FEEL: a French Expanded Emotion Lexicon

Amine ABDAOUI[1], Jérôme AZÉ[1], Sandra BRINGAY[1,2] and Pascal PONCELET[1]

(1) LIRMM UM B5, 860 St Priest Street, 34095 Montpellier, France
(2) MIAp UM3, Mende Road, 34199 Montpellier, France
{abdaoui, aze, bringay, poncelet}@lirmm.fr

Abstract. Sentiment analysis allows the semantic evaluation of a piece of text according to the expressed sentiments and opinions. While considerable attention has been given to the polarity (positive, negative) of English words, only few studies were interested in the conveyed emotions (joy, anger, surprise, sadness, etc.) especially in other languages. In this paper, we present the elaboration and the evaluation of a new French lexicon considering both polarity and emotion. The elaboration method is based on the semi-automatic translation and expansion to synonyms of the English NRC Word Emotion Association Lexicon (NRC-EmoLex). First, online translators have been automatically queried in order to create a first version of our new French Expanded Emotion Lexicon (FEEL). Then, a human professional translator manually validated the automatically obtained entries and the associated emotions. She agreed with more than 94% of the pre-validated entries (those found by a majority of translators) and less than 18% of the remaining entries (those found by very few translators). This result highlights that online tools can be used to get high quality resources with low cost. Annotating a subset of terms by three different annotators shows that the associated sentiments and emotions are consistent. Finally, extensive experiments have been conducted to compare the final version of FEEL with other existing French lexicons. Various French benchmarks for polarity and emotion classifications have been used in these evaluations. Experiments have shown that FEEL obtains competitive results for polarity, and significantly better results for basic emotions.

Keywords. Sentiment analysis, opinion mining, sentiment lexicon, polarity detection, emotion classification, semi-automatic translation.

1. Introduction

Automatic text analysis to detect the presence of subjective meanings, their polarity (positive, negative and neutral), the associated emotions (joy, anger, fear, etc.) as well as their intensity has been extensively investigated in the last decade. Called Sentiment or Opinion mining, they have a great deal of interest for real applications such as: managing customer relations (Homburg et al., 2015), predicting election results (Lewis-Beck and Dassonneville, 2015), etc. Actually, even dedicated API or applications have been proposed and included in well-known systems. For instance, Google Prediction API includes a sentiment analysis module[1] that can be used to build sentiment analysis models. Applied methods usually depends on the nature of the texts: tweets (Velcin et al., 2014), mails (Pestian et al., 2012), news headlines (Rao et al., 2013), etc., and obviously on the application domain: politics (Anjaria

---

[1] cloud.google.com/prediction/docs/sentiment_analysis

and Guddeti, 2014), environment (Hamon et al., 2015), health (Melzi et al., 2014), etc. They are often based on techniques from Statistics, Natural Language Processing and Machine Learning (ML). Supervised ML algorithms are frequently used to train text classifiers on tagged data sets. Their efficiency depends on the quality and size of the training data. However, it has been proved that the use of adapted sentiment lexicons can significantly improve the classification performances of bag of words classifiers (Hamdan et al., 2015). Indeed, recent studies suggest to include the words conveying each sentiment as descriptive features when learning text classification models (Mohammad et al., 2015).

Sentiment lexicons organize lists of words, phrases or idioms into predefined classes (polarities, emotions, etc.) (Devitt and Ahmad, 2013; Turney, 2002). For example, in NRC-EmoLex (Mohammad and Turney, 2013), starting point of this study, terms like *happy* and *heal* are labeled as positive, while terms like *abandon* and *hearse* are labeled as negative. Whereas each term has only one polarity, some terms may convey many emotions according to the used emotional typology. For example, in NRC-EmoLex, the word *happy* is associated with the emotions joy and trust, while the word *hearse* is associated with sadness and fear. Many emotion typologies exist in the literature (Ekman, 1992; Francisco and Gervás, 2006; Pearl and Steyvers, 2010; Plutchik, 1980). The most famous and at the same time the simplest typology among them is the one proposed by Ekman consisting in six basic emotions: joy, surprise, anger, fear, sadness and disgust. It has been considered in much of emotion classification studies (Mohammad and Kiritchenko, 2015; Roberts et al., 2012; Strapparava and Valitutti, 2004).

To date, most existing affect lexicons have been created for English and for polarity. In this paper, we describe the elaboration of a new French lexicon containing more than 14,000 terms according to their polarities (positive and negative) and their expressed emotions (we consider the Ekman basic emotions). The applied method is based on the automatic translation and expansion to synonyms of NRC-EmoLex, a publically available[2] emotion lexicon which has proven its performance in several sentiment and emotion classification tasks (Kiritchenko et al., 2014; Mohammad, 2012; Rosenthal et al., 2015). The translations have been obtained automatically by queering six online translators. An experienced human translator has validated the obtained entries as well as the associated emotions. She accepted more than 94% of the automatically pre-validated entries (those found by at least three online translators) and less than 18% of the remaining entries (those found by less than three online translators). Therefore, we believe that the proposed approach can be used to build high quality resources with low cost. Finally, in order to evaluate its quality, experiments for classification tasks (polarity and emotion) have been conducted with well-known French benchmarks. Results have shown that we obtain comparable scores for polarity classification comparing to the existing lexicons. More interestingly, we have shown that with FEEL clearly better results have been obtained for emotion classification when considering the available Ekman basic emotional classes. This result highlights that our resource is well adapted for polarity and emotion classifications. It can be accessed and downloaded publically on the internet[3] (Abdaoui et al., 2014).

The rest of the paper is organized as follows. Section 2 discusses a study of existing sentiment and emotion lexicons for both English and French. Section 3 describes our approach for automatically building a French lexicon as well as the manual validations. Section 4 compares FEEL with other existing French lexicons and shows their results in emotion and polarity classification tasks. Finally, Section 5 concludes and gives our main prospects.

---

[2] www.saifmohammad.com/WebPages/lexicons
[3] www.lirmm.fr/~abdaoui/FEEL

2. Related work

Sentiment lexicons can be constructed using three main approaches (Pang and Lee, 2008). First, they can be compiled manually by assigning the correct polarity or emotion conveyed by each word. Crowdsourcing tools and serious gaming are often used to get a large number of human annotations. (Mohammad and Turney, 2013) used the Amazon Mechanical Turk[4] service, while (Lafourcade et al., 2015a) designed an online Game With a Purpose (Like it![5]). Second, they can be compiled automatically using dictionaries. This approach uses a small set of seed terms for which the conveyed sentiments are known. Then, it grows the seed set by searching synonyms and antonyms using dictionaries (Strapparava and Valitutti, 2004). Finally, the third approach constructs sentiment lexicons automatically using corpora in two possible ways. On one hand, it can use annotated corpora of text documents and extract words that are frequent in a specific sentiment class and not in the other classes (Kiritchenko et al., 2014). On the other hand, it can use non-annotated corpora along with a small seed words list in order to discover new ones following their collocations (Harb et al., 2008) or using specifically designed rules (Neviarouskaya et al., 2011). However, each of these approaches has its own limitations. The manual approach is labor intensive and time consuming, while the automatic ones are error prone. In our case, we combine an automatic dictionary based approach with human manual annotation and supervision. Regarding the used sentiment and emotional typology, we have chosen the one proposed by (Ekman, 1992) consisting of two polarities (positive and negative) and six basic emotion classes (joy, surprise, sadness, fear, anger, disgust).

