Aspect Detection in Book Reviews: Experimentations

Jeanne Villaneau, Stefania Pecore, Farida Saïd

Aspect Detection in Book Reviews: Experimentations

Jeanne Villaneau[1], Stefania Pecore[1], and Farida Saïd[2]

[1] IRISA, Université de Bretagne Sud, France
[2] LMBA, Université de Bretagne Sud, France
{jeanne.villaneau,stefania.pecore,farida.said}@univ-ubs.fr

Abstract. Aspect Based Sentiment Analysis (ABSA) aims at identifying the aspects of entities and the sentiment expressed towards each aspect. Substantial work already exists in English language and in domains where aspects are easy to define such as restaurants, hotels, laptops, etc. This paper investigates detection of aspects in French language and in the books reviews domain where expression is more complex and aspects are less easy to characterize. On the basis of a corpus that we annotated, 21 aspects were defined and categorized into eight main classes including a catch-all class, *General*, which was found to be absorbent. Several methods were carried out to address this difficulty, with varying efficiency: Random Forest and SVM provided better results than kNN and Neural Net. Combining these methods with voting rules helped to improve noticeably the results. On another side, the difficulty of the task and the limits of a lexical approach were further explored with a qualitative analysis of errors and a topological mapping of the data using Self Organising Maps.

Keywords: Aspect Based Sentiment Analysis · aspect detection · opinion mining.

1 Introduction

Aspect Based Sentiment Analysis (ABSA) systems aim at detecting the main aspects (features) of an entity which are discussed in texts and at estimating the orientation of the sentiment expressed per aspect (how positive or negative the opinions are on each aspect) [7]. ABSA was first introduced as a shared task in SemEval-2014 [11], with datasets in English in two domains: laptops and restaurants. The task was repeated in SemEval-2015 and SemEval-2016, and extended to new entities (hotel, restaurant, telecom, consumer electronics) and to other languages (French, Dutch, Russian, Spanish and Turkish) [10].

ABSA is classically split into three subtasks: (i) extracting opinion expressions, (ii) determining the aspect of these expressions and (iii) determining their opinion value [4]. In SemEval 2016, determining the aspects was the subtask of ABSA (task 5) which called the largest number of contributions (216 over 245

submissions in total). As an example, French data sets were proposed in restaurant domain with 6 types of entities and 6 types of attributes [2]. On these data, the best system obtained a F_1 score of 0.612.

Despite challenges as SemEval, few studies were conducted in languages other than English and freely available data are scarse. We were interested in this work in investigating this task in French language and in a domain where aspects are more difficult to detect and where opinion is expressed in complex and varied forms. This paper presents a book reviews corpus which we collected and the work carried out to define aspects (Section 2) and to implement their automatic detection by using lexical statistical methods (Sections 3). It was found that these methods perform varyingly well and their performances can be improved when they are combined. Moreover, an analysis of the errors gives an idea of the difficulty of the task and the limits we have to go beyond to improve the results (Section 4).

2 Training and Test corpora - Task and Approach

2.1 Training Corpus and Annotation

We built a corpus of 900 reviews by concatenation of 450 book reviews from the French Sentiment Corpus (FSC), which was produced between 2009 and 2013 by Vincent and Winterstein (2013), and 450 more recent book reviews which we collected from the Amazon.fr website between 2016 and 2017 (NC).

The total number of words in the corpus is about 72,000 words.

We proposed an annotation schema suitable for all types of books, regardless of genre, which is based on 5 aspects and 20 attributes (see Table 1). The 21 resulting classes can be gathered into metaclasses to meet different needs.

Aspects	**Attributes**
General Feeling	-
Text	*General, Subject, Style, Characters, Pace/Narration, Readability, Translation/Adaptation,Interest/Accuracy*
Illustration	*General, Interest/Accuracy, Graphic quality*
Author	*General, Text Author, Translator, Illustration Author,*
Form	*General, Bookbinding, Typography, Inner structure, Distribution*

Table 1. Aspects and Attributes for book reviews

The complexity of the wording in book reviews makes difficult the task of allocating a unique aspect to an entity as usually done, for example in SemEval 2016 annotation task. The following examples, yet very simple, illustrate how entities, opinion phrases and context have to be taken into account to determine proper aspects.

