

Datasets for Aspect-Based Sentiment Analysis in French



Marianna Apidianaki¹, Xavier Tannier², Cécile Richart¹

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

Aspect Based Sentiment Analysis (ABSA)

Aspect Based Sentiment Analysis (ABSA) is the task of mining and summarizing opinions from text about specific entities and their aspects.

The ABSA task was first introduced for English in the SemEval 2014 evaluation campaign

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

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

In 2016, the SemEval ABSA task became multilingual: English, French, Dutch, Russian, Spanish, Arabic, Chinese and Turkish.

French training and test data

Corpus Statistics

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

- We collected an equivalent number of user reviews with different rankings based on a 5-star rating system.
- The retained reviews are thus uniformly distributed over all possible ranks.

Cross-lingual Guideline Applicability

To ensure the comparability of the multilingual datasets and ease the cross-lingual applicability of the systems, the same guidelines were used for all languages (Pontiki et al., 2015).

Additional guidelines were given in French for:

- recurrent expressions, e.g. rapport qualité/prix (*value for money*)
- MWEs where prepositional ambiguity makes complex the identification of the entity#attribute pair, e.g.

pâtes au foie gras et cèpes
 (*foie gras and porcini mushrooms pasta*)

vs. poissons à la plancha (*fish a la plancha*)

French Datasets

In-domain, supervised ABSA

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

Possible entity-attribute pairs in the Restaurant domain.

Out-of-domain ABSA

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

Possible entity-attribute pairs in the Museum domain.

Carte des vins sympa
 {DRINKS#STYLE_OPTIONS, TARGET:carte des vins}
 →POSITIVE
 ⇒ *Nice wine list.*

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.*

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.*

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

Tout est mis en œuvre pour mettre l'ensemble de ces trésors bien en valeur.
 {COLLECTIONS#SET_UP, TARGET:trésors}
 →POSITIVE
 {COLLECTIONS#INTEREST, TARGET:trésors}
 →POSITIVE
 ⇒ *Everything has been done to put forward all these treasures.*

Béton brut partout aucun revêtement sur les murs...
 {MUSEUM#ARCHITECTURE, TARGET:NULL}
 →NEGATIVE
 ⇒ *Raw concrete everywhere, no coating on the walls... treasures.*

Corpus Availability

<http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools>



The data and the annotation guidelines are publicly available under a non-commercial, no redistribution license through META-SHARE.

We would like to thank the organizers of the SemEval 2016 ABSA task for their help and support. The French data annotation was partly funded by the TransRead project (ANR-12-CORD-0015).

French ABSA Evaluation at SemEval-2016

The participating systems had to identify tuples containing three types of information:

Slot 1: the **aspect category**, i.e. an entity and attribute 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** (positive, negative, neutral) to be assigned to each identified <category, target> pair.

System	Slot 1 (F1)	Slot 2 (F1)	Slot 1 & 2 (F1)	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 (1 st run)	-	-	-	75.262
UWB/C (2 nd run)	-	-	-	64.319
Baseline/C	52.609	45.455	33.017	67.400

Results of the SemEval-2016 French ABSA Subtask 1.