ABCD Team at ASQP-PT 2025: Aspect Sentiment Quad Prediction in Portuguese as Text Generation

Bui Hong Son[1,2], Dang Van Thin[1,2,*]

[1]University of Information Technology-VNUHCM, Quarter 6, Linh Trung Ward, Thu Duc District, Ho Chi Minh City, Vietnam
[2]Vietnam National University, Ho Chi Minh City, Vietnam

Abstract

This paper presents our system for Aspect Sentiment Quad Prediction (ASQP) shared task in Portuguese hotel reviews. Our system leverages text generation-based language models to extract quadruples consisting of aspect categories, aspect terms, sentiment terms, and polarities from customer reviews. Our methodology focuses on a structured text generation paradigm that encodes the relationship between aspects and sentiments through a custom output format. Experimental results on the ABSAPT-2025 shared task dataset demonstrate the efficacy of our approach in handling the complexities of sentiment analysis in low-resource languages. The proposed model achieves competitive performance compared to the baseline and obtains an F1-score of 45.66% (rank 1) in the ASQP-PT 2025 [1, 2] shared task.

Keywords

Sentiment Analysis, Portuguese language, sentiment analysis, aspect-based sentiment analysis, aspect sentiment quad prediction, sequence-to-sequence, large-language-modelss

1. Introduction

Aspect-Based Sentiment Analysis (ABSA) has emerged as a critical component in understanding user-generated content, particularly in domains where detailed feedback analysis provides significant business value. While substantial progress has been made for high-resource languages like English and Chinese, low-resource languages such as Portuguese have received comparatively less attention. The ASQP task represents the most comprehensive form of ABSA, requiring systems to identify four interrelated elements: (1) aspect category (e.g., "service"), (2) aspect term (e.g., "atendimento"), (3) sentiment term (e.g., "excelente"), and (4) polarity (e.g., "POS"). This level of granularity provides actionable insights for businesses but introduces significant modeling challenges, especially for morphologically rich languages like Portuguese.

The ABSAPT-2025 shared task introduces the first annotated ASQP dataset for Portuguese hotel reviews, facilitating research in this understudied language context. In this paper, we present our approach to the ASQP-PT 2025 challenge, focusing on direct fine-tuning of a multilingual sequence-to-sequence model to generate structured outputs containing complete sentiment quadruples.

2. Related Work

ABSA has evolved from simple sentiment classification to increasingly fine-grained analysis. Early approaches such as those presented in SemEval tasks [3, 4, 5] addressed ABSA as separate subtasks: aspect extraction, aspect categorization, and sentiment classification. These approaches typically employed pipeline architectures where errors propagated through sequential components. More recent work has moved toward joint modeling of ABSA subtasks. [6] introduced the ASQP task, which unifies all ABSA subtasks into a single prediction challenge. [7] proposed a unified generative framework that

IberLEF 2025, September 2025, Zaragoza, Spain
*Corresponding author.
22521246@gm.uit.edu.vn (B. H. Son); thindv@uit.edu.vn (D. V. Thin)
https://nlp.uit.edu.vn/ (D. V. Thin)
0000-0001-8340-1405 (D. V. Thin)