- In the phrase *"le livre est bien mal écrit" [the book is very badly written]*, the part which expresses sentiment is *"bien mal écrit" (very badly written)* (value: -2) and the entity is *le livre [the book]*. The appropriate aspect is *Text* with *Style* for attribute, because of the verb *écrire [to write]*.
- In the review, *"la bobo au style frelaté" [the boho with degenerated style]*, the word *degenerated* expresses a very negative opinion (-2). It relates to the entity *Style* and it is classified in *Text#Style*. Because of the reference to the style, one can say that *bobo* refers to the author; *"la bobo"* represents both the entity and the opinion of the reviewer.

Since it often happens that entity and aspect do not coincide, it is essential to include an aspect detection phase in the annotation process. For that, we proceed in three steps:

- selection of a group of contiguous words which indicate an opinion (evaluated by an ordinal value),
- detection of the entity to which the opinion refers (when it is expressed),
- selection of an aspect and an attribute in the annotation schema.

The annotation task concerned about 4700 phrases related to 3300 opinion expressions. More information on the corpus (statistics, annotators, inter-annotators agreement) is given in [9].

2.2 Task, Test Corpus and Approach

Aspects were grouped into eight main classes because of the difficulty met by the annotators to separate certain aspects. More precisely, the following pairs of aspects were aggregated: *General* with *Text#General*, *Text#Readability* with *Text#Style* and *Text#Interest* with *Text#Subject*. The other considered aspects are *Text#Pace-Narration, Text#Characters*, *Illustrations, Form* and *Authors* regardless of attributes for the latter. Table 2 displays the relative importance of these classes in the training corpus. The large prevalence of the class *General* and the very limited size of the class *Illustrations* are to be mentioned.

The test corpus consists of 340 sentences or parts of text selected from the non-annotated part of the FSC corpus. The sentences were selected so as to present a unique aspect each and to cover all aspect classes, thereby reducing the prevalence of the class "*General*". The resulting distribution of the aspect classes is given in the last column of Table 2.

As mentioned above, sentences presenting more than one aspect were removed during the selection process, as in:

"Tant dans le contenu que dans l'écriture je n'ai pu trouver aucun intérêt à cet ouvrage" [Both in the contents and in the writing I was not able to find any interest in this work.]

Furthermore, it is whole sentences or their largest possible parts which were selected, as in:

Class	Aspect/Attribute	% Training	Test (Nb and %)
General (Ge)	{General Feel. - Text#General}	44.9%	91 27%
Pace (Pa)	Text#Pace-Narration	11.5%	64 19%
Interest (In)	Text#{Interest-Accur., Subject}	21.0%	53 16%
Characters (Ch)	Text#Characters	8.5%	53 16%
Style (St)	Text#{Style, Readibility}	3.2%	20 5.5%
Authors (Au)	Author#{*all_attributes*}	4.5%	20 5.5%
Illustrations (Il)	Illustration#{*all_attributes*}	0.7%	18 5%
Form (Fo)	Form #{*all_attributes*}	5.7%	21 6%

Table 2. Aspect classes and their distribution in both training and test corpora.

"Tout sonne faux, les relations entre les protagonistes, les dialogues qui semblent sortis de la bouche de mauvais acteurs, la psychologie des personnages." [Everything rings false, the relations between the protagonists, the dialogues which seem come out of the mouth of bad actors, the psychology of the characters.]

It should be noticed that some words which could seem to be key words in the determination of the target (Aspect#Attribute), can turn out to be false friends as in the previous sentence where the word *personnage [character]* can lead to misclassify the sentence in *Characters* while a human annotator would classify it in *Interest*.

Detecting opinion polarity meets several difficulties among which negation, use of humoristic or indirect expression, etc. On the other hand, the success of statistical methods based on simple bag of words (BoW) supports the hypothesis that determining aspects is essentially a lexical task. We investigated the efficiency of this approach on the corpus of book reviews.

Following lemmatization (with Treetagger), a list of lemmas (names, adjectives, verbs and adverbs excepting stop words) was selected according to their frequency in the corpus (i). Each annotaded expression in the training corpus is handled as a vector whose binary entries (0 or 1) code the co-occurrences of the expression with the lemmas (ii). A co-occurence matrix is built and then augmented with a column which specifies the *aspect#attribute* assigned to every annotated expression (iii).

