vecteur de traits *unigr.TF-IDF/avg.tf+lexique* permet une augmentation de +0,25 en mesure F1 pour la classe *PLAISIR* et le vecteur *bigr.TF-IDF/avg.tf+lexique* une augmentation +0,21 pour la classe *COLÈRE*. Cela prouve que le lexique que nous avons construit apporte un gain significatif pour la tâche de détection des émotions par rapport aux systèmes qui utilisent uniquement les N-grammes. En effet, les meilleurs scores en mesure F1 obtenu par le système à base de N-grammes sont : 0,38 en utilisant le vecteur *bigr.TF-IDF/avg.tf* et 0,35 en utilisant le vecteur de traits *unigr.TF-IDF/avg.tf* contre 0,53 avec le vecteur de traits *unigrammes+lexique* et 0,5 avec le vecteur *bigrammes+lexique*.

Nous avons comparé aussi nos résultats à ceux obtenus par (Roberts *et al.*, 2012) pour la détection des émotions dans les Tweets. Malgré le fait qu’il ne s’agit pas du même corpus de Tweets, cela nous permet de se situer par rapport à d’autres méthodes dans l’état de l’art. (Roberts *et al.*, 2012) utilisent un corpus de Tweets annoté automatiquement avec les sept émotions : PEUR, COLÈRE, SURPRISE, DÉGOÛT, JOIE, AMOUR et TRISTESSE. Afin de détecter les

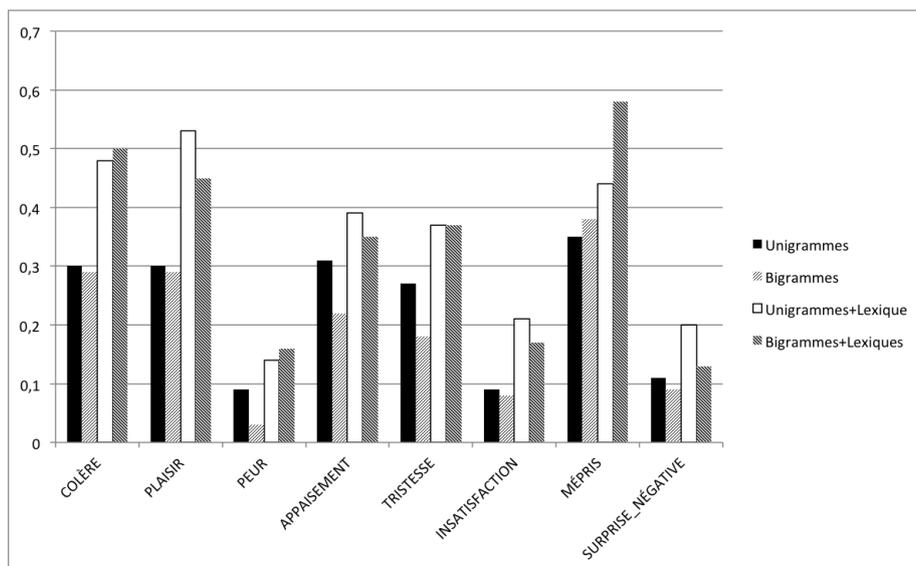


FIGURE 1 – Performance de différents types de traits utilisés pour la détection des émotions.

émotions présentes dans les tweets, les auteurs utilisent un classifieur par émotion en se basant sur différents types de traits : *unigrammes*, *bigrammes*, *trigrammes*, *synsets*, *contains!* (qui indique si le tweet contient un point d'exclamation), *contains?* (qui indique si le tweet contient un point d'interrogation) et *Topic* (qui indique le sujet du tweet). Malgré la différence de taille de corpus d'entraînement et la simplicité des traits utilisés pendant la phase d'apprentissage, nous obtenons, pour certaines émotions (MÉPRIS et PLAISIR) des scores comparables à ceux obtenus par (Roberts *et al.*, 2012).

Classes	unigrammes			unigr.+lexique			Δ	bigrammes			bigr.+lexique			Δ
	P.	R.	F1	P.	R.	F1		P.	R.	F1	P.	R.	F1	
COLÈRE	0,32	0,29	0,3	0,49	0,48	0,48	+0,18	0,29	0,31	0,29	0,56	0,46	0,5	+0,21
PLAISIR	0,39	0,27	0,28	0,56	0,51	0,53	+0,25	0,28	0,31	0,29	0,51	0,41	0,45	+0,19
PEUR	0,11	0,09	0,09	0,17	0,12	0,14	+0,05	0,07	0,02	0,03	0,13	0,21	0,16	+0,13
APAISEMENT	0,35	0,28	0,31	0,41	0,39	0,39	+0,08	0,29	0,18	0,2	0,35	0,37	0,35	+0,15
TRISTESSE	0,29	0,27	0,27	0,39	0,36	0,37	+0,1	0,17	0,21	0,18	0,41	0,34	0,37	+0,19
INSATISFACTION	0,11	0,09	0,09	0,17	0,29	0,21	+0,12	0,1	0,07	0,08	0,13	0,27	0,17	+0,09
MÉPRIS	0,41	0,32	0,35	0,47	0,42	0,44	+0,09	0,39	0,38	0,38	0,61	0,56	0,58	+0,2
SURPRISE NÉGATIVE	0,13	0,11	0,11	0,23	0,19	0,20	+0,09	0,12	0,08	0,09	0,18	0,11	0,13	+0,04

TABLE 12 – Résultats de la classification des émotions avec les différents types de traits (unigr.,bigr.,lexi.)

