Exploring Multimodal Models for Humor Recognition in Portuguese

Marcio Lima Inácio and **Hugo Gonçalo Oliveira**
Centre for Informatics and Systems of the University of Coimbra (CISUC)
Intelligent Systems Associated Laboratory (LASI)
Department of Informatics Engineering
Polo II, Pinhal de Marrocos, 3030-290
Coimbra, Portugal
{mlinacio, hroliv}@dei.uc.pt

Abstract

Verbal humor is commonly mentioned to be a complex phenomenon that requires deep linguistic and extralinguistic forms of knowledge. However, state-of-the-art deep learning methods rely exclusively on the input text, which motivates research on combining the power of LLMs with other types of information, namely humor-related numerical features. In this paper, we explore three methods of multimodal transformers to combine LLM representations with features from the literature and evaluate if they can improve baseline models for Humor Recognition in Portuguese. Our results show that, for BERTimbau-large, the inclusion of humor-related features increased F1-Score by 15.5 percentage points. However, this improvement was not observed for the other models tested. In this context, such an approach can be promising but might require better feature sets, feature combination methods, or some hyperparameter tuning.

1 Introduction

Incorporating creativity and, consequently, humor into computational systems is crucial to developing tools capable of handling complex and deep linguistic phenomena (Reyes et al., 2012). Especially when interpreting humor is considered a sign of fluency closer to that of native speakers, as it requires a profound knowledge of linguistic and extra-linguistic aspects of the text (Tagnin, 2005).

A traditional task in the computational processing of humor is that of Humor Recognition, whose goal is to classify a given text according to the presence of a humorous effect (Taylor and Mazlack, 2004; Potash et al., 2017; Chiruzzo et al., 2021). This is the task we tackle in this paper.

Our investigation focuses mainly on the Portuguese language, for which little research has been done. As most works in Humor Recognition are focused on English — as most research in Natural Language Processing (NLP) (Bender, 2019) — and Spanish, due to the massive efforts of the HAHA series of Shared Tasks (Castro et al., 2018; Chiruzzo et al., 2019, 2021; Rosso, 2023).

Traditionally, Humor Recognition has been tackled through the use of Machine Learning (ML) models with handcrafted feature sets including different kinds of information, such as measurements for ambiguity, imageability, concreteness, named entities, and others (Mihalcea and Strapparava, 2005; Gonçalo Oliveira et al., 2020). More recently, Deep Learning and Large Language Models (LLMs) have been taking over research on Humor Recognition, showing impressive results (Ren et al., 2021; Kumar et al., 2022; Rosso, 2023). However, such models usually only leverage the input text to perform the classification task, which can be limiting when dealing with such a complex phenomenon that is believed to require much extralinguistic knowledge. Thus, some researchers have been investigating ways of improving their performance by combining the power of LLMs with other forms of knowledge via Knowledge Injection (KI) (Zhang et al., 2023) or Multimodal Learning (Gu and Budhkar, 2021).

In this context, our work explores different strategies to combine LLMs with well-established features for Humor Recognition in an attempt to produce better results with such explicit knowledge. We classify a novel corpus of one-line punning jokes in Portuguese that contains 4,903 pairs of manually curated puns and their non-humorous counterparts created via micro-edition. The corpus is publicly available[1], alongside all experiments[2].

From our results, we observed that using such approaches for combining numerical data with LLMs improved the classification performance of

[1] https://anonymous.4open.science/r/Puntuguese-7B67/README.md

[2] https://github.com/Superar/multimodal-humor-recognition

BERTimbau-large, but not for other models. We thus discuss some possible reasons for this kind of approach and point out other kinds of knowledge that might be valuable. To our knowledge, this is the first attempt at tackling Humor Recognition by combining LLMs and humor-related features.