Table 1: Existing French resources for sentiment polarity and emotion

Resource	Description
Affects Lexicon (Augustyn et al., 2006)	Consists of about 1,200 French terms described by their polarity (positive and negative) and over 45 hierarchical emotional categories. It was automatically compiled and includes other information such as the intensity and the language level (common, literary).
CASOAR (Asher et al., 2008)	Contains polarized subjective terms in French. It consists of 270 verbs, 632 adjectives, 296 names, 594 adverbs and 51,178 expressions. It was manually constructed from several corpora (press articles, web comments, etc.). However, this resource is not publically available.
Polarimots (Gala and Brun, 2012)	Contains 7,483 French nouns, verbs, adjectives and adverbs whose polarity (positive, negative or neutral) has been semi-automatically annotated. 3,247 words have been added manually and 4,236 words has been created automatically by propagating the polarities.
Diko (Lafourcade et al., 2015a, 2015b)	Based on an online game with a purpose where players are asked to indicate the polarity and the emotion of the displayed expression. They can choose between three polarities (positive, negative and neutral), and 21 emotions. They can also enter a new emotion term when the exact emotion meaning of the displayed expression is not present between the 21 choices. Therefore, this lexicon associates 555,441 annotated expressions to almost 1,200 emotion terms.

[4] www.mturk.com/mturk/welcome
[5] www.jeuxdemots.org/likeit.php

Few French resources have been proposed, especially those dealing with emotions. Table 1 presents four French sentiment lexicons that we have found in the literature. If all of them offer the sentiment polarity, only two consider the exact emotional category. The Affects lexicon (Augustyn et al., 2006) which contains only around 1,200 terms associated with more than 45 hierarchical emotions and Diko (Lafourcade et al., 2015b) which contains about 450,000 non-lemmatized expression but associated with almost 1,200 emotion terms (many synonyms exist). The two remaining lexicons CASOAR (Asher et al., 2008) and Polarimots (Gala and Brun, 2012) consider only the polarity and not the emotion. Furthermore, CASOAR is not publically available making the number of truly exploitable French sentiment resources equal to three.

Table 2: Existing English resources for sentiment polarity and emotion

Resource	Description
General Inquirer (Stone et al., 1966)	Contains more than 10,000 English words labeled manually by 182 categories including polarity and some emotions.
WordNet Affect (Strapparava and Valitutti, 2004)	Contains only hundreds of English words labeled with their expressed polarity and emotion. It was created by manually identifying seeds (words whose associations with sentiments are known) and spreading these emotions to all their synonyms using WordNet.
MPQA (Wilson et al., 2005)	Contains 8,222 English subjectivity words associated with three polarities (positive, negative and neutral).
LIWC: Linguistic Inquiry and Word Count (Pennebaker et al., 2007)	Contains about 4,500 English words labeled by many categories including polarity and emotion. It was created by combining other existing resources and by validating the categories manually by human judges.
Bing Liu's Opinion Lexicon (Qiu et al., 2009)	Contains around 6,800 English opinion words associated with their polarities (positive and negative). It was created automatically using a corpus-based approach.
NRC-EmoLex (Mohammad and Turney, 2013)	Contains more than 14,000 English terms labeled by the expressed polarity (positive or negative) and emotion (joy, trust, anticipation, sadness, surprise, disgust, fear or anger). The authors used Amazon Mechanical Turk [6] in order to obtain a large number of manual annotations in order to compile their resource.
NRC Hashtag Emotion Lexicon (Mohammad and Kiritchenko, 2015)	Contains real valued English words between 0 (not associated) to infinity (maximally associated) for each sentiment polarity and emotion class. It gathers 16,862 unigrams (words) that have been created automatically using a corpus based approach. The corpus has been obtained from Twitter by extracting tweets that contains the following hashtags: #joy, #sadness, #surprise, #disgust, #fear and #anger.

More sentiment resources have been compiled for English terms. Table 2 shows seven English lexicons that we found in the literature. All of the English resources consider the sentiment polarity but only five offer the exact emotional category. As we want to build a sentiment lexicon that considers both emotion and polarity, we restrict our choice to the remaining five English lexicons. The most extensive English lexicons are NRC-EmoLex (Mohammad and Turney, 2013) and the NRC Hashtag Emotion lexicon (Mohammad and Kiritchenko, 2015). These lexicons have proven their performance in several sentiment and emotion classification tasks (Kiritchenko et al., 2014; Mohammad, 2012; Rosenthal et al., 2015). Indeed, their authors obtained remarkable results in the evaluation campaigns SEM-EVAL 2013 (Nakov et al., 2013) and SEM-EVAL 2014 (Rosenthal et al., 2014). Furthermore, NRC-EmoLex has been built on the General Inquirer (Stone et al., 1966) and the WordNet Affect (Strapparava and Valitutti, 2004) lexicons. Concretely, it corrects their terms and add new unigrams and bigrams using the wisdom of the crowds. For all these reasons, we decided to start from this resource in order to constitute a new comprehensive emotion resource for French.

3. Methods

In this section, we present the methods used for the automatic creation of FEEL. Then, we describe the manual validations by a professional human translator. Finally, we evaluate the sentiments associated with a subset of terms by three different human annotators.

3.1. Automatic Creation

After manually correcting some inconsistencies in NRC-EmoLex (words associated with all emotions and words associated with contradictory polarities), our aim was to automatically translate to French all of its English terms (14,182 terms). Automatic translation methods can be based on three types of resources: 1) aligned resources (Och and Ney, 2004); 2) comparable corpus (Sadat et al., 2003) and 3) multilingual encyclopedia (Erdmann et al., 2009). Since we do not have aligned resources nor comparable corpora in which we could find all the entries of the initial lexicon, we chose a different approach and used the wealth of automatic translators available online. For each entry of NRC-EmoLex, we automatically queried six online translators: Google Translate[7], Bing Translate[8], Collins Translator[9], Reverso Dictionary[10], Bab.la[11] and Word Reference[12]. Each English term may generate many French translations. The entries that have been obtained by at least three translators have been considered pre-validated.

In order to expand our resource we decided to include English and French Synonyms. Synonymy corresponds to a similarity in meaning between words or phrases in the same language. Therefore, synonyms should have the same emotion and polarity class. Antonyms have not been considered since our emotion model do not support contrary emotions. In the literature, synonymy has been used to build sentiment resources by expending seed words for which the polarity or the emotional class is already known (Strapparava and Valitutti, 2004). Here, we adopted a similar approach to expand both the English entries and the French translations. For all English entries of the original resource, we searched for synonyms using

---

[7] www.translate.google.fr

[8] www.bing.com/translator

[9] www.collinsdictionary.com

[10] www.reverso.net

[11] fr.bab.la/dictionnaire

[12] www.wordreference.com

eight online websites: Reverso Dictionary, Bab.la, Atlas [13], Thesaurus [14], Ortolang [15], SensAgent [16], The Free Dictionary [17] and the Synonym website [18]. The obtained English synonyms have been translated as previously described. Similarly, for all French entries, we searched for synonyms using two online websites: Ortolang and Synonymo [19]. Entries associated with contradictory polarities have been automatically removed. Finally, the automatically compiled resource contained 141,428 French entries (56,599 pre-validated entries and 84,829 non pre-validated entries).