Our attempts to enrich the model with linguistic parameters were not conclusive and the performances achieved were low below the results presented in the next section. Anyhow, the best results were obtained using lemmas rather than forms, possibly because of the modest size of our corpus.

3 Experiments and Results

Various experimentations were conducted using unsupervised and supervised classification approaches, namely SOM (Self-Organising Maps), kNN (k-Nearest Neighbours), NN (Neural Net), RF (Random Forest), SVM (Support Vector

Machine). Linguistic contexts of words were taken into account through the use of Word2vec. The well known language and environment for statistical computing, R, was used all along this work.

The results of our experimentations are presented below and they reflect well the difficulty of the task. In all tables, aspect classes are identified by the abbreviations given in Table 2.

3.1 SOM

Self Organising Maps is a competitive learning network based on unsupervised learning. It provides a low dimension representation of the input data and it serves for representation as well as for clustering. We used in our experimentations the *kohonen* R-package.

The topological map in Figure 1 was obtained by combining the observation-lemma matrix (weight of 5) with the vector of related aspect classes (weight of 1).

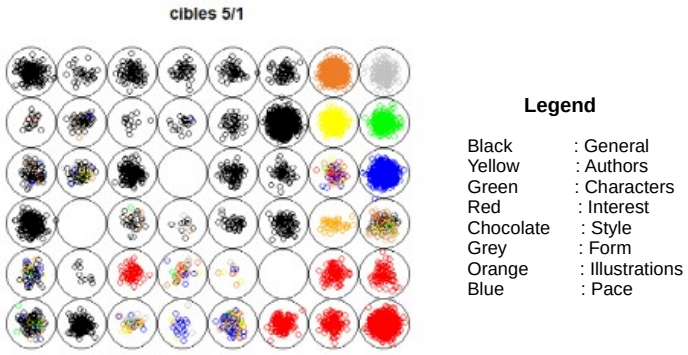

Fig. 1. SOM obtained by combining lexical data and classification into aspects in the proportions $(\frac{5}{6}, \frac{1}{6})$.

Figure 1 shows the extension of the aspect *General* (black) on more than half of the topological map. Aspects *Characters* (green), *Style* (chocolate) and *Form* (grey) appear well grouped, but to a lesser degree for the last two classes. In contrast, aspects *Illustrations* (orange) and *Authors* (yellow) are more heterogeneous with a sub-class closely grouped and a subset of data highly dispersed and mixed with other classes.

These observations have to be put into perspective with the outcomes per class shown in Table 8. If SOM gives interesting representations of the data, attempts to use it as a predictive tool (with *supersom* from kohonen R-package) gave poor results in terms of precision and recall.

3.2 k-Nearest Neighbors (kNN) - Neural Networks - Fuzzy classification

The best results with kNN are displayed in Table 3; they were obtained for $k = 2$. These performances are disappointing and reflect the difficulties encountered, especially the absorbtion capacity of the class *General*, the only class showing a precision score lower than the recall. As predicted by SOM, *Characters* is the class that obtains the best results.

	Ge	Pa	Ch	St	Au	In	Il	Fo
Ge	72	0	1	7	1	7	0	3
Pa	28	15	1	1	1	6	0	1
Ch	7	0	10	1	0	2	0	0
St	30	1	0	27	0	3	0	3
Au	14	1	0	3	2	0	0	0
In	32	3	1	3	0	14	0	0
Il	13	0	0	0	0	4	1	0
Fo	13	0	0	2	0	2	0	4

Class	Precision	Recall	F-measure
General	0.344	0.791	0.48
Pace	0.41	0.75	0.411
Characters	0.769	0.5	0.606
Style	0.614	0.422	0.5
Authors	0.5	0.1	0.167
Interest	0.368	0.264	0.308
Illustrations	1	0.056	0.105
Form	0.363	0.190	0.25

Table 3. Best result with kNN (for $k = 2$).

Fuzzy logic and Neural Networks already proved to be efficient in Sentiment analysis [1,5]. However, they provided very poor results when implemented on our data (R-package *frbs* and *neuralnet*), with almost all expressions classified in the class *General*.

