Procesamiento del Lenguaje Natural, Revista nº 69, septiembre de 2022, pp. 199-205 recibido 05-07-2022 revisado 22-07-2022 aceptado 25-07-2022

ABSAPT 2022 at IberLEF: Overview of the Task on Aspect-Based Sentiment Analysis in Portuguese

Resumen de la Tarea de Análisis de Sentimientos Basado en Aspectos en Portugués (ABSAPT) en IberLEF 2022

Felix L. V. da Silva[1], Guilherme da S. Xavier[1], Heliks M. Mensenburg[1], Rodrigo F. Rodrigues[1], Leonardo P. dos Santos[3], Ricardo M. Araújo[1 2], Ulisses B. Corrêa[1 2], Larissa A. de Freitas[1 2]

[1] Center for Technological Development (CDTec),
Federal University of Pelotas (UFPel), Pelotas, RS, Brazil
[2] Artificial Intelligence Innovation Hub (H2IA),
Federal University of Pelotas, Pelotas, RS, Brazil
[3] University of São Paulo, São Paulo, SP, Brazil
{flvdsilva, gdsxavier, hmmersenburg, rfrodrigues, ricardo, ub.correa, larissa}@inf.ufpel.edu.br,
leonardosantsper@gmail.com

Resumen: Este artículo presenta la Tarea sobre Análisis de Sentimientos basado en Aspectos en Portugués (ABSAPT), realizada en el IberLEF 2022. Les pedimos a los participantes que desarrollaran sistemas capaces de identificar aspectos (AE) y extraer la polaridad (ASC) en textos escritos en portugués. Doce equipos se inscribieron en la tarea, entre los cuales cinco presentaron predicciones e informes técnicos. El sistema con mejor rendimiento logró un valor de precisión (Acc) de 0,67 para la subtarea de AE (Equipo Deep Learning Brasil) y un valor de precisión equilibrada (Bacc) de 0,82 para la subtarea de ASC (Equipo Deep Learning Brasil).
Palabras clave: Análisis de Sentimiento basado en Aspectos, Portugués, Reseñas de Hoteles.

Abstract: This paper presents the task on Aspect-Based Sentiment Analysis in Portuguese (ABSAPT), held within Iberian Languages Evaluation Forum (IberLEF 2022). We asked the participants to develop systems capable of extracting aspects (AE) and classifying sentiment of aspects (ASC) in texts. We created one *corpora* containing reviews about hotels. Twelve teams registered to the task, among which five submitted predictions and technical reports. The best performing system achieved an Accuracy (Acc) value of 0.67 in AE sub-task (Team Deep Learning Brasil) and a Balanced Accuracy (Bacc) value of 0.82 in ASC sub-task (Team Deep Learning Brasil).
Keywords: Aspect-Based Sentiment Analysis, Portuguese, Hotel Reviews.

1 Introduction

Sentiment Analysis (SA) is the field of Natural Language Processing (NLP) that automatically analyzes people's sentiments or opinions towards some entity. These sentiments can be valuable sources of information about the consumer's feelings about a particular product or idea, which can help in decisions by companies or governments (Liu, 2015).

SA can be done on different levels, focusing mainly on three possible granularity levels: document level, sentence level, and aspect level (de Freitas, 2015). At the aspect level, it is possible to analyze different opinions held towards different aspects of some entity or different entities in the same document or sentence.

The ABSAPT 2022 task aims to challenge different teams to propose techniques capa-

ISSN 1135-5948. DOI 10.26342/2022-69-17

ble of extracting aspects and classifying sentiment of aspects in the hotel reviews. The present paper presents an overview of the task. First, we briefly present some theoretical reflections Aspect-Based Sentiment Analysis (ABSA) (Section 2) and describe the proposal of our task (Section 3). Section 4 presents the *corpora* description and the annotation process. In Section 5, we describe the evaluation measures. Participant systems and the results are discussed in Section 6. Finally, the final remarks are done in Section 7.

