

FrenchSentiClass : un Système Automatisé pour la Classification de Sentiments en Français

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

► **To cite this version:**

Mike Donald Tapi Nzali, Amine Abdaoui, Jérôme Azé, Sandra Bringay, Christian Lavergne, et al..
FrenchSentiClass : un Système Automatisé pour la Classification de Sentiments en Français. Actes de
l'atelier DEFT de la conférence TALN 2017, Jun 2017, Orléans, France. 2017. <lirmm-01563411>

HAL Id: lirmm-01563411

<https://hal-lirmm.ccsd.cnrs.fr/lirmm-01563411>

Submitted on 17 Jul 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.

FrenchSentiClass : un Système Automatisé pour la Classification de Sentiments en Français

Mike Donald Tapi Nzali^{1, 2} Amine Abdaoui¹ Jérôme Azé¹ Sandra Bringay¹
Christian Lavergne² Caroline Mollevi³ Pascal Poncelet¹

(1) LIRMM, 860 Rue Saint Priest 34095 Montpellier, France

(2) IMAG, Place Eugene Bataillon 34095 Montpellier, France

(3) ICM - Unité de biostatistique, 208 Avenue des Apothicaires 34298 Montpellier, France

mike-donald.tapi-nzali@lirmm.fr, a_abdaoui@esi.dz

RÉSUMÉ

Ce papier décrit le système *FrenchSentiClass* que nous avons utilisé pour les tâches du défi de fouilles de texte (DEFT 2017). Cette treizième édition du défi a porté sur l'analyse de l'opinion et du langage figuratif dans des tweets rédigés en Français. Le défi propose trois tâches : (i) la première concerne la classification des tweets non figuratifs selon leur polarité; (ii) la deuxième concerne l'identification du langage figuratif et (iii) la troisième concerne la classification des tweets figuratifs et non figuratifs selon leur polarité. Nous avons proposé un système automatisé basé sur les Machines à Vecteurs de Support (SVM). Le système choisit automatiquement à chaque niveau les meilleurs prétraitements, descripteurs syntaxiques et lexiques de sentiments en validation croisée sur l'ensemble d'apprentissage. Il effectue aussi une évaluation de l'apport de la sélection d'attributs et un tuning du paramètre de complexité du modèle SVM. Par conséquent, ce système permet de réduire considérablement le temps d'exploration des données et du choix de la meilleur représentation de descripteurs.

ABSTRACT

FrenchSentiClass : an Automated System for French Sentiment Classification

This paper describes the system we used on the tasks of the text mining challenge (DEFT 2017). This thirteenth edition of this challenge concerned the analysis of opinions and figurative language in French tweets. Three tasks have been proposed : (i) the first one concerns the classification of non-figurative tweets according to their polarity ; (ii) the second one concerns the identification of figurative language, while (iii) the third one concerns the classification of figurative and non-figurative tweets according to their polarity. We proposed an automated system based on Support Vector Machines (SVM). The system automatically chooses on each step the best preprocessing, syntactic features and sentiment lexicons by cross validation on the training set. Furthermore, it performs an evaluation of feature subset selection and a tuning SVM complexity parameter. Therefore, this system can significantly reduce the time necessary to explore the data and choose the best feature representation.

MOTS-CLÉS : Analyse d'opinions, détection de polarité, langage figuratif.

KEYWORDS: Opinion analysis, polarity detection, figurative language.