3.3 Random Forest

The statistical approach using Random Forest ($ntree = 500$) gives encouraging results. The class *General* is still absorbent but all classes have their precision and recall scores greatly improved. In accordance with SOM (Figure 1), class *Characters* performs well. The results of class *Author* remain mediocre and those of class *Form* are poor, while the recall of class *Illustrations* is very low.

While names can be sufficient for the determination of aspects in certain domains, four parts of Speech are highlighted in our experimentation. Indeed, the top twenty words in Random Forest consist in 9 names, 5 adjectives, 5 verbs and 1 adverb and among them, the adjectives *interesting, clear* and *likeable* which are respectively associated with classes *Interest, Style/Readibility, Characters* and the adverb *facilement [easily]* which is associated with the class *Style/Readibility*.

3.4 SVM

In the field of ABSA, SVM classifiers made their proof for both aspect and polarity detection [6,13]. The classic approach by SVM with linear kernel outclasses

	Ge	Pa	Ch	St	Au	In	Il	Fo
Ge	77	3	1	6	0	3	0	1
Pa	16	31	1	0	0	4	0	1
Ch	0	0	20	0	0	0	0	0
St	15	1	0	42	3	3	0	0
Au	9	0	0	3	8	0	0	0
In	20	1	1	3	0	27	0	1
Il	12	0	0	0	0	0	6	0
Fo	14	2	0	1	0	0	0	4

Class	Precision	Recall	F-measure
General	0.472	0.846	0.606
Pace	0.816	0.585	0.681
Characters	0.870	1	0.930
Style	0.764	0.656	0.706
Authors	0.727	0.4	0.516
Interest	0.730	0.509	0.6
Illustrations	1	0.333	0.5
Form	0.571	0.190	0.286

Table 4. Results with Random Forest ($ntree = 500$).

Random Forests, however the improvement is not general: classes *Pace, Characters, Interest* obtain poorer results. By contrast, the improvement of the results of the class *Form* is particularly remarkable.

Besides, we still observe the trend to an overuse of the class *General*.

	Ge	Pa	Ch	St	Au	In	Il	Fo
Ge	75	3	2	6	1	2	0	2
Pa	20	27	1	2	2	1	0	0
Ch	0	0	19	0	1	0	0	0
St	13	1	0	45	3	2	0	0
Au	9	0	0	2	7	2	0	0
In	25	1	0	1	2	24	0	1
Il	8	0	0	1	0	1	8	0
Fo	5	2	0	0	0	0	0	14

Class	Precision	Recall	F-measure
General	0.484	0.824	0.610
Pace	0.794	0.509	0.621
Characters	0.864	0.95	0.905
Style	0.789	0.703	0.744
Authors	0.467	0.35	0.4
Interest	0.75	0.453	0.565
Illustrations	1	0.444	0.615
Form	0.824	0.667	0.737

Table 5. SVM Results (linear kernel).

3.5 SVM+Word2Vec (SVMW2V)

Many of the words used in the test corpus do not appear in the training corpus because of its small size. To deal with this difficulty, the last approach makes use of Word2Vec to enrich the space of words in the test corpus. Word2Vec was trained with the corpora FSC, NC and Wikipedia.

The training corpus remains unchanged and only the co-occurrence matrix is modified: the entry in the co-occurence matrix of every name, adjective, verb or adverb that does not appear in the training corpus, is replaced by its similarity score with its closest lemma in the training corpus.

The results were globally below expectations (except for class *General*) (cf. Table 6); a reason for that could be that the noise brought by W2V limited the global gain.

	Ge	Pa	Ch	St	Au	In	Il	Fo
Ge	74	2	2	5	1	4	0	3
Pa	20	26	1	2	2	2	0	0
Ch	0	0	17	0	3	0	0	0
St	10	4	0	43	4	2	0	1
Au	8	0	0	2	7	3	0	0
In	23	3	1	0	1	22	0	1
Il	9	0	0	0	0	1	8	0
Fo	6	2	0	0	0	0	0	13

Class	Precision	Recall	F-measure
General	0.493	0.813	0.614
Pace	0.703	0.491	0.578
Characters	0.810	0.85	0.829
Style	0.827	0.672	0.741
Authors	0.389	0.35	0.368
Interest	0.647	0.431	0.518
Illustrations	1	0.444	0.615
Form	0.722	0.619	0.667

Table 6. SVMW2V Results.

