Datasets for Aspect-Based Sentiment Analysis in French

Marianna Apidianaki[1]**, Xavier Tannier**[2]**, Cécile Richart**[1]

1. LIMSI, CNRS, Université Paris-Saclay, Rue John von Neumann, 91403 Orsay, France
2. LIMSI, CNRS, Univ. Paris-Sud, Université Paris-Saclay, Rue John von Neumann, 91403 Orsay, France
`firstname.surname@limsi.fr`

Abstract

Aspect Based Sentiment Analysis (ABSA) is the task of mining and summarizing opinions from text about specific entities and their aspects. This article describes two datasets for the development and testing of ABSA systems for French which comprise user reviews annotated with relevant entities, aspects and polarity values. The first dataset contains 457 restaurant reviews (2365 sentences) for training and testing ABSA systems, while the second contains 162 museum reviews (655 sentences) dedicated to out-of-domain evaluation. Both datasets were built as part of SemEval-2016 Task 5 "Aspect-Based Sentiment Analysis" where seven different languages were represented, and are publicly available for research purposes. This article provides examples and statistics by annotation type, summarizes the annotation guidelines and discusses their cross-lingual applicability. It also explains how the data was used for evaluation in the SemEval ABSA task and briefly presents the results obtained for French.

Keywords: French Aspect-Based Sentiment Analysis, opinion mining, user reviews.

1. Introduction

Aspect Based Sentiment Analysis (ABSA) is the task of mining and summarizing opinions from text about specific entities and their aspects (Hu and Liu, 2004; Ganu et al., 2009). We have created and present two new datasets for the development and testing of ABSA systems for French which have been released in the framework of the SemEval-2016 Task 5 "Aspect-Based Sentiment Analysis". The first dataset comprises French restaurant user reviews annotated with relevant entities, aspects and polarity values. For this domain, both training and test data was provided. The second dataset is a smaller set of museum reviews for which no training data is available, dedicated to out-of-domain ABSA evaluation.

The ABSA task was first introduced for English in the SemEval 2014 evaluation campaign (Pontiki et al., 2014), where restaurant and laptop reviews annotated with aspect terms, categories and their polarity were provided for training and testing ABSA systems. The task was repeated in SemEval 2015 with a different, more unified, framework where aspect categories were defined as combinations of an entity type, an attribute type and a polarity value (Pontiki et al., 2015):

(1)   The fajitas were delicious, but expensive.
{FOOD#QUALITY, TARGET: *fajitas*}→POSITIVE
{FOOD#PRICES, TARGET: *fajitas*}→NEGATIVE

(2)   Great for a romantic evening.
{AMBIENCE#GENERAL, TARGET: NULL}→ POSITIVE

An out-of-domain subtask was also proposed where annotated hotel reviews were provided for testing but no training data was released. In 2016, the SemEval ABSA task became multilingual (Pontiki et al., 2016). New datasets were released for English allowing systems to be tested on the same domains as in previous years (laptops, restaurants and hotels), but datasets were also developed in new languages using a common set of annotation guidelines. More precisely, the multilingual

dataset created for this year's task comprises 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. All but the last dataset were released for Subtask 1 (Sentence-level ABSA) and part of the data was annotated at the text level for Subtask 2 (Text-level ABSA). The French Museum reviews dataset was released for Subtask 3 "Out-of-domain ABSA" where only test and no training data was provided.[1] In what follows, we describe the data collection procedure and the annotation guidelines that were developed for the two domains addressed in French. The data and the annotation guidelines are publicly available under a non-commercial, no redistribution license[2] through META-SHARE (Piperidis, 2012),[3] a repository devoted to the sharing and dissemination of language resources, and on the SemEval-2016 ABSA task website.[4]

2. Datasets and Annotation

2.1. Data Collection

French datasets were developed for two of the SemEval-2016 ABSA subtasks. For in-domain sentence-level ABSA (Subtask 1) the dataset comprises annotated restaurant reviews while for out-of-domain ABSA (Subtask 3) annotated museum reviews were released. For the first subtask, the restaurant domain, both training and test data was provided. In Subtask 3, the participating teams had the opportunity to test their systems in a previously unseen

---

[1] More details on the configuration of the different subtasks are given on the task's webpage: `http://alt.qcri.org/semeval2016/task5/`.