1 Introduction

L'analyse de sentiment est l'étude computationnelle et sémantique des parties de textes en fonction des opinions, des sentiments et des émotions exprimés dans le texte (Liu, 2010, 2015). Le plus souvent, l'expression « analyse des sentiments » est utilisée pour désigner la tâche de classification automatique des unités de texte en fonction de leur polarité. Cependant, cette expression couvre un plus grand nombre de tâches relatives à l'attitude générale de l'auteur du texte vers une cible particulière (Liu, 2012). En effet, l'attitude de l'auteur peut être observée à travers de multiples dimensions : sa polarité (positive, négative ou neutre) (Pang *et al.*, 2002), sa subjectivité (objective ou subjective) (Riloff *et al.*, 2005), l'émotion exprimée (joie, surprise, colère, etc.) (Mohammad & Kiritchenko, 2015), son intensité (soit discrète (Pang & Lee, 2005) ou les valeurs réelles des sentiments (Kiritchenko *et al.*, 2016)), etc. D'autre part, l'attitude de l'auteur présentée (polarité, subjectivité, émotion, etc.) peut être étudiée à différents niveaux de granularité : au niveau du document (Turney, 2002), au niveau de la phrase (Wilson *et al.*, 2005) et au niveau aspectuel (Pontiki *et al.*, 2014). Dans cet article, nous décrivons *FrenchSentiClass* un système automatisé pour la classification de sentiment en langue française au niveau document.

Les méthodes d'analyse des sentiments et d'opinions sont généralement basées sur des techniques statistiques, de traitement automatique du langage et d'apprentissage supervisé. Elles sont souvent distinguées en méthodes supervisées nécessitant des données d'entraînement (Pang *et al.*, 2002) et méthodes non supervisées souvent basées sur des lexiques (Turney, 2002). Récemment, il a été démontré que combiner l'utilisation des lexiques de sentiments avec des systèmes supervisés permet d'améliorer les résultats (Nakov *et al.*, 2013; Rosenthal *et al.*, 2014). Dans ce papier nous décrivons un système basé sur les Machines à Vecteurs de Support (SVM) utilisant diverses caractéristiques syntaxiques (n-grammes, POS, etc.) et lexicaux (lexiques de sentiments et d'émotions). La particularité de ce système est qu'il réduit considérablement le temps passé à choisir la meilleure représentation de descripteurs selon le type de données utilisé (tweets, commentaires, etc.) ou la tâche considérée (polarité, subjectivité, etc.). Le code source du système est disponible sur le dépôt Github¹.

Un des challenges actuels de l'analyse de sentiment se trouve au niveau du langage traité. En effet, les subtilités du langage humain ne permettent pas aux algorithmes actuels d'apprendre efficacement à prédire le sentiment exprimé. Contrairement au langage usuel, le langage figuratif détourne le sens propre pour lui conférer un sens dit figuré ou imagé, comme l'ironie, le sarcasme, l'humour, la métaphore ou encore les jeux de mots. Quelques travaux se sont intéressés à cette problématique pour le Français (Karoui *et al.*, 2015). Dans cet article, nous évaluons notre système sur des tâches d'analyse d'opinions dans un contexte de langages figuratif et non-figuratif. Nous présentons les résultats que nous avons obtenus sur toutes les tâches proposées à la treizième édition du défi de fouille de textes (DEFT 2017). La première tâche (tâche 1) à laquelle nous avons participé consiste à classer les tweets non-figuratifs selon leur polarité (*positif*, *négatif*, *neutre* ou *mixte*). Dans la seconde tâche (tâche 2), il s'agit d'identifier si un tweet contient du langage figuratif ou non (*figuratif* ou *non figuratif*). La dernière tâche (tâche 3) du défi consiste à classer les tweets figuratifs et non figuratifs selon leur polarité (*positif*, *négatif*, *neutre* ou *mixte*).

Le reste de l'article sera organisé comme suit. La section 2 décrit les méthodes utilisées dans *FrenchSentiClass* (prétraitements, descripteurs syntaxiques, lexiques ainsi que le processus de sélection du paramètre de complexité du modèle). La section 3 présente les configurations choisies et les résultats obtenus pour chacune des tâches. Enfin, la section 4 conclut ce travail.