3.6 Combining approaches

The last three approaches (Random Forest, SVM, SVMW2V) obtain encouraging results and their performances are globally close. We combined them by adopting a majority vote with a special handling of the class *General*. The voting rules are presented in Table 7

The second rule states that if at least one system out of three chooses a class other than *General*, this class is favoured. The underlying purpose of this rule is to reduce the absorbing bias of class *General* which was observed in all single systems.

The third rule specifies that in case of total disagreement between the three systems, class *General* is chosen. This rule aims to avoid a random draw when there is no well defined class.

Results	**Choice**
(1) Three equal results: $r_1 = r_2 = r_3 = C$	C
(2) Two equal results (example: $r_1 = r_2 = C$, $r_3 = C'$, $C' \neq C$)	
if $C \neq$ *General*	C
if $C =$ *General*	C'
(3) Three distinct results	*General*

Table 7. The rules of choice.

The outcomes of the combined system are given in Table 8. In global or on average across all classes, we notice that combining the three approaches leads to a slight reduction in the precision compared with SVM, which is widely compensated with an increase in the recall.

Table 9 gives the results of the combined system by class. Before combination, Random Forest outperformed the 2 other systems in 4 of the 8 classes, SVM in 3 classes and SVMW2V in the class General. Random Forest outperforms the combined system in 3 classes and SVM in the class Form. Seen from this perspective, no system outclasses totally the others.

	Ge	Pa	Ch	St	Au	In	Il	Fo
Ge	71	3	2	6	1	5	0	3
Pa	13	31	1	2	2	3	0	1
Ch	0	0	19	0	1	0	0	0
St	7	2	0	49	3	3	0	0
Au	6	0	0	2	9	3	0	0
In	14	2	0	2	1	31	0	1
Il	6	0	0	1	0	1	10	0
Fo	5	2	0	0	0	0	0	14

Class	Precision	Recall	F-measure
General	0.582	0.780	0.667
Pace	0.775	0.585	0.667
Characters	0.864	0.95	0.905
Style	0.790	0.766	0.778
Authors	0.529	0.45	0.486
Interest	0.674	0.608	0.639
Illustrations	1	0.556	0.714
Form	0.737	0.667	0.7

Table 8. Results of the combined system.

Class	RF	SVM	SVMW2V	Final
General	0.606	0.610	0.614	**0.667**
Pace	**0.681**	0.621	0.578	0.667
Characters	**0.930**	0.905	0.829	0.905
Style	0.706	0.744	0.741	**0.778**
Authors	**0.516**	0.4	0.368	0.486
Interest	0.6	0.565	0.518	**0.639**
Illustrations	0.5	0.615	0.615	**0.714**
Form	0.286	**0.737**	0.667	0.7

	System	Prec.	Recall	F1
Macro	RF	0.744	0.565	0.642
	SVM	**0.747**	0.613	0.673
	SVMW2V	0.699	0.584	0.636
	Final	0.744	**0.670**	**0.705**
Micro	RF	0.847	0.554	0.670
	SVM	**0.852**	0.578	0.689
	SVMW2V	0.795	0.551	0.651
	Final	0.832	**0.660**	**0.736**

Table 9. F_1-measure per aspect and system (left table) and Precision, Recall and F_1-measure per system (right table). Final corresponds to the combined system.

4 BoW approach: efficiency and limits

In 78 out of 340 tests, none of the statistical systems selected the same aspect as the human annotators. A qualitative analysis of the disagreements allows to go deeper in the understanding of the limits of the lexical approach.

Disagreements can be classified into three classes:

1. There are 19 *"false errors"* for which the human annotation may be questioned. For example :
   (a) *"C'est drôle et enlevé, puissant et sensuel?: un chef-d'oeuvre de vie, dédié à la vie d'une ville incomparable." [It is funny and spirited, powerful and sensual?: a masterpiece of life, dedicated to the life of an incomparable city.]*
   This test is classified as *General* by human annotators and as *Style* by the three statistical systems. Both choices are justified: the first choice is understandable if we consider the whole sentence and the second choice is essentially motivated by the first part of the sentence. This example shows the limits of a strict classification since classes are not necessarily mutually exclusive.

