Feature engineering for tweet polarity classification in the 2015 DEFT challenge

François Morlane-Hondère Eva D'hondt
LIMSI, CNRS, Rue John Von Neumann, 91400 Orsay, France
francois.morlane-hondere@limsi.fr, eva.dhondt@limsi.fr

Résumé. Dans cet article, nous présentons notre participation à la tâche 1 du Défi Fouille de Textes (DEFT) 2015. Cette dernière consiste à identifier la polarité de tweets en français. Notre système de classification s'appuie sur des traits de nature variée tels la présence des mots du tweet dans les lexiques, leurs propriétés typographiques, la façon dont sont utilisés les éléments de la syntaxe de Twitter (hashtags, mentions) ou encore le fait qu'un tweet ait été généré automatiquement ou produit par un humain. Nos deux soumissions ont respectivment obtenu une macro-précision de 0.687 and 0.688. Elles se situent au-dessus de la moyenne de l'ensemble des participants (0.582) mais légèrement en dessous de la médiane (0.693).

Abstract.

Feature engineering for tweet polarity classification in the 2015 DEFT challenge
In this paper we present our contribution to the first task of the 2015 DEFT challenge which dealt with polarity classification of French tweets. We explored the impact of a large number of different types of features, such as lexicon-based features, typography-based features, Twitter-specific features and features that incorporate external (world) knowledge. We submitted two runs and achieved macro-averaged precision scores of 0.687 and 0.688 respectively, which is above the average of all submitted runs (0.582) and slightly below the median (0.693).

Mots-clés : détection de polarité, analyse de sentiments, DEFT, Twitter, réseaux sociaux.

Keywords: polarity classification, sentiment analysis, DEFT, Twitter, social media.

1 Introduction

Over the last eight years microblogging sites such as Twitter have become a powerful and inlfuential means of communication on a global scale. As Twitter is increasingly regarded as the digital voice of public opinion, there is a high demand for automated tools that can analyze the content of tweets for the purposes of sentiment classification or knowledge extraction. This is a challenging task, however : Twitter's 140-character limit forces the user to express their message in a terse, compact manner. Moreover, the language use in tweets is very informal, with creative spelling and punctuation, emoticons, typos, slang, new words, abbreviations, and the inclusion of URLs and #hashtags. Twitter language changes fairly quickly as well : hashtags used to be added to the end of a message as a sort of label independent of the main 'tweet text', but are now increasingly used within sentences, e.g. *On passe à la #chasse aux #loups sans le dire*. Over the last few years, twitter (language) analysis has gathered a lot of attention from the Natural Language Processing and Machine Learning research communities (Kong *et al.*, 2014), but the vast majority of existing resources and systems are limited to English. In the 2015 DEFT challenge, three different twitter analysis tasks were proposed for a given corpus of French tweets on the theme of ecology and climate change. This article presents our contribution to the first task : *polarity classification of French tweets* into three different categories (positive, negative and neutral). We explore the impact of a large number of different types of features, such as lexicon-based features, typography-based features, Twitter-specific features and features based on world knowledge.

2 DEFT 2015 task and corpus

The first task of the 2015 DEFT challenge constituted *polarity classification of French tweets*. In the task description, *polarity* was defined as expressing either an opinion, sentiment or emotion [1]. A given tweet had to be assigned one label, either *positive*, *negative*, *neutral*, and there was no *mixed* or other residual category. Please note that the polarity was assigned to the tweet as a whole, i.e. on message level (cf. task B of the SemEval 2013 challenge [2]).

The DEFT 2015 corpus consisted of about 15,000 tweets on the topic of climat change, which were collected in the context of the uComp project [3]. The training and test set were released separately and contained 7,929 and 3,383 tweet IDs, respectively. We used the downloader program made available by the track organizers to download the individual tweets. For some IDs, however, the corresponding tweet was no longer available. As a result we trained and evaluated our system on a training and test set of 7,830 and 3,381 tweets respectively.