1. <https://github.com/amineabdaoui/french-sentiment-classification>

2 Méthodes

L'idée de base du système *FrenchSentiClass* consiste à utiliser le corpus d'apprentissage afin de choisir automatiquement les meilleurs prétraitements, descripteurs syntaxiques et ressources lexicales selon une mesure d'évaluation choisie. Ce processus s'effectue sur plusieurs étapes regroupant une ou plusieurs caractéristiques semblables. À chaque étape le système sélectionne la meilleure combinaison de caractéristiques en prenant comme baseline ceux choisies à l'étape précédente. Le corpus d'apprentissage est séparée une seule fois au départ du processus et toutes les évaluations sont effectuées sur les mêmes plis de validation croisée. La figure 1 présente l'architecture du système.

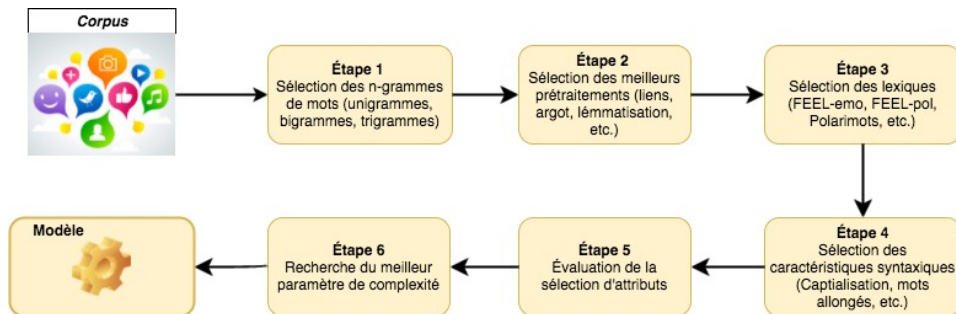


FIGURE 1: Architecture du système.

Dans cette section, nous détaillons chacune de ces étapes :

2.1 Étape 1 : N-grammes de mots

Les n-grammes de mots sont considérés comme les caractéristiques de base dans la classification de textes, y compris l'analyse des sentiments. Dans la littérature, il a été rapporté que l'utilisation de la représentation binaire fonctionne mieux que les représentations basées sur la fréquence pour l'analyse des sentiments (Pang *et al.*, 2002; Liu, 2012). Par conséquent, nous considérons la présence ou l'absence des unigrammes (U), des bigrammes (B) et des trigrammes (T). Ici le système va donc évaluer les résultats obtenus par toutes les combinaisons possibles et ne garder que celle qui donne le meilleur résultat (U, B, T, U+B, U+T, B+T et U+B+T).

Il est important de mentionner qu'à l'issue de cette première étape le système effectue un premier tuning du paramètre de complexité du SVM en utilisant une recherche exhaustive (entre 0 et 1 avec un pas de 0.1). Un deuxième tuning plus fin sera appliqué à la dernière étape.

2.2 Étape 2 : Prétraitements

Comme mentionné dans (Haddi *et al.*, 2013), les textes des médias sociaux ont des particularités linguistiques qui peuvent affecter la performance des classificateurs. Pour cette raison, les prétraitements suivants ont été mis en oeuvre :

1. Normalisation des liens, emails et pseudonymes ;

2. Remplacement des mots d'argot avec le texte correspondant en utilisant une liste préétablie ;
3. Mise en minuscule ;
4. Lemmatisation en utilisant TreeTagger (Schmid, 1994) ;
5. Suppression des mots vides.

Chaque combinaison de ces prétraitements est évaluée et celle qui donne le meilleur résultat sera retenue.

2.3 Étape 3 : Ressources lexicales

Les sentiments sont principalement véhiculés par des mots. Par conséquent, de nombreuses études ont construit des ressources de sentiments qui se composent de listes de mots, de phrases ou d'expressions idiomatiques en classes prédéfinies (polarité, émotion, etc.). Les lexiques que nous avons utilisés sont décrits ci-dessous :

