ADVANSE: Sentiment, Opinion and Emotion Analysis in French Tweets

Amine Abdaoui, Mike Donald Tapi Nzali, Jérôme Azé, Sandra Bringay, Christian Lavergne, Caroline Mollevi, Pascal Poncelet

ADVANSE : Analyse du sentiment, de l'opinion et de l'émotion sur des Tweets Français

Amine Abdaoui[1] Mike Donald Tapi Nzali[1, 2] Jérôme Azé[1] Sandra Bringay[1] Christian Lavergne[2] Caroline Mollevi[3] Pascal Poncelet[1]
(1) LIRMM UM CNRS, UMR 5506, 161 Rue Ada, 34095 Montpellier, France
(2) Institut de Mathématiques et de Modélisation de Montpellier, Université Montpellier, France
(3) Unité de biostatistique, Institut de Cancérologie de Montpellier, France
amin.abdaoui@lirmm.fr, mike-donald.tapi-nzali@lirmm.fr

Résumé. Ce papier décrit les systèmes que nous avons soumis au défi DEFT 2015 (Défi Fouille de Texte). Cette onzième édition a porté sur l'analyse de l'opinion, du sentiment et de l'émotion dans des tweets rédigés en Français. Le défi propose trois tâches, nous avons participé à la tâche 1 qui concerne la classification des tweets selon leur polarité, à la tâche 2.1 qui concerne l'identification de la classe générique de l'information exprimée dans les tweets et enfin à la tâche 2.2 qui concerne l'identification de la classe spécifique de l'opinion, du sentiment ou de l'émotion présente dans les tweets. Nous avons proposé des méthodes supervisées basées sur les machines à vecteurs de support (SVM) utilisant plusieurs types d'attributs comme les n-grammes de mots, les n-grammes de caractères, les patrons syntaxiques les plus fréquents, etc. Nous avons également construit et utilisé des lexiques de sentiments et d'émotions spécifiques pour le français.

Abstract.
ADVANSE : Sentiment, Opinion and Emotion Analysis in French Tweets
This paper describes the methods we submitted to the DEFT 2015 Challenge (Text Mining Challenge). This eleventh edition concerned the analysis of opinions, sentiments and emotions expressed in French tweets. Three tasks have been proposed, we participated to task 1 which concerned the classification of tweets according to their polarities, to task 2.1 concerning the identification of the generic class of information expressed in the tweet, and finally to task 2.2 that concerned the identification of the specific class of opinion, sentiment or emotion. We proposed supervised methods based on support vector machines (SVM) using several types of attributes such as word n-grams, character n-grams, most common syntactic patterns, etc. Moreover, we constructed and used two French lexicons of sentiments and emotions.

Mots-clés : Analyse de sentiments, analyse d'opinions, analyse d'émotions, analyse de subjectivité, fouille de textes.

Keywords: Sentiment analysis, opinion analysis, emotion analysis, subjectivity analysis, text mining.

1 Introduction

L'analyse automatique de textes pour y détecter la présence d'états affectifs, leur polarité, les émotions associées et les opinions exprimées a suscité de nombreux travaux ces dernières années. Cet intérêt toujours croissant peut s'expliquer, en partie, par les perspectives d'applications qu'elle ouvre dans de nombreux domaines tels que : les systèmes de recommandation en ligne (Tatemura, 2000), l'intelligence économique (Melnik & Alm, 2002), la veille politique et gouvernementale (Laver *et al.*, 2003; Mullen & Malouf, 2006), etc. Les méthodes appliquées sont généralement spécifiques aux types de textes traités : aux tweets (Roberts *et al.*, 2012), aux titres de presse (Strapparava & Mihalcea, 2008), aux commentaires des internautes (Vincent & Winterstein, 2013), etc. Si la majorité des méthodes proposées ont été créées pour l'Anglais et pour la polarité, quelques travaux plus récents existent aussi pour le Français (Grouin *et al.*, 2009) et pour les émotions (Mohammad, 2012). Dans cet article, nous nous intéressons à la classification de tweets en langue française. Nous présentons des systèmes que nous avons soumis à la onzième édition du défi de fouille de textes (DEFT 2015). La première tâche (tâche 1) à laquelle nous avons participé consiste à classifier les tweets selon leur polarité (*positif, négatif* ou *neutre*). Dans la seconde tâche (tâche 2.1), il s'agit de détecter la classe générique exprimée dans un tweet (*information, opinion,*

sentiment ou *émotion*). La dernière tâche (tâche 2.2) à laquelle nous avons participé consiste à détecter la classe spécifique de l'opinion, du sentiment ou de l'émotion exprimée dans le tweet (18 classes ont été considérées telle que : *colère, peur, tristesse, amour, plaisir surprise, accord, etc.*). Nous avons obtenu 73,33% de macro-précision pour la tâche 1 à 0,27% du meilleur système. Pour la tâche 2.1 nous avons obtenu le meilleur système avec 61,29% de macro-précision. Cependant, nous n'avons pas considéré les bonnes classes lors de notre soumission pour la tâche 2.2, nous présentons ici nos résultats avec les bonnes classes. Les données d'apprentissage et de test ont été fournies par les organisateurs de la compétition ; les tableaux 1 et 2 montrent la distribution des classes sur le corpus d'apprentissage et de test.