8 Conclusion

Dan ce papier, nous avons étudié et évalué l'apport des interjections pour les systèmes d'analyse de sentiments et de fouille d'opinions. Malgré leur statut grammaticale et leur richesse sémantique à expliciter un état émotionnel, les interjections sont restées marginalisées par les systèmes d'analyse de sentiments. Nous avons proposé une méthode automatique pour collecter et étiqueter les messages de Twitter en considérant les interjections présentes dans les messages comme annotations bruitées. Nous avons ainsi créé un corpus émotionnel de 19061 tweets repartis sur 8 classes d'émotions. Afin, d'évaluer la pertinence du corpus collecté, nous l'avons utilisé comme corpus d'apprentissage pour la tâche de détection des émotions. Nos résultats ont montré que nous obtenons des performances comparables à d'autres systèmes entraînés sur des corpus annotés manuellement par des experts. Ensuite, nous avons proposé une méthode permettant d'extraire, de façon automatique, à partir du corpus collecté un lexique affectif fin. Enfin, nous avons évalué la qualité du lexique produit sur une tâche de détection des émotions, qui a montré un gain en mesure F1 allant selon les émotions de +0,04 à +0,21. L'approche proposée dans cet article, affine et complète les méthodes existantes basées sur l'utilisation des émoticônes

et des mots-dièses pour la classification des textes subjectifs. Pour la suite de nos travaux, nous souhaitons évaluer la généralité de notre approche en l'appliquant à d'autres langues.

Références

- BARBERIS J.-M. (1992). Onomatopée, interjection : un défi pour la grammaire. *L'information grammaticale*, **53**, p.52–57.
- BONNARD H. (1971). Interjections. *Grand Larousse de la langue française, tome IV*, p. p.2758.
- CAMBRIA E., LIVINGSTONE A. & HUSSAIN A. (2012). The hourglass of emotions. *Cognitive behavioural systems*.
- CHURCH K. W. & HANKS P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, **16**, p.22–29.
- EKMAN P. (1970). Universal facial expressions of emotions. In *Digest, California mental Health Research*.
- FALAISE A. (2005). Constitution d'un corpus de français tchaté. In *actes de RÉCITAL*, Dourdan.
- FAN R.-E., CHANG K.-W., HSIEH C.-J., WANG X.-R. & CHIH-JEN (2008). Liblinear : A library for large linear classification. *Journal of Machine Learning Research*, **9**, p.1871–1874.
- FANO R. (1961). Transmission of information : A statistical theory of communications. *MIT Press, Cambridge*.
- FRAISSE A. & PAROUBEK P. (2014a). Toward a unifying model for opinion, sentiment and emotion information extraction. In *the 9th International Conference on Language Resources and Evaluation*, Reykjavik, Iceland.
- FRAISSE A. & PAROUBEK P. (2014b). Twitter as a comparable corpus to build multilingual affective lexicons. In *the 7th Workshop on Building and Using Comparable Corpora*.
- GARCIA-FERNANDEZ A., VASILESCU I. & ROSSET S. (2010). Euh as cue for speaker confidence and word searching in human spoken answers in french. In *DISS-LPSS Joint Workshop/Disfluency*.
- GONÇALVES M. (2008). Sur le statut linguistique de l'interjection. In *Actas del VIII congreso de Lingüística General*, p. p.14, Universidad Autónoma de Madrid.
- GREVISSE M. (1969). *Le bon usage*. Duculot et Gembloux, Haltier.
- GUILLAUME G. (1973). Principes de linguistiques théoriques. *Presses de l'université Laval, Québec et Klincksieck, Paris*, p. p.146.
- HALTÉ P. (2013). *Les marques modales dans les chats : étude sémiotique et pragmatique des interjections et des émoticônes dans un corpus de conversations synchrones en ligne*. Universités de Luxembourg et de Lorraine.
- KLEIBER G. (2006). Sémiotique de l'interjection. *Langages*, **161**.
- LEVENSON R. W. (2011). Basic emotion questions. *Emotion Review*, **3**(4), p.379–386.
- MATSUMOTO D. (2009). Spontaneous facial expressions of emotion of blind individuals. *Journal of Personality and Social Psychology*, **96**(1), p.1–10.
- MOHAMMAD S. M. (2012). Emotional tweets. In *proceedings of First Joint Conference on Lexical and Computational Semantics*, p. p.246–255.
- PAK A. & PAROUBEK P. (2010). Construction d'un lexique affectif pour le français à partir de twitter. In *proceedings of TALN (Traitement Automatique des Langues Naturelles) 2010*, Montréal, Canada.
- PAK A., PAROUBEK P., FRAISSE A. & FRANCOPOULO G. (2014). Normalization of term weighting scheme for sentiment analysis. *Human Language technology Challenges for Computer Science and Linguistics. Series : Lecture Notes in Artificial Intelligence*, **8387**.
- PAROUBEK P., PAK A. & MOSTEFA D. (2010). Annotations for opinion mining evaluation in the industrial context of the doxa project. In *the Seventh International Conference on Language Resources and Evaluation*.
- QADIR A. & RILOFF E. (2013). Bootstrapped learning of emotion hashtags hashtags4you. In *the 4th Workshop on Computational Approaches to Subjectivity Sentiment and Social Media Analysis*, Atlanta.
- READ J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *proceedings of The 43rd Annual Meeting of the Association for Computational Linguistics*, p. p.43–48.
- ROBERTS K., ROACH M. A., JOHNSON J., GUTHRIE J. & HARABAGIU S. M. (2012). Empatweet : Annotating and detecting emotions on twitter. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey.
- TESNIÈRE L. (1959). Éléments de syntaxe structurale. *Langue française*, **1/1**, p.36–40.