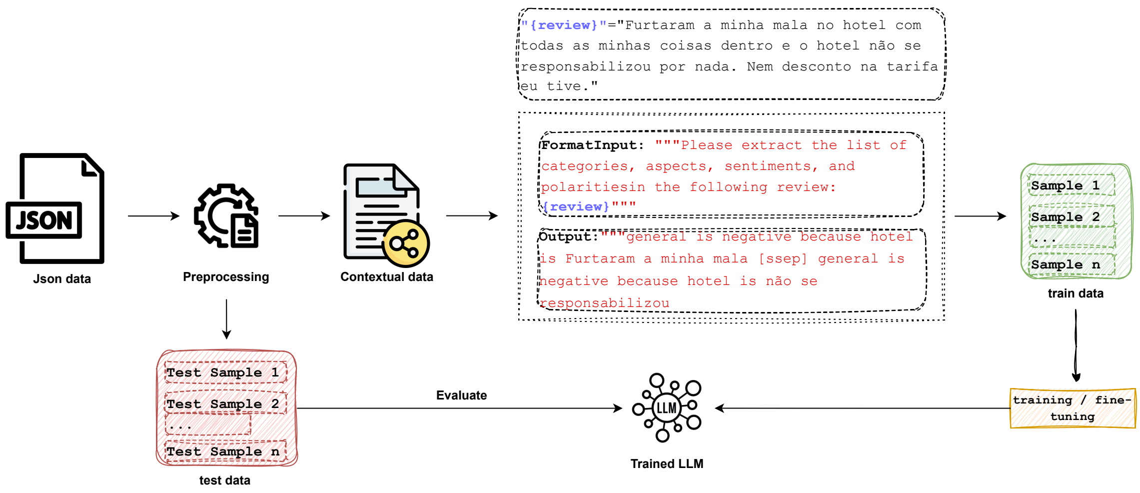

Figure 1: Our overall pipeline for the ASQP-PT 2025 shared task - Task 4.

converts all ABSA subtasks into text generation problems. Similarly, [8] demonstrated the effectiveness of sequence-to-sequence approaches for aspect sentiment triplet extraction.

Multilingual approaches to sentiment analysis have gained popularity with the advent of cross-lingual pre-trained language models. Pretrained models such as mBERT [9], XLM-R [10], and mT5 [11] have demonstrated remarkable cross-lingual transfer capabilities. For Portuguese specifically, prior work has been limited. The ABSAPT-2022 and ABSAPT-2024 shared tasks [12, 13] focused on aspect term extraction and sentiment orientation but did not address the complete ASQP challenge. To our knowledge, the ABSAPT-2025 task represents the first benchmark for complete ASQP in Portuguese.

Recent advances in NLP have demonstrated the effectiveness of framing structured prediction tasks as sequence-to-sequence problems. [14] showed that T5 models can effectively handle diverse NLP tasks through text-to-text transfer. [15] introduced BART, which achieved state-of-the-art performance on several generation tasks. For structured prediction specifically, [16] demonstrated that sequence-to-sequence models can effectively generate outputs with complex structure when provided with appropriate output templates. This approach has been successfully applied to information extraction tasks [17] and semantic parsing [18].

3. Methodology

As illustrated in Figure 1, our approach implements an end-to-end sequence-to-sequence workflow for Aspect Sentiment Quad Prediction. The process begins with raw JSON data containing Portuguese hotel reviews and their annotations, which undergoes preprocessing to normalize text and structure the training examples. We transform this data into contextual examples where each input-output pair follows a specific pattern. As shown in the workflow diagram, this creates paired training samples where the model learns to generate structured outputs from unstructured review text. The example demonstrates how the Portuguese review **"Furtaram a minha mala no hotel com todas as minhas coisas dentro e o hotel não se responsabilizou por nada. Nem desconto na tarifa eu tive."** is transformed into the structured output **"general is negative because hotel is não se responsabilizou,"** capturing the aspect category, polarity, aspect term, and sentiment term in a single coherent structure.

The processed data is used to fine-tune a pre-trained language model (mT5), which is then evaluated against test samples following the same format. This unified approach offers several advantages over pipeline methods, particularly in capturing the interdependencies between aspect categories, terms, sentiments, and polarities.

3.1. Problem Formulation

We conceptualize Aspect Sentiment Quad Prediction (ASQP) as a structured sequence transduction problem. Given an input review text $R = \{w_1, w_2, ..., w_n\}$ consisting of n tokens, our objective is to generate a set of quadruples $Q = \{q_1, q_2, ..., q_m\}$, where each quadruple $q_i = (c_i, a_i, s_i, p_i)$ comprises an aspect category $c_i \in C$, an aspect term a_i (a contiguous span in R), a sentiment term s_i (another span in R), and a sentiment polarity $p_i \in \{POS, NEG, NEU\}$.

The fundamental challenge lies in capturing the complex interdependencies among these four elements while maintaining computational efficiency. Rather than decomposing ASQP into subtasks—as has been common in prior ABSA research—we adopt a holistic generative approach that leverages the semantic understanding capabilities of large pre-trained language models.