Please note that the features described in section 3.2 are only generated over the actual text of the tweet and that we did not use any twitter meta-data such as timestamps, username, ... as additional features.

While the tweet corpus was clean (i.e. no formatting errors), the annotations in the training set were rather inconsistent at times. For example, the following tweet text appeared 9 times in the training corpus (each time retweeted to another user but containing the exact same message) but it was classified five times as positive, three times as negative, and one time as having a neutral polarity.

@lalibrebe 10° à Werchter. 35° à Werchter. Le réchauffement climatique accélère. `https://t.co/x1JS7Tox3C` *Merci de RT.*

3 System description

3.1 Classifier

Like Pak & Paroubek (2010), we build a polarity classifier using the multinomial Naïve Bayes classification algorithm (as implemented in the Weka 3.6 toolkit). This algorithm yielded similar performance to SVM models but was much faster.

3.2 Features

3.2.1 Unigrams

Word n-grams are sequences of n words extracted from a given text. Baseline classification systems generally use n-gram features, as they generally yield good performance and are computationally cheap to compute. A classic approach consists in combining them with other features to achieve higher accuracy.

In our system, we used (the presence of) the 450 most discriminating unigrams – 1-grams – in the training corpus, as calculated by the InfoGainAttributeEval function implemented in Weka 3.6.12 (Hall *et al.*, 2009), as the basis classification features. As its name suggests, this function computes the information gain of each feature in respect to the polarity of the tweets. Following Pang *et al.* (2002), we defined these features as binary, that is, the feature capture the presence of a unigram term in a given tweet, irrespective of its frequency.

3.2.2 Lexicons

A traditional approach in polarity (or sentiment) classification is that of dictionary look-up methods using lists of *positive* and *negative* words, usually nouns, verbs and adjectives. Such lexicons can be used in many ways. We chose to compile several lexicons and generate features that denote the presence of a term in each lexicon as a binary feature.

1. https://deft.limsi.fr/2015/guideAnnotation.fr.php
2. http://www.cs.york.ac.uk/semeval-2013/task2/
3. http://www.ucomp.eu/

For this task we chose to work with relatively small lexicons so that we would be able to manually check their content. By doing so, we could identify and remove polysemous words, thus eliminating potential sources of noise (see examples below). Table 1 indicates the number of words in each lexicon. Overall, we have a higher coverage of negative words.

Lexicon name		# of positive words	# of negative words
Polarimots		54	363
Dictionnaire électronique des Synonymes	nouns	242	355
	verbs	210	384
	adjectives	377	489
Swear words		-	102
Complementary		10	17
total		893	1,710

TABLE 1 – Number of words included in each lexicon.

Polarimots Polarimots[4] is a lexicon containing 7,483 French nouns, verbs, adjectives and adverbs whose polarity – positive, negative or neutral – has been semi-automatically annotated (Gala & Brun, 2012). There are three degrees of annotation confidence. We built a positive and a negative lexicon from the positive and negative words whose annotation confidence is the highest, i.e. when all the annotators agreed (Gala & Brun (2012) showed that including annotations that have a lower agreement score slightly degraded the performance of a polarity classification system).

Dictionnaire électronique des Synonymes A second series of six lexicons was built using the *Dictionnaire électronique des Synonymes*[5] – DES –, a French thesaurus containing 203,311 synonyms (Manguin *et al.*, 2004). We manually built sets of ten seed words for nouns, adjectives and verbs, in the positive and negative polarity, respectively. Two of the six resulting seed sets can be seen below :

— Positive adjectives : *beau*, *gentil*, *intelligent*, *utile*, *agréable*, *sympathique*, *honnête*, *prudent*, *propre*, *bon*
— Negative nouns : *scandale*, *désastre*, *violence*, *mensonge*, *douleur*, *agression*, *tristesse*, *peur*, *haine*, *mort*

Then, assuming that the polarity of a word is propagated through its synonyms (Rao & Ravichandran, 2009), we extracted all the synonyms of the seeds. Below are some of the synonyms of the seed words listed above :

— Synonyms of positive adjectives : *consciencieux*, *euphorique*, *digne*, *peinard*, *humanitaire*, *moral*, *reposant*, ...
— Synonyms of negative nouns : *colère*, *terreur*, *mystification*, *agitation*, *lâcheté*, *rancune*, *inquiétude*, ...