3.2. Validating the translations

In order to obtain a high quality resource and to evaluate the quality of the automatic process, we hired a human professional translator. All the automatically obtained entries have been presented to her via a web interface. For each English term, she can validate or not the automatically obtained translations, manually add a new translation and change the associated polarities and emotions. Examples of sentences using the current term have been presented in order to better understand its meaning. These sentences have been generated from the Linguee website[20]. Our professional translator worked full-time for two months. She validated less than 18% of the entries that have been obtained by less than three translators (15,091 terms), against more than 94% of ones that have been found by at least three online translators (53,277 terms). This result shows that it is possible to use online translators in order to uncostly compile good quality resources. In addition to the validated entries based on the automatic translators, our human translator manually added 10,431 new French translations based on the displayed English terms. Finally, our resource contained 81,757 French entries (lemmas and flexed forms), which have been lemmatized using the TreeTagger tool (Schmid, 1994). This process generated 14,127 distinct lemmatized terms consisting in 11,979 words and 2,148 compound terms. The lemmatized terms have been associated with all the emotions of their inflected forms. Terms associated with contradictory polarities have been removed (81 terms). We considered that these terms dot not convey sentiments by their own and may be positive or negative according to their context. For example, the word "to vote" may be used either in a positive context "to vote for" or in a negative one "to vote against". Table 3 shows the repartition of the final lemmatized terms between the two considered polarities and the six basic emotions, and the intersections between them. It appears that most positive entries are associated with the emotion joy. However, some positive entries are associated with the emotions surprise, fear, sadness, anger and disgust. For example, the human translator validated the word *plonger (dive)* as positive but associated with the emotion fear. On the hand, most negative entries are associated with the emotions surprise, fear, sadness, anger and disgust. Nevertheless, very few negative entries are associated with the emotion joy. For example, the word *capiteux* (*heady*) is negative but has been associated with the emotion joy. We decided not to consider these associations as inconsistent since our human translator validated them. Similarly, emotions may have common terms especially negative ones. For example, the word *accuser* (*accuse*) is associated with the emotions anger and disgust. Finally, joy is the most pure emotion since it does not have any common entry with the remaining Ekman basic emotions.

---

[13] dico.isc.cnrs.fr

[14] www.thesaurus.org

[15] www.cnrtl.fr/synonymie/

[16] dictionnaire.sensagent.com/synonyme/en-fr/

[17] www.thefreedictionary.com

[18] www.synonym.com

[19] www.synonymo.fr

[20] www.linguee.fr

Table 3: The intersections between the polarities and emotions in FEEL

	Positive	Negative	Joy	Surprise	Anger	Disgust	Sadness	Fear
Positive	5,704							
Negative	0	8,423						
Joy	513	7	521					
Surprise	435	747	0	1,182				
Anger	120	1,983	0	355	2,103			
Disgust	92	1,922	0	133	889	2,014		
Sadness	132	2,381	0	291	932	837	2,513	
Fear	223	2,976	0	657	1,335	909	1,532	3,199

3.3. Evaluating the sentiments

While the professional manual translations can be considered reliable, the associated sentiments and emotions may be subjective (only one annotator). In order to evaluate the quality of our resource, the sentiments and emotions associated with a subset of FEEL terms have been evaluated manually by three new annotators. In order to compile this subset, we selected terms that are frequent in four French benchmarks. These benchmarks will be used later in order to test whether FEEL can improve sentiment and emotion classification. Three of these benchmarks have been produced for the third edition of the French Text Mining challenge (DEFT'07)[21]. The task was the classification of text documents from various sources according to their polarity. The fourth benchmark has been produced for the 11th edition of the same challenge (DEFT'15)[22], where the task was the classification of tweets according to their polarity, subjectivity and expressed emotions. Table 4 presents the nature and the subject of each benchmark and the considered classification task(s). If all the benchmarks consider the polarity of French texts, only the fourth one considers the exact emotional class.

Table 4: Details about the used benchmarks

Benchmark	Description	Task
See and Read	Movie, book and show reviews from the *avoir-alire* website[23]	Polarity
Political Debate	Debate reports in the French National Assembly (2002 – 2007)[24]	Polarity
Videos Games	Video games reviews from the *jeux-videos* website[25]	Polarity
Climate	Tweets about Climate change annotated during the *ucomp* project[26]	Polarity/ emotion

---

[21] www.deft.limsi.fr/2007
[22] www.deft.limsi.fr/2015
[23] www.avoir-alire.com
[24] www.assemblee-nationale.fr/12/debats
[25] www.jeuxvideo.com
[26] www.ucomp.eu

Terms that appear at least 10 times in the training set and at least 10 times in the testing set of each benchmark have been selected. Figure 1 shows the frequency of FEEL terms in the training set of the *Climate* benchmark (shown in a $\log_{10}$ scale). The horizontal line (y=1) corresponds to our frequency threshold ($\log_{10}(10)=1$). Finally, 120 terms have been selected which represents less than 1% of FEEL terms. However, this subset of terms represents almost a third of FEEL terms occurrences in the presented benchmarks. Regarding their division between the two polarities, 109 terms were initially assigned to the positive polarity against 11 terms associated with the negative one. On the other hand, each emotion of the Ekman typology has only seven terms except the emotion "Anger" that has four terms. Most of the terms are not associated with any emotion.
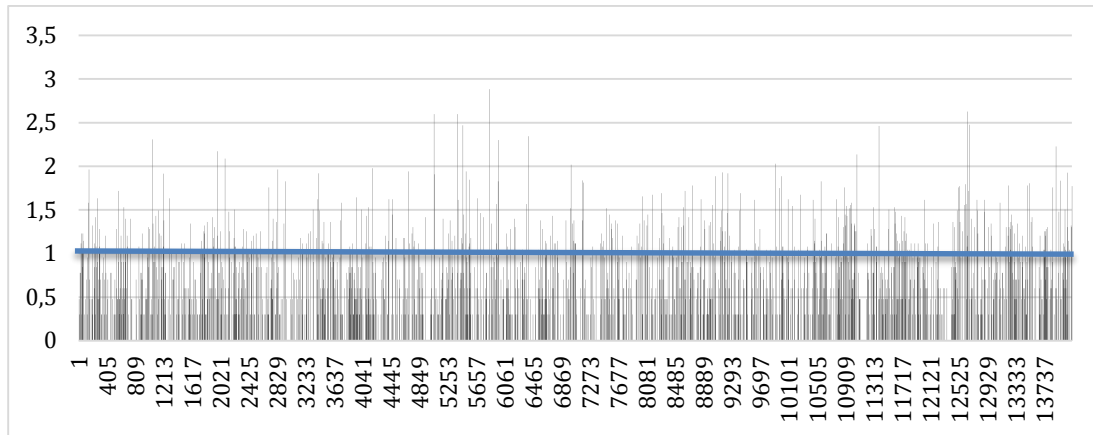

Figure 1 : The distribution (in a $\log_{10}$ scale) of FEEL terms in the training set of the Climate benchmark

These terms have been presented to three new annotators in order to check the associated polarities and emotions. In order to handle polysemy, two types of annotation have been performed:

- **Annotation without context:** the annotators are asked to choose the associated polarities and emotions without presenting any example to them.
- **Annotation in context:** the annotators are asked to choose the associated polarities and emotions according to its sense in the displayed sentence. Four contexts have been considered corresponding to the four used benchmarks. From each benchmark, we selected the first sentence containing the corresponding term and present it as an example to the annotators.

