Procesamiento del Lenguaje Natural, Revista nº 67, septiembre de 2021, pp. 269-276 recibido 05-07-2021 revisado 12-07-2021 aceptado 16-07-2021

Overview of the IDPT Task on Irony Detection in Portuguese at IberLEF 2021

Resumen de la tarea de detección de ironía en portugués (IDPT) en IberLEF 2021

Ulisses B. Corrêa[2,3], Leonardo Coelho[1], Leonardo Santos[4], 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] Sul-Rio-Grandense Federal Institute of Education, Science, and Technology (IFSul), Charqueadas, RS, Brazil
[4] University of São Paulo, São Paulo, SP, Brazil
{larissa, lgcoelho}@inf.ufpel.edu.br, ulissescorrea@ifsul.edu.br, leonardosantos@usp.br

Abstract: This paper presents the Task on Irony Detection in Portuguese (IDPT), held within Iberian Languages Evaluation Forum (IberLEF 2021). We asked the participants to develop systems capable of identifying irony in texts. We created two *corpora* containing tweets and news articles. Twelve teams registered to the task, among which six submitted both predictions and technical reports. The best performing system achieved a Balanced Accuracy (Bacc) value of 0.52 for tweets (Team PiLN) and 0.92 for news (Team BERT4EVER).
Keywords: Irony Detection, Portuguese, Tweets, News.

Resumen: Este artículo presenta la Tarea sobre Detección de Ironías en Portugués (IDPT), realizada en el IberLEF 2021. Les pedimos a los participantes que desarrollaran sistemas capaces de identificar la ironía en los textos. Creamos dos *corpora* que contienen tweets y artículos de noticias. Doce equipos se inscribieron en la tarea, entre los cuales seis presentaron predicciones e informes técnicos. El sistema con mejor rendimiento logró un valor de precisión equilibrada (Bacc) de 0,52 para los tweets (Equipo PiLN) y 0,92 para las noticias (Equipo BERT4EVER).
Palabras clave: Detección de Ironía, Portugués, Tweets, Noticias.

1 Introduction

This is nothing new that, in recent decades, a large part of human communication has taken place on the internet. This way, social networks are essential for disseminating opinions, positions, reflections, debates, and many types of manifestation. As a result, these platforms are also a valuable source of information about public opinion. Therefore, a target of interest for companies, advertising, politics, and research, as pointed out by (Pang and Lee, 2008). In this context, there has been an increase in research interest in text mining on social networks in recent years.

Social networks coexist with other means of online communication and information. If, on the one hand, the immediate communication which characterizes social networks such as Twitter is an important way of disseminating opinions, the contents published by traditional journalism vehicles are also the subject of studies. Journalistic portals publish, daily, thousands of news, reports, reviews, and analyses, among other materials, which circulate quickly in society. All this diverse content is subject to investigation and analysis by researchers and other interested parties.

Manifestations published on more robust platforms such as news portals are broadcast in natural language, as are texts on social networks. In this sense, they include

ISSN 1135-5948. DOI 10.26342/2021-67-23

several linguistic resources and communicational strategies. Therefore, it is common to find sensational statements conveyed in these spaces, even based on irony and other types of figurative language.

As a characteristic of human language, communication on social networks and journalistic portals presents a complexity that challenges investigation techniques and methods. Notably, for research in Artificial Intelligence, the processing of figurative language represents a relevant challenge. Specifically, the frequent use of irony in this genre has important implications for tasks such as sentiment analysis and opinion mining that aim to extract positive and negative opinions automatically from texts.

Interpreting ironic messages is a relatively easy task for humans. However, some *de facto* acts involved in this communicational operation can confuse. Developing Natural Language Processing (NLP) resources that aim to identify irony is crucial to improving several NLP tasks performance, such as Sentiment Analysis and Hate Speech Detection. Therefore, the IDPT – Irony Detection in Portuguese is a task held within the Iberian Languages Evaluation Forum (IberLEF 2021), a comparative evaluation campaign for Natural Language Processing Systems in Iberian languages co-located with the *Sociedad Española de Procesamiento del Lenguaje Natural* (SEPLN) conference.