3.2. Model Architecture

3.2.1. Portuguese T5

The T5-portuguese model [19] represents a language-specific adaptation of the original T5 architecture, specifically pre-trained on a large corpus of Portuguese text. Unlike multilingual variants, this model concentrates its entire parameter capacity on a single language, potentially offering more nuanced linguistic representations for Portuguese-specific phenomena. The model follows the encoder-decoder architecture of the original T5:

- **Encoder:** Transforms the input review text into contextualized representations
- **Decoder:** Autoregressively generates the structured output containing sentiment quadruples

The T5-portuguese model benefits from:

- Focused pre-training on Portuguese linguistic patterns
- Enhanced handling of Portuguese morphology and syntax
- Better representation of Portuguese-specific semantic nuances
- Domain adaptation to Portuguese web content

This language-specific pre-training theoretically provides advantages in capturing the subtleties of sentiment expression in Portuguese, which often employs complex verbal constructions and rich adjectival morphology that differ significantly from other Romance languages.

3.2.2. Multilingual T5

In parallel, we evaluate the multilingual T5 model [20], specifically the base variant containing approximately 580M parameters. The mT5 architecture follows the same encoder-decoder paradigm but has been pre-trained on mC4, a massive multilingual corpus covering 101 languages including Portuguese.

The decision to include mT5 is motivated by several factors:

- Its substantial exposure to Portuguese data during pre-training
- Its proven cross-lingual transfer capabilities, allowing it to leverage patterns learned from high-resource languages
- The potential for more robust representations through cross-lingual knowledge sharing

We hypothesize that while mT5 allocates only a fraction of its parameter capacity to Portuguese, the cross-lingual transfer capabilities may compensate by adapting knowledge from related high-resource languages like Spanish and French.

3.3. Input and Output Formulation

For input encoding, we prepend a task-specific prompt to each review:

```
Please extract the list of categories, aspects, sentiments and polarities in the
following comment: "{review_text}".
```

For output encoding, we design a structured format that explicitly captures the relationships between elements in each quadruple:

```
{category} is {polarity} because {aspect} is {sentiment} [ssep] ...
```

Where `[ssep]` is a special separator token used to delineate multiple quadruples. This format offers several advantages:

1. It expresses the logical relationship between aspects and sentiments
2. It maintains a consistent structure that the model can learn to reproduce
3. It allows for variable numbers of quadruples per review
4. It reduces the need for complex output parsing mechanisms

This formulation transforms ASQP into a direct text generation problem, allowing us to leverage the language model's implicit knowledge without introducing task-specific architecture modifications

3.4. Model Training

For model optimization, we employ a cross-entropy loss function at the token level:

$$\mathcal{L} = -\sum_{i=1}^{|Y|} \log P(y_i | y_{<i}, X; \theta)$$

where Y represents the target output sequence, X represents the input sequence, and θ represents the model parameters.

We fine-tune all parameters of the pre-trained mT5 model using the AdamW optimizer with a learning rate of 3×10^{-4} and a linear warmup schedule over 10 percent of training steps. This approach allows the model to adapt its pre-trained representations to the specific linguistic patterns of Portuguese sentiment expression.

4. Experimental Settings

4.1. Dataset

We conduct experiments on the ABSAPT-2025 dataset, which consists of hotel reviews written in Portuguese. The dataset contains annotations for aspect categories, aspect terms, sentiment terms, and polarities. Each review may contain multiple aspect-sentiment quadruples, with an average of 2.7 quadruples per review.

The reviews cover various aspects of hotel accommodations, including service, cleanliness, location, and value. The dataset is particularly challenging due to:

- The linguistic complexity of Portuguese, with its rich morphology
- The domain-specific terminology related to hospitality
- The implicit sentiment expressions common in review language
- The variable length and structure of user-generated content

Table 1
The experimental result for ASQP Task on the development set.

ASQP as Paraphrase	T5-portuguese			mT5		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Without instruction	0.54	0.49	0.51	0.53	0.49	0.51
English instruction	0.54	0.50	0.52	0.55	0.49	0.52
Portuguese instruction	0.50	0.47	0.49	0.47	0.44	0.47
Data Augmentation with LLMs (with En instruction)						
Without source	0.49	0.49	0.49	0.48	0.47	0.48
With source	0.52	0.49	0.51	0.51	0.47	0.49

Table 2
System rankings for the ASQP-PT 2025 Task 4.