Tâche 1					**Tâche 2.1**				
Classes	**Apprentissage**		**Test**		**Classes**	**Apprentissage**		**Test**	
	#	%	#	%		#	%	#	%
Positif	2448	31	1057	31	Émotion	820	12	351	10
Négatif	1875	24	804	24	Information	3571	53	1518	45
Neutre	3544	45	1518	45	Sentiment	2275	34	537	16
					Opinion	82	1	973	29
Total	**7867**	**100**	**3379**	**100**	**Total**	**6754**	**100**	**3379**	**100**

TABLE 1 – Distribution des classes pour la tâche 1 et la tâche 2.1 sur le corpus d'apprentissage et de test. Les données ont été accessibles par les identifiants des tweets et un script de téléchargement (fourni par les organisateurs)

Tâche 2.2				
Classes	**Apprentissage**		**Test**	
	#	%	#	%
déplaisir	47	67,72	21	1,54
dérangement	13	0,4	6	0,44
mépris	176	5,53	75	5,51
surprise négative	10	0,31	4	0,29
peur	274	8,60	114	8,37
colère	210	15,16	87	6,39
ennui	4	0,12	2	0,14
tristesse	36	1,13	16	1,17
plaisir	35	1,10	15	1,10
apaisement	9	0,28	5	0,36
amour	8	0,25	4	0,29
surprise positive	4	0,12	2	0,14
satisfaction	73	2,29	32	2,35
insatisfaction	9	0,28	5	0,36
accord	154	4,83	67	4,92
valorisation	1504	47,25	644	20,23
désaccord	216	6,77	92	6,76
dévalorisation	401	12,60	170	12,49
Total	**3183**	**100**	**1361**	**100**

TABLE 2 – Distribution des classes pour la tâche 2.2 sur le corpus d'apprentissage et de test. Les données ont été accessibles par les identifiants des tweets et un script de téléchargement (fourni par les organisateurs)

Les méthodes d'analyse de sentiments, d'opinions et d'émotions sont généralement basées sur des techniques statistiques, de traitement automatique du langage et d'apprentissage supervisé. Les plus performantes se basent souvent sur des systèmes supervisés qui utilisent des lexiques adaptés (Nakov *et al.*, 2013; Rosenthal *et al.*, 2014). Dans ce travail, nous avons choisi d'utiliser les machines à vecteurs de support (SVM) tout en exploitant deux lexiques de sentiments (polarités) et d'émotions que nous avons construit au préalable. Le premier est un lexique de sentiments et d'émotions qui a été construit en traduisant le lexique anglais NRC (Mohammad & Turney, 2010) d'une manière semi-automatique, supervisée par un traducteur humain expérimenté. Ce lexique a été étendu en anglais (avant traduction) et en français (après traduction) via l'étude des synonymes et des antonymes. Le deuxième est un lexique de sentiments (polarités seulement) construit d'une

manière automatique en utilisant le corpus d'apprentissage fourni par l'équipe d'organisation de DEFT 2015. Nous avons ensuite construit des attributs spécifiques pour prendre en considération ces deux lexiques dans l'apprentissage de nos modèles de classification. En plus des attributs exploitant les lexiques, nous en avons testé d'autres : les unigrammes de mots, les n-grammes de caractères (de longueur 5 et 6), le nombre d'émoticônes, le nombre de mots en majuscules, le nombre de lettres répétées, le nombre de hashtags, la présence de négateurs et les patrons syntaxiques les plus fréquents. Enfin, une étape de sélection d'attributs a été appliquée pour sélectionner les attributs les plus pertinents pour chaque tâche en ne conservant que les attributs pour lesquels le gain d'information est positif.

Le reste de l'article sera organisé comme suit : la section 2 décrit la création des ressources de sentiments et d'émotions utilisées. La section 3 présente les méthodes proposées : prétraitements, constructions et sélection d'attributs, classification. La section 4 présente les configurations choisies pour chaque tâche et les résultats obtenus. Enfin, la section 5 conclut et donne nos principales perspectives.

2 Création des ressources

Beaucoup de travaux liés à l'analyse de sentiments se basent sur des lexiques d'expressions de sentiments (liste de mots, phrases, idiomes, etc.). À ce titre, (Mohammad *et al.*, 2013) ont souligné l'importance des lexiques dans la classification des tweets suivant leur polarité. En effet, d'auteurs ont observé une augmentation qui dépasse les 8% de leurs macro F-scores en utilisant des lexiques qu'ils ont construit de manière manuelle et automatique. Cependant, la majorité de ces ressources a été construite pour l'anglais et la polarité. Très peu de lexiques existent pour le français et quand ils existent, ils ne contiennent pas beaucoup de termes. Pour cela, nous avons créé nous-mêmes deux ressources pour le français : le lexique de sentiments et d'émotions FEEL (Abdaoui *et al.*, 2014) et le lexique de sentiments basé sur l'information mutuelle (Church & Hanks, 1990).

