Complex and Precise Movie and Book Annotations in French Language for Aspect Based Sentiment Analysis

Stefania Pecore, Jeanne Villaneau

Complex and Precise Movie and Book Annotations in French Language for Aspect Based Sentiment Analysis

Stefania Pecore, Jeanne Villaneau

IRISA-D6 Media et Interactions

Université de Bretagne Sud, France

{stefania.pecore,jeanne.villaneau}@univ-ubs.fr

Abstract

Aspect Based Sentiment Analysis (ABSA) aims at collecting detailed opinion information according to products and their features, via the recognition of targets of the opinions in text. Though some annotated data have been produced in challenges as SemEval, resources are still scarce, especially for languages other than English. We are interested in enhancing today's mostly statistical text classification with the use of linguistics tools, in order to better define and analyze what has been written. The work presented in this paper focuses on two French datasets of movies and books online reviews. In reviews, text length is much higher compared to a tweet, giving us the opportunity to work on a challenging and linguistically interesting dataset. Moreover, movies and books are products that make classifying opinions into aspects quite complex. This article provides an analysis of the particularities of the two domains during the process of collecting and annotating data, a precise annotation scheme for each domain, examples and statistics issued from the annotation phase, and some perspectives on our future work.

1. Introduction

The Web makes it relatively easy to collect a large amount of textual data. Researchers, as well as enterprises, politicians and governments are very interested in analyzing them, especially to learn more about people's opinions. This is due to many reasons: to improve services and products, to know the public opinion, to avoid terrorism acts. For researchers, opinion analysis is a multifaceted problem, with different names according to the studied aspect: subjectivity analysis (Wiebe et al., 2004), opinion mining (Pang and Lee, 2008), sentiment extraction (Das and Chen, 2007). While the subject is mature, as proved by the many published surveys (Pang and Lee, 2008; Liu, 2012), there is still room for improvement, demonstrated by the interest for the yearly NLP conferences and workshops as SemEval or Wassa(Mohammad and Bravo-Marquez, 2017) and by the challenges that opinion analysis still offers (Breck and Cardie, 2016; Mohammad, 2016).

Aspect Based Sentiment Analysis (ABSA) aims at "determining the orientation of sentiment expressed on each aspect" (Liu, 2012). ABSA was introduced as a shared task for the first time in SemEval-2014 (Pontiki et al., 2014), with datasets in English language for two domains: laptops and restaurants. In SemEval-2015, the task was repeated and extended, adding the hotel domain with a dataset of whole reviews and not just isolated sentences (Pontiki et al., 2015). In SemEval-2016 (Pontiki et al., 2016), new and multilingual datasets were provided: restaurant reviews in six languages (English, French, Dutch, Russian, Spanish and Turkish), hotel reviews in Arabic, consumer electronics reviews in three languages (English, Dutch and Chinese), telecom reviews in Turkish and museum reviews in French.

In this paper we propose an annotated French dataset for different domains compared with SemEval ones: movies and books. We have chosen to create a corpus of reviews and not tweets because we are interested in combining statistical and linguistical approaches, therefore we need to work on texts longer than tweets. As far as we know, there are few works in French for these domains (Hamdan et al., 2016), as a consequence it is not so easy to find this kind of data. Using ABSA is a difficult task on such reviews, because opinion is expressed in complex and various forms. Unlike other kind of reviews, limited in total amount of usable characters, these reviews are non-predictable in terms of length, and may carry opinions about other products related to the reviewed one, which are used as comparison. They can also merge in a same paragraph user's opinion and description of the evaluated product.

The paper describes our corpus (see Section 2.) and the specific form of annotation we have chosen in order to cope with comparisons and language complexity (Section 3.). Finally, we briefly explain how our future work will exploit these annotated data (Section 4.).