(b) *"Attention: livre impossible à lâcher avant la dernière page?!" [Attention: book impossible to put down before the last page?!]*
BoW systems classified the sentence as *Pace* (SVMW2V) or as *Interest* (SVM and RF), while human annotators chose the class *General*, possibly because the choice is unclear.

The significant number of *false errors* points out the difficulty of the task in the field of book reviews and the fuzzy outlines between the defined classes.

2. Another group of errors (about 12) can be related to training bias due to new words appearing in the tests. For example:
   (a) *"Pouchkine est un écrivain au style sûr, simple et envoûtant" [Pushkin is a writer with a sure, simple and mesmerizing style]*
   This sentence should be classified as *Author* since it expresses a general opinion on Puchkin's style. However, all systems classified it as *Style* because "Puchkin" did not appear in the training corpus. It is likely that a list of authors' names would improve the results of the class *Author*.
   (b) *"Un très joli livre, avec de très belles peintures chinoises à l'intérieur." [A very attractive book, with very beautiful Chinese paintings inside.]*
   This sentence related to *Illustrations* is miclassified as *General* by the systems. This error can be explained by the low occurrence of the keyword *"peinture" [painting]* in the training corpus.
   One would hope that using Word2Vec would allow to go beyond the limits of training corpus' vocabulary by extending it. However, in our experiments, the noise introduced by the similarity scores negated the expected improvement.
3. Lastly, the vast majority of errors is related to the limits of BoW approach. Firstly, representing a sentence as a bag of lemmas is very simplistic; on the other hand, the understanding of the reader uses contexts of various types: temporal, cultural, pragmatic, textual, of common sense, etc. [3].
   (a) *"il manque l'essentiel, les bonnes adresses, les accès, les plages, bref, aucun détail, c'est un TOP 10 sans le moindre intérêt." [The main part, the good addresses, the accesses, the beaches, in brief, no detail is missing, it is a PIP 10 without the slightest interest.]*
   Here, the aspect is expressed in the word *"Interest"* and yet, the test is classified by the systems as *General*, probably because the word is buried in many others, as *"essential"*. It can be assumed that linguistic context, especially the adverb *bref [in short]* which introduces a conclusion, could make it possible to give more importance to this keyword.
   (b) *"Je n'ai pas accroché à l'histoire, il convient sûrement à toutes les petites et jeunes filles ans des poupées, mais la trame est cousue de fils blancs*[3] *[I did not stick to the story, it is certainly advisable to all the girls and the girls the years of dolls, but the framework is a blatant lie...]*
   The test is classified as *General* (instead of *Pace*) by all the systems despite the keyword *"histoire" ["story"]*. The word *trame [framework]*,

[3] In French language, there is a play on words between *trame*, which means "weft" or "framework" depending on the context, and phrase *cousue de fils blancs* litteraly "sewn of white threads".

almost synonymic but much less common, was probably not taken into account by the systems, including W2V.

(c) *"le livre reste un catalogue d'interprétations déjà connues." [the book remains a catalog of already known interpretations]*
BoW systems classify this test in *General* instead of *Interest*. The adverb *"déjà"* plays a key role to show the lack of interest of the book.

(d) *"L'auteur abuse de mots aussi savants qu'inutiles qui détournent du sujet traité?; un défaut difficilement pardonnable." [The author makes excessive use of words as fancy as they are useless which divert from the handled subject?; a fault hardly overlooked]*
The word *"auteur"* makes the test classified in class *Authors*, while the sentence relates to the style of the book and not to its author in general.

(e) *"Sauter de la page 288 à la 337 n'aide pas du tout à apprécier un roman, notamment si celui-ci doit tre le dernier d'une série." [Jumping from page 288 to 337 does not help at all to appreciate a novel, especially if it is the last of a series...]*
"l'auteur oublie ici et là des mots qui AIDE à comprendre les phrase." [the author forgets here and there words which help to understand the sentences]
Both tests express negative sentiments with a certain sense of humour (irony or sarcasm), which is a real challenge for automatic systems. For instance, a specific session of SemEval was devoted to sarcastic tweets [8] and numerous works addressed this topic (see for example [3]) .