The automatic expansion was followed by a manual filtering of the extracted synonyms in which we removed polysemous terms, like *salade* ('salad'), which can be used as a synonym of *mensonge* ('lie').

Although we took care to choose relatively monosemous seeds, some synonyms tend to have multiple meanings. Thus, the extraction of second degree synonyms – synonyms of the synonyms of the seeds – was found to be too noisy and was therefore abandoned.

One of the limitations of the use of the DES is that it was built from traditional dictionaries and thesauri written between 1864 and 1992 (François *et al.*, 2003). Thus, the DES reveals some discrepancies with today's French usage – especially on Twitter – that we had to rectify. For example, the word *bath* – a synonym of *beau* ('beautiful') – is not used in French since the 1970's.

Swear words and insults A list of swear words and insults was compiled from the Web[6] with the assumption that these words tend to be associated uniquely with a negative mood.

Interestingly, polysemy is also a problem here. For example, the words *fumier* ('manure') and *ordure* ('garbage') can both be figuratively used to refer to a despicable person. But in the corpus, they occur in their proper meaning, in positive or neutral tweets :

4. http://polarimots.lif.univ-mrs.fr/
5. http://www.crisco.unicaen.fr/des/
6. http://francais-oral.wikispaces.com/Lexique+des+insultes

— *Quand le fumier de cheval sert à se chauffer, un beau projet de #methanisation à @Caenlamer* `http://t.co/rRZiKQLePQ` (+)
— *@DrDree_ non non je parlais en fait de l'ecologie. Puis je suis arrive aux problemes de la gestion des ordures* (=)

We manually went through the list to discard such words.

Complementary lexicon These are two small lists of additional positive and negative words we found in the training corpus and that were not already included in other lexicons.

3.2.3 Handling negation

The assumption that positive words occur in positive tweets and negative words in negative ones is not as straightforward as it may sound : Many contextual phenomenons or stylistic factors can affect the meaning – and, thus, the polarity – of words. Benamara *et al.* (2012) showed that different types of negation can affect polarity in different ways.

We handled this highly complex problem with a simple – simplistic – polarity shifting system consisting in regular expressions checking for the following words : *pas* ('not'), *aucun* ('none'), *jamais* ('never'), *non* ('no'), *peu* ('few'), *ni* ('nor') and *rien* ('nothing'), in a window of two words before and after occurrences in tweets of words found in our lexicons. Like the previous features, the polarity shifter feature is binary : If a polarity shifter is found in the context of a word included in a positive (resp. negative) lexicon, then the value of its negative (resp. positive) counterpart is set to 1.

3.2.4 Term extraction on the training corpus

Using the Alchemy [7] keyword and entity extraction software, we processed the positive, negative and neutral subsets of the training corpus to obtain the most discriminating (multiword) features for each subset. Alchemy uses deep learning to find dependencies between words over large corpora. We used the extracted words and phrases as additional weighted binary features in our second submitted run : While the presence (resp. absence) of an extracted term in a given tweet resulted in a 1- (resp. 0-) value, the feature weight was a normalized version of the extraction score that Alchemy returned for that term. This way the presence of an extracted term for which Alchemy had a low confidence score had less impact on the classification process than that of a term with a high confidence score.

3.2.5 Twitter-specific features

Following Arakawa *et al.* (2014), we call *Twitter-specific features* the commands and conventions used by Twitter users in their posts. Two recurring commands are the hashtag (#), used to turn words they are added to into clickable tags, and the *at* sign (@), used to mention or to reply to another user. We used the presence of hashtags and mentions and their location in tweets (i.e. in the beginning of the tweet or in the end end) as binary features. The number of hashtags or mentions was not found to be relevant, as well as the presence of the mention *RT* (*retweet*), which is used to share a tweet with a users followers.