The remainder of the paper is organized as follows. In section 2, we present previous works that motivated this paper. The corpus, feature set, and methods used in our experiments are then described in section 3. Finally, section 4 presents the main results of the experiments followed by the conclusions in section 5. We also acknowledge some limitations of the work in section 6.

2 Related Work

Our work mostly relates to previous research on Humor Recognition in Portuguese, first explored by Clemêncio (2019) and Gonçalo Oliveira et al. (2020), who not only published a corpus of Portuguese jokes but also developed a set of humor-related features based on relevant literature. The authors reached an F1-Score of 80% for one-liners and 76% for satiric headlines.

Subsequently, Inácio et al. (2023) did further experiments on the same corpus achieving 99.6% F1-Score with a fine-tuned BERTimbau model (Souza et al., 2020). The authors also performed a Machine Learning Explainability analysis, which indicated that the model was considering unrelated details, such as punctuation and question words, to perform the classification, i.e., it was not really learning humor or humor-related characteristics. These observations highlight the necessity for a new benchmark for Humor Recognition in Portuguese, that considers such aspects of learning.

This paper is also related to past research on how to merge different pieces of information with LLMs for classification. For instance, Gu and Budhkar (2021) present a toolkit[3] with various techniques to combine categorical and numerical features with transformer-based representations. They also show how including this extra information impacts results in various ML tasks, namely: regression of Airbnb listing prices (Xie, 2019), binary classification of female clothing recommendations (Brooks, 2018), and multiclass classification of pet adoption speeds (Addison et al., 2018). Specifically to the binary classification task (the one that is mostly similar to ours), the authors were able to obtain better results by leveraging both the text and tabular features, even if not with a large margin: from 95.7% when using only the input text to 96.8% with the multimodal approach.

For the sake of completeness, we also mention related initiatives in Portuguese, such as the creation and description of corpora for satire and irony (Carvalho et al., 2009, 2020; Wick-Pedro and Vale, 2020; Wick-Pedro and Santos, 2021).

3 Methods

Our work relies on three main aspects: a novel corpus of Humor Recognition in Portuguese, the feature set used, and the exploration of multimodal approaches to combine texts with explicit features for detecting humor.

3.1 Corpus

As previously mentioned, the currently existing corpus by Gonçalo Oliveira et al. (2020) has some specific details that made ML models associate unrelated aspects of the text with the presence or absence of humor. This motivated the creation of a new corpus of jokes in Portuguese, in which we focused exclusively on a specific format of humor: puns (Miller et al., 2017). To this extent, we manually gathered and curated 4,903 punning texts from multiple sources in Brazilian and European Portuguese.

To overcome the previous concerns described by Inácio et al. (2023), we carried out a process of micro-edition in the gathered jokes, similar to the approach of Hossain et al. (2019) in the creation of their Humicroedit corpus. In our case, the puns were manually edited by 18 fluent speakers of Portuguese from Brazil and Portugal, so that they lost their humorous effect with the least number of modifications. In this sense, the whole classification corpus consists of 9,806 instances and is naturally balanced: for each humorous text, there is a correspondent non-funny text.

With this micro-edition methodology, we expect that learning and evaluation processes better capture the phenomenon of humor, as funny and non-funny pairs differ exclusively in the sense of the witty effect, preserving most surface-level characteristics. An example of a joke in the corpus and its edited counterpart can be seen below:

- Original joke: Um parto não costuma demorar muito tempo. Mas para as grávidas parece

[3] https://github.com/georgian-io/Multimodal-Toolkit

maternidade. (*A childbirth doesn't usually take long. But for pregnant women, it feels like motherhood.*)

- Edited joke: Um parto não costuma demorar muito tempo. Mas para as grávidas parece uma eternidade. (*A childbirth doesn't usually take long. But for pregnant women, it feels like an eternity.*)

We discuss further details about the corpus creation and curation in a separate paper entirely dedicated to this matter. (Inácio et al., 2024)