2 Aspect-Based Sentiment Analysis

On this granularity level, all opinions expressed towards any aspect of the entity are analyzed individually. This level allows a better understanding of the opinions and entities in the text. To accomplish the analysis on this level, the task used to be broken into two sub-tasks: Aspect Extraction (AE) and Aspect Sentiment Classification (ASC).

2.1 Aspect Extraction

This sub-task determines which aspects of a given entity are considered in a text.

For example, in the sentence "Hotel com boa **localização**" ["Hotel with good **localization**"], the goal of AE would be to identify the aspect '**localização**'.

2.2 Aspect Sentiment Classification

This sub-task consists of the classification of the polarity for each aspect that has been identified in the text.

For example, in the sentence "Hotel com boa **localização**" ["Hotel with good **localization**"], the goal of ASC would be to classify the aspect '**localização**' as positive.

3 Task Description

People's opinions are a great source of information for other people and organizations, public or private. Typically works focused on Portuguese perform document level SA. It is hard to find ABSA approaches or datasets available for Portuguese.

We propose to create an ABSA for TripAdvisor reviews written in Portuguese. Two sub-tasks will be available: AE and ASC. The first sub-task comprehends the identification of aspects in the reviews, and the second subtask proposes to extract the sentiment orientation (polarity) of the review about a single aspect mentioned in it.

The availability of corpora written in Portuguese is scarce, which limits the amount of research done for this language.

This task will contribute to the progress of Portuguese NLP, as there is a demand for developing new methods and tools.

Previous ABSA competitions, such as SemEval 2014 (Pontiki et al., 2014), SemEval 2015 (Pontiki et al., 2015), SemEval 2016 (Pontiki et al., 2016) and EVALITA (Mattei et al., 2020) inspired us to develop a specific task for Portuguese.

4 Corpora Description and Annotation Process

This section describes the *corpora* proposed for evaluation and the annotation process (annotation guidelines and inter-annotator agreement).

4.1 *Corpora* Description

The corpora contain travellersŕeviews about accommodation services companies, written in Portuguese. In ABSAPT 2022, we used corpora developed previously by Freitas (de Freitas, 2015) and Corrêa (Corrêa, 2021). Freitasćorpus is publicly available, so it will be used only in the training dataset (3111 samples from 847 reviews). Corrêa's corpus is private and will be split into training and test dataset (257 samples to AE and 686 samples to ASC). The full dataset will be available after the event on the website http://absapt2022.tk/.

Both datasets were annotated following the same annotation guidelines (de Freitas, 2015). The concepts on the Accommodation Services Domain Ontology, HOntology (Chaves, de Freitas, and Vieira, 2012) were aspects annotated. HOntology contains 282 concepts categorized into 16 top-level concepts. The concept hierarchy has a maximum depth of 5.

In Figure 1, one can see an overview of the training dataset. The full training dataset contains 77 aspects. Due to space limitation, we present the 40 most frequent aspects and the polarities distribution. In Figure 2, we present an overview of the test dataset.

4.2 Annotation Process

The manual annotation of Freitas' corpus (training dataset) was conducted by two an-