3.2.6 Extracting information from tinyURLs

Presence of tinyURL We observed that the majority (5849 out of 7830) of tweets in the training set contain at least one tinyURL. While not a very strong feature, a tweet without a tinyURL has a relative higher probability of belonging to the positive or negative category than the neutral one. Taking the presence of tinyURLs into account lead to small but significant improvements.

Generation history of tweet For a secondary feature based on the tinyURLs, we explored how the actual text in the tweet is generated. We found that for a substantial number of tweets in the training set, the tweet text was either the title or introductory sentence of the online article or post it referred to. These tweets are the result of (semi-)automatic sharing of online content with minimal human interaction. We hypothesized that such tweets are more likely to belong to the neutral

7. `http://www.alchemyapi.com/`

category, and that tweets which express a positive or negative opinion on a subject would contain more information written by the user (either in the form of adding hashtag to specify the information, or by an accompanying sentence that comments on the content of the article or post). We therefore added a feature that categorized a tweet as either "human" (written by a human and containing novel information), "automatic" (the result of automatic sharing of existing content with minimal manual editing) or "unknown" (a surplus category of tweets for which we lacked information to determine the level of human interaction). The features were created as follows : For each tweet, we extracted the tinyURL (if present) and downloaded the title and introductory sentence(s) from the corresponding webpage. If the text in the tweet matched with (part of) the title and introductory sentences, the tweet was categorized as "automatic". If not the case, for example, because extra information was added in the form of extra hashtags, or the tweet text was an own comment or summary of the referenced article, the tweet was categorized as "human" [8]. Please note that we used fuzzy matching as implemented in the `FuzzyWuzzy` Python package [9] to account for small edits in the original texts. For example in the following tweet

"On passe à la #chasse aux #loups sans le dire" via Pescalune `http://t.co/KiD1N1Oq7v`,

certain words in the tweet have been converted into hashtags by the user while the phrase still corresponds to the title of the referenced article. By allowing fuzzy matching with a moderately high threshold (>70%) we can still identify these "reposted" tweets. For those tweets for which we were not able to extract information on the article or post (either because of time-out errors, or difficulties in parsing the returned html), the label "unknown" was given. This category is fairly small : 525 out of 7830 tweets.

Content classification of the referenced webpages We also experimented with a third feature which was based on the content of the referenced webpage. We manually categorized a subset of the URLs in the training set into the following 6 categories : *combo* (websites such as `www.scoop.it` where users can share and publish content from other sources), *ecoBlog* (websites dedicated to ecology), *news* (news sites), *polBlog* (political blogs and websites from political parties), *science* (webpages from universities or research facilities), *other*. For each website we extracted the domain name as well as the website description and keywords from its main page. This data was then used to train a separate classifier that would classify an unseen url (and extracted information from the associated website) into the relevant category. The lack of coherence in website description and overall quality of the meta description of the websites lead to a very sparsely trained and unreliable classifier. We therefore opted not to use this feature in the submitted runs.

3.2.7 Smileys

We observed three kinds of smileys in the training corpus :
- typographic smileys. They are *compositional* smileys built with letters, numbers and ponctuation marks used to mimic eyes, noses and mouths. We found two different types of typographic smileys :
  - Western-style smileys : `:-)` `:D` `:p`
  - Japanese-style smileys (or *kaomojis*) : `O_o` `O___O` `-_-'`
- graphic smileys. They are Unicode characters : 😁😈😱

We built two separate sets of regular expressions to check for the presence of positive and negative typographic smileys in the tweet text. Likewise, two lists of graphic smileys were built using web sources [10]. For a given tweet, the value of the features containsPosSmiley and containsNegSmiley is 1 if one or several positive – resp. negative – smiley(s) is (are) found in the message.