As tabular features for our methods, we took advantage of the list used by Clemêncio (2019) and Inácio et al. (2023) in their previous works, as the scripts to calculate them are publicly available[4]. The set comprises 27 features, namely: sentiment polarity (3 features) (Silva et al., 2012); number of slangs (1 feature); alliteration as character n-grams (4 features); number of antonymy pairs (1 feature) (Gonçalo Oliveira, 2018); ambiguity as the number of senses of words (2 features) (de Paiva et al., 2012); number of out-of-vocabulary words (1 feature) (Hartmann et al., 2017); incongruity as the semantic similarity of pairs of words (2 features); number of named entities by category and in total (11 features) (Freitas et al., 2010); and imageability and concreteness (2 features) (Soares et al., 2017).

For more extensive details on how exactly the features are defined, we recommend reading the paper by Gonçalo Oliveira et al. (2020), as we do not elaborate on them further due to space limitations.

3.2 Multimodal Large Language Models

For the classification process, we explore the general architecture of multimodal transformers presented by Gu and Budhkar (2021), which is shown in Figure 1.

Within this architecture, we evaluated three different strategies for the combination module. Given that the transformer provides a text representation $\mathbf{x}$ and that the input feature vector is represented by $\mathbf{v}$, the general strategies can be described as: Concatenation ($\mathbf{x}\|\mathbf{v}$); Feature pooling ($\mathbf{x}\|f(\mathbf{v})$); and Shared representation ($f(\mathbf{x}\|f(\mathbf{v}))$).

In the formulas, $f(\cdot)$ represents merely a linear layer with the GELU activation function (Hendrycks and Gimpel, 2016). When used to pool the input features, the dimensionality is maintained, i.e. since we use 27 features, the pooled representation $f(\mathbf{v})$ will have 27 positions as well. For the shared representation, we decided to maintain the same dimension of the underlying transformer model: 768 or 1536 depending on the model.

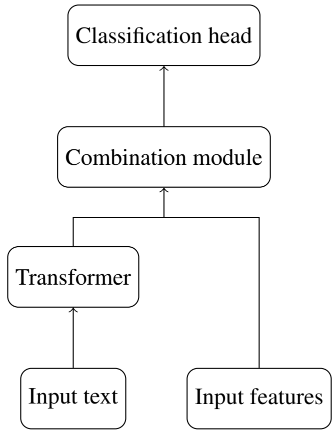

Figure 1: General model of multimodal transformers. Source: adapted from Gu and Budhkar (2021).

For the underlying transformer model, we tested four different pre-trained LLMs for the Portuguese language: BERTimbau-base, BERTimbau-large (Souza et al., 2020), Albertina-900M PT-BR, and Albertina-900M PT-PT (Rodrigues et al., 2023).

For every experiment, to fairly compare the models under the same training conditions, we decided to use the same starting learning rate of 5×10^{-5} with a linear decay. They were fine-tuned for 5 epochs each, identical to Gu and Budhkar (2021). The classification head is the same across every test: a Linear layer with two outputs (as we are in a binary classification scenario) and a softmax function to define the classification probabilities.

4 Results

The models were fine-tuned and tested using 10-fold cross-validation. The splits are stratified concerning the two classes. The results of the classification task are shown in Table 1.

In general, results in the new corpus are not up to par with those obtained by Inácio et al. (2023) in their work (99.6% using BERTimbau base), which is expected, since they used a dataset that had some data leakage flaws, as discussed in their paper. Compared to systems for other languages, our models still underperformed; for Spanish, for instance, Deep Learning systems obtained a median F1-Score of approximately 75% in the HUHU shared task (Rosso, 2023). It is worth mentioning that the data for HUHU was collected from tweets

