AttentionApp: An Interactive Tool for Analyzing Transformer Attention Patterns in Portuguese

Ricardo G. Oliveira[1], **Daniela Barreiro Claro**[1]

[1]FORMAS Research Center on Data and Natural Language
Institute of Computing – Federal University of Bahia (UFBA) – Salvador - Bahia - Brazil
{gomesricardo, dclaro}@ufba.br

Abstract

This paper presents *AttentionApp*, an interactive demonstration system designed to support the inspection and linguistic analysis of attention mechanisms in Transformer-based language models for Portuguese. The tool allows users to input sentences in Portuguese and visualize attention distributions across layers and heads, enabling fine-grained qualitative analysis of syntactic and semantic patterns captured by the model. AttentionApp is intended as a research-oriented tool, facilitating exploratory analysis, hypothesis generation, and interpretability studies for Portuguese Natural Language Processing.

1 Introduction

The introduction of the Transformer architecture based on self-attention mechanisms (Vaswani et al., 2017) have become a dominant paradigm in Natural Language Processing (NLP) domain. This paradigm has been successfully adopted in large pre-trained models such as BERT (Devlin et al., 2019), including language-specific variants for Portuguese, such as BERTimbau (Souza et al., 2020).

Despite their empirical success, understanding how these models encode linguistic structure remains an open research problem. Attention mechanisms, while not explanations per se, provide a useful lens for exploratory analysis and qualitative inspection of model behavior, as demonstrated in prior analyses of attention distributions in BERT-based models (Clark et al., 2019; Voita et al., 2019), including recent studies focusing on syntactic patterns in Portuguese (Oliveira et al., 2025).

For Portuguese, this challenge is amplified by linguistic phenomena such as rich verbal morphology, flexible word order, clitic placement, and complex agreement patterns. However, most interpretability tools and visualization frameworks have been developed with English-centric assumptions, limiting their applicability to Portuguese.

In this demonstration, we present *AttentionApp*, an interactive tool designed to visualize and analyze attention patterns produced by Transformer models when processing Portuguese text. The system aims to support researchers, students, and practitioners interested in interpretability, syntactic analysis, and information extraction for Portuguese.

2 System Overview

AttentionApp is implemented as a modular Python-based system with a web-oriented interactive interface. The architecture is composed of three main components:

- **Preprocessing and Tokenization Module**, responsible for sentence segmentation, subword tokenization, and token alignment, following the standard tokenization schemes used in BERT-style models (Devlin et al., 2019).
- **Model and Attention Extraction Module**, which loads Transformer-based language models for Portuguese, such as BERTimbau (Souza et al., 2020), and extracts attention weights across layers and heads.
- **Visualization and Interaction Module**, which renders attention distributions in an interactive format, enabling exploration of token-to-token relations.

The system was designed with extensibility in mind, allowing future integration of additional models, linguistic annotations, or downstream tasks.

3 Main Functionalities

AttentionApp provides the following core functionalities:

- Sentence-level input in Portuguese, allowing users to analyze arbitrary examples.

Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026)
Vol. 2, pages 18–20

- Extraction of attention weights from all layers and attention heads of the selected Transformer model.
- Interactive visualization of attention matrices, supporting token-level inspection and comparison across heads, as commonly adopted in attention analysis studies (Clark et al., 2019).
- Layer and head selection, enabling fine-grained analysis of attention specialization, in line with findings on head-level functional differentiation (Voita et al., 2019).
- Exploratory linguistic analysis, allowing users to qualitatively assess whether attention patterns align with syntactic or semantic relations, following methodologies previously applied to syntactic analysis through attention heads (Oliveira et al., 2025).

These features make the tool suitable for exploratory research, pedagogical use, and qualitative evaluation of model behavior.

4 Demonstration Setup

During the demonstration session, participants will interact directly with AttentionApp through a live interface. Users will input Portuguese sentences and dynamically explore the resulting attention visualizations by selecting layers and heads.

The demonstration will focus on illustrating how different attention heads capture distinct relational patterns and how these patterns vary across layers, as previously observed in empirical analyses of Transformer attention (Clark et al., 2019; Voita et al., 2019). The system runs locally and does not require specialized hardware beyond a standard laptop, ensuring robustness in a conference environment.

5 System Availability

AttentionApp is an open-source research tool and its complete source code is publicly available on GitHub. The repository includes installation instructions, dependency specifications, and example configurations that allow the system to be executed locally.

The project is available at:

`https://github.com/rgoliveirati/attention_app`

This availability ensures transparency, reproducibility, and facilitates further extensions by the research community.

AttentionApp is an open-source research tool and its complete source code is publicly available on GitHub (Oliveira, 2025).

6 Linguistic Relevance and Applications

AttentionApp is particularly relevant for research on Portuguese NLP, as it enables direct inspection of model behavior on language-specific constructions processed by Transformer-based architectures (Vaswani et al., 2017; Devlin et al., 2019). Potential applications include:

- Analysis of syntactic dependencies and word order effects.
- Support for research on attention-based information extraction.
- Qualitative evaluation of model interpretability.
- Pedagogical use in computational linguistics and deep learning courses.

By focusing explicitly on Portuguese and leveraging language-specific pre-trained models (Souza et al., 2020), the tool contributes to reducing the methodological gap between high-resource and less-resourced languages in interpretability research.

7 Conclusion

This demonstration presents *AttentionApp*, an interactive tool for analyzing attention mechanisms in Transformer models applied to Portuguese. By enabling direct and fine-grained inspection of attention patterns, the system supports exploratory linguistic analysis and interpretability-oriented research. AttentionApp is intended as a practical resource for the Portuguese NLP community and aligns with the goals of the PROPOR Demonstration Track.

8 Acknowledgments

This work was supported by FAPESB (TIC002/2015 and CCE022/2023), CAPES, CNPQ, and INCT-TILDIAR/CNPq (408490/2024-1).

References

Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. 2019. What does BERT look at? an analysis of BERT's attention. In *Proceedings of the BlackboxNLP Workshop at EMNLP 2019*, pages 276–286. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, pages 4171–4186, Minneapolis, MN, USA. Association for Computational Linguistics.

Ricardo G. Oliveira. 2025. Attention app: Interactive visualization and analysis of attention heads in transformer models. `https://github.com/rgoliveirati/attention_app`. Source code repository.

Ricardo G. Oliveira, Daniela B. Claro, and Rerisson Cavalcante. 2025. Syntactic analysis in transformers through attention heads. In *Anais do XVI Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL 2025)*, pages 295–306, Fortaleza, CE, Brazil. Sociedade Brasileira de Computação.

Fabio Souza, Rodrigo F. Nogueira, and Roberto A. Lotufo. 2020. BERTimbau: Pretrained BERT models for brazilian portuguese. In *Intelligent Systems: 9th Brazilian Conference on Intelligent Systems (BRACIS 2020), Part I*, volume 12066 of *Lecture Notes in Computer Science*, pages 403–417, Rio Grande, Brazil. Springer.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008. Curran Associates, Inc.

Elena Voita, Jean Talbot, Fedor Moiseev, Rico Sennrich, and Ivan Titov. 2019. Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. In *Proceedings of ACL 2019*, pages 5797–5808.