Please note that we disregarded *neutral* smileys. Smileys are by their very nature means of expressing emotion, so the existence – and actual usage – of neutral ones seems unlikely : This assumption was confirmed by an analysis of the corpus as we did not find any occurrence of what might be considered a neutral smiley, i.e. `:-|`, in either the train and test corpora.

8. If the tweet did not contain an tinyURL, i.e. was written from scratch, is was classified as "human" as well.
9. `https://github.com/seatgeek/fuzzywuzzy`
10. `http://unicode-table.com/fr/search/?q=emoticons`

3.2.8 Punctuation marks

Like smileys, punctuation is a common means of expressing emotion or intention in textual content. For the task, we considered five types of punctuation marks : exclamation marks, question marks, ellipsis, comma and quotation marks. The presence (or absence) of each punctuation mark resulted in a binary feature. Although exclamation marks are somewhat ambiguous – they can be associated to both joy and anger –, they can be useful to discriminate between positive/negative tweets and neutral ones. Question marks are relevant for polarity classification in that they can be used in rhetorical questions. Such questions often carry a negative polarity, as in the two examples below :

— *Bravo @RoyalSegolene ;Encore merdé, encore cédé :-(Après tout, l'écologie, c'est un truc de bobo, n'est-ce pas ?* `http://t.co/7hqIzMjYKf` (-)
— *Comment imaginer une pareille chose ? ? ? ? ? ? ? ? La SQ démantèle un réseau de voleurs d'huile de friture* `http://t.co/7IOQVrCO9o` (-)

Ellipsis is also interesting in that it can denote sarcasm :

— *L'écologie, cette valeur de gauche... – Ecotaxe : la carte des projets locaux menacés* `http://t.co/HBARJ3AyyT` *via @lemondefr* (-)
— *Abandonner l'écotaxe le jeudi et recevoir Schwarzenegger le lendemain pour parler de lutte contre le réchauffement climatique... Logique.* (+) [11]

The detection of the presence of quotation marks is more stringent than that of question or exclamation marks. This binary feature is only set to 1 if one or two – consecutive – words are quoted. By adding this constraint, we wanted to focus on sarcastic usage of quotations marks :

— *Chaleur et électricité "propre" ?,quelle idée saugrenue,la géothermie .....*`http://t.co/cLpnsFLqLX` (-)
— *On se fait maintenant expliquer pourquoi on se fait ROULER pour "notre bien" avec les #éoliennes #HydroQuébec à #rdieconomie #RadioCanada* (-)

3.2.9 Miscellaneous features

Interrogative markers This binary feature indicates the presence of an interrogative marker out of a manually compiled list, i.e. *quel* ('which'), *quoi* ('what'), *comment* ('how'), *combien* ('how much'), *pourquoi* ('why'), in the tweet text. The aim of this feature is to improve the identification of rhetoric questions (cf. 3.2.8).

Case This feature indicates the presence of a series of 50 characters in uppercase :

— *REFUS DE S'ATTAQUER AUX CAUSES DU PROBLÈME, INTRINSÈQUES AU CAPITALISME..* `http://t.co/NVfSeu9TL9` (-)

We assume that this is a mark of emotion and that it will not be found in neutral tweets.

Repetition This feature is set to 1 if the tweet includes a sequence of 3 identical characters :

— *@Lorenzo75019 Ah ouiiiiii c'est vrai mdddddrr c'est développement durable qui me fait rire :) mais bon je me moque pas* (-)

As for the Case feature, we assume that repetitions are emotional markers.

Separators These three binary features indicate the presence of a vertical bar (|), a square (■) or a right-pointing triangle (▶) in the tweet. These characters are exclusively used as separators in automatically-generated tweets. However, they were not found discriminating and, therefore, removed from the final set of features.

11. Although being positively annotated, the polarity of this tweet is definitely negative.

4 Submitted runs

We submitted two runs to the official evaluation. An overview of the features used in each run can be found in Table 2.