[4] https://github.com/Superar/HumorRecognitionPT

Model	Multimodality	F1-Score
Albertina-900M PT-BR	$\mathbf{x}$	49.2%
	$\mathbf{x}\Vert\mathbf{v}$	51.1%
	$\mathbf{x}\Vert f(\mathbf{v})$	51.4%
	$f(\mathbf{x}\Vert f(\mathbf{v}))$	51.5%
Albertina-900M PT-PT	$\mathbf{x}$	49.3%
	$\mathbf{x}\Vert\mathbf{v}$	50.0%
	$\mathbf{x}\Vert f(\mathbf{v})$	51.1%
	$f(\mathbf{x}\Vert f(\mathbf{v}))$	52.1%
BERTimbau-base	$\mathbf{x}$	67.0%
	$\mathbf{x}\Vert\mathbf{v}$	67.0%
	$\mathbf{x}\Vert f(\mathbf{v})$	67.8%
	$f(\mathbf{x}\Vert f(\mathbf{v}))$	67.3%
BERTimbau-large	$\mathbf{x}$	53.2%
	$\mathbf{x}\Vert\mathbf{v}$	67.1%
	$\mathbf{x}\Vert f(\mathbf{v})$	68.7%
	$f(\mathbf{x}\Vert f(\mathbf{v}))$	67.2%

Table 1: Results of Humor Recognition systems with different multimodal strategies (average across folds)

and the negative instances were not obtained using the same micro-edition methodology.

Moreover, we observe that, by using only the models with no form of multimodality, the BERTimbau-base (67.0%) is considerably better in this task than every other model. This observation might be because the other models are larger and more complex, so they might require more data or better hyperparameter tuning.

Even though BERTimbau-base was the most successful model on its own, the strategies for combining its knowledge with other kinds of numerical features did not help much in the task.

Regarding Albertina, both PT-BR and PT-PT versions reached similar F1 around 49% to 52%. Introducing humor-related features does not seem to provide much improvement, with a maximum increase of $\approx$5% (from 49.3% to 52.1%).

On the other hand, BERTimbau-large is the most benefited model regarding the introduction of multimodality. Its baseline setup ($\mathbf{x}$) did not perform as well as BERTimbau-base, maybe for the same reasons we mentioned for Albertina. However, by including the numerical features, it can get up to par with its base version (an increase of 29% from 53.2% to 68.7% F1-Score using the features pooling method). It is possible that, with more hyperparameter tweaking, it can even outperform BERTimbau-base.

4.1 Explainability Analysis

To better analyze if the new corpus is more robust against data leakage than the previous one by Gonçalo Oliveira et al. (2020), we used SHAP — the same Machine Learning Explainability tool used by Inácio et al. (2023) — to understand which information the model uses when classifying the texts. To this extent, we ran SHAP on the best model that uses exclusively the textual input ($\mathbf{x}$): BERTimbau-base. Since we carried out our experiments using cross-validation, we selected the model trained in the split that produced the best results for this specific model (70.3% F1); as input examples for SHAP, we used the remaining test fold in this specific split.

From these results, we observed that there are no tokens that concentrate much importance (high absolute value) as in the previous corpus, in which punctuation and question words were clearly more important. In Figure 2, we present the 150 most important tokens (or subtokens, as they are given by BERTimbau's tokenizer) as a word cloud; the more important a token, the larger the word[5]. We note that the importance of a specific token is given by the average absolute Shapley value across every instance in which it occurs.

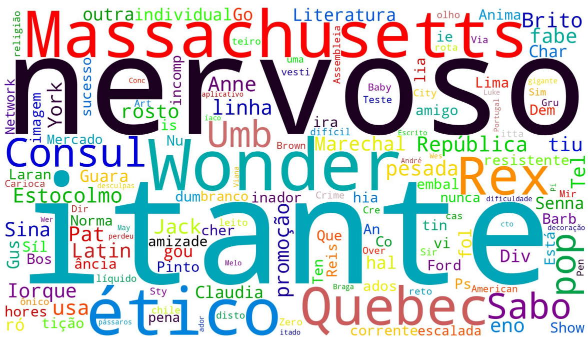

Figure 2: Word cloud with 150 most important tokens for classification

In the cloud, we can see that the most important tokens are usually general words ("nervoso", "Massachusetts", "Quebec", "literatura") and word parts ("itante", "incomp", "ãncia", "tição", "teiro"). From this observation, we can at least state that the model is taking into consideration more textual aspects than with the previous corpus since it is no