[2] The Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Licence.

[3] `http://metashare.ilsp.gr:8080`

[4] `http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools`

	Restaurants	Museums
Training data		
Review texts	337	–
Sentences	1669	–
Test data		
Review texts	120	162
Sentences	696	655

Table 1: French datasets released for ABSA 2016.

Entity	Attribute
RESTAURANT	GENERAL
	PRICES
	MISCELLANEOUS
FOOD	QUALITY
	PRICES
	STYLE & OPTIONS
DRINK	QUALITY
	PRICES
	STYLE & OPTIONS
SERVICE	GENERAL
LOCATION	GENERAL
AMBIENCE	GENERAL

Table 2: Possible entity-attribute pairs in the Restaurant domain.

domain for which no training data was available. The interest of the out-of-domain subtask was to see whether a system trained on some domain (e.g. restaurants) could be applied to a different domain with minimum adaptation. Data was thus provided for evaluation in the museum domain but no training data was made available. In Table 1, we give details on the corpus statistics.

The restaurant and museum user reviews were collected online and manually annotated. Participants in the French ABSA task were given a tool for downloading the unannotated reviews and were asked to assign the relevant annotations following the guidelines described in the next section. Each system's output was compared to the manually assigned ('gold') annotations during evaluation.

2.2. Annotation Schema and Guidelines for Restaurant Reviews

The French data was annotated by a native speaker linguistics student, then revised and corrected by the task organizers (a native and a non-native French speaker). Given a review text about a restaurant, the task of the annotator was to identify opinions expressed towards specific entities and their attributes and to assign the respective aspect category and polarity labels. The category values (entity-attribute pairs, E#A) had to be chosen from the inventory of entities and attributes provided for annotation of restaurant reviews in the SemEval-2015 task. To increase the comparability of the multilingual datasets and the applicability of the systems, the same guidelines were applied for annotation in this domain in different languages. In the French guide, the English annotation examples were replaced with examples in French. In total, six entity labels are available in the restaurant domain and can be combined with specific attribute labels. The possible combinations are described in Table 2.

Annotation was done at the sentence level. If the sentence contained an explicit mention to the revised entity (e.g. a specific food or restaurant), the label was assigned to that specific occurrence of the expression in the text. Otherwise, if an entity was implicitly referred (for example, through pronouns) or inferred in a sentence, the label was assigned to the whole sentence and no specific target expression was marked in the text, as illustrated in the following examples.

(3) Carte des vins sympa.
{DRINKS#STYLE&OPTIONS, TARGET: *carte des vins*}→POSITIVE
⇒ *Nice wine list.*

(4) Nous nous sommes faits livrer plusieurs fois et nous n'avons jamais été déçus, mais le service sur

place est déplorable.
{RESTAURANT#MISC, TARGET: NULL}→POSITIVE
{SERVICE#GENERAL, TARGET: *service*}→NEGATIVE
⇒ *We have used the delivery service several times and have never been disappointed, but on-site service is horrible.*

(5) Pourtant les plats sont bons et la déco est sympa.
{FOOD#QUALITY, TARGET: *plats*}→POSITIVE
{AMBIENCE#GENERAL, TARGET: *déco*}→POSITIVE
⇒ *However, the food is good and the decoration is nice.*

2.3. Cross-lingual Applicability of the Restaurant Guidelines

The ABSA-2016 task on Restaurant reviews involved six languages (English, French, Dutch, Russian, Spanish and Turkish). To ensure the comparability of the datasets and facilitate the cross-lingual applicability of the systems, the guidelines used for annotation were the same for all languages and corresponded to the guidelines provided for English in ABSA-2015 (Pontiki et al., 2015). This did not pose any particular problems in French at the level of category values, as the entities and attributes identified in the texts are similar in these closely related languages. We have however spotted the recurrent use of some expressions in the French reviews for which additional guidelines had to be given as these were not covered in the English guide. An example is the expression "*rapport qualité/prix*" in French (*value for money*), which assesses the quality of a restaurant or of the food/drinks served, compared to the cost. Sentences containing this expression are generally assigned two labels in the French dataset: one expressing the polarity regarding the prices (e.g. RESTAURANT#PRICES, FOOD#PRICES or DRINK#PRICES) and a second one describing the quality of the food/drinks, or the general quality of the restaurant. To decide whether the entity addressed in a sentence is the restaurant or the food/drinks, it is often necessary to look at the preceding sentences in the text. In example 6 the entity referred to is the RESTAURANT, while in example 7 it is the FOOD because the preceding sentence in the text was: *Perso j'avais pris le tournedos à 16 euros.* So, although this is a sentence-level annotation task, the selection of the most