2.1 Lexique de sentiments et d'émotions FEEL

Ce lexique a été construit de manière semi-automatique en traduisant et en étendant le lexique de sentiments et d'émotions anglais NRC (Mohammad & Turney, 2010). D'autres lexiques anglais existent à l'image de WordNet Affect (Strapparava *et al.*, 2004) mais à notre connaissance, NRC-2010 est le plus complet. Il inclut à la fois les polarités et les émotions ; surtout donne de bons résultats pour des tâches similaires aux nôtres (Kiritchenko *et al.*, 2014a). NRC-2010 associe à chaque terme une polarité (*positive* ou *négative*) et une ou plusieurs émotions parmi les six émotions proposées par (Ekman, 1992), à savoir : *joie, colère, tristesse, dégoût, surprise* et *peur*. Il a été construit manuellement en utilisant le service Amazon Mechanical Turk [1]. Afin de construire une ressource similaire pour le français, nous avons adopté une approche de traduction et d'extension semi-automatique. D'abord, pour chaque entrée de la ressource initiale, nous avons interrogé six traducteurs automatiques (Google, Bing, Collinsdictionary, Reverso, Babla et Wordreference). Les traductions ont été validées ou pas par un traducteur humain expérimenté. Le traducteur avait la possibilité de changer les polarités et les émotions associées via une interface graphique. Ensuite, nous avons émis l'hypothèse que la polarité était conservée par la synonymie. Nous avons donc étendu notre lexique aux synonymes. Huit outils en ligne ont été utilisés pour la recherche des synonymes (Babla, le dico de l'Institut des Sciences Cognitives, reverso, sensagent, cnrtl, synonym, thefreedictionary et thesaurus). Enfin, nous avons étendu notre lexique aux antonymes (en inversant les polarités) et en utilisant encore une fois deux outils en ligne (antonyme et cnrtl). Finalement, nous avons obtenu une ressource avec plus de 14 000 termes distincts (lemmatisés), chacun associé à une polarité et à une ou plusieurs émotions [2].

2.2 Lexique de sentiments basé sur l'information mutuelle

Étant donné que les résultats des systèmes basés sur les lexiques dépendent du type de données et du domaine d'application (Pang & Lee, 2008), nous avons décidé de construire automatiquement un lexique de sentiments en utilisant le corpus d'apprentissage fourni pour ce défi. Nous avons implémenté et adapté l'approche de (Kiritchenko *et al.*, 2014b). Au lieu de calculer un seul score pour chaque mot *w*, nous avons calculé trois scores (un par polarité) pour permettre la prise en compte de la classe neutre. Les trois scores sont présentés dans les formules (1), (2) et (3). Chaque score est basé sur la

1. https://www.mturk.com/mturk/welcome
2. Ce lexique est en téléchargement libre sur le lien : http://www.lirmm.fr/ abdaoui/FEEL.html

PMI (Pointwise Mutual Information (Bouma, 2009)) du mot dans un ensemble de tweets comme présenté dans la formule (4).

$$ScorePositif(w) = PMI(w, \{positif\}) - PMI(w, \{negatif\} \cup \{neutre\}) \quad (1)$$

$$ScoreNegatif(w) = PMI(w, \{negatif\}) - PMI(w, \{positif\} \cup \{neutre\}) \quad (2)$$

$$ScoreNeutre(w) = PMI(w, \{neutre\}) - PMI(w, \{positif\} \cup \{negatif\}) \quad (3)$$

$$PMI(w, \{NomDeLaClasse\}) = \log_2 \frac{freq(w, \{NomDeLaClasse\}) * N}{freq(w) * freq(\{NomDeLaClasse)\}} \quad (4)$$

où *freq(w, {NomDeLaClasse)}* est le nombre de fois où le mot *w* apparaît dans un tweet appartenant à *{NomDeLaClasse}*, *freq(w)* est le nombre d'apparitions du mot *w* dans le corpus, *freq ({NomDeLaClasse)}* est le nombre de tweets appartenant à la classe *NomDeLaClasse* et *N* est le nombre total de termes dans le corpus.

3 Méthodes

Pour chacune de nos trois tâches, nous appliquons d'abord des prétraitements. Ensuite, nous construisons des attributs pertinents pour la tâche en question. Une fois les attributs construits, nous ne sélectionnons que les plus pertinents. Enfin, nous apprenons un modèle de classification sur le corpus d'apprentissage et nous l'appliquons sur le corpus de test. Chacune de ces étapes est détaillée dans cette section.

3.1 Prétraitements

Comme indiqué par (Balahur, 2013), les textes issus des réseaux sociaux ont des particularités linguistiques qui peuvent influencer les performances de la classification. Pour cette raison, nous avons appliqué les prétraitements suivants : 1) le remplacement de tous les liens hypertextes qui figurent dans les tweets par 'lienHTPP' ; 2) le remplacement de toutes les adresses mails présentes dans les tweets par 'mail' ; 3) le remplacement de tous les tags utilisateur par '@tag' ; 4) la lemmatisation de tous les mots en utilisant l'outil TreeTagger (Schmid, 1994).

En outre, nous avons remarqué que certains prétraitements tels que le remplacement des mots d'argots, la mise en minuscules et le remplacement des mots allongés font chuter les macro-précisions (voir section 4). De ce fait, ces derniers n'ont pas été considéré dans nos soumissions.