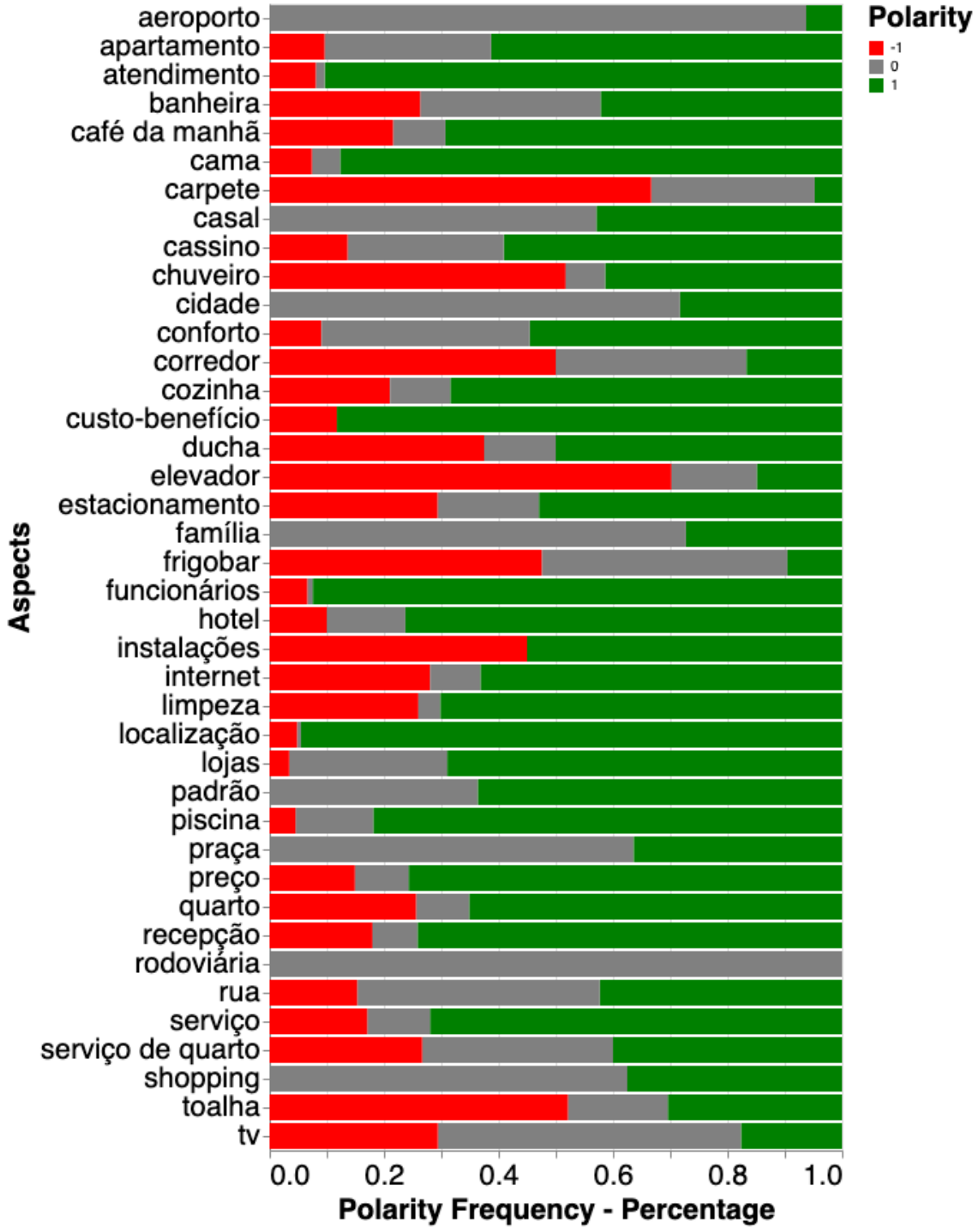

Figure 1: Training dataset: polarity frequency for aspects with at least 10 samples.

notators, both native speakers of Portuguese, one linguist, and one computer scientist. And the manual annotation of Corrêa's corpus (training and test dataset) was conducted by twelve students and professors of computer science and engineering annotators. Both used the tool developed in the context of (de Freitas, 2015).