Table 5 : Annotators agreement for polarity and emotions (arithmetic mean) in each annotation type. We present the Fleiss' Kappa and the percentage of terms for which all annotators chose the same sentiment.

	Fleiss' Kappa		Percentage of terms for which all annotators agreed	
	Without context	In context	Without context	In context
Polarity (positive/negative)	0.68	0.56	92.5%	85.4%
Emotions (yes/no) - mean	0.22	0.18	95.4%	95.6%

Table 5 presents the agreement between the three annotators in each annotation type. First, Fleiss' kappa shows good polarity agreement and bad emotion agreement in both annotation types. These results are similar to those obtained in (Mohammad and Turney, 2013) when building the original English NRC-EmoLex. However, Fleiss' kappa does not take into account the number of items per category. Since we have very unbalanced categories (much more terms associated with the category "no" than terms associated with the category "yes" for a given emotion), we also present the percentage of terms for which the three annotators have chosen the same category. Indeed, our three annotators agreed for most of the terms (more than 85% in each task and annotation type). Finally, our annotators suggested to include the polarity "neutral" in our future work.

Table 6 : Evaluating the sentiments of the chosen subset of terms

	P_{mi}	R_{mi}	F_{mi}
Polarity (positive/negative)	0.99	0.99	0.99
Emotions (yes/no) – arithmetic mean	0.96	0.99	0.98

Finally, the annotations without context have been used to evaluate the initial sentiments and emotions. A majority vote has been considered in order to extract the reference annotations. Table 6 presents the micro averaged precisions, recalls and F1-measures for polarity and emotions. Micro averaging is used to deal with unbalanced data sets. In our case, we used the label-frequency-based micro-averaging (Van Asch, 2012). It weighs each class results with its proportion of documents in the test set. The emotions evaluation metrics have been averaged by arithmetic mean between the six emotions. The presented results show very high consistency between the initial sentiments and those selected by at least two new annotators (majority vote).

4. Evaluations

In this section, we compare FEEL with existing French resources using various French benchmarks for polarity and emotion classifications.

4.1. Lexicons

Here, we present the lexicons used in our evaluations. Among the four French lexicons listed in section 2, only CASOAR has not been included here since it is not publically available. The remaining three French lexicons have been downloaded and used in our evaluations. All of it contain lemmatized terms excepting Diko. The expressions of this last lexicon have been cleaned and grouped into lemmatized terms. Figure 2 presents the percentage of terms in each lexicon according to their number of words. It appears that almost all Affects and Polarimots terms are composed of only one word (100% for Polarimots and over 99% for Affects). More than 85% of FEEL terms are words and almost 15% are compound terms. Among the compound terms, 9% are composed of two words and 5% are composed of three words. Finally, only 33% of Diko terms are words. The rest are devided as follow: 31% are composed of two words, 22% are composed of three words, 8% are composed of four words, 3% are composed of five words and the remaining 3% are composed of more than five words.
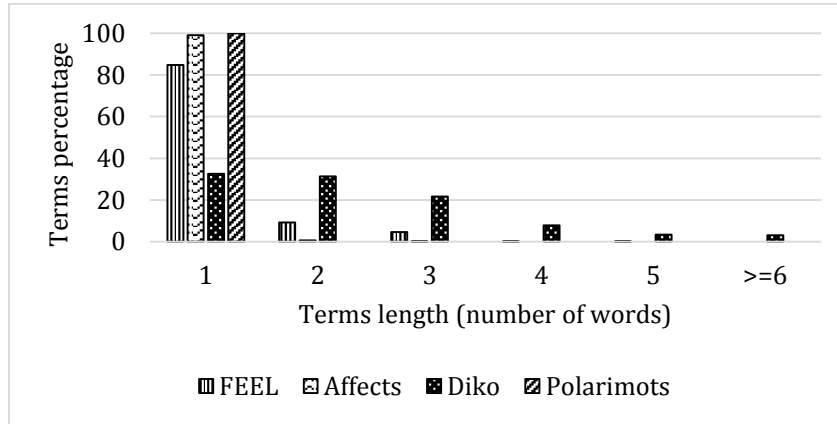
Figure 2: The percentage of terms in each lexicon according to their length (number of words)

Table 7 presents the number of terms in each lexicon and the number of common terms between each couple of lexicons. Diko is the largest resource with 382,817 lemmatized French entries. FEEL is the second largest with 14,127 terms. Polarimots and Affects lexicon contain 7,483 and 1,348 terms respectively. Diko covers almost 97% of FEEL terms (13,681 out of 14,127), almost 88% (1,182 out of 1,348) of Affects terms and more than 98% of Polarimots terms (7,359 out of 7,483). Therefore, Diko is clearly the most extensive resource but we do not have information about the proportion of noisy terms that it may contains (non-affective terms).

Table 7: The intersections between the terms on each couple of lexicons

	FEEL	*Affects*	*Diko*	*Polarimots*
FEEL	14,127			
Affects	559	1,348		
Diko	13,681	1,182	382,486	
Polarimots	2,747	237	7,359	7,483

Table 8 shows the number of positive, negative and neutral terms in each lexicon. FEEL is the only lexicon that do not consider the neutral polarity. We notice that all lexicons have more negative terms than positive ones except Diko. The algorithm used for selecting the candidate terms may explain this observation (Lafourcade et al., 2015c).

Table 8: The number of positive, negative and neutral terms in each lexicon

	FEEL	*Affects*	*Diko*	*Polarimots*
Positive	5,704	437	224,832	1,315
Negative	8,423	790	55,593	1,464
Neutral	0	121	102,061	4,704

Regarding the agreement between each couple of lexicons about the associated polarities, Table 9 presents the percentage of common terms having the same polarity. Neutral terms have

not been considered in these calculations. Table 9 shows that for all couples of lexicons, more than 80% of their common positive and negative terms are associated with the same polarity. The highest agreement is observed between Diko and Polarimots with 91% of common terms associated with the same polarity.

Table 9: Percentage of common terms between each couple of lexicons having the same polarity

Lexicons	FEEL	Affects	Diko
Affects	89%		
Diko	83%	89%	
Polarimots	80%	86%	**91%**

Finally, all the used lexicons consider the polarity of French terms but only three give the exact emotion class (Polarimot do not consider emotions). Each one of the remaining lexicons follows its own emotional typology (FEEL: 6 emotions, Affects Lexicon: 45 emotions, Diko: more than 1,200 emotion terms).