2. Corpus

2.1. Collecting text reviews

In our first experiments, the corpus was only composed by a collection of 450 books reviews and 450 movies reviews from the French Sentiment Corpus (FSC) produced between 2009-2013 by Vincent and Winterstein (2013). According to that paper, FSC was not originally used for Aspect-Based Sentiment Analysis, but to study opinion polarities. We decided to extend the corpus by adding more recent reviews - the NC Corpus, created between 2016 and 2017. The final corpus is the result of the union of these two corpora, both with equal dimensions. In total it is composed by 1800 reviews, i.e 4113 sentences for the books domain and 5222 for the movies domain. The corpus comes from two websites: Amazon.fr for books, and Allocine.fr for movies. By taking into account word and line counts of similar projects, it is possible to state that the corpus used in this project is suitable for being used in the context of Aspect-Based Sentiment Analysis tasks. Compared to other corpora in the field, the one used in this project has

a higher-than-average number of sentences. Some examples are (Hamdan et al., 2016) with 200 books reviews in French, (Alvarez-Lopez et al., 2017) with 2977 sentences from English books reviews, (Thet et al., 2010) with 1000 sentences from English movies reviews, (Sorgente et al., 2014) with 2648 sentences from movies reviews in Italian. The corpus deals with books and movies using three types of rating: 1 for extremely negative reviews, 3 for the moderate ones, and 5 for the extremely positive ones. We decided to include 3 stars rating reviews because we noticed that, in such reviews, reviewers do not express a strong opinion: they are therefore induced to justify their balanced opinion to precise what they consider to be the negative and positive aspects of the book or movie.

The corpus provides the title of the reviewed product, to detect if the movie or book analyzed is the same of the title, or if it is another one used as term of comparison.

While making NC corpus, we noticed that collecting negative books reviews was not so easy, because there was an overwhelmingly higher amount of positive reviews compared with the negative ones. We suppose that, unlike people reviewing movies, people who don't like the book that they are reading prefer to close it and not to write a review - unless they have read others author's works and they want to criticize the author's choices compared to the previous ones.

Concerning the linguistics point of view we noticed that, more than in movies reviews, the book genre constitutes a linguistics sub-domain of books, i.e. user's style changes based on the genre of the book that is reviewed. For this reason we chose, when possible, to collect reviews of different types of books, in order to have a more complete annotation. In effect we noticed two phenomena: books' reviewers seem to make fewer misspellings than movies' reviewers; in addition to this, it seems that there is a difference between the diversity of words used in movies and books reviews, even when they share the same topic. The difference is to be found in the fact that, in movies reviews, users do an evaluation by just labeling the movies as good or bad (following a description of the several aspects of the product). In books reviews, instead, users try also to develop a critique that uses a variety of context-specific words.

In Table 1 a comparison between a movie review and a book review sharing the same topic: the consequences of immigration in France and the presence of Islam.

We can notice that the movie review shows several **misspellings** (*déçut* [for déçu - disappointed], *ou* [for où - where], *vraimant — *vraiament* [for vraiment - really], *kritic* [for critique - critics], *parceuq'lle* [for parce qu'elle - because she]), **slang** (*kiffé* [love],*Pourkoi* [for Pourquoi - why],*je vois pas* [for je ne vois pas - I don't see (the reasons why...)], *ke* [for que - what]) and a **general vocabulary** to evaluate a movie and not specially designed for the topic (*un beau film, j'ai aimé les images magnifiques, les décors somptueux (...) y avait la musique vraimant belle* [a good film, I loved the wonderful images, the gorgeous sets (...) the music was really good]). On the other hand, in books reviews sometimes it is harder to find misspellings. Moreover, it is not just an evaluation about the strengths and weaknesses of the product (*L'habilité de*

l'écrivain nous permet de voir la société dans laquelle nous vivons [*the writer's talent* lets us perceive the society in which we live]). The reviewer, in fact, develops a critique about the topic of the book using more context-related expressions (*excès individualistes* [individualistic excess], *facilité par lâcheté*[facilitated by cowardice], *remords historiques*[historical remorse], *repentances incessantes* [continuous repentance], *être chassé de ses terres* [to be chased out of his lands], *politique d'immigration* [immigration policy], *pamphlet de la haine* [pamphlet of hatred], *paix sociale* [social peace], *viols, vols, trafics et agressions* [rape, robbery, traffic and assault])