- **FEEL** : contient plus de 14 000 termes pour le français. Il associe chaque terme à deux polarités (positive, négative) et à six émotion (joie, colère, peur, tristesse, dégoût et surprise). Le lexique a été obtenu de manière semi-automatique. D'abord en traduisant et en étendant aux synonymes le lexique anglais NRC EmoLex. Ensuite, en faisant valider chacune des entrées par un annotateur humain. Une partie des sentiments associés à chaque terme a été annoté par plusieurs annotateurs humains (Abdaoui *et al.*, 2016) ;
- **Affect** : se compose d'environ 1 300 termes français décrits par leur polarité (positive et négative) et plus de 45 catégories hiérarchiques émotionnelles. Il a été construit automatiquement et comprend d'autres informations telles que l'intensité et le niveau de langue (commun, littéraire) (Augustyn *et al.*, 2006) ;
- **Diko** : est basé sur un jeu en ligne avec un but où les joueurs sont invités à indiquer la polarité et l'émotion de l'expression affichée. Ils peuvent choisir entre trois polarités (positive, négative et neutre) et 21 émotions. Ils peuvent également entrer un nouveau type d'émotion lorsque le sens de l'émotion exacte de l'expression affichée n'est pas présent entre les 21 choix. Par conséquent, ce lexique associe 555 441 expressions annotées à près de 1 200 termes d'émotion (Lafourcade *et al.*, 2015) ;
- **Polarimots** : contient 7 483 noms, verbes, adjectifs et adverbes français dont la polarité (positive, négative ou neutre) a été semi-automatiquement annotée. 3 247 mots ont été ajoutés manuellement et 4 236 mots ont été créés automatiquement en propageant les polarités (Gala & Brun, 2012).

En utilisant ces ressources, nous avons construit les attributs suivants :

1. Nombre de termes exprimant chaque polarité et chaque émotion selon les 4 lexiques (FEEL-pol : 2 attributs, FEEL-emo : 6 attributs, Affects-pol : 3 attributs, Affects-emo : 45 attributs, Diko-pol : 3 attributs, Diko-emo : 1,198 attributs and Polarimots-pol : 3 attributs).
2. Le nombre d'incongruité de la polarité (représenté par un seul attribut qui calcul le nombre de fois ou un terme positif est suivi par un terme négatif ou qu'un terme négatif est suivi par un terme positif).

2.4 Étape 4 : Caractéristiques syntaxiques

Les caractéristiques syntaxiques présentées ci-dessous et inspirées de (Mohammad *et al.*, 2013) ont été implémentées :

1. Mots allongés : nombre de mots contenant des caractères répétés (plus de trois caractères consécutifs identiques) ;
2. Ponctuation : présence ou absence d'un point d'exclamation ou d'un point d'interrogation ;
3. Capitalisation : nombre de mots avec tous les caractères en majuscules ;
4. Smileys : présence ou absence de smileys positifs et négatifs ;
5. Hashtags : nombre de hashtags ;
6. Négation : nombre de termes de négation ;
7. Étiquette morphosyntaxique : présence ou absence de chaque partie de la balise vocale.

Le système va donc évaluer toute les combinaisons possibles de ces attributs et ne garder que celle qui donne le meilleur résultat.

2.5 Étape 5 : Sélection d'attributs

Afin de sélectionner les attributs les plus discriminants pour chaque tâche, une étape de sélection d'attributs a été effectuée pour mesurer le gain d'information de chaque attribut par rapport à la classe (Mitchell, 1997). Après avoir calculé le gain d'information pour chaque attribut, nous ne gardons que ceux pour lesquels ce gain est supérieur à 0. Cependant, il n'est pas garanti que la sélection d'attributs améliore les résultats. De ce fait, le système *FrenchSentiClass* évalue l'apport de la sélection d'attributs. Par conséquent, cette étape n'est retenue que si elle améliore les résultats en validation croisée.

2.6 Étape 6 : Tuning du paramètre de complexité

Comme algorithme de classification, nous avons choisi d'utiliser les SVM (Support Vector Machine) avec la méthode SMO (Sequential Minimal Optimization) (Platt, 1999) implémentée 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 des sentiments et d'émotions. Il est robuste sur les grands espaces de caractéristiques. En effet, notre modèle de classification exploite une variété de caractéristiques syntaxiques, sémantiques et lexicales.