3.2 Attributs

Pour la construction de nos attributs, nous avons d'abord testé ceux utilisés par (Mohammad *et al.*, 2013) lors du défi SemEval-2013 sur des tweets en langue anglaise en validations croisées sur le corpus d'apprentissage. Ensuite, nous les avons adaptés pour le français et nous en avons rajouté d'autres. Chaque tweet a donc été représenté par un sous ensemble des attributs suivants :

— **Unigrammes de mots** : présence ou absence des mots lemmatisés dans le tweet.
— **Caractères n-grammes** : présence ou absence des séquences continues de 5 et de 6 caractères. Nous n'avons sélectionné que les séquences les plus fréquentes par classe (celles qui apparaissent au moins 550 fois dans une seule classe). Ce seuil a été fixé par plusieurs validations croisées sur le corpus d'apprentissage.
— **Les patrons syntaxiques les plus fréquents** : pour construire ces attributs, nous avons d'abord utilisé l'outil TreeTagger pour remplacer chaque mot par son étiquette morphosyntaxique. Ensuite, nous avons utilisé l'algorithme de (Fournier-Viger *et al.*, 2008) pour l'extraction de motifs séquentiels fréquents. Nous en avons extrait les 1% les plus fréquents pour chaque classe. Nous n'avons gardé que les patrons discriminants, fréquents dans une classe et pas dans les autres. Ces derniers ont été rajoutés comme attributs de type « booléen ». Chaque attribut prendra la valeur *vraie* si l'étiquetage morphosyntaxique du tweet correspond au patron syntaxique en question.

— **Lexique de sentiments FEEL** : les attributs suivants ont été considérés : 1) le nombre de mots positifs ; 2) le nombre de mots négatifs ; 3) la somme des scores de tous les mots ; 4) le score le plus élevé ; 5) le score du dernier mot comme cela a été fait par (Mohammad *et al.*, 2013).
— **Lexique d'émotions FEEL** : les attributs suivants ont été considérés : 1) le nombre de mots exprimant la confiance ; 2) le nombre de mots exprimant la joie ; 3) le nombre de mots exprimant la colère ; 4) le nombre de mots exprimant la tristesse ; 5) le nombre de mots exprimant le dégoût ; 6) le nombre de mots exprimant la surprise ; 7) le nombre de mots exprimant la peur.
— **Lexique basé sur l'information mutuelle** : les attributs suivants ont été considérés pour ce lexique : 1) le nombre de mots positifs ; 2) le nombre de mots négatifs ; 3) le nombre de mots neutres ; 4) la somme de tous les scores positifs ; 5) la somme de tous les scores négatifs ; 6) la somme de tous les scores neutres ; 7) le score positif le plus élevé ; 8) le score négatif le plus élevé ; 9) le score neutre le plus élevé ; 10) le score positif du dernier mot ; 11) le score négatif du dernier mot ; 12) le score neutre du dernier mot.
— **Ponctuation** : deux attributs ont été considérés : 1) le nombre de séquences continues de points d'exclamation, de points d'interrogation et des deux ; 2) la présence ou l'absence d'un point d'exclamation ou d'un point d'interrogation dans le dernier terme dans n'importe quel position.
— **Émoticônes** : présence ou non d'un émoticône dans le tweet.
— **Les mots allongés** : nombre de mots avec au moins 3 caractères répétés séquentiellement (par exemple : *mdrrrr*).
— **Négation** : présence ou l'absence d'un négateur dans le tweet.

Concernant la négation, nous avons aussi essayé d'implémenter la méthode proposée dans (Kiritchenko *et al.*, 2014a) en rajoutant un suffixe aux mots qui se trouvent sous la portée d'un négateur. Les résultats de cette approche sont décrits dans la section 4.

3.3 Sélection d'attributs

Afin de sélectionner les attributs les plus discriminants pour chaque tâche, nous avons appliqué une étape de sélection d'attributs en mesurant le gain d'information de chaque attribut par rapport à la classe (Mitchell, 1997). L'équation 5 présente la formule utilisée :

$$GainInformation(attribut, classe) = Entropie(classe) - Entropie(classe|attribut) \quad (5)$$

Après avoir calculé le gain d'information pour chaque attribut, nous ne gardons que ceux pour lesquels ce gain est supérieur à 0. Pour la tâche 1, 762 attributs ont été sélectionnés dont les plus discriminants sont ceux obtenus à partir du lexique de la PMI suivis par ceux obtenus à partir du lexique FEEL. Pour la tâche 2.1, 460 attributs ont été sélectionnés dont les plus discriminants sont des mots comme : *menacer, contre, lien,* etc. Pour la tâche 2.2, 75 attributs ont été sélectionnés dont les plus discriminants sont ceux issus du lexique d'émotions FEEL et des mots comme : *menacer, espèce, lien,* etc.