Entity	Attribute
MUSEUM	GENERAL
	PRICES
	COMFORT
	ACTIVITIES
	ARCHITECTURE
	MISCELLANEOUS
LOCATION	GENERAL
SERVICE	GENERAL
FACILITIES	GENERAL
	PRICES
TOUR_GUIDING	GENERAL
COLLECTIONS	INTEREST
	SET_UP

Table 3: Possible entity-attribute (E#A) combinations in the Museum domain.

appropriate label sometimes requires taking into account information at the text level.

(6) Un moment agréable avec des amis pour un rapport qualité/prix imbattable.
{RESTAURANT#GENERAL, TARGET: NULL} →POSITIVE
{RESTAURANT#PRICES, TARGET: NULL}→POSITIVE
⇒ *A good time with friends and a great value for money.*

(7) Perso j'avais pris le tournedos à 16 euros.
⇒ *Me, I took the tournedos at 16 euros.*
Très bon rapport qualité prix !
{FOOD#PRICES, TARGET: NULL}→POSITIVE
{FOOD#QUALITY, TARGET: NULL}→POSITIVE
⇒ *Great value for money!*

The MISCELLANEOUS label is assigned to attributes not covered by the other labels. In the French data, we have assigned this label to sentences describing the hygienic conditions in a restaurant, its style and the services provided (take-away, delivery, web reservations, etc.).

(8) A emporter uniquement.
{RESTAURANT#MISCELLANEOUS, TARGET: NULL} → NEUTRAL
⇒ *Takeaway only.*

(9) Déjà il y a un gros loupé niveau hygiène : le cuisinier sort dehors avec ses gants puis retourne en cuisine sans les changer...
{RESTAURANT#MISCELLANEOUS, TARGET: NULL} →NEGATIVE
⇒ *There is a serious problem with hygiene: the cook comes out wearing his gloves and then goes back to the kitchen without changing them...*

Although the applicability of the English category and attribute labels in French was quite straightforward, the identification of the entities referred to in the text turned out to be a not so easy task, especially in cases of multi-word expressions. This task is easier in English where dish names often correspond to nominal compounds, for example *lobster ravioli*, *mashed potatoes* or *porcini mushroom pasta special*. In French, dish names are often described by multi-word expressions containing prepositional phrases. The ambiguity of the French prepositions makes more complex the identification of the entity towards which an opinion is expressed in the text and the selection of the category attribute that is being addressed. In the following example,

(10) *Nous avons pris les pâtes au foie gras et cèpes, celles-ci baignaient dans de la crème et de la crème balsamique!*
{FOOD#QUALITY, TARGET: *pâtes au foie gras et cèpes*} →NEGATIVE
⇒ *We took the foie gras and porcini mushrooms pasta, which was floating in cream and balsamic cream!*

the prepositional phrase *au foie gras et cèpes* describes the ingredients of the recipe (*foie gras and porcini mushrooms*), so the whole expression *pâtes au foie gras et cèpes* is the target. In other cases, the prepositional phrase describes the preparation or serving mode and might be considered as a part of the target or not. For instance, *poissons à la plancha* (*fish a la plancha*) is a target of the category FOOD as it denotes a specific way of cooking fish with no polarity value, but in *les frites maison sont à volonté* (*homemade french fries at will*) the target is *frites* (*french fries*) and gets the label {FOOD#STYLE&OPTIONS → POSITIVE} because the preparation (*maison* (*homemade*)) and serving mode (*à volonté* (*at will*)) are attributes with positive polarity.