Movies	Ben moi j'ai vu le film aujourd'hui à 16h à Torcy (...) Eh ben j'ai pas été ***déçut** (...) C'est juste un beau film, j'ai aimé les images magnifiques, les décors somptueux, rien que pour ça déjà j'ai trop **kiffé**. (...) ***Pourkoi** cette hostilité alors qu'il est bien filmé, ***je vois pas ou** il s'est senti trahi??? bon voilà, c'est que mon avis, pas une ***kritic** mytho. Ey puis y avait la musique ***vraimant** belle. j'ai passé un bon moment et faut que j'y retourne avec ma mère ***parceuq'lle** voulait pas y aller a cause des polémiques. Moi je trouve que ça parle ***vraiament** de ce ***ke** se passe dans les quartiers.
Books	L'habilité de l'écrivain nous permet de voir la société dans laquelle nous vivons, avec ses **excès individualistes** (...). N'importe quel peuple qui baisse les bras et cède à la **facilité par lâcheté** (...), **remords historiques** et **repentances incessantes** finit par **être chassé de ses terres**. Cela n'a rien de surprenant en soi (...) seulement il y a une sorte de couvercle posé sur l'Islam en France, parce qu'il est le résultat d'une **politique d'immigration** (...). Ce livre n'est pas un **pamphlet de la haine** (...) Ceux qui y voient une incitation à la haine sont les mêmes qui, au nom de la **paix sociale**, acceptent les milliers de **viols**, **vols**, **trafics** et **agressions** que subissent depuis trop longtemps les Français(es) honnêtes.

Table 1: Two examples of movies and books reviews showing some differences concerning writing style and misspellings.

2.2. Corpus statistics

Each part of the corpora (divided per rating) contains 150 reviews. Table 2 gives the number of words for each of them. The total number of words is 169,333 distributed over 9335 sentences.

Rating	Rat. 1	Rat. 3	Rat. 5	Years
FSC movie	11621	10682	11319	2009-13
NC movie	19974	31668	11339	2016-17
FSC book	14498	11259	8816	2009-13
NC book	14847	9732	13578	2016-17
Total	60940	63341	45052	

Table 2: Number of words of the corpora.

Statistical data (see Table 3) show that the number of words of the reviews is very variable (some reviews have more than 2000 words). The great σ indicates a great dispersion of the values and the difference between *mean* and *median* is the clue of an asymetrical distribution.

	Mean	σ	median	min	max
Movies	103.02	139.64	60	1	2094
Books	81.20	118.74	48	12	2052

Table 3: Statistic data related to the number of words of the reviews.

The histogram showing the number of words of movies and books reviews (see Figure 1) indicate that most of the reviews are composed by less than 200 words. A box-plot reveals that, for the movies domain, reviews with more than 254 words are to be considered outliers, and the average amount of words per review is 127.5. On the other hand, for the books domain, reviews with more than 181 words are to be considered outliers, and the average amount of words per review is 96.5.
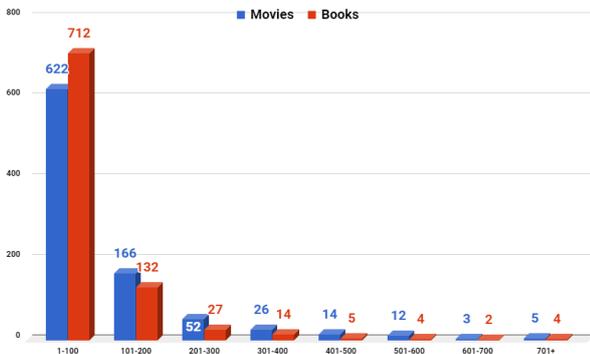

Figure 1: Number of words of reviews: histograms.

Another difference between the two parts of the corpus is that, even if movies reviews are generally longer than books reviews, they are less dense. In other words, we noticed that book reviewers are more creative in language expression than movie reviewers, as it is possible to see from the table 4 that shows the ratio between the total number of words (*token*) and the unique words (*types*) on the different parts of the corpus divided per type and rating. Movies' users tend to write longer reviews when they are giving a three-star review to a product (M3 has 43289 tokens, compared with negative reviews - 31595 tokens, and positive - 22658 tokens).

FSC corpus, despite the domain, shows a number of tokens lower than the new one (exception made by the FSC B3 reviews). We suppose it is due to the increasing comfort for the user in being critical compared with the past. In fact, it is common to find very long reviews reposted from blog articles.