3.4 Classification

Pour nos trois tâches, nous employons des méthodes d'apprentissage supervisé. Nous avons choisi d'utiliser les machines à vecteur de support SVM (Support Vector Machine) avec la méthode SMO (Sequential Minimal Optimization) (Platt, 1999) implémenté dans Weka (Hall *et al.*, 2009). D'après l'état de l'art, cet algorithme d'apprentissage s'est avéré efficace sur des tâches de catégorisation de textes et spécifiquement d'analyse de sentiments et d'émotions. Il est robuste sur les grands espaces de caractéristiques. En effet, notre modèle de classification exploite une variété de surface de forme, sémantique et des sentiments caractéristiques issus des lexiques.

Afin de choisir le meilleur paramètre de complexité « C » de SVM, nous avons effectué 20 validations croisées à 10 plis sur le corpus d'apprentissage. Pour chaque tâche, nous avons testé toutes les valeurs comprises entre 0,1 et 2,0 avec un pas de 0,1. Pour les trois tâches, la meilleure macro précision en validation croisée a été obtenue avec la valeur 0,4. Nous avons donc choisi cette valeur pour le paramètre de complexité « C » dans les expérimentations décrites ci-dessous.

4 Expérimentations

Dans cette section, nous présentons les configurations choisies pour chaque tâche, les résultats obtenus et leurs discussions.

4.1 Tâche 1 : Détection de la polarité d'un tweet

L'objectif de cette tâche est de déterminer la polarité d'un tweet. Il s'agit de prédire si un tweet donné est *positif, négatif* ou *neutre*. Les attributs utilisés pour cette tâche sont les suivants : les n-grammes de mots, les n-grammes de caractères, les patrons syntaxiques, le lexique de sentiments FEEL, le lexique de sentiments de la PMI, la ponctuation, les émoticônes, les mots allongés et la négation (attribut booléen). Le tableau 3 présente les macro-précisions de plusieurs expérimentations que nous avons effectué pour déterminer l'effet des attributs que nous avons construit et des étapes que nous avons effectué. Nous testons plusieurs configurations en rajoutant ou en enlevant des attributs ou des étapes.

Expérimentations	**Macro-précision**
Baseline (prédire la classe majoritaire)	33,33%
Système proposé	73,33%
- Unigrammes de mots lemmatisés	42,39% (-30,94%)
- N-grammes de caractères de longueur 5 et 6	73,46% (+0,13%)
- Lexique de polarités FEEL	72,45% (-0,88%)
- Lexique de polarités PMI	73,32% (-0,01%)
- Patrons syntaxiques les plus fréquents	72,32%(-0,01%)
- Négation (attribut booléen)	73,33% (-0,00%)
- Ponctuation, émoticônes, hashtags et mots allongés	73,45% (+0,12%)
- Sélection d'attributs	68,92% (-4,41%)
- Prétraitements {Lemmatisation, lien, mail, tag}	70,71% (-2,62%)
+ Prétraitements {Argots, minuscules, mots allongés}	73,00% (-0,33%)
+ Négation (rajouter un suffixe)	73,15% (-0,18%)

TABLE 3 – Macro-Précisions obtenues pour la tâche 1

Nous avons entraîné des classifieurs de types SVM sur un ensemble de 7 867 tweets, puis nous avons appliqué les modèles appris sur les 3 379 tweets qui nous ont été fournis pour la phase de test. La mesure choisie par les organisateurs du défi est la macro- précision. Comme baseline, nous considérons un système qui prédit toujours la classe majoritaire (ici la classe neutre). La macro-précision obtenue par un tel système est de 33,33%. Concernant le système que nous avons soumis au défi, la macro-précision obtenue est de 73,33%. Nous constatons que les attributs donnant le plus de gain sont les *unigrammes de mots lemmatisés*, ils fournissent un gain de plus de 30,94%. Par contre, le fait d'enlever les n-grammes caractères fait augmenter la macro-précision de 0,13%. Ces derniers ne permettent donc pas d'améliorer les résultats, voir ils les font même chuter un peu. Le lexique de polarités FEEL permet d'obtenir un gain de 0,88%. Concernant la négation en utilisant un attribut booléen, les patrons syntaxiques les plus fréquents et le lexique de polarités de la PMI, ces attributs n'ont pas vraiment eu d'impact sur les performances de notre système. La ponctuation, les émoticônes, le nombre de hashtags et de mots allongés font chuter la macro-précision du système de 0,12%, alors que la sélection d'attributs parait une étape cruciale puisqu'elle permet d'obtenir un gain de 4,41%. Les prétraitements que nous avons appliqués nous ont permis d'obtenir un gain non négligeable de 2,62%. Finalement, le fait de rajouter les prétraitements que nous avons décidé d'écarter et la méthode de traitement de la négation par ajout de suffixe fait chuter la macro-précision de 0,33% et de 0,18% respectivement.

4.2 Tâche 2.1 : Identification de la classe générique

Il s'agit ici d'identifier la classe générique de l'information exprimée dans un tweet. Les 4 classes génériques proposées dans le cadre de cette tâche sont : *information, opinion, sentiment* et *émotion*. Les attributs utilisés pour cette tâche sont les suivants : les n-grammes de mots, les n-grammes de caractères, la ponctuation, les émoticônes, les mots allongés et la négation. Le tableau 4 présente les macro-précisions de plusieurs expérimentations que nous avons effectué pour

déterminer l'effet des attributs que nous avons construit et des étapes que nous avons proposé pour la tâche 2.1. Nous testons plusieurs configurations en rajoutant ou en enlevant des attributs ou des étapes.