La valeur donnée au paramètre de complexité du SVM peut altérer l'efficacité du modèle appris (Mohammad *et al.*, 2013). Afin de choisir le meilleur paramètre de complexité C , le système *FrenchSentiClass* effectue une recherche exhaustive de la meilleure valeur entre 0 et 1 avec un pas de 0.05. Pour chaque valeur, 10 validation croisées différentes à 10 plis chacune sont générées et la moyenne est considérée pour trouver la valeur qui donne le meilleur résultat.

3 Expérimentations

Dans cette section, nous présentons les données utilisées, les tâches considérées, les configurations choisies pour chaque tâche, les *runs* soumis et les résultats obtenus.

3.1 Description des tâches et des données

Les différentes tâches de DEFT17 sont les suivantes :

- Tâche 1 : déterminer la polarité (*positif, négatif, neutre* ou *mixte*) d'un tweet non figuratif ;
- Tâche 2 : identifier le langage figuratif (*figuratif* ou *non figuratif*) ;
- Tâche 3 : déterminer la polarité (*positif, négatif, neutre* ou *mixte*) d'un tweet figuratif et non figuratif.

Les tables 1, 2 et 3 présentent la distribution des données par classe sur chaque tâche. Pour plus de détail sur les tâches et les données proposés dans ce défi, veuillez vous référer au papier de (Farah *et al.*, 2017).

Tâche 1				
Classes	Apprentissage		Test	
	#	%	#	%
Positif	494	13	123	12,6
Négatif	1268	32	318	32,6
Objectif	1643	42	411	42,1
Mixte	501	13	124	12,7
Total	3906	100	976	100

TABLE 1: Distribution des classes pour la tâche 1 sur le corpus d'apprentissage et de test.

Tâche 2				
Classes	Apprentissage		Test	
	#	%	#	%
Figuratif	1947	33	488	33
Non figuratif	3906	67	976	67
Total	5853	100	1464	100

TABLE 2: Distribution des classes pour la tâche 2 sur le corpus d'apprentissage et de test.

Tâche 3				
Classes	Apprentissage		Test	
	#	%	#	%
Positif	504	10	125	10
Négatif	2263	44	568	44
Objectif	1718	34	430	34
Mixte	633	12	158	12
Total	5118	100	1281	100

TABLE 3: Distribution des classes pour la tâche 3 sur le corpus d'apprentissage et de test.

Pour évaluer nos systèmes, nous avons utilisés la micro f-mesure et la macro f-mesure. La mesure officielle utilisée dans le défi est la macro f-mesure.

3.2 Configurations sélectionnées

Dans la table 4, nous présentons les caractéristiques et les paramètres sélectionnés par validation croisée sur les données d'entraînement de chaque tâche. Nous utilisons trois configurations :

- Configuration 1 : utiliser la micro f-mesure pour toutes les étapes ;
- Configuration 2 : utiliser la micro f-mesure les 5 premières étapes et la macro f-mesure pour l'étape 6 ;
- Configuration 3 : utiliser la macro f-mesure pour toutes les étapes.

		Tâche 1			Tâche 2			Tâche 3		
		C1	C2	C3	C1	C2	C3	C1	C2	C3
Étape 1	Unigrammes	✓	✓	✓	✓	✓	✓	✓	✓	✓
	Bigrammes	✓	✓	✓				✓	✓	✓
	Trigrammes	✓	✓					✓	✓	
Étape 2	Liens					✓	✓			
	Emails	✓	✓		✓	✓	✓	✓	✓	
	Pseudonymes	✓	✓							
	Argot	✓	✓			✓	✓			✓
	Lemmatisation	✓	✓					✓	✓	✓
	Miniscule			✓	✓		✓			
	Mots vides									
Étape 3	FEEL-pol			✓	✓	✓	✓	✓	✓	
	FEEL-emo									✓
	Affects-pol	✓	✓	✓				✓	✓	✓
	Affects-emo	✓	✓	✓				✓	✓	✓
	Diko-pol			✓	✓	✓	✓	✓	✓	✓
	Diko-emo	✓	✓	✓						
	Polarimots			✓			✓	✓	✓	✓
	Incongruité		✓		✓	✓		✓	✓	✓
Étape 4	Capitalisation			✓						
	Mots allongés			✓	✓	✓				
	Hashtags	✓	✓		✓	✓	✓			✓
	Termes négatifs				✓	✓	✓			✓
	Ponctuation			✓	✓	✓	✓	✓	✓	
	POS tags			✓			✓			✓
	Smileys							✓	✓	
Étape 5	Sélection d'attributs									
Étape 6	SVM Tuning	0,05	0,4	0,4	0,1	0,1	0,15	0,1	0,4	0,25