Actual mistakes point clearly toward the need to take into account multiple contexts and knowledges to improve systems, as emphasized by Benamara and Co [2017]. Within our study, the most relevant aspects relate to the choice of wording and its structure, and to take into account the linguistic context of the words : expressions varyingly litteral (*"cousu de fil blanc" [blindingly obvious]*), linkage of adverbs and qualificatives... not to mention the detection of irony, a full study program in itself.

5 Conclusion

In a complex field where aspects are sometimes hard to sort out, even for a human annotator, a simple SVM approach with a linear kernel on words (lemmas in this instance) is, despite its lackings, relatively efficient. Regardless, the combination with other statistical approaches, especially with Random Forest, noticeably improves the attained results. Furthermore, an intake in lexical resources, like the list of authors, could help to better circumvent some classes.

Despite this, the analysis of errors brings to light the limits of the BoW approach. An improvement of the results inevitably requires a better analysis of contexts with the problems that come with the use of a language all-in-all lacking in normalization on one hand and, on the other hand in French language which proves much poorer in resources than English language.

At present, our research concerns polarity determination. Besides BoW approaches, we also take into account the linguistic context by implementing a surface analysis. First results seem to evidence that the use of linguistic parameters can allow to outclass widely a simple BoW approach in this task.

References

1. Afzaal, M., Usman, M., Fong, A.C.M., Fong, S., Zhuang, Y.: Fuzzy aspect based opinion classification system for mining tourist reviews. Advances in Fuzzy Systems **2016** (2016)
2. Apidianaki, M., Tannier, X., Richart, C.: Datasets for aspect-based sentiment analysis in french. In: Proceedings of LREC 2016. European Language Resources Association (ELRA), Paris, France (may 2016)
3. Benamara, F., Taboada, M., Mathieu, Y.: Evaluative language beyond bags of words: Linguistic insights and computational applications. Comput. Linguist. **43**(1), 201–264 (Apr 2017)
4. De Clercq, O., Lefever, E., Jacobs, G., Carpels, T., Hoste, V.: Towards an integrated pipeline for aspect-based sentiment analysis in various domains. In: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis - WASSA 2017. pp. 136–142. EMNLP 2017 (2017)
5. Khalil, T., El-Beltagy, S.R.: Niletmrg at semeval-2016 task 5: Deep convolutional neural networks for aspect category and sentiment extraction. In: Proceedings of SemEval-2016. ACL, San Diego, California (June 2016)
6. Kiritchenko, S., Zhu, X., Cherry, C., Mohammad, S.: Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In: Proceedings of SemEval 2014. ACL and Dublin City University, Dublin, Ireland (August 2014)
7. Liu, B.: Sentiment Analysis and Opinion Mining. Morgan and Claypool Publishers (2012)
8. Nakov, P., Rosenthal, S., Kiritchenko, S., Mohammad, S.M., Kozareva, Z., Ritter, A., Stoyanov, V., Zhu, X.: Developing a successful semeval task in sentiment analysis of twitter and other social media texts. Lang. Resour. Eval. **50**(1), 35–65 (Mar 2016)
9. Pecore, S., Villaneau, J.: Complex and precise movie and book annotations in french language for aspect based sentiment analysis. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018. (2018)
10. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al.: Semeval-2016 task 5: Aspect based sentiment analysis. In: ProWorkshop on Semantic Evaluation (SemEval-2016). pp. 19–30. Association for Computational Linguistics (2016)
11. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: Semeval-2014 task 4: Aspect based sentiment analysis. In: Proceedings of SemEval 2014. Dublin, Ireland (August 2014)
12. Vincent, M., Winterstein, G.: Construction et exploitation d'un corpus français pour l'analyse de sentiment. In: TALN-RÉCITAL 2013. Les Sables d'Olonne, France (2013)
13. Wagner, J., Arora, P., Cortes, S., Barman, U., Bogdanova, D., Foster, J., Tounsi, L.: Dcu: Aspect-based polarity classification for semeval task 4. In: Proceedings of SemEval 2014. pp. 223–229. Association for Computational Linguistics and Dublin City University, Dublin, Ireland (August 2014)