ASQP – Task 4				
Team	**Precision**	**Recall**	**F1 Score**	**Rank**
ABCD	0.4435	0.4706	0.4566	1
Baseline	0.4276	0.3678	0.3955	2

4.2. Implementation Details

We implement our approach using the `Hugging Face Transformers` library. The model configuration is based on the `google/mt5-base` architecture, which contains approximately 580 million parameters. Both input and output sequences are limited to a maximum length of 512 tokens. We train the model using a batch size of 8 for 10 epochs. Optimization is performed using the AdamW optimizer with a weight decay of 0.01, and a learning rate of 3×10^{-4}, incorporating 10% warmup steps. To improve memory efficiency, we use mixed precision training with the `bfloat16` format. All training is conducted on an NVIDIA GPU with CUDA support. Additionally, dropout is applied at a rate of 0.1 for regularization, and gradient clipping with a maximum norm of 1.0 is employed to prevent exploding gradients.

4.3. Inference and Post-processing

At inference time, we employ beam search decoding with beam width 3, temperature 0.1, and top-p sampling with p=0.9. This configuration balances between deterministic output (necessary for consistent structured prediction) and diversity (useful for capturing varied expressions).

The model's generated text is post-processed using a custom parser that:

1. Splits the sequence on the `[ssep]` token to identify individual quadruples
2. Extracts the category, polarity, aspect, and sentiment components using regular expressions
3. Locates the exact span positions of aspect and sentiment terms in the original text
4. Handles special cases such as implicit sentiments (marked as "NULL") and generic aspect references

The resulting structured data is converted to the competition's required JSON format, including position information for aspects and sentiment terms.

5. Results & Discussions

Tables 1 and Table 2 present the performance of our system with different backbone architectures for the quadrupled extraction task. Table 1 shows our internal experimental results on a validation split derived from the training data, allowing us to compare different model configurations and training strategies. Table 2 presents the official competition results as evaluated by the organizers on the held-out test set, demonstrating our system's performance in the competitive context.

As shown in Table 1, our internal validation experiments explored various configurations of model architecture and training strategies. We evaluated both T5-portuguese and mT5 models across different instruction prompting approaches and data augmentation techniques. For instruction prompting, English instructions consistently outperformed both no instructions and Portuguese instructions across both model variants. The T5-portuguese model achieved its highest F1-score of 0.52 with English instructions, matching the best performance of the mT5. This counterintuitive finding—that English instructions work better than Portuguese instructions for a Portuguese language task—suggests that the models' pre-training regimes may have better prepared them to follow English instructions. Our data augmentation experiments showed that augmentation strategies preserving the structural characteristics of the source data ("With source") outperformed less constrained augmentation methods ("Without source"). However, neither augmentation approach improved upon the best non-augmented configuration, indicating that the quality of augmented data requires further refinement to provide meaningful benefits.

Table 2 reveals that our submission (Team ABCD) using the method with main configurations including English intruction prompt designs with T5-portuguese model achieved the top ranking in the official ASQP task evaluation, with an F1-score of 0.4566 compared to the baseline's 0.4055. The performance gap is particularly pronounced in recall (0.4706 vs. 0.3678), suggesting our approach's superior ability to identify relevant quadruples across diverse review contexts.

The difference between our internal validation scores (up to 0.52 F1) and the official competition score (0.4566 F1) highlights the challenges of domain adaptation and the potential presence of distributional shifts between training and test data. This phenomenon is common in shared tasks and underscores the importance of robust evaluation practices that account for potential overfitting to validation data.

6. Conclusion

In this paper, we present our system, which leverages the power of sequence-to-sequence modelling with pre-trained language models to extract quadruples consisting of aspect categories, aspect terms, sentiment terms, and polarities from customer reviews. Experimental results on the ABSAPT-2025 shared task dataset demonstrate the efficacy of our approach in addressing the complexities of sentiment analysis in low-resource languages. The proposed model achieves competitive performance without requiring complex architectural modifications or parameter-efficient fine-tuning techniques, establishing a strong baseline for future research in multilingual ASQP. For future work, we plan to fine-tune large language models to enhance overall performance and explore data augmentation techniques to expand the training set.

Acknowledgements

This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number C2024-26-02.