4.2. Evaluation Benchmarks

Table 10 presents the repartition of positive and negative text documents for training and testing in each benchmark. It shows that the benchmark *Political Debate* contains the largest number of documents. It also shows that there is an acceptable number of documents for training and for testing in each benchmark.

Table 10: The repartition of training and testing documents for polarity in each benchmark

Benchmark	Training			Testing		
	positive	negative	total	positive	negative	total
See and Read	1,150	309	1,459	768	207	975
Political Debate	6,899	10,400	17,299	4,961	6,572	11,533
Videos Games	874	497	1,371	583	332	915
Climate	2,448	1,875	4,323	1,057	804	1,861

Regarding the reparation of text documents into the emotion classes, the only considered benchmark is *Climate*. This benchmark distinguishes 18 emotion classes, which are presented in Figure 3. For better visualization, the number of tweets is shown in logarithmic scale (base 10). Only four among the six Ekman basic emotion classes are present in this emotional typology. Figure 4 shows the repartition of tweets between these four emotions for training and testing sets (positive surprise and negative surprise have been grouped in one class). In both figures, it appears that the emotion classes are very unbalanced. For example, only 6 tweets are associated with Boredom, while 2,148 tweets are labeled with Valorization. The complete table presenting the repartition of *Climate* training and testing tweets between the 18 original emotions is presented in the appendices.
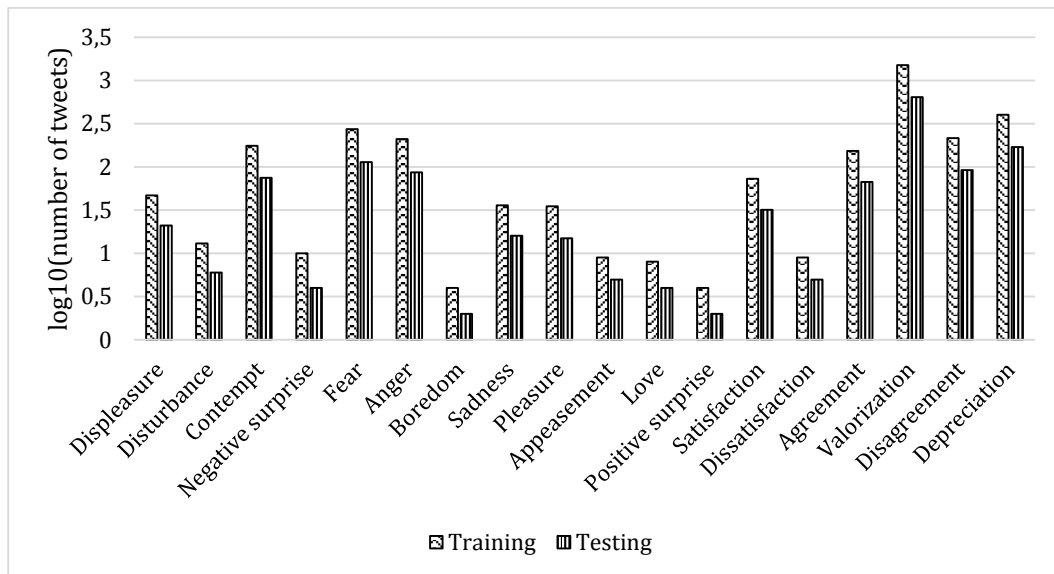
Figure 3: The repartition of Climate training and testing tweets between the original 18 emotion classes (logarithmic scale)
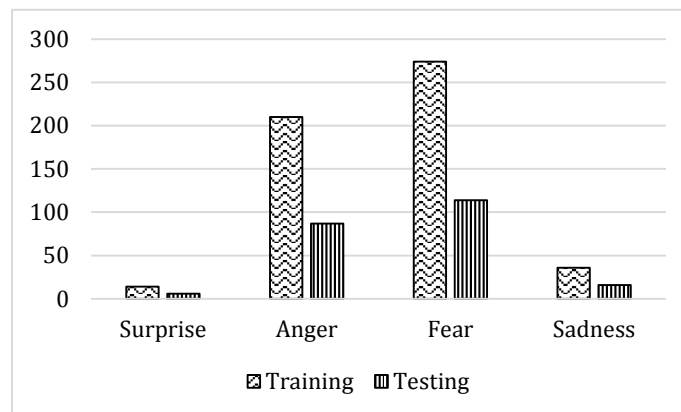

Figure 4: The repartition of Climate training and testing tweets between the available Ekman basic emotions

4.3. Evaluation in a Polarity Classification Task

Our aim is to evaluate the classification gain when using features extracted from different lexicons compared to bag of words classifiers. First, Support Vector Machines (SVM) have been trained on each data set with the Sequential Minimal Optimization method (Platt, 1999). The Weka data-mining tool (Hall et al., 2009) have been used to train these classifiers with default settings on lemmatized and lowercased text documents. A feature selection step has been performed using the Information Gain filter (words having positive Information Gain have been selected). In our experiments, we call this configuration Bag_Of_Words. Then we add to this configuration, two features from each lexicon. Indeed, we compute the number of positive words and the number of negative words according to each lexicon. These two features have been added before applying the Information Gain filter. Six other configurations have been evaluated for each data set corresponding to the four tested lexicons and the two additional FEEL variations: FEEL with replacement of the 120 terms from the annotation without context (FEEL_WiCxt) and in the corresponding context (FEEL_InCxt). The macro (arithmetic mean) and micro (weighted mean) precisions, recalls and F1-measures of these configurations applied on each corpus are presented in Tables 11, 12, 13 and 14.

Table 11: Polarity classification results on the See and Read data set

	P_{macro}	R_{macro}	F_{macro}	P_{micro}	R_{micro}	F_{micro}
Bag_Of_Words	83.5	74.2	77.4	86.2	86.9	85.8
Bag_Of_Words + FEEL	**84.5**	**76.6**	**79.5**	**87.2**	**87.8**	**87.0**
Bag_Of_Words + FEEL_WiCxt	**84.5**	**76.6**	**79.5**	**87.2**	**87.8**	**87.0**
Bag_Of_Words + FEEL_InCxt	**84.5**	**76.6**	**79.5**	**87.2**	**87.8**	**87.0**
Bag_Of_Words + Affects	84.2	75.0	78.3	86.7	87.3	86.3
BW + Diko	84.0	75.6	78.7	86.8	87.4	86.5
Bag_Of_Words + Polarimots	83.5	74.2	77.4	86.2	86.9	85.8

Table 12: Polarity classification results on the Political Debate data set

	P_{macro}	R_{macro}	F_{macro}	P_{micro}	R_{micro}	F_{micro}
Bag_Of_Words	70.2	70.2	70.0	70.6	70.8	70.7
Bag_Of_Words + FEEL	**70.6**	**70.2**	**70.3**	**71.0**	**71.1**	**71.0**
Bag_Of_Words + FEEL_WiCxt	70.5	70.1	70.1	70.9	71.1	70.9
Bag_Of_Words + FEEL_InCxt	70.4	70	70.2	70.8	71	70.8
Bag_Of_Words + Affects	70.4	70.0	70.2	70.8	71.0	70.9
BW + Diko	70.4	70.0	70.1	70.8	71.0	70.8
Bag_Of_Words + Polarimots	70.2	69.9	70.0	70.6	70.8	70.7