Feature Group	Feature Name	Run 1	Run 2
Unigrams	-	✓	✓
Lexical Features	inLexiquePosPolarimots	✓	✓
	inLexiqueNegPolarimots	✓	✓
	inLexiqueNegDESadj	✓	✓
	inLexiqueNegDESnom	✓	✓
	inLexiqueNegDESver	✓	✓
	inLexiquePosDESadj	✓	✓
	inLexiquePosDESnom	✓	✓
	inLexiquePosDESver	✓	✓
	inLexique d'injures	✓	✓
	InLexiquePosManuel	✓	✓
	inLexiqueNegManuel	✓	✓
	inLexiquePosSite	✓	✓
Negation	-	✓	✓
TermExtraction	inTermsPosTrainingSet		✓
	inTermsNegTrainingSet		✓
	inTermsNeutralTrainingSet		✓
Twitter-specific Features	@inTweet	✓	✓
	@inBeginTweet	✓	✓
	@atEndTweet	✓	✓
	#inTweet	✓	✓
	#inBeginTweet	✓	✓
	#atEndTweet	✓	✓
	containsRT		
	numberOf@		
	numberOf#		
tinyURL	containsTinyURL	✓	✓
	writtenByHuman	✓	✓
	catOfTinyURL		
Smileys	containsPosSmiley	✓	✓
	containsNegSmiley	✓	✓
Punctuation	containsExcl	✓	✓
	containsMultiExcl		
	containsQuestionMark	✓	✓
	containsQuotation	✓	✓
	containsElipsis	✓	✓
	containsSemiColon	✓	✓
Miscellaneous Features	containsInterrogativeMarker	✓	✓
	containsUpperCase	✓	✓
	containsRepitition (>3)	✓	✓
	containsSeparators		

TABLE 2 – Overview of features used in two official submissions

5 Results

The classification scores of the two submitted runs can be found in Table 3. We find that adding the terms extracted by the third-party Alchemy software has positive effect on classification, particularly in identifying negative tweets, which is nevertheless so small to be insignificant. Compared to the other submitted runs our system performed slightly below average and is bulked with the main group of participants. The three top-scoring systems achieved macro-averaged scores of near 73%.

Precision	Run1	Run2	average (of all submissions for 12 groups)	median (of all submissions for 12 groups)
micro	0.676	0.672	-	-
macro	0.687	0.688	0.582	0.693

TABLE 3 – Results for submitted runs (expressed in micro- and macro-averaged precision

6 Discussion

We performed a subtractive analysis to investigate the (relative) influence of each set of features used in Run1. Table 4 shows the result of the removal of each feature set from the entire set of features. We see that the removal of the unigrams has the biggest influence on the general macro-precision. This is not surprising as the unigrams are by far the biggest set of features – 450 – and that they have been selected according to their discriminative potential. Some of the most discriminative words according to Weka's InfoGainAttributeEval function are *contre* ('against'), *menacée* ('endangered'), *espèce* ('species'), *solaire* ('solar'), *fromage* ('cheese') and *banque* ('bank'). Although *contre* ('against') and *menacée* ('endangered') – in the phrase *espèce menacée* ('endangered species') – seem to be negatively-valenced words, the presence of a word like *fromage* ('cheese') is quite surprising. This is actually due to the fact that there are, in the corpus, more than 35 retweets of a website article entitled *Le fromage, une espèce menacée ?*. These tweets being negatively annotated, the presence of the word *fromage* in a tweet has been identified as a good indicator for negative polarity. Thus, the repetition of tweets in the corpus is a bias that may lead to overfitting : Performing a simple information gain computation – as we did – does not seem to be a robust strategy.

The lexicon-based features are the second most influential features. Although most of the words in these lexicons have been chosen according to their polarity regardless of the corpus theme, and despite the many contextual phenomena – like irony – which can shift a words polarity, the simple assumption that positive and negative words are used in positive and negative tweets seems to hold.