The IDPT task aims to challenge different teams to propose techniques capable of identifying ironic utterances automatically in news *corpora* published in journalistic portals and tweets published in the microblogging platform Twitter. The present paper presents an overview of the task. First, we briefly present some theoretical reflexions on irony concepts (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 On Irony

Even in works that focus on system development, it is important to review concepts related to the linguistic phenomena explored in NLP research. In this section, we propose briefly present some important reflections on irony. It should be noted that theoretical insights are not our focus in this article.

The irony concept is usually understood as a linguistic resource used with the purpose of expressing the opposite to the literal meaning of an utterance (Cignarella et al., 2018). As discussed in Freitas (Freitas, dos Santos, and Deon, 2020), although many researchers project their efforts on irony studies, there is no consensus on the definition of this linguistic phenomenon. Searle (Searle, 1969) and Grice (Grice, 1975) propose an irony definition as an apparent violation of pragmatic principles. Kreuz and Glucksberg (Kreuz and Glucksberg, 1989) understand that the presence of irony conveys a pragmatic meaning by alluding to expectations (failures or not). According to Reyes (Reyes, Rosso, and Buscaldi, 2012), irony consists of a contradictory property in a given context or event.

Other linguistic phenomena are often associated with the concept of irony. It is common to find discussions about the differences (or not) between irony, sarcasm, satire, and other terms in several areas. Theorists such as Grice (Grice, 1975) and Sperber and Wilson (Sperber and Wilson, 1981) propose different definitions for each phenomenon, while others consider sarcasm and irony the same phenomenon, making no distinctions. They are part of the second study group, such as those by Attardo (Attardo, 2000), Reyes (Reyes, Rosso, and Veale, 2013), and Hee (Van Hee, Lefever, and Hoste, 2016). According to what Gibbs (Gibbs and Colston, 2001) defends, he argues that sarcasm, combined with other linguistic resources such as humor, hyperbole, and rhetoric, can be considered a type of irony. Furthermore, Marchetti (Marchetti, Massaro, and Valle, 2007) argues that irony is an 'umbrella concept' that encompasses other concepts, such as sarcasm and satire.

Although discussions in the field of philosophy of language tend to differentiate phenomena, it is observed that, in Artificial Intelligence, the tendency is to group concepts. Considering that NLP tools aimed at detecting irony, in general, cannot consider the context of production of an utterance, the area focuses on identifying irony from elements internal to the text.

3 Task Description

Inferring ironic meanings is an easy task to humans, yet some of the speech acts involved in this operation might still cause communicational misunderstandings (Freitas et al., 2014). Creating methods to understand ironic text can be a challenging task automatically. Still, it is crucial to improve the performance of other NLP's tasks, e.g., Sentiment Analysis (Gupta and Yang, 2017), and Hate Speech Detection (Bosco et al., 2018).

This task aims to instigate participants to apply their solutions for Irony Detection in Portuguese. Unfortunately, the availability of *corpora* written in Portuguese is scarce, limiting the amount of research done for this language.

This task will contribute to the progress of Portuguese NLP, as there is a demand in the area for the development of new methods and tools. Previous irony detection competitions, such as IDAT (Ghanem et al., 2019), IroSvA (Bueno et al., 2019), IronITA 2018 (Cignarella et al., 2018), and SemEval 2018 Task 3 (Hee, Lefever, and Hoste, 2018), inspired us to develop a specific task for Portuguese.

The task consists of automatically classify the texts (tweets and news) for irony. We propose two independent subtasks:

- Subtask A: Irony detection in Portuguese tweets;
- Subtask B: Irony detection in Portuguese news.

The two subtasks have the same objective: systems should determine whether a message is ironic or not according to a specified context (assigning a binary value 1 or 0).