Table 13: Polarity classification results on the Videos Games data set

	P_{macro}	R_{macro}	F_{macro}	P_{micro}	R_{micro}	F_{micro}
Bag_Of_Words	93.6	93.4	93.5	94	94	94
Bag_Of_Words + FEEL	93.5	93.5	93.5	94	94	94
Bag_Of_Words + FEEL_WiCxt	93.5	93.5	93.5	94	94	94
Bag_Of_Words + FEEL_InCxt	93.5	93.5	93.5	94	94	94
Bag_Of_Words + Affects	93.5	93.5	93.5	94	94	94
BW + Diko	93.8	93.7	93.8	94.2	94.2	94.2
Bag_Of_Words + Polarimots	**94.0**	**94.0**	**94.0**	**94.4**	**94.4**	**94.4**

Table 14: Polarity classification results on the Climate data set

	P_{macro}	R_{macro}	F_{macro}	P_{micro}	R_{micro}	F_{micro}
Bag_Of_Words	72.8	69.1	69.2	72.4	71.6	70.3
Bag_Of_Words + FEEL	76.1	74.8	75.1	76.1	76.1	75.8
Bag_Of_Words + FEEL_WiCxt	76.4	75.6	75.8	76.5	76.6	76.4
Bag_Of_Words + FEEL_InCxt	76.4	75.6	75.8	76.5	76.6	76.4
Bag_Of_Words + Affects	73.3	72.4	70.2	73.3	72.4	71.3
BW + Diko	**77.8**	**76.0**	**76.4**	**77.6**	**77.4**	**77.1**
Bag_Of_Words + Polarimots	74.2	70.7	71.0	73.7	73.0	72.0

The Bag_Of_Words configuration with lemmatization, lowercasing and especially feature subset selection represents a highly efficient baseline. Indeed, this configuration obtained high micro and macro precisions, recalls and F-measures on all benchmarks. Moreover, the Information Gain filter selected between 63 and 390 lemmatized words for every benchmark. Therefore, it is difficult to observe a significant gain only by adding two new features. Still, the performance gain is noticeable in all benchmarks. Almost all the lexicons induce a gain that varies from 0.1% to 7.1% in the considered evaluation metrics. If the use of lexicons obtains a little gain on the three first benchmarks (*See and Read*, *Political Debate* and *Videos Games*), their use induce a 7% gain on the fourth benchmark (*Climate*). This observation may be related to the text nature, since the fourth benchmark is the only one that contains tweets. Indeed, tweets are very short text documents (less than 140 characters) while product reviews or debate reports can contain hundreds of words. Regarding the performance of each lexicon, we notice that it depends on the benchmark. There is no lexicon that obtains the best results in all the used benchmarks. However, FEEL obtains the best results on two benchmarks (online reviews and debate transcriptions), Polarimots obtains the best results on *Video Games* and Diko on tweets. Globally, FEEL obtains very competitive results being the best on two benchmarks and second on a third one (*Climate*). The difference between FEEL and the best configuration is always less than 1%. Regarding the two derivations of FEEL from the re-annotation, we observe a small change in the results in comparison the original resource. This observation may be explained by the very high consistency between FEEL_WiCxt and FEEL as presented in table 6. On the other, the choice of the example sentence in the annotation with a context may be unrepresentative of the term use whole benchmark.

4.4. Evaluation in an Emotion Classification Task

Only the fourth benchmark provides emotion classes for its text documents (tweets). It uses an emotional typology divided into 18 classes as presented in Figure 3. As mentioned before, these emotional classes are very unbalanced. For example, only six tweets are associated with the emotion Boredom, while 2,148 tweets are labeled with the emotion Valorization. Therefore, macro averaging is not adapted in this case. Here, we only consider the label-frequency-based micro averaging. Regarding the lexicons, Polarimots is the only resource that do not consider emotions. We perform our evaluations using the remaining lexicons. FEEL proposes six emotion classes, Affects has 45 emotions and Diko associates its terms with 1,198 emotion expressions. We use the same baseline as in the polarity classification task (Bag_Of_Words).

To this configuration, we evaluate the add of features extracted from each emotion lexicon. These features represent the number of terms expressing each emotion. Therefore, six features are added for FEEL, FEEL_WiCxt and FEEL_InCxt, 45 features are added for Affects and 1,198 features are added for Diko. The feature selection step is applied after adding these features. Lemmatization and lowercasing are also performed when searching the emotion terms inside the tweets. Table 13 presents the emotion classification results when considering the 18 original emotion classes.

Table 15: The emotion classification results when considering 18 emotional classes

	P_{micro}	R_{micro}	F_{micro}
Bag_Of_Words	46.9	49.7	39.7
Bag_Of_Words + FEEL	50.8	53.6	44.7
Bag_Of_Words + FEEL_WiCxt	51.1	53.9	45.1
Bag_Of_Words + FEEL_InCxt	50.9	53.7	45
Bag_Of_Words + Affects	50.9	53.8	45.4
Bag_Of_Words + Diko	**52.6**	**55.0**	**46.8**

As shown in table 15, all emotion lexicons improve significantly the classification results. The gain is between 5.7% and 12.9% in micro precision, between 3.9% and 5.3% in micro recall and between 5% and 7.1% in micro F-measure. Diko obtains the highest micro recall but the lowest micro precision (due to its large number of entries). FEEL is ranked third but close to the best configuration for each evaluation metric. FEEL_WiCxt and FEEL_InCxt improve slightly the classification results. However, the emotional typology of the *Climate* corpus (18 classes) do not refer to a well-known classification. We are evaluating FEEL on classes that it does not consider. In order to have an estimation of each lexicon performance according to the Ekman emotional classes, we perform the same experiments but when considering only the four Ekman emotions that are present is the *Climate* corpus. The repartition of the considered tweets between the emotions (surprise, anger, fear and sadness) are presented in Figure 4. In addition to the bag of words configuration, we evaluate the add of six features for FEEL, FEEL_WiCxt and FEEL_InCxt, 45 features for Affects and 1,198 features Diko.

Table 16: The emotion classification results when considering Ekman emotional classes

	P_{micro}	R_{micro}	F_{micro}
Bag_Of_Words	74	70	68.2
Bag_Of_Words + FEEL	**74.3**	**74.4**	**72.8**
Bag_Of_Words + FEEL_WiCxt	73.6	73.5	72.2
Bag_Of_Words + FEEL_InCxt	73.6	73.5	72.2
Bag_Of_Words + Affects	69.1	69.5	69.2
Bag_Of_Words + Diko	71.7	68.6	66

Table 16 shows that FEEL obtained the best results. It generates a gain of 0.3% in micro precision, 4.4% in micro recall and 4.6% micro F1-measure in comparison to the bag of words configuration. FEEL_WiCxt and FEEL_InCxt come second with close precisions, recalls and F1-measures. Finally, Affects and Diko generate a decrease in the evaluation metrics, which suggests that these lexicons are not adapted to the Ekman emotions. Since Affects and Diko propose a finer emotional typology, we may think that this should not influence the classification performance with less emotional classes. Even though, FEEL significantly outperforms these two lexicons for the available Ekman emotions (four out of six). Since *Climate* is the only available French benchmark for emotion classification, we could not test FEEL on the Ekman emotions: joy and disgust.