The twitter features and the use of tinyURLs have a smaller influence, but a positive one. On the other hand, we can see that the last three features have a slightly negative influence. The fact that the presence of smileys is not discriminative is quite surprising in that they are explicit indicators of the writer's mood. We interpret the negative influence of the last two features as being due to the ambiguity of the punctuation marks, repetitions and case shifting : The hypothesis that these properties would help to discriminate positive/negative tweets and neutral ones does not seem to hold. We found that removing the last three sets of features, i.e. smileys, punctuation marks and miscellaneous features, results in a macro-averaged precision score of 0.690% over the test set.

feature	macro-precision	difference with the entire set
all features	0.687	-
- unigrams	0.596	-0.091
- lexicons	0.661	-0.026
- twitter feats.	0.682	-0.005
- tinyurl	0.679	-0.008
- smileys	0.688	+0.001
- punctuation	0.689	+0.002
- misc. feats.	0.691	+0.004

TABLE 4 – Subtractive analysis of the features used in Run1.

Analysis of the results on the test set for the highest-scoring run (Run2) shows that the main error of our classifier is overgeneration of neutral labels, which is not surprising as this category constituted the majority of the training corpus, and is therefore the best trained classifier. Of all three classifiers the positive classifier has the worst performance, particularly in distinguishing between the neutral and positive labels. We remarked a similar trend when evaluating on the training corpus with cross-validation.

Reference / Run2	=	-	+
=	597	121	196
-	36	204	37
+	67	51	274

TABLE 5 – Confusion Matrix of submitted results in Run2

7 Conclusion

This paper describes our participation to the tweet polarity classification task that was organized as part of the DEFT 2015 competition. In our approach we explored a variety of features, ranging from traditional dictionary look-up methods to twitter-specific features such as the presence and location of hashtags, as well as some features that were based on more external knowledge such as the source of the tinyURL in the tweet. We found that the traditional features such as unigrams and presence in lexicon had the most impact. Interestingly, the features that focused on Twitter-specific characteristics and on micro-blogging language (smileys, repetition of characters, ...) had little to no impact. Our systems achieved scores of 0.687 and 0.688 macro-averaged accuracy.

Références

ARAKAWA Y., KAMEDA A., AIZAWA A. & SUZUKI T. (2014). Adding twitter-specific features to stylistic features for classifying tweets by user type and number of retweets. *Journal of the Association for Information Science and Technology*, **65**(7), 1416–1423.

BENAMARA F., CHARDON B., MATHIEU Y. Y., POPESCU V. & ASHER N. (2012). How do negation and modality impact on opinions ? Jeju Island, Korea.

FRANÇOIS J., MANGUIN J. L. & VICTORRI B. (2003). La réduction de la polysémie adjectivale en contexte nominal : une méthode de sémantique calculatoire. In *Cahiers du CRISCO*, volume 14. Université de Caen : CRISCO.

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 de sentiments (spreading polarities among word families : Impact of morphology on building a lexicon for sentiment analysis) [in french]. In *Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 2 : TALN*, p. 495–502, Grenoble, France : ATALA/AFCP.

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

KONG L., SCHNEIDER N., SWAYAMDIPTA S., BHATIA A., DYER C. & SMITH N. A. (2014). A dependency parser for tweets. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, to appear*.

MANGUIN J. L., FRANÇOIS J., EUFE R., FESENMEIER L., OZOUF C. & SÉNÉCHAL M. (2004). Le dictionnaire électronique des synonymes du CRISCO : un mode d'emploi à trois niveaux. In *Cahiers du CRISCO*, volume 34. Université de Caen : CRISCO.

PAK A. & PAROUBEK P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In N. C. C. CHAIR), K. CHOUKRI, B. MAEGAARD, J. MARIANI, J. ODIJK, S. PIPERIDIS, M. ROSNER & D. TAPIAS, Eds., *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta : European Language Resources Association (ELRA).

PANG B., LEE L. & VAITHYANATHAN S. (2002). Thumbs up ? : sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, p. 79–86 : Association for Computational Linguistics.

RAO D. & RAVICHANDRAN D. (2009). Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '09, p. 675–682, Stroudsburg, PA, USA : Association for Computational Linguistics.