TABLE 4: Les caractéristiques et paramètres sélectionnés par validation croisée pour chaque configuration sur les données d'entraînement de chaque tâche (C1, C2 et C3 sont des abréviations respectives des configurations 1, 2 et 3).

3.3 Résultats

Pour chacune des tâches, nous avons soumis les *runs* suivants :

- *Run 1* : le modèle appris avec la configuration 1 ;
- *Run 2* : le modèle appris avec la configuration 2 ;
- *Run 3* : le vote entre le modèle appris avec la configuration 1, celui appris avec la configuration 2 et celui appris avec la configuration 3.

	Tâche 1		Tâche 2		Tâche 3	
	F_{mi}	F_{ma}	F_{mi}	F_{ma}	F_{mi}	F_{ma}
<i>Run 1</i>	64	55.5	78.8	75	63.4	53.1
<i>Run 2</i>	62.8	53.9	78.8	74.9	62	51.5
<i>Run 3</i>	65.3	55.7	79	75	63.9	52.2

TABLE 5: Résultats des différents *runs* soumis au défi pour chaque tâches (F_{mi} : micro f-mesure, F_{ma} : macro f-mesure)

Sur la tâche 1, le *run 3* (le vote) fournit les meilleurs résultats avec une macro f-mesure de 55,7% et une micro f-mesure de 65,3%. La médiane de la macro f-mesure de l'ensemble des systèmes soumis est 52,2%. Cependant, le meilleur système du défis sur cette tâche a obtenu une macro f-mesure de 64,9%.

Sur la tâche 2, c'est toujours le *run 3* qui fournit les meilleurs résultats avec une macro f-mesure de 75% et une micro f-mesure de 79%. On constate également que la macro f-mesure du *run 3* est égale à celle du *run 1*. La médiane de la macro f-mesure de l'ensemble des systèmes soumis est 52,3%. Cependant, le meilleur système du défis sur cette tâche a obtenu une macro f-mesure de 78,3%. Sur cette tâche, les résultats de notre système *FrenchSentiClass* sont proches du meilleur résultat en considérant la macro f-mesure.

Sur la tâche 3, le *run 1* fournit la meilleure macro f-mesure (53,1%), et le *run 2* fournit la meilleure micro f-mesure (63,9%). La médiane de la macro f-mesure de l'ensemble des systèmes soumis est 51,9%. Cependant, le meilleur système du défis sur cette tâche a obtenu une valeur de 59,3%.

Globalement, ces résultats montrent que le système *FrenchSentiClass* permet de donner des résultats compétitifs tout en réduisant considérablement le temps passé dans l'exploration des données et le choix de la meilleure représentation de descripteurs et paramètres.