Expérimentations	**Macro-précision**
Baseline (prédire la classe majoritaire)	25,00%
Système proposé	61,29%
- Unigrammes de mots lemmatisés	11,23% (-50,06%)
- N-grammes de caractères de longueur 5 et 6	61,45% (-0,16%)
- Négation (attribut booléen)	61,32% (+0,03%)
- Ponctuation, émoticônes, hashtags et mots allongés	61,17% (-0,12%)
- Sélection d'attributs	55,67% (-5,62%)
- Prétraitements {Lemmatisation, lien, mail, tag}	62,27% (+0,98%)
+ Prétraitements {Argots, minuscules, mots allongés}	62,16% (+0,87%)
+ Négation (rajouter un suffixe)	61,26% (-0,03%)

TABLE 4 – Macro-Précisions obtenues pour la tâche 2.1

Nous avons entraîné des classifieurs de types SVM sur un ensemble de 6 784 tweets, puis nous avons appliqué nos modèles sur les 3 379 tweets qui nous ont été fournis après pour la phase de test. Nous utilisons la même baseline qui consiste à prédire la classe majoritaire ce qui donnera cette fois 25% de macro-précision (puisqu'il y a quatre classes). La macro-précision obtenue par le système soumis au défi est de 61,29% (meilleur résultat pour cette tâche). Cependant, le tableau 4 montre qu'il est possible d'améliorer encore ce résultat pour dépasser les 62%. Les unigrammes de mots lemmatisés donnent le meilleur gain (plus de 50%). Ensuite et comme pour la première tâche, en faisant une sélection d'attributs on augmente les résultats de plus de 5%, ce qui rend cette étape aussi cruciale pour la tâche 1 que pour la tâche 2.1. Les n-grammes de caractères donnent de leur côté un faible de gain de 0,16%. Par ailleurs, la négation par attribut booléen et par ajout de suffixe n'a pas vraiment d'impact sur la macro-précision. La ponctuation, les émoticônes, le hashtags et les mots allongés font chuter la macro-précision de 0,12%. Les prétraitements que nous avons choisi d'appliquer (à savoir : la lemmatisation, le remplacement des liens, des mails, et des tags) semblent être inadaptés pour cette tâche puisqu'en les appliquant nous perdons 0,92%, alors qu'en appliquant les autres prétraitements (remplacement des mots d'argots, la mise en minuscules et le traitement des mots allongés) qui permettent d'obtenir un gain de 0,87%.

4.3 Tâche 2.2 : Identification de la classe spécifique de l'opinion, du sentiment ou de l'émotion

Il s'agit d'identifier la classe de l'opinion, sentiment ou émotion. Étant donné un tweet, il faudrait reconnaître l'opinion/sentiment/émotion principal(e) exprimé(e) explicitement dans ce tweet. Pour cela, 18 classes sont proposées : *colère, peur, tristesse, dégoût, ennui, dérangement, déplaisir, surprise négative, apaisement, amour, plaisir, surprise positive, insatisfaction, satisfaction, accord, valorisation, désaccord et dévalorisation*. Lors de notre soumission nous n'avons pas considérés ces classes-là, nous avons donc obtenu des macro-précisions très faibles. Nous présentons ici les résultats du même système soumis mais en considérant les bonnes classes. Les attributs utilisés pour cette tâche sont les suivants : les n-grammes de mots, les n-grammes de caractères, le lexique d'émotions FEEL, la ponctuation, les émoticônes, les mots allongés et la négation. Le tableau 5 présente les macro-précisions de plusieurs expérimentations que nous avons effectué pour déterminer l'effet des attributs que nous avons construit et des étapes que nous avons effectué pour la tâche 2.2. Nous testons plusieurs configurations en rajoutant ou en enlevant des attributs ou des étapes au système proposé.

Nous avons entraîné des classifieurs de types SVM sur un ensemble de 3 162 tweets, puis nous avons appliqué nos modèles sur les 3 379 tweets qui nous ont été fournis après pour la phase de test. Nous utilisons la même baseline qui consiste à prédire la classe majoritaire ce qui donnera cette fois 5,56% de macro-précision (puisqu'il y a 18 classes). La macro-précision obtenue par le système soumis au défi et appris sur les bonnes classes est de 31,72%. Cependant, le tableau 5 montre qu'il est possible d'améliorer encore ce résultat pour dépasser les 37%. Encore une fois, les unigrammes de mots lemmatisés donnent le meilleur gain (25,72%). Ensuite, le lexique de polarités FEEL, la ponctuation, les émoticônes, le hashtags et les mots allongés n'ont pas d'impact sur les résultats. Par contre, lexique d'émotions FEEL (qui contient 7 émotions seulement) et la négation par attributs booléen ont considérablement diminué la macro-précision. Le lexique a causé une perte de 2,66%, alors que la négation par attribut booléen a causé une perte de 5,28%. À l'opposé des deux tâches précédentes, la sélection d'attributs a causé une perte 0,66% pour la tâche 2.2. Par ailleurs, les prétraitements que nous avons choisis d'appliquer (à savoir : la lemmatisation, le remplacement des liens, des mails, et des tags) semblent être