Movies	Token	Type	Ratio
m1	31595	6624	20.9%
m3	43289	8244	19%
m5	22658	5443	24%
Books	Token	Type	Ratio
b1	29345	7492	**25.5%**
b3	21002	4916	**23.4%**
b5	22394	5895	**26.3%**

Table 4: Density of the text: differences between corpora and domain.

3. Annotation

The objective of the annotation is to provide a both generic and precise tool, which makes it possible to:

1. study (or automatically learn) the vocabulary used to express a negative or positive opinion, generally or in a given aspect;

2. know indications of the aspect related to a given positive or negative sentiment.

At the same time, we wanted the annotation to offer a generic aspect classification, which can be applied to various kinds of books or movies.

3.1. Annotation scheme and guidelines

We tried to create annotation schemes suitable for all type of movies or books, regardless of genre. We did not provide an annotation scheme specially designed for e-books. This is due to two reasons: first, we did not find many big differences between a book review and an e-book review; second our corpus does not have e-books reviews. The annotation scheme is composed by aspects and attributes (see Table 5 and 6). The annotation scheme for the book domain is composed by 5 aspects and 19 attributes. The movie one is composed by 7 aspects and 28 attributes. These classifications may be viewed as very precise, but grouping classes is easy, depending on the expected use of the corpus.

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

Table 5: Aspects and Attributes for books reviews.

Aspects	Attributes
General Feeling	_
Direction	*General, Director, Point of view, Direction of actors, Shooting, Sound recording*
Acting	*General, Actor, Stuntman*
Visual	*General, Sets, Costumes, Special FX*
Sound	*General, Music, Sound Effect, Songwriter*
Script	*General, Subject, Plot, Dialogues, Characters, Pace/Narration, Remake/Adaptation/Reboot*
Distribution	*General, Type of Data Storage, Original Version/French Version*

Table 6: Aspects and Attributes for movies reviews.

3.2. Opinion annotation

Each opinion expression is annotated in three steps.

1. The first step is to select a group of contiguous words that indicate a positive or negative opinion. Opinion is evaluated by an ordinal value: -1 or -2 for a negative sentiment, according to its intensity; 1 or 2 if the sentiment is positive.

2. The second step is to detect the entity to which the opinion has to be reported. This entity is not always expressed, especially if it is the movie or book that is evaluated. When it is expressed, it is most of the time a name or a nominal group. Since including coreference resolution is beyond the subject of this work, pronouns are not selected as entities. Whenever opinion expression refers to a pronoun, the entity is reported of its previous closest reference. If the entity is detected, a relation is created, which joins opinion expression with entity phrase.

3. In the third step, an aspect and an attribute are chosen in the annotation scheme.

The many forms of opinion expressivity are related to various relations between entity, opinion word and aspect, as pointed out by the following examples.

- In the very simple phrase *"Un bon film" [a good movie]*, *"bon" [good]* indicates a positive sentiment (value: 1), the entity is *film [movie]* and the aspect *General Feeling* is chosen because of the entity.

- In *"c'est un navet" [it's a rubbishy movie]* the word *"navet" [rubbishy movie]* indicates a very negative sentiment (value: -2) and refers to the entity at the same time. So, the aspect, given by the entity, is *General Feeling*.

- In the phrase *"Le style est très agréable" [the style is very pleasant]*, extracted from the book corpus, the phrase *"very pleasant"* indicates a very positive sentiment (value: 2). The entity is *"style"*. It indicates that the category is *Text* with *Style* attribute.

- In the same corpus, the phrase *"le livre est bien mal écrit" [the book is very badly written]*, the phrase related to the sentiment is *"bien mal écrit" [very badly written]* (value: -2), the entity is *"livre" [book]*. The aspect is *Text* with *Style* attribute, because of the verb *écrire [to write]*.

- In a very negative book review, *"la bobo au style frelaté" [the boo-boo with degenerated style]*, the word *degenerated* refers to a very negative opinion (-2). It can be reported to the entity *Style* and classified in *Text#Style*). Because of the reference to the style, one can say that *bobo* refers to the author; like in *"un navet"*, *"la bobo"* expresses in a single word the entity and the opinion of the reviewer.