4 Conclusion

Pour notre participation aux tâches 1, 2 et 3 du défi de fouille de texte DEFT 2017, nous avons utilisé notre système *FrenchSentiClass* qui est basé sur un classifieur SVM et qui exploite diverses caractéristiques lexicales et syntaxiques. Dans ce travail, nous avons mis en avant l'automatisation de notre système. En effet, il peut choisir automatiquement les meilleurs prétraitements, descripteurs et ressources sur un corpus donné. Le principal avantage du système est qu'il peut réduire considérablement le temps dédiée à l'exploration des données et aux choix des descripteurs. Les résultats obtenus sur chaque tâche sont compétitifs, car ils dépassent toujours la médiane des macro f-mesure de l'ensemble des systèmes soumis pour ce défi. Une de ces limites est liée aux caractéristiques

et descripteurs utilisées. En effet, très peu de caractéristiques sont liés à la détection d'ironie et au sarcasme, lesquelles sont l'objet de ce défi.

Références

- ABDAOUI A., AZÉ J., BRINGAY S. & PONCELET P. (2016). Feel : a french expanded emotion lexicon. *Language Resources and Evaluation*, p. 1–23.
- 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.
- FARAH B., CYRIL G., JIHEN K., VÉRONIQUE M. & ISABELLE R. (2017). Analyse d'opinion et langage figuratif dans des tweets : présentation et résultats du défi fouille de textes deft2017. In *Actes de l'atelier DEFT de la conférence TALN 2017.*
- GALA N. & BRUN C. (2012). Propagation de polarités dans des familles de mots : impact de la morphologie dans la construction d'un lexique pour l'analyse d'opinions. In *Actes de Traitement Automatique des Langues Naturelles, Grenoble*, p. 495–502.
- HADDI E., LIU X. & SHI Y. (2013). The Role of Text Pre-processing in Sentiment Analysis. *Procedia Computer Science*, **17**, 26–32.
- 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.
- KAROUI J., ZITOUNE F. B., MORICEAU V., AUSSENAC-GILLES N. & BELGUITH L. H. (2015). Détection automatique de l'ironie dans les tweets en français. In *22eme Conference sur le Traitement Automatique des Langues Naturelles (TALN 2015)*, p. 1–6.
- KIRITCHENKO S., MOHAMMAD S. M. & SALAMEH M. (2016). Semeval-2016 task 7 : Determining sentiment intensity of english and arabic phrases. In *Proceedings of the International Workshop on Semantic Evaluation (SemEval), San Diego, California, June.*
- LAFOURCADE M., JOUBERT A. & LE BRUN N. (2015). *Games with a Purpose (GWAPS)*. John Wiley & Sons.
- LIU B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, **2**, 627–666.
- LIU B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, **5**(1), 1–167.
- LIU B. (2015). *Sentiment analysis : Mining opinions, sentiments, and emotions*. Cambridge University Press.
- MITCHELL T. M. (1997). *Machine learning.*, volume 45. Burr Ridge, IL : McGraw Hill.
- MOHAMMAD S. M. & KIRITCHENKO S. (2015). Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, **31**(2), 301–326.
- 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 7th International Workshop on Semantic Evaluation (SemEval 2013)*, p. 321–327.
- 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. (2005). Seeing stars : Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd annual meeting on association for computational linguistics*, p. 115–124 : Association for Computational Linguistics.
- PANG B., LEE L. & VAITHYANATHAN S. (2002). Thumbs up ? : sentiment classification using machine learning techniques. In *Proceedings of the ACL conference on Empirical methods in natural language processing*, p. 79–86 : Association for Computational Linguistics.
- 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.
- PONTIKI M., GALANIS D., PAVLOPOULOS J., PAPAGEORGIU H., ANDROUTSOPOULOS I. & MANANDHAR S. (2014). Semeval-2014 task 4 : Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, p. 27–35 : Citeseer.
- RILOFF E., WIEBE J. & PHILLIPS W. (2005). Exploiting subjectivity classification to improve information extraction. In *Proceedings of the national conference of the american association for artificial intelligence*, volume 20, p. 1106–1111 : Menlo Park, CA ; Cambridge, MA ; London ; AAAI Press ; MIT Press ; 1999.
- 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.
- TURNER P. D. (2002). Thumbs up or thumbs down ? : semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, p. 417–424.
- WILSON T., WIEBE J. & HOFFMANN P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, p. 347–354 : Association for Computational Linguistics.