Expérimentations	Macro-précision
Baseline (prédire la classe majoritaire)	5,56%
Système proposé	31,72%
- Unigrammes de mots lemmatisés	6,00% (-25,72%)
- Lexique de polarités FEEL	31,72% (0,00%)
- Lexique d'émotions FEEL	34,38% (+2,66%)
- Négation (attribut booléen)	37,00% (+5,28%)
- Ponctuation, émoticônes, hashtags et mots allongés	31,72% (0,00%)
- Sélection d'attributs	32,38% (+0,66%)
- Prétraitements {Lemmatisation, lien, mail, tag}	29,94% (-1,78%)
+ Prétraitements {Argots, minuscules, mots allongés}	35,93% (+4,21%)
+ Négation (rajouter un suffixe)	31,65% (-0,07%)

TABLE 5 – Macro-Précisions obtenues pour la tâche 2.2

adaptés pour cette tâche puisqu'en les appliquant nous obtenons un gain 1,78%. Les autres prétraitements (remplacement des mots d'argots, la mise en minuscules et le traitement des mots allongés) permettent d'obtenir un gain de 4,21%. Finalement et comme pour les deux premières tâches, la méthode de négation en rajoutant un suffixe à tous les mots qui se trouvent sous la portée d'un négateur fait chuter légèrement la macro-précision du système (0,07%).

5 Conclusion

Nous avons présenté les systèmes soumis à la onzième édition du défi de fouille de texte (DEFT 2015). Les systèmes proposés sont basés sur des classifieurs de types SVM exploitant des traits d'ordre morphologique, syntaxique et sémantique. De plus, nous avons construit et exploité deux lexiques de sentiments et d'émotions. Le premier est un lexique de sentiments et d'émotions obtenu en traduisant semi-automatiquement le lexique NRC alors que le deuxième est un lexique de sentiments obtenu d'une manière automatique en utilisant le corpus d'apprentissage. Le système soumis pour la tâche 1 a obtenu une macro-précision de 73,33% à 0,27% du meilleur système. Le système que nous avons proposé pour la tâche 2.1 a obtenu une macro-précision de 61,29% soit la meilleure macro-précision pour cette tâche. Le système soumis pour la tâche 2.2 et appris sur les bonnes classes a obtenu une macro-précision de 31,72%. À noter que pour cette tâche, nous obtenons une macro-précision de 37% pour une de nos expérimentations (en enlevant la négation), alors que le meilleur système pour cette tâche soumis au défi a obtenu 34.68%. À travers les résultats obtenus, nous préconisons l'utilisation de SVM comme classifieur et des unigrammes comme attributs pour les trois tâches proposées dans ce défi. De plus, nous constatons que la méthode de traitement de la *négation* en rajoutant un suffixe à tous les mots qui se trouvent sous la portée d'un négateur ne donne pas de meilleures macro-précisions pour les trois tâches. On pourrait donc rejoindre la conclusion de (Vincent & Winterstein, 2013) qui montre que cette approche qui a été créée pour l'anglais tend à ne pas fonctionner aussi bien pour le français. Enfin, nous soulignons l'importance de la sélection d'attributs pour les tâches 1 et 2.1 qui améliorent les macro-précisions de plus de 4%.

Comme perspectives, nous prévoyons de tester plusieurs autres configurations que nous n'avons pas eu le temps de tester pour ce défi. Une perspective à court terme est de rajouter comme attributs les identifiants des clusters de Brown obtenus sur le corpus d'apprentissage (Brown *et al.*, 1992; Blitzer & Zhu, 2008). Le Clustering de Brown permet de représenter des textes sous un format moins sparse, et cela en regroupant les mots selon leurs contextes. Nous avons déjà implémenté cette méthode mais la construction des attributs prenant beaucoup de temps, nous n'avons donc pas pu les intégrer durant la période de test qui a duré 3 jours seulement. Une deuxième perspective est d'estimer le seuil du gain d'information de l'algorithme de sélection d'attributs en validation croisée comme cela a été fait pour le paramètre de complexité de SVM. En effet, ces estimations peuvent se faire en même temps comme décrits dans (Li *et al.*, 2015). Pour pallier au problème des classes non équilibrées dans la tâche 2.2, nous comptons aussi tester une technique de sur-échantillonnage implémentée dans Weka et appelée SMOTE (Synthetic Minority Oversampling TEchnique) (et. al., 2002). Cette technique crée de nouvelles instances pour les classes minoritaires en se basant sur les instances existantes. Finalement, il serait intéressant de tester nos méthodes sur d'autres domaines que l'environnement et d'autres types de données que les tweets et d'essayer d'en faire des méthodes génériques pour le français.

Références

ABDAOUI A., JÉRÔME A., BRINGAY S. & PONCELET P. (2014). Feel : French extended emotional lexicon. volume ISLRN : 041-639-484-224-2. ELRA Catalogue of Language Resources.