Previous examples, though being very simple, show how entities, opinion phrases and context have to be combined, to determine the aspect to which they have to be reported. The complexity of expression in the corpus makes it difficult to allocate aspects only to entities, as it is classically done, for example in SemEval 2016 annotation (Apidianaki et al., 2016).

3.3. Entities related to other products

Some phrases indicate a positive or negative sentiment related to another book or movie, most of the time to be compared with the reviewed one. For example, in
"rien à voir avec le seigneur des anneaux, carrément passionnant" [(this film has) nothing to do with the Lord of the rings, (that is) downright fascinating], (1)
the phrase *"downright fascinating"* indicates a very positive opinion, but it is applied to another movie: *"the Lord of the rings"*. On the contrary, the full phrase indicates a negative feeling about the movie.

Such comparisons are frequent and can be a problem for an automatic opinion detection; that is why we wanted the possibility for the annotation to report them precisely. To cope with the problem, the annotation of the entities indicates whether they are or not related to the evaluated product, a product of the same series, another product, etc. So, in the previous example (1), the very positive phrase *"downright fascinating"* is reported to *"the Lord of the rings"* classified as an entity which refers to *another product*. The phrase *"(this film has) nothing to do with the Lord of the rings"* is annotated as a negative opinion, reported to an entity which refers to the evaluated product.

3.4. Annotation Process

The annotation has been done by two experts: a native speaker and a non-native speaker. After the choice of the annotation form and the redaction of the guidelines, experiments have been conducted to estimate inter-annotator agreement. The most difficult task was the selection of the phrases related to an opinion, with particular attention to the determination of their scope. As per word selection, Cohen's κ was equal to 0.71, an acceptable result given the difficulty and subjectivity of the task. However, to improve the reliability of the corpus, we decided to perform a cross-reading of the annotations between the two annotators.

The annotation was performed via Glozz software (Wid-locher and Mathet, 2012). Glozz is a multi-purpose text annotation tool, which comes with a fully WYSIWYG interface. It makes it possible to create units, defined as contiguous span of texts and relations between them. Annotations may be exported in several file formats and especially as *SQL data*.

3.5. Annotation Results

We annotated 5001 opinion phrases on movies (M1, M3, M5) and 3274 on books (B1, B3, B5). Annotations on negative reviews outnumber annotations on positive reviews, with circa 1992 annotations on M5 and B5 corpora against around 2899 on M1 and B1 corpora.

Figure 3 shows how annotations are distributed between the main classes. Nearly half of annotations are classified as *General_Feeling* in both corpora.
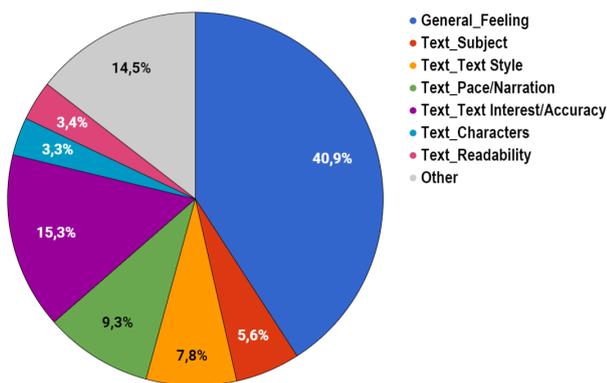

Figure 2: Books reviews: annotations collected on specific aspects.

In the book corpus, the most important six classes related to specific aspects are *Text#Interest and Accuracy*, *Text#Pace-Narration*, *Text#Style*, *Text#Characters*, *Text#Readability* and *Text#Subject*. All of them are related to the aspect *Text* and they collect 44.7% of the annotations not classified as *General_Feeling*. In the very wide variety of the assessed books, the textual aspect represents therefore a very large majority, with great importance given to the interest in the content.