This task is similar with previous tasks on irony detection at IDAT (Arabic) (Ghanem et al., 2019), IronITA 2018 (Italian) (Cignarella et al., 2018), and SemEval 2018 Task 3 (English) (Hee, Lefever, and Hoste, 2018).

4 Corpora Description and Annotation Process

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

4.1 Datasets Description

The *corpora* proposed for evaluation contain texts (tweets and news) about different topics written in Portuguese. In this task, we used *corpora* previously developed by Freitas (Freitas et al., 2014), Silva (Silva, 2018), and Schubert (Schubert and de Freitas, 2020) for training purposes.

Freitas (Freitas et al., 2014) extracted 2,779 tweets with "Fim do mundo" ("End of the world") expression, between December 19th and 23rd. Silva (Silva, 2018) extracted 12,700 tweets labeled with ironic hashtags (#ironia and #sarcasmo) and 2,700 tweets about the economy, politics, and education - without ironic hashtags, between October 8th and June 10th. Still, in this collection, we removed all retweets. Schubert (Schubert and de Freitas, 2020) extracted ironic news articles from Sensacionalista[1] and The Piauí Herald[2]. Non-ironic news articles came from Estadão[3].

Training data was drawn from public datasets of tweets[4] and news articles[5]. In summary, this training data contain a set of 15,212 tweets (12,736 ironic and 2,476 non-ironic) and 18,494 news articles (7,222 ironic and 11,272 non-ironic).

The test *corpora* were created for this competition through manual annotation of 300 tweets and 300 news articles.

The test dataset is composed of tweets with the ironic hashtags #ironia or #sarcasmo. The remaining tweets talk about the reality show Big Brother Brasil[6], referred through the hashtag #bbb.

Both sets of tweets were joined and shuffled in a single *corpus*. Then the *corpus* was split into three subsets, and each subset was assigned to three annotators.

The test dataset for news comprises 118 ironic news articles from the Diário de Barrelas[7] and 182 non-ironic news articles from the R7 newspaper[8]. Diário de Barrelas is a fictitious newspaper created to satirize the news. However, the R7 newspaper is well

[1] https://www.sensacionalista.com.br/
[2] https://piaui.folha.uol.com.br/herald/
[3] https://www.estadao.com.br/
[4] https://github.com/fabio-ricardo/deteccao-ironia
[5] https://github.com/schuberty/PLNCrawler
[6] https://gshow.globo.com/realities/bbb/
[7] https://www.diariodebarrelas.com.br/category/noticias/
[8] https://www.r7.com/

known nationally and is a source of real news.

The extraction was divided into two steps, the collection of news articles from the Diário de Barrelas and the collection of the latest news articles from R7. Then both sets, ironic and non-ironic, were joined and shuffled using the Python Random package[9].

The shuffled dataset was split into three sets with 100 samples each. We did that to distribute among annotators. Three volunteers annotated each subset.

Table 1 presents statistics about the size of the documents from news and tweets test *corpora*.

The following section presents the Annotation Guidelines that instructed volunteers during the manual annotation process of the tweets and news test sets.

4.2 Annotation Guidelines

IDPT proposed to the participants to classify texts into two categories: ironic and non-ironic.

We provided an Annotation Guide for a set of volunteers annotators (linguists and computer science students) that were responsible for manually classify the testing samples. In this guide, we developed a discussion on irony concepts and gathered some examples (extracted from (Wick-Pedro and Vale, 2020)) of irony-annotated tweets. The volunteers, as the resources developed by the task participants, should classify the texts as Ironic or Non-ironic.

- Non-ironic text: sentences that do not contain linguistic mechanisms that alternate their meaning should be considered non-ironic.

  1 Eu sou a favor da saída da atual Presidente. [I am in favor of the departure of the current President.]

- Ironic text: one must consider ironic the text where there is an opposition of meaning between what is intended and written.

  2 São muito nobres, afinal, a chapa usou dinheiro de corrupção. [They are very noble, after all, the ticked have used money from corruption.]