The agreement between annotators (Freitas' corpus) was measured with Kappa Statistics (Landis and Koch, 1977). The annotators agreement about ABSA from do-

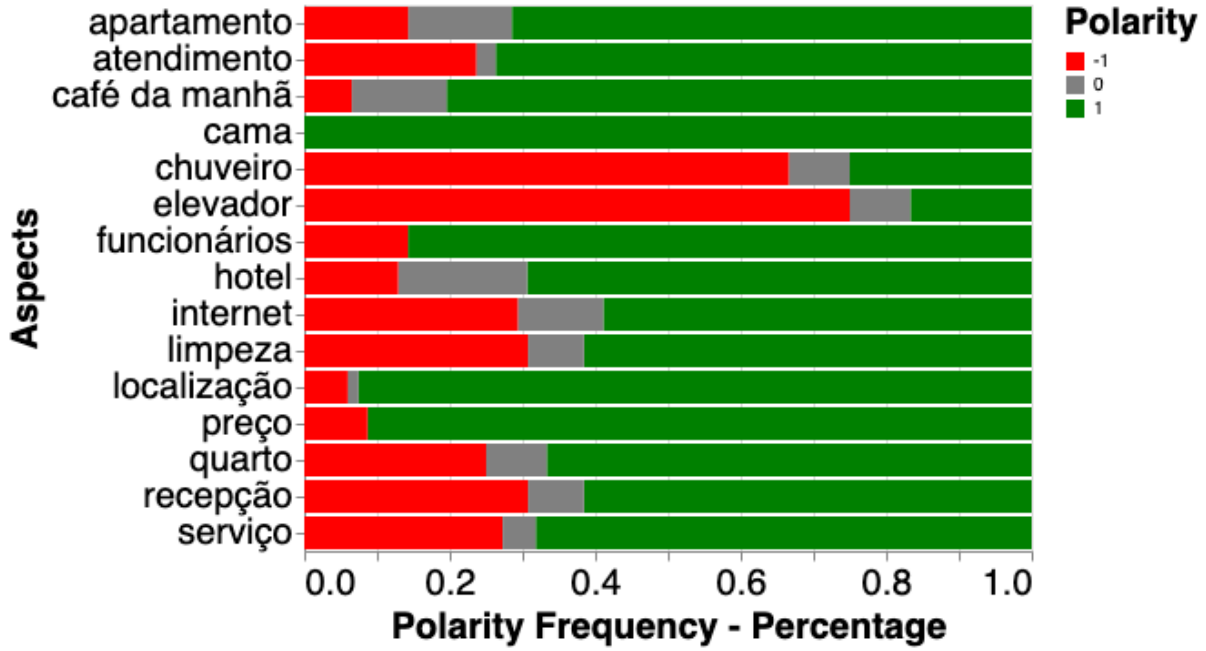

Figure 2: Test dataset: polarity frequency for aspects with at least 10 samples.

main ontology using Kappa was 0.58 for explicit aspects, which is considered a moderate agreement. We believe that the annotation has an acceptable Kappa value. It is also important to note that only in a few cases the annotators disagreed between negative and positive polarities, the majority of disagreements was about positive and neutral polarities, or negative and neutral polarities.

The agreement between annotators (Corrêa's corpus) was measured with Fleiss Kappa (Fleiss, 1971) suitable for more than two annotators. The majority annotator group share k $>$ 0.4 (moderate agreement).

5 Evaluation Measures

The training set was released on April 08th, and participants had sixteen days to train their systems. The test set was released on April 24th, and each participant had twelve days to submit one run.

Participating teams will receive training and test datasets. The latter was sent without the label of the samples.

We evaluated the predictions sent by the participants using several metrics: Acc (Eq. 1), Precision (Eq. 2), Recall (Eq. 3), F1 (Eq. 4), and Bacc (Eq. 6). Bacc to rank competitors. The submissions will be ranked according to Acc in AE sub-task and Bacc in ASC sub-task.

$$Acc = \frac{True\ Positives + True\ Negatives}{Total\ Number\ of\ Instances} \quad (1)$$

$$Precision = \frac{True\ Positives}{True\ Positives + True\ Negatives} \quad (2)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3)$$

$$F1 = 2 \times \frac{Precision \cdot Recall}{Precision + Recall} \quad (4)$$

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \quad (5)$$

$$Bacc = \frac{(Recall + Specificity)}{2} \quad (6)$$

6 Participants Systems and Discussion of the Results

Twelve teams registered for the task, among which five submitted predictions and technical reports. Participants are from universities

and institutes in Brazil (UFG, UFPI, IFPI, UFSCAR, USP, UFPR and UESC).