AUGUSTYN M., BEN HAMOU S., BLOQUET G., GOOSSENS V., LOISEAU M. & RINCK F. (2006). Lexique des affects : constitution de ressources pédagogiques numériques. In *Colloque International des étudiants-chercheurs en didactique des langues et linguistique.*, p. 407–414, Grenoble, France.

BALAHUR A. (2013). Sentiment analysis in social media texts. In *4th workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, p. 120–128 : Citeseer.

BLITZER J. & ZHU X. J. (2008). Semi-supervised learning for natural language processing. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies : Tutorial Abstracts*, p. 3–3 : Association for Computational Linguistics.

BOUMA G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, p. 31–40.

BROWN P. F., DESOUZA P. V., MERCER R. L., PIETRA V. J. D. & LAI J. C. (1992). Class-based n-gram models of natural language. *Computational linguistics*, **18**(4), 467–479.

CHURCH K. W. & HANKS P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, **16**(1), 22–29.

EKMAN P. (1992). An argument for basic emotions. *Cognition & emotion*, **6**(3-4), 169–200.

ET. AL. N. V. C. (2002). Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, **16**, 321–357.

FOURNIER-VIGER P., NKAMBOU R. & NGUIFO E. M. (2008). A knowledge discovery framework for learning task models from user interactions in intelligent tutoring systems. In *MICAI 2008 : Advances in Artificial Intelligence*, p. 765–778. Springer.

GROUIN C., HURAULT-PLANTET M., PAROUBEK P. & BERTHELIN J.-B. (2009). Deft'07 : une campagne d'évaluation en fouille d'opinion. *Fouille de données d'opinion*, **17**, 1–24.

HALL M., FRANK E., HOLMES G., PFAHRINGER B., REUTEMANN P. & WITTEN I. H. (2009). The weka data mining software : an update. *ACM SIGKDD explorations newsletter*, **11**(1), 10–18.

KIRITCHENKO S., ZHU X., CHERRY C. & MOHAMMAD S. (2014a). Nrc-canada-2014 : Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, p. 437–442, Dublin, Ireland : Association for Computational Linguistics and Dublin City University.

KIRITCHENKO S., ZHU X. & MOHAMMAD S. M. (2014b). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, p. 723–762.

LAVER M., BENOIT K. & GARRY J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, **97**(02), 311–331.

LI X., LI J. & WU Y. (2015). A global optimization approach to multi-polarity sentiment analysis. *PLoS ONE 10(4)*.

MELNIK M. I. & ALM J. (2002). Does a seller's ecommerce reputation matter ? evidence from ebay auctions. *The journal of industrial economics*, **50**(3), 337–349.

MITCHELL T. M. (1997). *Machine learning.*, volume 45. Burr Ridge, IL : McGraw Hill.

MOHAMMAD S. (2012). Portable features for classifying emotional text. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, p. 587–591, Montréal, Canada : Association for Computational Linguistics.

MOHAMMAD S. M., KIRITCHENKO S. & ZHU X. (2013). Nrc-canada : Building the state-of-the-art in sentiment analysis of tweets. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics (SEMSTAR13)*, p. 321.

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*, CAAGET '10, p. 26–34, Stroudsburg, PA, USA : Association for Computational Linguistics.

MULLEN T. & MALOUF R. (2006). A preliminary investigation into sentiment analysis of informal political discourse. In *AAAI Spring Symposium : Computational Approaches to Analyzing Weblogs*, p. 159–162.

NAKOV P., KOZAREVA Z., RITTER A., ROSENTHAL S., STOYANOV V. & WILSON T. (2013). Semeval-2013 task 2 : Sentiment analysis in twitter.

PANG B. & LEE L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, **2**(1-2), 1–135.

PLATT J. C. (1999). Advances in kernel methods. chapter Fast Training of Support Vector Machines Using Sequential Minimal Optimization, p. 185–208. Cambridge, MA, USA : MIT Press.

ROBERTS K., ROACH M. A., JOHNSON J., GUTHRIE J. & HARABAGIU S. M. (2012). Empatweet : Annotating and detecting emotions on twitter. In *LREC*, p. 3806–3813.

ROSENTHAL S., NAKOV P., KIRITCHENKO S., MOHAMMAD S., RITTER A. & STOYANOV V. (2015). Semeval-2015 task 10 : Sentiment analysis in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, p. 451–463, Denver, Colorado : Association for Computational Linguistics.

ROSENTHAL S., RITTER A., NAKOV P. & STOYANOV V. (2014). Semeval-2014 task 9 : Sentiment analysis in twitter. p. 73–80.

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, p. 44–49 : Citeseer.

STRAPPARAVA C. & MIHALCEA R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, p. 1556–1560 : ACM.

STRAPPARAVA C., VALITUTTI A. *et al.* (2004). Wordnet affect : an affective extension of wordnet. In *LREC*, volume 4, p. 1083–1086.

TATEMURA J. (2000). Virtual reviewers for collaborative exploration of movie reviews. In *Proceedings of the 5th international conference on Intelligent user interfaces*, p. 272–275 : ACM.

VINCENT M. & WINTERSTEIN G. (2013). Construction et exploitation d'un corpus français pour l'analyse de sentiment. *TALN-RÉCITAL 2013*, p. 764.