  3 Chora petezada Chora Venezuela... minha morada (marquise ou viaduto) é de luxo... tá ok?? [Cry Workers Party and Venezuela supporters... my house (marquee or viaduct) is a luxury one... ok???]

In the sentence (1) the speaker expresses his opinion without presenting, in the text, any elements that indicate a contradiction between the explicit opinion and the intended message.

On the other hand, we observe in sentences (2) and (3) elements that contradict the explicit message and the intended one. Thus, in (2), it is possible to infer an ironic meaning because of the opposition between "being noble" and "using money from corruption". Since we know that being noble is good quality and using money from corruption is illegal, we can infer that the sentence is ironic. In (3), there are at least two elements that indicate irony: (i) the opposition between the idea of 'living under a marquee or viaduct' and 'living in a luxury house' and (ii) the use of three trite expressions used in Brazil by president Bolsonaro and his supporters ('*petezada*' and 'Venezuela' for referring to left-wing people, and '*ta ok?*').

The annotation procedure consists of marking each of the statements as ironic or non-ironic. We have not requested teams or even our volunteers' team to note the opposition found.

4.3 Inter-annotator Agreement

Each subset of 100 samples was annotated by three annotators. Based on the annotations of each subset we assessed the inter-annotator agreement using the Fleiss Kappa (Fleiss, 1971). The value of Kappa for each subset is:

- Tweets #0: 0.32
- Tweets #1: 0.36
- Tweets #2: 0.25
- News #0: 0.80
- News #1: 0.94
- News #2: 0.50

As one can see, based in Table 2 interpretation intervals, the inter-annotator agreement for tweets subsets is considered *fair*. In

[9]https://docs.python.org/3/library/random.html

		Characters			Tokens		
		Mean	Min	Max	Mean	Min	Max
News	**Ironic**	1,650	60	2,548	270	10	435
	Non-ironic	1,813	305	10,362	299	49	1,810
Tweets	**Ironic**	124	22	300	19	5	51
	Non-ironic	97	4	259	15	1	44

Table 1: Documents Statistics per *corpus*.

the case of the news articles subsets, we obtained results between *moderated* and *almost perfect*.

Values of κ	Interpretation
$\kappa < 0$	*poor*
$0 < \kappa < 0,2$	*slight*
$0,21 < \kappa < 0,4$	*fair*
$0,41 < \kappa < 0,6$	*moderate*
$0,61 < \kappa < 0,8$	*substantial*
$0,81 < \kappa < 1$	*almost perfect*

Table 2: Fleiss Kappa.

Finally, we considered ironic or non-ironic instances in which at least two annotators agreed, respectively. Considering this criterion, we obtained a corpus with 123 ironic and 177 non-ironic tweets and 115 ironic and 185 non-ironic news.

5 Evaluation Measures

The training set has been released on March 30th, and participants had sixteen days to train their systems. The test set has been released on April 16th, and each participant had twenty days to submit a maximum of three runs.

Participating teams will received 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: Accuracy (Eq. 1), Precision (Eq. 2), Recall (Eq. 3), F1 (Eq. 4), and Bacc (Eq. 6). However, due to the unbalanced datasets we choose the Balanced Accuracy to rank competitors.

$$Accuracy = \frac{True\ Pos + True\ Neg}{Total\ \#Instances} \quad (1)$$

$$Precision = \frac{True\ Pos}{True\ Pos + True\ Neg} \quad (2)$$

$$Recall = \frac{True\ Pos}{True\ Pos + False\ Neg} \quad (3)$$

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

$$Specificity = \frac{True\ Neg}{True\ Neg + False\ Pos} \quad (5)$$

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

6 Participants Systems and Discussion of the Results

Twelve teams registered to the task, among which six submitted both predictions and technical reports. Participants are from universities and companies from four different countries: Brazil, China, Portugal, and Spain.

Participants used either traditional machine learning approaches (Support Vector Machine, MultiLayer Perceptron, Logistic Regression, Naïve Bayes, Random Forest, and others) and/or deep learning methods (Transformers).