[5]We advise caution when interpreting the scale of words. For instance, "nervoso" has a score of 0.485 and "Consul" of 0.464.

longer relying exclusively on punctuation and question words. We also highlight that some of these words are part of the punchline of a joke ("Qual cantor virou desenho da Disney? Stitch Wonder." / "O que escrevem no placar quando o Elvis joga fora de casa? Elvisitante") or are part of the editions made during corpus creation ("Qual é o estado americano que não cai duas vezes no mesmo lugar? Massachusetts." / "Qual é a marca de eletrodomesticos de que os políticos mais gostam? Consul."). However, we expect more deep research to confirm these observations.

5 Conclusions and Future Work

In this paper, we explored methods from multimodal transformers for combining LLMs with numerical features from the literature on Humor Recognition, to take advantage of multiple points of view about the input text. The results show that using such strategies can be fruitful for some underlying models, but are not consistently better across the board. For example, for BERTimbau-large, we improved the 53.2% F1-Score to 68.7% using the feature pooling method, but for other models, this specific method did not result in large improvements.

We also briefly presented a new corpus for Humor Recognition that intends to solve some problems with the currently available corpora. This new dataset consists of 4,903 one-line puns in Brazilian and European Portuguese, each of which is paired with a micro-edited non-humorous counterpart so that we have examples of funny and non-funny texts that differ little in their surface form.

This work opens up various paths for future research. One could investigate if different feature sets are more suitable for this kind of approach, as it seems promising in some contexts or for some models, for instance, an interesting type of knowledge to consider is phonetic transcriptions, especially for punning humor. In the same line of investigating different points of view on the input text, joint learning might be a suitable approach, e.g. a model that jointly learns to identify humor and transcribe it phonetically.

Another natural path for future research is to explore other approaches for the combination module, such as attention and gated strategies (Rahman et al., 2020; Gu and Budhkar, 2021).

Finally, we believe that the new corpus provides a strong benchmark in Portuguese for the evaluation of Humor Recognition systems, as the micro-edition approach makes it more difficult for approaches based solely on surface-level information compared to previous corpora.

6 Limitations

As mentioned during our analysis in section 4, we believe that, for some systems, better hyperparameter tuning can make a difference in the models' performance. In our experiments, however, we tested a single set of parameters that may not be the best possible for either fine-tuning the baseline model or training the whole multimodal pipeline.

Acknowledgements

This work is funded by national funds through the FCT – Foundation for Science and Technology, I.P. (grant number UI/BD/153496/2022), within the scope of the project CISUC (UID/CEC/00326/2020); and also supported by the Portuguese Recovery and Resilience Plan (PRR) through project C645008882-00000055, Center for Responsible AI.

References

Howard Addison, Michael Apers, and Jedi Mongrel. 2018. Petfinder.my adoption prediction.

Emily M. Bender. 2019. The #BenderRule: On Naming the Languages We Study and Why It Matters.

Nick Brooks. 2018. Women's e-commerce clothing reviews.

Paula Carvalho, Bruno Martins, Hugo Rosa, Silvio Amir, Jorge Baptista, and Mário J. Silva. 2020. Situational Irony in Farcical News Headlines. In Paulo Quaresma, Renata Vieira, Sandra Aluísio, Helena Moniz, Fernando Batista, and Teresa Gonçalves, editors, *Computational Processing of the Portuguese Language*, volume 12037, pages 65–75. Springer International Publishing, Cham.