2.4. Annotation Schema and Guidelines for Museum Reviews

The Museum domain had not been addressed in previous years or in other languages, so we had to compile a new set of guidelines for this domain. Annotated museum reviews were provided as the gold standard for evaluation in the out-of-domain Subtask 3 where no training data was released. Six entity labels are available for this domain in the annotation guidelines and can be combined with specific attribute labels. Some of them are specific to the museum domain (e.g. TOUR_GUIDING, COLLECTIONS) while others are also found in the restaurant domain (SERVICE, LOCATION). The possible entity-attribute combinations are described in Table 3.

The annotation for the out-of-domain substask was also done at the sentence level. Similarly to Subtask 1, if an explicit mention to the revised entity E (e.g. a museum) is contained in the text, the label is assigned to that specific occurrence of the expression. If an entity is implicitly referred or inferred, the label is assigned to the whole sentence, as shown in the following examples.

(11) Très beau musée.
{MUSEUM#GENERAL, TARGET: *musée*} →POSITIVE
⇒ *Very nice museum.*

(12) Tout est mis en œuvre pour mettre l'ensemble de ces trésors bien en valeur.

Entity	Attribute	Train				Test				Train
		POS	NEG	NEU	Total	POS	NEG	NEU	Total	+ Test
	GENERAL	175	214	14	403	58	63	4	125	528
RESTAURANT	PRICES	30	60	12	102	9	10	4	23	125
	MISCELLANEOUS	27	62	17	106	9	19	6	34	140
	QUALITY	332	340	43	715	137	128	22	287	1002
FOOD	PRICES	31	45	10	86	6	23	6	35	121
	STYLE & OPTIONS	118	97	30	245	49	45	15	109	354
	QUALITY	28	20	3	51	15	5	1	21	72
DRINK	PRICES	2	14	1	17	1	4	1	6	23
	STYLE & OPTIONS	12	9	1	22	4	6	3	13	35
SERVICE	GENERAL	218	305	14	537	70	111	6	187	724
LOCATION	GENERAL	34	3	3	40	20	1	4	25	65
AMBIENCE	GENERAL	157	43	6	206	63	18	7	88	294

Table 4: Statistics by annotation type in the Restaurant domain.

Entity	Attribute	POS	NEG	NEU	Total
	GENERAL	122	64	15	201
	PRICES	16	24	1	41
	COMFORT	13	27	6	46
MUSEUM	ACTIVITIES	4	3	2	9
	ARCHITECTURE	47	11	4	62
	MISCELLANEOUS	10	24	7	41
LOCATION	GENERAL	24	2	1	27
SERVICE	GENERAL	11	25	0	36
FACILITIES	GENERAL	14	16	4	34
	PRICES	1	6	0	7
TOUR GUIDING	GENERAL	9	4	3	16
COLLECTIONS	SETUP	28	62	6	96
	INTEREST	172	81	22	275

Table 5: Statistics by annotation type in the Museum domain.

{COLLECTIONS#SET_UP, TARGET: *trésors*}
→POSITIVE
{COLLECTIONS#INTEREST, TARGET: *trésors*}
→POSITIVE
⇒ *Everything has been done to put forward all these treasures.*

(13)   béton brut partout aucun revêtement sur les murs...
{MUSEUM#ARCHITECTURE,   TARGET: NULL}
→NEGATIVE
⇒ *raw concrete everywhere, no coating on the walls...*

3.   Composition of the Datasets

In an effort to keep the training and test datasets as balanced as possible with respect to the polarity of the opinions expressed in the texts, we collected an equivalent number of user reviews with different rankings based on a 5-star rating system (excellent, very good, average, poor, terrible). As a consequence, the retained reviews are uniformly distributed over all possible ranks. Of course, a review may convey opposing sentiment for different aspect categories (i.e. E#A pairs). For example, a restaurant with good service might be considered too expensive or badly located, and a museum that houses interesting collections might provide poor facilities. Tables 4 and 5 contain statistics by annotation type for each domain. We show the number of

occurrences of an E#A pair with positive (POS), negative (NEG) or neutral (NEU) polarity in the data. For the Restaurant domain, we provide counts for the training and test sets separately and combined.