Participants used rules and lexicon (Team UFSCAR), traditional machine learning approaches as CRF (Team NILC and Team UFPR), and deep learning methods as Transformers (Team Deep Learning Brasil, Team PiLN, and Team UFPR).

Tables 1 and 2 present participants' results for each sub-task submitted run. The results are ranked according to the Acc in AE and Bacc in ASC.

Accuracy	Team
0.67	**TeamDeepLearningBrasil**
0.65	TeamPiLN
0.59	TeamUFSCAR
0.22	TeamNILC
0.17	TeamUFPR

Table 1: Participants results ranked in terms of Acc in AE sub-task.

For each system, best run is highlighted in bold. Team Deep Learning Brasil, used transformers to achieve an Acc of 0.67 in AE and a Bacc of 0.82 in ASC.

Below we summarize the proposed approach of each team:

- **Team Deep Leaning Brasil:** The authors proposed different methodologies for both sub-tasks of ABSA. The AE used a single sentence tagging approach, and the ASC tested with two different strategies, one as a Sentence Pair Classification and the other as a Conditional Text Generation. In addition, also were used other ABSA datasets such as Evalita, MAMs, Semeval 2014, 2015, and 2016 competition, and Bidirectional Encoder Representations from Transformers (BERT) pre-trained models on the Portuguese language and multilingual. The proposed approach reached the best-performing systems, achieving new state-of-the-art results on both subtasks (Gomes et al., 2022).
- **Team PiLN:** The authors proposed simple and well-known approaches to the sub-tasks of ABSA. The AE used a string-match strategy using a multilingual ontology for the accommodation sector named HOntology. The ASC used a BERTimbau (BERT pre-trained model on the Portuguese language), approaching the reviews as a Sentence Pair Classification. The proposed approach reached the second best-performing system, achieving the following results: Acc of 0.65 in AE, Bacc of 0.78, F1 of 0.77, Precision of 0.76, Recall of 0.78 in ASC (Neto et al., 2022).
- **Team UFSCAR:** The AE sub-task includes preprocessing, tokenization, feature extraction, a lexicon, and rule-based aspect identification. The ASC sub-task has two main steps: meaningful surroundings extraction and sentiment extraction using GoEmotions (Demszky et al., 2020), followed by polarity extraction. The proposed approach reached the third best-performing system, achieving the following results: Acc of 0.59 in AE, Bacc of 0.62, F1 of 0.61, Precision of 0.65, Recall of 0.62 in ASC (Assi et al., 2022).
- **Team NILC:** The authors participated only in AE sub-task. Its approach is based on the Conditional Random Fields (CRF) machine learning algorithm combined with a post-processing step. After applying the method, the authors performed an error analysis of detected and non-detected aspects (Machado and Pardo, 2022).
- **Team UFPR:** In the AE used CRF, the dataset was adapted with tokenization and the POS Tagging technique, and a pre-trained model for Portuguese was used for the POS Tagging process. If the POS Tagging process has a better adaptation for Portuguese, there will be a gain in results for the CRF performance. In the ASC used BERTimbau (Heinrich and Marchi, 2022). The proposed approach reached the worst run, achieving the Acc of 0.17 in AE.

7 Final Remarks

Motivated by the necessity of improvements in the ABSA task focused on Portuguese, we proposed a task within the IberLEF 2022. This paper overviews the first task on ABSA in Portuguese to identify aspects and extract the polarity in hotel reviews.

The datasets (training and test) have been manually annotated. The inter-annotator

Bacc	F1	Precision	Recall	Team
0.82	**0.81**	**0.81**	**0.82**	**TeamDeepLearningBrasil**
0.78	0.77	0.76	0.78	TeamPiLN
0.62	0.61	0.65	0.62	TeamUFSCAR
0.62	0.61	0.65	0.62	TeamUFPR

Table 2: Participants results ranked in terms of Bacc in ASC sub-task.

agreement for training and test dataset is considered moderate.