Paula Carvalho, Luís Sarmento, Mário J. Silva, and Eugénio de Oliveira. 2009. Clues for detecting irony in user-generated contents: Oh...!! it's "so easy" ;-). In *Proceeding of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion - TSA '09*, page 53, Hong Kong, China. ACM Press.

Santiago Castro, Luis Chiruzzo, and Aiala Rosá. 2018. Overview of the HAHA Task: Humor Analysis based on Human Annotation at IberEval 2018. In *Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018)*, pages 187–194, Sevilla. CEUR-WS.org.

Luis Chiruzzo, Santiago Castro, Mathias Etcheverry, Diego Garat, Juan José Prada, and Aiala Rosá. 2019. Overview of the HAHA Task: Humor Analysis based on Human Annotation at IberEval 2019. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019)*, pages 132–144, Bilbao. CEUR-WS.org.

Luis Chiruzzo, Santiago Castro, Santiago Góngora, Aiala Rosá, J. A. Meaney, and Rada Mihalcea. 2021. Overview of HAHA at IberLEF 2021: Detecting, Rating and Analyzing Humor in Spanish. *Procesamiento del Lenguaje Natural*, 67:257–268.

André Clemêncio. 2019. *Reconhecimento Automático de Humor Verbal*. MSc, Universidade de Coimbra, Coimbra.

Valeria de Paiva, Alexandre Rademaker, and Gerard de Melo. 2012. OpenWordNet-PT: An open brazilian wordnet for reasoning. In *Proceedings of COLING 2012: Demonstration Papers*, pages 353–360, Mumbai. The COLING 2012 Organizing Committee.

Cláudia Freitas, Cristina Mota, Diana Santos, Hugo Gonçalo Oliveira, and Paula Carvalho. 2010. Second HAREM: Advancing the State of the Art of Named Entity Recognition in Portuguese. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Hugo Gonçalo Oliveira. 2018. A survey on Portuguese lexical knowledge bases: Contents, comparison and combination. *Information*, 9(2).

Hugo Gonçalo Oliveira, André Clemêncio, and Ana Alves. 2020. Corpora and baselines for humour recognition in Portuguese. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1278–1285, Marseille, France. European Language Resources Association.

Ken Gu and Akshay Budhkar. 2021. A Package for Learning on Tabular and Text Data with Transformers. In *Proceedings of the Third Workshop on Multimodal Artificial Intelligence*, pages 69–73, Mexico City, Mexico. Association for Computational Linguistics.

Nathan Hartmann, Erick Fonseca, Christopher Shulby, Marcos Treviso, Jéssica Rodrigues, and Sandra Aluísio. 2017. Portuguese Word Embeddings: Evaluating on Word Analogies and Natural Language Tasks. In *Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*, pages 122–131, Uberlândia.

Dan Hendrycks and Kevin Gimpel. 2016. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*.

Nabil Hossain, John Krumm, and Michael Gamon. 2019. "President Vows to Cut Hair": Dataset and Analysis of Creative Text Editing for Humorous Headlines. In *Proceedings of the 2019 Conference of the North*, pages 133–142, Minneapolis, Minnesota. Association for Computational Linguistics.

Marcio Inácio, Gabriela Wick-pedro, and Hugo Gonçalo Oliveira. 2023. What do humor classifiers learn? An attempt to explain humor recognition models. In *Proceedings of the 7th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 88–98, Dubrovnik, Croatia. Association for Computational Linguistics.

Marcio Inácio, Gabriela Wick-pedro, Renata Ramisch, Luís Espírito Santo, Xiomara S. Q. Chacon, Roney Santos, Rogério Sousa, Rafael Anchiêta, and Hugo Gonçalo Oliveira. 2024. Puntuguese: A corpus of puns in Portuguese with micro-editions. Submitted to LREC-COLING 2024.

Vijay Kumar, Ranjeet Walia, and Shivam Sharma. 2022. DeepHumor: A novel deep learning framework for humor detection. *Multimedia Tools and Applications*, 81(12):16797–16812.