4.   ABSA Evaluation at SemEval-2016

The SemEval-2016 ABSA evaluation is presented in detail in the task description paper (Pontiki et al., 2016). In this section, we provide the ranking of the systems that participated in the French ABSA task along with some information on the way that the gold review annotations were used for evaluation. More information about the systems can be found in the system description papers of the SemEval task.

In the sentence-level subtasks (1 & 3), given an unannotated review text, the participating systems had to identify tuples containing three types of information:

- Slot 1: the aspect category, i.e. an entity and attribute (E#A) pair towards which an opinion is expressed in the text

- Slot 2: an opinion target expression possibly used in the text to refer to the reviewed entity E of a pair E#A

- Slot 3:   the sentiment polarity to be assigned to each identified <category, target> pair, from the set

1125

System	Slot 1 (F-1)	Slot 2 (F-1)	Slot 1&2 (F-1)	Slot 3 (Acc.)
XRCE/C	**61.207**	65.316	**47.721**	**78.826**
IIT-TUDA/U	57.875	**66.667**	-	72.222
IIT-TUDA/C	57.033	-	-	-
INSIGHT-1/C	53.592	-	-	73.166
UFAL/U	49.928	-	-	-
UWB/C (1st run)	-	-	-	75.262
UWB/C (2nd run)	-	-	-	74.319
baseline/C	52.609	45.455	33.017	67.4

Table 6: Results of the SemEval-2016 French ABSA Subtask 1.

$$P = \{positive, negative, neutral\}$$

In Slot 1, the evaluation assesses whether a system correctly identifies the aspect categories towards which an opinion is expressed in the texts. The categories returned by a system are compared to the corresponding gold annotations and precision (P), recall (R) and F-1 scores are calculated. In Slot 2, the evaluation assesses whether a system correctly identifies the expressions used in a sentence to refer to the reviewed entities. The returned list of target expressions is compared to the gold annotations to calculate P, R and F-1 scores. For Slot 3, the total accuracy score for polarity classification is retained. System performance for all slots is compared to a baseline score. The baseline system for Subtask 1 selects the categories and polarity values to be assigned to the test sentences using a Support Vector Machine (SVM) based on bag-of-words (BoW) features and trained on the manually annotated training data.

Participants were free to decide the domain(s), subtask(s) and slot(s) they wished to participate in. Five systems participated in Subtask 1 for French. The out-of-domain Subtask 3 had no participants. However, the data developed for this subtask remains available and can be used for future research. The results in F-1 score (Slots 1 and 2) and the accuracy scores (Slot 3) obtained in Subtask 1 are reported in Table 6. The submissions were marked as constrained (C) or unconstrained (U) depending on whether the participants used only the provided training data or if they also exploited information from additional resources (e.g. lexicons or additional training data). We give in boldface the score of the system that performed best in the evaluation for each slot. The baseline systems and evaluation scripts are available for download from the SemEval-2106 ABSA website. More details on the evaluation procedure and the results are provided in the task description paper (Pontiki et al., 2016).

5. Conclusion

We have presented two new datasets for Aspect-Based Sentiment Analysis in French. These are the first French datasets annotated with this type of information and were initially developed and released in the framework of the multilingual SemEval-2016 ABSA task. Two subtasks were addressed, namely in-domain sentence-level ABSA and out-of-domain ABSA (Subtasks 1 and 3) where user-generated reviews in two domains were provided. Although the datasets were released for training and evaluation in the framework of the SemEval ABSA task, they remain available and can be used for research purposes.

6. References

Ganu, G., Elhadad, N., and Marian, A. (2009). Beyond the Stars: Improving Rating Predictions using Review Text Content. In *12th International Workshop on the Web and Databases, WebDB 2009*, Providence, Rhode Island.

Hu, M. and Liu, B. (2004). Mining and Summarizing Customer Reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177, New York, NY.

Piperidis, S. (2012). The META-SHARE Language Resources Sharing Infrastructure: Principles, Challenges, Solutions. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey.

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.

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.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., Clercq, O. D., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S. M., and Eryiğit, G. (2016). SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*, San Diego, California.