Declaration on Generative AI

During the preparation of this work, we used ChatGPT and Grammarly to check grammar and spelling and edit the content for clarity and coherence. After using these tools, we reviewed and edited the content as needed and took full responsibility for the publication's content.

References

[1] J. Á. González-Barba, L. Chiruzzo, S. M. Jiménez-Zafra, Overview of IberLEF 2025: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with the 41st Conference of the Spanish Society for Natural Language Processing (SEPLN 2025), CEUR-WS. org, 2025.

[2] E. P. Lopes, G. Gomes, A. T. Bender, R. M. Araújo, L. A. de Freitas, U. B. Corrêa, Overview of asqp-pt at iberlef 2025: Overview of the task on aspect-sentiment quadruple prediction in portuguese, Procesamiento del Lenguaje Natural 75 (2025).

[3] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, SemEval-2014 task 4: Aspect based sentiment analysis, in: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), 2014, pp. 27–35.

[4] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. Androutsopoulos, SemEval-2015 task 12: Aspect based sentiment analysis, in: P. Nakov, T. Zesch, D. Cer, D. Jurgens (Eds.), Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), 2015, pp. 486–495.

[5] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, G. Eryiğit, SemEval-2016 task 5: Aspect based sentiment analysis, in: Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), 2016, pp. 19–30.

[6] L. Xu, L. Bing, W. Lu, F. Huang, Aspect sentiment classification with aspect-specific opinion spans, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 3561–3567.

[7] W. Zhang, Y. Deng, X. Li, Y. Yuan, L. Bing, W. Lam, Aspect sentiment quad prediction as paraphrase generation, in: M.-F. Moens, X. Huang, L. Specia, S. W.-t. Yih (Eds.), Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, pp. 9209–9219.

[8] I. Naglik, M. Lango, Aste transformer modelling dependencies in aspect-sentiment triplet extraction, 2024. URL: https://arxiv.org/abs/2409.15202. arXiv:2409.15202.

[9] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, pp. 4171–4186.

[10] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 8440–8451.

[11] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, C. Raffel, mT5: A massively multilingual pre-trained text-to-text transformer, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 483–498.

[12] F. L. da Silva, G. d. S. Xavier, H. M. Mensenburg, R. F. Rodrigues, L. P. dos Santos, R. M. Araújo, U. B. Corrêa, L. A. de Freitas, Absapt 2022 at iberlef: Overview of the task on aspect-based sentiment analysis in portuguese, Procesamiento del Lenguaje Natural 69 (2022) 199–205.

[13] A. T. Bender, G. A. Gomes, E. P. Lopes, R. M. Araujo, L. A. de Freitas, U. B. Corrêa, Overview of absapt at iberlef 2024: Overview of the task on aspect-based sentiment analysis in portuguese, Procesamiento del Lenguaje Natural 73 (2024) 315–322.

[14] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, Journal of Machine Learning Research 21 (2020) 1–67. URL: http://jmlr.org/papers/v21/20-074.html.

[15] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, in: Proceedings of the 58th Annual Meeting of the Association for

Computational Linguistics, 2020, pp. 7871–7880.

[16] G. Paolini, B. Athiwaratkun, J. Krone, J. Ma, A. Achille, R. ANUBHAI, C. N. dos Santos, B. Xiang, S. Soatto, Structured prediction as translation between augmented natural languages, in: International Conference on Learning Representations, 2021. URL: https://openreview.net/forum?id=US-TP-xnXI.

[17] Y. Hao, S. He, W. Jiao, Z. Tu, M. Lyu, X. Wang, Multi-task learning with shared encoder for non-autoregressive machine translation, arXiv preprint arXiv:2010.12868 (2020).

[18] R. Shin, C. Lin, S. Thomson, C. Chen, S. Roy, E. A. Platanios, A. Pauls, D. Klein, J. Eisner, B. Van Durme, Constrained language models yield few-shot semantic parsers, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, pp. 7699–7715.

[19] D. Carmo, M. Piau, I. Campiotti, R. Nogueira, R. Lotufo, Ptt5: Pretraining and validating the t5 model on brazilian portuguese data, arXiv preprint arXiv:2008.09144 (2020).

[20] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, C. Raffel, mt5: A massively multilingual pre-trained text-to-text transformer, 2021. URL: https://arxiv.org/abs/2010.11934. arXiv:2010.11934.