Tables 3 and 4 present participants' results for each dataset submitted run. The results are ranked according to the Bacc. For each system, best run is highlighted in bold. Team BERT4EVER, from China, used transformers to achieve a Bacc of 0.92 for news dataset. For the tweets dataset, Team PiLN, from Brazil, used superficial features and SVM to achieve Bacc of 0.52.

Below we summarize the proposed approach of each team:

- **BERT4EVER:** the authors use the BERT model pre-trained by Souza et al. (Souza, Nogueira, and Lotufo, 2020)

Bacc	**Accuracy**	**F1**	**Precision**	**Recall**	**Team**	**Run**
0.92	0.91	0.89	0.83	0.95	**TeamBERT4EVER**	news_3.csv
0.91	0.90	0.88	0.81	0.95	TeamBERT4EVER	news_1.csv
0.90	0.89	0.87	0.81	0.94	TeamBERT4EVER	news_2.csv
0.89	0.89	0.86	0.82	0.91	TeamSiDi-NLP	news_1.csv
0.83	0.82	0.78	0.71	0.87	TeamUFPR	news_2.csv
0.81	0.81	0.77	0.72	0.81	TeamUFPR	news_1.csv
0.80	0.77	0.76	0.64	0.93	TeamPiLN	news_2.csv
0.80	0.79	0.75	0.68	0.85	TeamCISUC	news_3.csv
0.78	0.79	0.73	0.72	0.74	TeamUFPR	news_3.csv
0.78	0.74	0.74	0.59	0.98	TeamGuillemGSubies	news_2.csv
0.78	0.73	0.73	0.59	0.98	TeamGuillemGSubies	news_3.csv
0.76	0.71	0.72	0.57	0.98	TeamGuillemGSubies	news_1.csv
0.71	0.74	0.63	0.69	0.58	TeamPiLN	news_1.csv
0.52	0.51	0.48	0.40	0.60	TeamCISUC	news_2.csv

Table 3: Participants results ranked in terms of Bacc for news dataset.

Bacc	**Accuracy**	**F1**	**Precision**	**Recall**	**Team**	**Run**
0.52	0.47	0.55	0.42	0.80	**TeamPiLN**	tweets_2.csv
0.51	0.46	0.55	0.41	0.80	TeamPiLN	tweets_1.csv
0.51	0.42	0.58	0.41	1.00	TeamCISUC	tweets_1.csv
0.50	0.42	0.58	0.41	0.99	TeamCISUC	tweets_2.csv
0.50	0.41	0.58	0.41	1.00	TeamUFPR	tweets_1.csv
0.50	0.41	0.58	0.41	0.99	TeamCISUC	tweets_3.csv
0.50	0.41	0.58	0.41	1.00	TeamGuillemGSubies	tweets_3.csv
0.50	0.41	0.58	0.41	1.00	TeamGuillemGSubies	tweets_2.csv
0.50	0.41	0.58	0.41	1.00	TeamGuillemGSubies	tweets_1.csv
0.50	0.41	0.58	0.41	1.00	TeamSiDi-NLP	tweets_1.csv
0.49	0.41	0.57	0.40	0.99	TeamUFPR	tweets_2.csv
0.49	0.41	0.57	0.40	0.98	TeamBERT4EVER	tweets_3.csv
0.49	0.40	0.57	0.40	0.99	TeamBERT4EVER	tweets_2.csv
0.48	0.40	0.56	0.40	0.93	TeamBERT4EVER	tweets_1.csv
0.42	0.38	0.46	0.36	0.64	TeamUFPR	tweets_3.csv

Table 4: Participants results ranked in terms of Bacc for tweets dataset.

with three different strategies. Strategy 1: fine-tune the BERT model separately for the training set in each field. Strategy 2: adopt the Loss Weight strategy for the training set in each field to solve data imbalance. Strategy 3: combine both datasets.