The deep learning methods based on Transformers performed better than approaches based on rules and lexicons. The Deep Learning Brasil Team achieved a Bacc of 0.67 for the AE and Acc of 0.82 for the ASC, while the UFSCAR Team achieved a Bacc of 0.59 for the AE and Acc of 0.62 for the ASC.

Acknowledgments

Thank you to all annotators for their essential work. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan X Pascal GPU used for this research. This work was financed in part by the following Brazilian research agencies: CAPES and CNPq.

References

Assi, F. M., G. B. Candido, L. N. dos Santos Silva, D. F. Silva, and H. de Medeiros Caseli. 2022. Ufscar's team at absapt 2022: Using syntax, semantics and context for solving the tasks. In *Proceedings of the Iberian Languages Evaluation Fórum (IberLEF 2022), co-located with the 38th Conference of the Spanish Society for Natural Language Processing (SEPLN 2022), Online. CEUR.org.*

Chaves, M. S., L. A. de Freitas, and R. Vieira. 2012. Hontology: A multilingual ontology for the accommodation sector in the tourism industry. In J. Filipe and J. L. G. Dietz, editors, *KEOD*, pages 149–154. SciTePress.

Corrêa, U. B. 2021. *Análise de sentimento baseada em aspectos usando aprendizado profundo: uma proposta aplicada à língua portuguesa.* Ph.D. thesis, Universidade Federal de Pelotas, Pelotas.

de Freitas, L. A. 2015. *Feature-level sentiment analysis applied to Brazilian Portuguese reviews.* Ph.D. thesis, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre.

Demszky, D., D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online, July. Association for Computational Linguistics.

Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Gomes, J. R. S., R. C. Rodrigues, E. A. S. Garcia, A. F. B. Junior, D. F. C. Silva, and D. F. Maia. 2022. Deep learning brasil at absapt 2022: Portuguese transformer ensemble approaches. In *Proceedings of the Iberian Languages Evaluation Fórum (IberLEF 2022), co-located with the 38th Conference of the Spanish Society for Natural Language Processing (SEPLN 2022), Online. CEUR.org.*

Heinrich, T. and F. Marchi. 2022. Teamufpr at absapt 2022: Aspect extraction with crf and bert. In *Proceedings of the Iberian Languages Evaluation Fórum (IberLEF 2022), co-located with the 38th Conference of the Spanish Society for Natural Language Processing (SEPLN 2022), Online. CEUR.org.*

Landis, J. R. and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1).

Liu, B. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions.* Cambridge University Press.

Machado, M. T. and T. A. S. Pardo. 2022. Nilc at absapt 2022: Aspect extraction for portuguese. In *Proceedings of the Iberian Languages Evaluation Fórum (IberLEF 2022), co-located with the 38th Conference*

of the Spanish Society for Natural Language Processing (SEPLN 2022), Online. CEUR.org.

Mattei, L. D., G. D. Martino, A. Iovine, A. Miaschi, M. Polignano, and G. Rambelli. 2020. Ate absita@ evalita2020: Overview of the aspect term extraction and aspect-based sentiment analysis task. *Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020), Online. CEUR.org.*

Neto, F. A. R., R. F. de Sousa, R. L. de S. Santos, R. T. Anchiêta, and R. S. Moura. 2022. Piln at absapt 2022: Lexical and bert strategies for aspect-based sentiment analysis in portuguese. In *Proceedings of the Iberian Languages Evaluation Fórum (IberLEF 2022), co-located with the 38th Conference of the Spanish Society for Natural Language Processing (SEPLN 2022), Online. CEUR.org.*

Pontiki, M., D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. D. Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryiğit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California, June. Association for Computational Linguistics.

Pontiki, M., D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos. 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., D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar. 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.