Rada Mihalcea and Carlo Strapparava. 2005. Making computers laugh: Investigations in automatic humor recognition. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 531–538, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Tristan Miller, Christian Hempelmann, and Iryna Gurevych. 2017. SemEval-2017 Task 7: Detection and Interpretation of English Puns. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 58–68, Vancouver, Canada. Association for Computational Linguistics.

Peter Potash, Alexey Romanov, and Anna Rumshisky. 2017. SemEval-2017 Task 6: #HashtagWars: Learning a Sense of Humor. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 49–57, Vancouver, Canada. Association for Computational Linguistics.

Wasifur Rahman, Md Kamrul Hasan, Sangwu Lee, AmirAli Bagher Zadeh, Chengfeng Mao, Louis-Philippe Morency, and Ehsan Hoque. 2020. Integrating Multimodal Information in Large Pretrained Transformers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2359–2369, Online. Association for Computational Linguistics.

Lu Ren, Bo Xu, Hongfei Lin, and Liang Yang. 2021. ABML: Attention-based multi-task learning for jointly humor recognition and pun detection. *Soft Computing*, 25(22):14109–14118.

Antonio Reyes, Paolo Rosso, and Davide Buscaldi. 2012. From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74:1–12.

João Rodrigues, Luís Gomes, João Silva, António Branco, Rodrigo Santos, Henrique Lopes Cardoso, and Tomás Osório. 2023. Advancing neural encoding of portuguese with transformer albertina pt-*.

Roberto Labadie Tamayo y Berta Chulvi y Paolo Rosso. 2023. Everybody hurts, sometimes overview of HUrtful HUmour at IberLEF 2023: Detection of humour spreading prejudice in twitter. *Procesamiento del Lenguaje Natural*, 71(0):383–395.

Mário J. Silva, Paula Carvalho, and Luís Sarmento. 2012. Building a Sentiment Lexicon for Social Judgement Mining. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Helena Caseli, Aline Villavicencio, António Teixeira, and Fernando Perdigão, editors, *Computational Processing of the Portuguese Language*, volume 7243, pages 218–228. Springer Berlin Heidelberg, Berlin, Heidelberg.

Ana Paula Soares, Ana Santos Costa, João Machado, Montserrat Comesaña, and Helena Mendes Oliveira. 2017. The Minho Word Pool: Norms for imageability, concreteness, and subjective frequency for 3,800 Portuguese words. *Behavior Research Methods*, 49(3):1065–1081.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In *Intelligent Systems: 9th Brazilian Conference, BRACIS 2020, Rio Grande, Brazil, October 20–23, 2020, Proceedings, Part I*, pages 403–417, Berlin, Heidelberg. Springer-Verlag.

Stella E. O. Tagnin. 2005. O humor como quebra da convencionalidade. *Revista Brasileira de Linguística Aplicada*, 5(1):247–257.

J.M. Taylor and L.J. Mazlack. 2004. Humorous wordplay recognition. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, pages 3306–3311, The Hague, Netherlands. IEEE.

Gabriela Wick-Pedro and Roney L. S. Santos. 2021. Complexidade textual em notícias satíricas: uma análise para o português do Brasil. In *Anais do XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL 2021)*, pages 409–415, Brasil. Sociedade Brasileira de Computação.

Gabriela Wick-Pedro and Oto Araújo Vale. 2020. Comentcorpus: descrição e análise de ironia em um corpus de opinião para o português do Brasil. *Cadernos de Linguística*, 1(2):01–15.

Tyler Xie. 2019. Melbourne Airbnb Open Data.

Zhengyan Zhang, Zhiyuan Zeng, Yankai Lin, Huadong Wang, Deming Ye, Chaojun Xiao, Xu Han, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2023. Plug-and-Play Knowledge Injection for Pre-trained Language Models.