5. Conclusion

Due to its huge number of applications, sentiment analysis received much attention in the last decade. Most studies dealt with polarity detection in English texts. Whereas emotion detection have many applications (such as detecting angry customers and directing them to upper hierarchy), only few studies considered it especially in French. In this work, we presented the elaboration and the evaluation of a new French sentiment lexicon. It considers both polarity and emotion following the Ekman emotional typology. It has been compiled by translating and expanding to synonyms the English lexicon NRC-EmoLex. A human professional translator supervised all the automatically obtained terms and enriched them with new manual terms. She validated more than 94% of the entries that have been found by at least three online translators, and less than 18% of the ones that have been obtained by less than three translators. This result shows that online translators can be used to inexpensively compile such resources using appropriate heuristics and thresholds. The final resource contains 14,127 French entries where around 85% are single words and 15% are compound words. While the professional manual translations can be considered reliable, the associated sentiments and emotions may be subjective. Therefore, three new annotators re-evaluated the polarities and emotions associated with a subset of 120 terms. This step showed high consistency between the initial sentiments and the new ones. Then, we performed exhaustive evaluations on all the French benchmarks that we found in the literature for polarity and emotion classifications. We compared our results with the existing French sentiment lexicons. In order to represent each lexicon we used the number of terms expressing each sentiment as a new feature, but other configurations may be evaluated. The obtained results highlight that our new French Expanded Emotion Lexicon improves the classification performances on various benchmarks dealing with very different topics. Indeed FEEL obtained competitive results for polarity (being first and two benchmarks and always very close to the best configuration) and the best results for emotion (when considering the Ekman emotional typology). It could be noticed that the classification gain is more important for short text documents such as tweets. Finally, this work shows that automatic translation can be used in order to compile resources having different emotional typologies with low cost.

The first perspective to this work is to compile a benchmark of French text documents tagged with the six basic Ekman emotions. Similar benchmarks have been compiled for English (Strapparava and Mihalcea, 2008) following the Ekman typology. Crowdsourcing tools can be used to obtain large number of manual annotations. We can also scroll the Twitter API with the following hashtags: #joy, #surprise, #anger, #sadness, #fear and #disgust. Indeed, (Mohammad and Kiritchenko, 2015) show that this process has led to a good quality English benchmark. The second perspective focuses on the use of FEEL in order to build sentiment analysis systems. Using FEEL, we built a complete sentiment classification system that participated to the evaluation campaign DEFT 2015. Among 22 teams that have registered to the challenge, we

were ranked first in subjectivity classification, third in polarity classification and fifth in emotion classification (when considering 18 classes). The proposed system is also based on SVM classifiers but with more elaborated features. A publically available version of this system can be downloaded on GitHub[27]. Furthermore, a sentiment classification platform is now under development. Users will have the possibility to use this system online or as an external API. Similar tools exist for English such as Sentiment Treebank[28] or Semantria[29]. Finally, the proposed method can be used in order to uncostly compile French lexicons for other applications. On one hand, we want to detect agreement and disagreement in online forum discussions. The objective is to compute a user reputation value based on the replies addressed to him (Abdaoui et al., 2015). Agreement and disagreement lexicons can be used to evaluate the trust or distrust expressed inside the textual content of replies. We suggest using the proposed method in order to translate to French English resources that have been compiled for agreement and disagreement (Wang and Cardie, 2014). On the other hand, we are working on a project that aims to prevent suicide using social networks (Facebook, Twitter, forums, etc.). Cases of suicides have been reported in recent years as people have posted on social networks expressing their thought or addressing messages to their families (Cherry et al., 2012). We believe that sentiment and emotion analysis can be adapted to detect dysphoric states. Specific lexicons for depression symptoms have been created for English (Karmen et al., 2015). Similarly, automatic translation can be used to create depression symptoms lexicons for French.

Acknowledgment

References

Abdaoui, A., Azé, J., Bringay, S., Poncelet, P., 2015. Collaborative Content-Based Method for Estimating User Reputation in Online Forums, in: 16th Web Information Systems Engineering Conference, Part II. pp. 292–299.

Abdaoui, A., Azé, J., Bringay, S., Poncelet, P., 2014. FEEL : French Extended Emotional Lexicon, in: ELRA Catalogue of Language Resources. ISLRN: 041-639-484-224-2.

Anjaria, M., Guddeti, R.M.R., 2014. Influence factor based opinion mining of Twitter data using supervised learning, in: 6th International Conference on Communication Systems and Networks, pp. 1–8.

Asher, N., Benamara, F., Mathieu, Y.Y., 2008. Distilling Opinion in Discourse: A Preliminary Study., in: the International Conference on Computational Linguistics. pp. 7–10.

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. Grenoble, France, pp. 407–414.

---

[27] github.com/amineabdaoui/SentimentClassification
[28] nlp.stanford.edu/sentiment
[29] www.lexalytics.com/demo
[30] www.lirmm.fr/patient-mind/pmwiki/pmwiki.php
[31] www.mesrs.dz

Cherry, C., Mohammad, S.M., de Bruijn, B., 2012. Binary Classifiers and Latent Sequence Models for Emotion Detection in Suicide Notes. *Biomed Inform Insights* 5, 147–154.

Devitt, A., Ahmad, K., 2013. Is there a language of sentiment? An analysis of lexical resources for sentiment analysis. *Lang Resources & Evaluation* 47, 475–511.

Ekman, P., 1992. An argument for basic emotions. *Cognition & emotion* 6, 169–200.

Erdmann, M., Nakayama, K., Hara, T., Nishio, S., 2009. Improving the extraction of bilingual terminology from Wikipedia. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 5, 31:1–31:17.

Francisco, V., Gervás, P., 2006. Exploring the compositionality of emotions in text: Word emotions, sentence emotions and automated tagging, in: AAAI-06 Workshop on Computational Aesthetics: Artificial Intelligence Approaches to Beauty and Happiness.

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, pp. 495–502.

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, 10–18.

Hamdan, H., Bellot, P., Bechet, F., 2015. Sentiment Lexicon-Based Features for Sentiment Analysis in Short Text, in: 16th International Conference on Intelligent Text Processing and Computational Linguistics, pp. 1–10.

Hamon, T., Fraisse, A., Paroubek, P., Zweigenbaum, P., Grouin, C., 2015. Analyse des émotions, sentiments et opinions exprimés dans les tweets : présentation et résultats de l'édition 2015 du défi fouille de texte (DEFT), in: 11eme Défi Fouille de Texte. Association pour le Traitement Automatique des Langues, pp. 1–11.

Harb, A., Plantié, M., Dray, G., Roche, M., Trousset, F., Poncelet, P., 2008. Web Opinion Mining: How to Extract Opinions from Blogs?, in: 5th International Conference on Soft Computing As Transdisciplinary Science and Technology, pp. 211–217.

Homburg, C., Ehm, L., Artz, M., 2015. Measuring and Managing Consumer Sentiment in an Online Community Environment. *Journal of Marketing Research* 52, 629–641.

Kiritchenko, S., Zhu, X., Mohammad, S.M., 2014. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research* 723–762.