In the movie corpus, the most important five classes related to specific aspects are *Acting#General*, *Script#Plot*, *Direction#General*, *Script#Pace/Narration* and *Script#General*. They collect 47.6% of the annotations not classified as *General_Feeling*. Apart from *Distribution*, all the aspects collect a significant number of annotations: the cinema is a multi-modal media, which combines sound and images to tell a story played by actors.
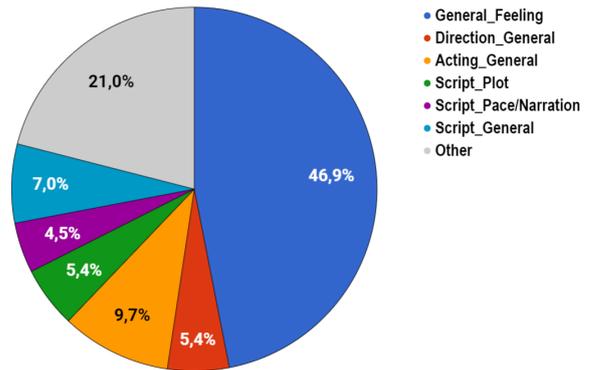

Figure 3: Movies reviews: annotations collected on specific aspects.

6392 entities have been selected; 5693 of them are related to the book or movie on which the review was written; this means that around 5% of the opinion expressions are related to another book or movie: a little more than 3.8% for the movie corpus and slightly less than 6.9% for the book corpus. Since most of them are very different from the general opinion of the reviewed book or movie, it is interesting to detect them.

	No. Relations	No. Aspects	No. Entities
Movies	5001	6235	3903
Books	3274	4787	2489

Table 7: Number of relations, aspects and entities annotated.

4. Perspectives and conclusion

The goal of the annotation is to give to the research community a dataset composed of annotated French opinion expressions about movies and books. The dataset has been created using the Aspect-Based Sentiment Analysis method, which is, based on our experience, the best fitted method to analyze these domains. Our future work will use our annotations on a system merging linguistics analysis and statistics. We will use a shallow parser and an opinion lexicon that takes advantage of the dataset to track complex expressions.

5. Bibliographical References

Alvarez-Lopez, T., Fernandez-Gavilanes, M., Costa-Montenegro, E., Juncal-Martiez, J., Garcia-Mendez, S., and Bellot, P. (2017). A book reviews dataset for aspect based sentiment analysis. In *8th Language Technology Conference*.

Apidianaki, M., Tannier, X., and Richart, C. (2016). Datasets for aspect-based sentiment analysis in french. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may. European Language Resources Association (ELRA).

Breck, E. and Cardie, C., (2016). *Oxford Handbook of Computational Linguistics.*, chapter Opinion mining and sentiment analysis. Oxford University press, 2nd edition. *to appear*.

Das, S. R. and Chen, M. Y. (2007). Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9):1375–1388.

Hamdan, H., Bellot, P., and Bechet, F. (2016). Sentiment analysis in scholarly book reviews. In *arXiv preprint arXiv:1610.03106*.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan and Claypool Publishers.

Mohammad, S. M. and Bravo-Marquez, F. (2017). WASSA-2017 shared task on emotion intensity. In *Proceedings of the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA)*, Copenhagen, Denmark.

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In Herb Meiselman, editor, *Emotion Measurement*. Elsevier.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., and Manandhar, S. (2014). Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., and Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado, June. Association for Computational Linguistics.

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. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *ProWorkshop on Semantic Evaluation (SemEval-2016)*, pages 19–30. Association for Computational Linguistics.

Sorgente, A., Vettigli, G., and Mele, F. (2014). An italian corpus for aspect based sentiment analysis of movie reviews. In *CLICIT*.

Thet, T. T., Na, J.-C., and Khoo, C. S. (2010). Aspect-based sentiment analysis of movie reviews on discussion boards. *J. Inf. Sci.*, 36(6):823–848, December.

Vincent, M. and Winterstein, G. (2013). Construction et exploitation d'un corpus français pour l'analyse de sentiment. In *TALN-RÉCITAL 2013*, Les Sables d' Olonne, France.

Widlocher, A. and Mathet, Y. (2012). The glozz platform: A corpus annotation and mining tool. In *Proceedings of the 2012 ACM symposium on Document engineering*, pages 171–180.

Wiebe, J., Wilson, T., Bruce, R., Bell, M., and Martin, M. (2004). Learning subjective language. *Comput. Linguist.*, 30(3):277–308, September.