- **PiLN:** the authors use superficial features and Support Vector Machine (SVM); embeddings and MultiLayer Perceptron (MLP). In this work, the superficial features are linguistic features such as number of named entities, presence/absence of some symbols, expressions, number of emojis, frequent words, among others. And, the embedding used is Distributed Bag of Words (DBOW).
- **SiDi-NLP:** the authors use BERT model pre-trained by Souza et al. (Souza, Nogueira, and Lotufo, 2020) to classify irony sentences and five machine learning classifiers – SVM, Logistic Regressor (LR), MLP, Random Forest (RF), and Naïve Bayes (NB).
- **CISUC:** the authors use three machine learning classifiers LR, NB, and RF.
- **UFPR:** the authors explore a total of nine strategies in the preprocessing step, four on the feature extraction step (CountVectorizer, TfidfVectorizer, HashingVectorizer, and Word2Vec), and ten algorithms in the learning step (RF,

MLP, SGD, linearSVC, SVC, decisionTree, perceptron, k-nearest neighbors, multinomialNB, and gaussianNB).

- **GuillemGSubies**: the author use HashingVectorizer and RF to classify the documents.

From the six systems presented, four use classical machine learning methods. Just two systems use deep learning methods, BERT4EVER and SiDi-NLP, based on transformers architecture.

7 Final Remarks

Motivated by the necessity of improvements in Irony Detection task focused on Portuguese, we proposed a task within the IberLEF 2021. This paper overviews the first task on irony detection in Portuguese to classify tweets or news as ironic or non-ironic. The datasets have been manually annotated, and the inter-annotator agreement for tweets subsets is considered fair. In the case of the news articles subsets, we obtained results between moderated and almost perfect.

Twelve teams registered to the task, among which six teams, from four countries, submitted both predictions and technical reports.

Classical feature-based models outperformed deep learning methods when applied to the tweets IDPT dataset, achieving a Bacc of 0.52. It is important to note that several systems performed with Bacc values around 0.50 for the tweets task. Without a robust statistical framework, we can not assure the superiority of any of them.

The deep learning methods, based on Transformers and BERT, performed better when dealing with bigger inputs. The BERT4EVER Team achieved a Bacc of 0.92 for the news dataset, while approaches based on classical methods performed, at best, with Bacc of 0.83.

Acknowledgments

The annotators for their essential work. Thanks to Amanda Oliveira, Carolina Gadelha, Felix Silva, João Eduardo Soares, Johnny Ferreira Birck, Luiz Otávio Alves Hammes, and Rodrigo Rodrigues.

Gabriel Schubert and Fábio Ricardo Araújo da Silva by the *corpora* used to create the training dataset.

Johnny Ferreira Birck and Leonardo Coelho by the collection of the samples used in the test dataset.

References

Attardo, S. 2000. Irony as relevant inappropriateness. *Pragmatics*, 32(6):793–826.

Bosco, C., F. Dell'Orletta, F. Poletto, M. Sanguinetti, and M. Tesconi. 2018. Overview of the EVALITA 2018 hate speech detection task. In T. Caselli, N. Novielli, V. Patti, and P. Rosso, editors, *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Turin, Italy, December 12-13, 2018*, volume 2263 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Bueno, R. O., F. M. R. Pardo, D. I. H. Farías, P. Rosso, M. Montes-y-Gómez, and J. Medina-Pagola. 2019. Overview of the task on irony detection in spanish variants. In M. Á. G. Cumbreras, J. Gonzalo, E. M. Cámara, R. Martínez-Unanue, P. Rosso, J. Carrillo-de-Albornoz, S. Montalvo, L. Chiruzzo, S. Collovini, Y. Gutiérrez, S. M. J. Zafra, M. Krallinger, M. Montes-y-Gómez, R. Ortega-Bueno, and A. Rosá, editors, *Proceedings of the Iberian Languages Evaluation Forum co-located with 35th Conference of the Spanish Society for Natural Language Processing, IberLEF@SEPLN 2019, Bilbao, Spain, September 24th, 2019*, volume 2421 of *CEUR Workshop Proceedings*, pages 229–256. CEUR-WS.org.