Lafourcade, M., Joubert, A., Brun, N.L., 2015a. Games with a Purpose (GWAPS). *John Wiley & Sons*. ISBN: 978-1-84821-803-1.

Lafourcade, M., Le Brun, N., Joubert, A., 2015b. Collecting and Evaluating Lexical Polarity with a Game with a Purpose, in: the International Conference on Recent Advances in Natural Language Processing, pp. 329–337.

Lafourcade, M., Le Brun, N., Joubert, A., 2015c. Vous aimez ?...ou pas ? LikeIt, un jeu pour construire une ressource lexicale de polarité, in: Actes de La 22e Conférence Sur Le Traitement Automatique Des Langues Naturelles. Association pour le Traitement Automatique des Langues, Caen, France, pp. 330–336.

Lewis-Beck, M.S., Dassonneville, R., 2015. Forecasting elections in Europe: Synthetic models. *Research & Politics* 2, 1–11.

Melzi, S., Abdaoui, A., Azé, J., Bringay, S., Poncelet, P., Galtier, F., 2014. Patient's rationale: Patient Knowledge retrieval from health forums, in: 6th International Conference on eHealth, Telemedicine, and Social Medicine. pp. 140–145.

Mohammad, S., 2012. Portable Features for Classifying Emotional Text, in: the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 587–591.

Mohammad, S.M., Kiritchenko, S., 2015. Using Hashtags to Capture Fine Emotion Categories from Tweets. *Computational Intelligence* 31, 301–326.

Mohammad, S.M., Turney, P.D., 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence* 29, 436–465.

Mohammad, S.M., Zhu, X., Kiritchenko, S., Martin, J., 2015. Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management* 51, 480–499.

Nakov, P., Kozareva, Z., Ritter, A., Rosenthal, S., Stoyanov, V., Wilson, T., 2013. Semeval-2013 task 2: Sentiment analysis in twitter, in: 7th International Workshop on Semantic Evaluation.

Neviarouskaya, A., Prendinger, H., Ishizuka, M., 2011. SentiFul: A lexicon for sentiment analysis. *IEEE Transactions on Affective Computing* 2, 22–36.

Och, F.J., Ney, H., 2004. The alignment template approach to statistical machine translation. Computational linguistics 30, 417–449.

Pang, B., Lee, L., 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval* 2, 1–135.

Pearl, L., Steyvers, M., 2010. Identifying emotions, intentions, and attitudes in text using a game with a purpose, in: the Naacl Hlt 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, pp. 71–79.

Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., Booth, R.J., 2007. The development and psychometric properties of LIWC2007. http://www.liwc.net/LIWC2007LanguageManual.pdf. Accessed on 16 December 2015.

Pestian, J.P., Matykiewicz, P., Linn-Gust, M., South, B., Uzuner, O., Wiebe, J., Cohen, K.B., Hurdle, J., Brew, C., 2012. Sentiment analysis of suicide notes: A shared task. *Biomedical informatics insights* 5, 3.

Platt, J.C., 1999. Advances in Kernel Methods, in: Schölkopf, B., Burges, C.J.C., Smola, A.J. (Eds.), . MIT Press, Cambridge, MA, USA, pp. 185–208.

Plutchik, R., 1980. A general psychoevolutionary theory of emotion. *Theories of emotion* 1. ISBN: 978-0-12-558701-3.

Qiu, G., Liu, B., Bu, J., Chen, C., 2009. Expanding Domain Sentiment Lexicon through Double Propagation., in: the International Joint Conference on Artificial Intelligence, pp. 1199–1204.

Rao, Y., Lei, J., Wenyin, L., Li, Q., Chen, M., 2013. Building emotional dictionary for sentiment analysis of online news. *World Wide Web* 17, 723–742.

Roberts, K., Roach, M.A., Johnson, J., Guthrie, J., Harabagiu, S.M., 2012. EmpaTweet: Annotating and Detecting Emotions on Twitter., in: the Language Resources and Evaluation Conference. pp. 3806–3813.

Rosenthal, S., Nakov, P., Kiritchenko, S., Mohammad, S., Ritter, A., Stoyanov, V., 2015. Semeval-2015 task 10: Sentiment analysis in Twitter, in: 9th International Workshop on Semantic Evaluation, pp. 451–463.

Rosenthal, S., Ritter, A., Nakov, P., Stoyanov, V., 2014. SemEval-2014 Task 9: Sentiment Analysis in Twitter, in: the 8th International Workshop on Semantic Evaluation, pp. 73–80.

Sadat, F., Yoshikawa, M., Uemura, S., 2003. Bilingual terminology acquisition from comparable corpora and phrasal translation to cross-language information retrieval, in: 41st Annual Meeting on Association for Computational Linguistics-Volume 2, pp. 141–144.

Schmid, H., 1994. Probabilistic part-of-speech tagging using decision trees, in: the International Conference on New Methods in Language Processing, pp. 44–49.

Stone, P.J., Dunphy, D.C., Smith, M.S., 1966. The General Inquirer: A Computer Approach to Content Analysis, *Journal of Regional Science* 8(1), 113–116.

Strapparava, C., Mihalcea, R., 2008. Learning to identify emotions in text, in: the ACM Symposium on Applied Computing, pp. 1556–1560.

Strapparava, C., Valitutti, A., 2004. WordNet Affect: an Affective Extension of WordNet., in: the Language Resources and Evaluation Conference, pp. 1083–1086.

Tsoumakas, G., Katakis, I., Vlahavas, I., 2009. Mining Multi-label Data, in: Maimon, O., Rokach, L. (Eds.), *Data Mining and Knowledge Discovery Handbook*, pp. 667–685.

Turney, P.D., 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews, in: 40th Annual Meeting on Association for Computational Linguistics, pp. 417–424.

Van Asch, V., 2012. Macro-and micro-averaged evaluation measures. Belgium: University of Antwerp. http://www.cnts.ua.ac.be/~vincent/pdf/microaverage.pdf. Accessed on 16 December 2015.

Velcin, J., Kim, Y., Brun, C., Dormagen, J., SanJuan, E., Khouas, L., Peradotto, A., Bonnevay, S., Roux, C., Boyadjian, J., others, 2014. Investigating the image of entities in social media: Dataset design and first results, in: 9th Language Resources and Evaluation Conference, pp. 818–822.

Wilson, T., Wiebe, J., Hoffmann, P., 2005. Recognizing contextual polarity in phrase-level sentiment analysis, in: the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347–354.

Appendices

Table 17: The repartition of Climate training and testing tweets between the 18 emotions

Emotion	Training	Testing	Ekman
Displeasure	47	21	No
Disturbance	13	6	No
Contempt	176	75	No
Negative surprise	10	4	Yes
Fear	274	114	Yes
Anger	210	87	Yes
Boredom	4	2	No
Sadness	36	16	Yes
Pleasure	35	15	No
Appeasement	9	5	No
Love	8	4	No
Positive surprise	4	2	Yes
Satisfaction	73	32	No
Dissatisfaction	9	5	No
Agreement	154	67	No
Valorization	1,504	644	No
Disagreement	216	92	No
Depreciation	401	170	No
Total	**3,183**	**1,361**	