Cignarella, A. T., S. Frenda, V. Basile, C. Bosco, V. Patti, and P. Rosso. 2018. Overview of the EVALITA 2018 task on irony detection in italian tweets (ironita). In T. Caselli, N. Novielli, V. Patti, and P. Rosso, editors, *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Turin, Italy, December 12-13, 2018*, volume 2263 of *CEUR Workshop Proceedings*. CEUR-WS.org.

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

Freitas, L., L. dos Santos, and D. Deon. 2020. Padrões linguísticos para detecção de ironia em múltiplos idiomas. *Revista Eletrônica de Iniciação Científica em Computação*, 18(4), nov.

Freitas, L. A., A. A. Vanin, D. N. Hogetop, M. N. Bochernitsan, and R. Vieira. 2014. Pathways for irony detection in tweets. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, SAC '14, page 628–633, New York, NY, USA. Association for Computing Machinery.

Ghanem, B., J. Karoui, F. Benamara, V. Moriceau, and P. Rosso. 2019. Idat at fire2019: Overview of the track on irony detection in arabic tweets. In *Proceedings of the 11th Forum for Information Retrieval Evaluation*, FIRE '19, page 10–13, New York, NY, USA. Association for Computing Machinery.

Gibbs, R. W. and H. L. Colston. 2001. *The risks and rewards of ironic communication*. IOS Press.

Grice, H. P. 1975. *Logic and conversation*. Academic Press.

Gupta, R. K. and Y. Yang. 2017. CrystalNest at SemEval-2017 task 4: Using sarcasm detection for enhancing sentiment classification and quantification. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 626–633, Vancouver, Canada, August. Association for Computational Linguistics.

Hee, C. V., E. Lefever, and V. Hoste. 2018. SemEval-2018 task 3: Irony detection in English tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 39–50, New Orleans, Louisiana, June. Association for Computational Linguistics.

Kreuz, R. J. and S. Glucksberg. 1989. How to be sarcastic: The choice reminder theory of verbal irony. *Experimental Psychology*, 118(4):374–386.

Marchetti, A., D. Massaro, and A. Valle. 2007. *Non dicevo sul serio. Riflessioni su ironia e psicologia*. Franco Angeli.

Pang, B. and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135.

Reyes, A., P. Rosso, and D. Buscaldi. 2012. From humor recognition to irony detection: The figurative language of social media. *Data and Knowledge Engineering*, 74:2–12.

Reyes, A., P. Rosso, and T. Veale. 2013. A multidimensional approach for detecting irony in twitter. *Language Resources and Evaluation*, 47(1):239–268.

Schubert, G. and L. A. de Freitas. 2020. A construção de um corpus para detecção de ironia e sarcasmo em português. In *Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional*, pages 709–717, Porto Alegre, RS, Brasil. SBC.

Searle, J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

Silva, F. R. 2018. Detecção de ironia e sarcasmo em língua portuguesa: uma abordagem utilizando deep learning. https://github.com/fabio-ricardo/deteccao-ironia. Acessado em: 25-05-2019.

Souza, F., R. Nogueira, and R. Lotufo. 2020. BERTimbau: pretrained BERT models for Brazilian Portuguese. In *9th Brazilian Conference on Intelligent Systems, BRACIS, Rio Grande do Sul, Brazil, October 20-23*.

Sperber, D. and D. Wilson. 1981. *Irony and the Use - Mention Distinction*. Academic Press.

Van Hee, C., E. Lefever, and V. Hoste. 2016. Guidelines for annotating irony in social media text, version 2.0. Technical report, LT3 Technical Report Series.

Wick-Pedro, G. and O. A. Vale. 2020. Comentcorpus: Description and analysis of irony in a corpus of opinion for brazilian portuguese. *Cadernos De Linguística*, 1(2):1–15.