Sequence-to-sequence and transformer approaches to Portuguese text style transfer

Pablo Botton da Costa and **Ivandré Paraboni**
University of São Paulo (EACH-USP)
Av Arlindo Bettio 1000, São Paulo, Brazil
pablo.costa@usp.br, ivandre@usp.br

Abstract

In Natural Language Generation, text style transfer is the task of rewriting a given source text according to a target style of interest while preserving as much as possible of its meaning. As a means to foster research in this field, this paper presents a range of style transfer models using sequence-to-sequence and transformer architectures alike. In doing so, we would like to compare alternative approaches for the task, and identify opportunities to move towards more robust style transfer in Portuguese.

1 Introduction

Natural language generation (NLG) has experienced considerable progress in recent years with the aid of deep neural network methods applied to sequence learning. Among these, the use of attention mechanisms (Vaswani et al., 2017) in both sequence-to-sequence and transformed-based architectures has been shown to improve the state-of-the-art in a wide range of NLG tasks and applications (Krishna et al., 2022; Garcia et al., 2021; Luo et al., 2019; Wu et al., 2019).

Of particular interest to the present work, in what follows we discuss the issue of *text style transfer*, that is, the data-driven task of rewriting a given source text according to a particular target style of interest whilst preserving as much as possible of its meaning (Jin et al., 2022)[1] The task is usually regarded as an instance of text-to-text generation (Shen et al., 2017; Li et al., 2018) and studies of this kind include, for instance, formality (Wang et al., 2019), sentiment (Luo et al., 2019; Li et al., 2018; Wu et al., 2019), and arbitrary (or non style-specific) transfer (Krishna et al., 2020; Reif et al., 2022).

As elsewhere in the NLG field, research in style transfer is well-developed for the English and a few other languages, with a number of relevant resources (e.g. aligned style corpora, language models, etc.) made available for this purpose. We notice, however, that our target language – Portuguese – still lacks behind in this respect. Based on these observations, this paper uses a purpose-built aligned corpus for style transfer to investigate a range of sequence-to-sequence and transformer models of Portuguese. In doing so, we would like to compare alternatives for the present task, and identify opportunities to move towards more robust style transfer in these scenarios.

The rest of this paper is structured as follows. Section 2 reviews recent work in text style transfer. Section 3 describes how our present aligned corpus has been created. Section 4 introduces the computational models taken into consideration, and Section 5 reports results of our experiment. Finally, Section 6 summarises our present results and suggests future work.

2 Related work

Table 1 summarises recent work in the field of text style transfer, organised according to the kind of transfer task under consideration, the use of parallel corpus, computational approach (s2s = sequence-to-sequence, meta-learning, autoencoder) and evaluation method (I=intrinsic, H=human). Further details are discussed below.

Formality style transfer – the task of rewriting an input text as a more or less formal version – has been addressed in Xu et al. (2012); Rao and Tetreault (2018); Wang et al. (2019) by making use of supervised sequence-to-sequence models. Models of this kind generally follow a similar approach by taking as an input an aligned corpus of sentence pairs (x, y) in which x is the source text and y is the target text rendered in the target style.

Arbitrary style transfer consists of rewriting an input text by modifying any stylistic aspect with the

[1] Thus, we follow Jin et al. (2022) in that style is presently understood as any attribute that varies from source to target texts, and not in its strict linguistic sense.

Study	Transfer type	Parallel?	Method	Evaluation
(Xu et al., 2012)	formality	y	s2s	I, H
(Rao and Tetreault, 2018)	formality	y	s2s	I, H
(Wang et al., 2019)	formality	y	s2s	I, H
(Krishna et al., 2020)	arbitrary	N	s2s	I, H
(Reif et al., 2022)	arbitrary	N	meta-learning	I, H
(Riley et al., 2021)	arbitrary	N	meta-learning	I, H
(Krishna et al., 2022)	multilingual	N	s2s	I, H
(Garcia et al., 2021)	multilingual	y	s2s	I, H
(Hu et al., 2017)	sentiment	N	autoencoder	I, H
(Shen et al., 2017)	sentiment	N	autoencoder	I
(John et al., 2018)	sentiment	N	autoencoder	I
(Fu et al., 2018)	sentiment	N	autoencoder	I, H
(Xu et al., 2018)	sentiment	N	autoencoder	I, H
(Luo et al., 2019)	sentiment	N	autoencoder	I, H
(Li et al., 2018)	sentiment	y	autoencoder	I, H
(Wu et al., 2019)	sentiment	y	autoencoder	I, H

Table 1: Existing work in text style transfer

aid of paraphrases or other non-style specific methods. Studies of this kind, as in Krishna et al. (2020); Riley et al. (2021); Reif et al. (2022), have been mainly applied to scenarios lacking sufficient data in the intended style, and usually make use of large language models (LLMs) in supervised or semi-supervised fashion to create synthetic datasets, in some cases implementing a zero-shot strategy.

Multilingual style transfer focuses on resource-rich languages to perform style transfer in a second, resource-poor alternative, in supervised fashion. For instance, the work in Krishna et al. (2022) introduces a two-stage neural architecture for this purpose. The first stage makes us of an LLM to extract a style vector from the input texts as proposed in Garcia et al. (2021). The second stage generates the text according to a target style based on the differences between style vector pairs according to a GPT model (Brown et al., 2020).

Finally, sentiment transfer consists of rewriting an input text according to a target (e.g., positive or negative) sentiment. Studies as in Hu et al. (2017); John et al. (2018); Fu et al. (2018); Xu et al. (2018); Bao et al. (2019); Luo et al. (2019); Wu et al. (2019) perform the task in unsupervised fashion, once again as a means to overcome the lack of suitable training data for the task.

3 A corpus for style transfer

The kind of style transfer experiment envisaged in our current work requires parallel corpora in the Portuguese language representing two aligned styles, that is, a set of texts in the source style to be modified, and a second set of texts with the same meanings, but written in another target style of interest. Given the difficulties in obtaining a linguistic resource of this type with adequate quality and size, we created, purely for illustration purposes, a synthetic dataset in which source texts are taken from the corpus *UstanceBR* (Pavan and Paraboni, 2022; Pereira et al., 2023), and target texts are obtained by back translation. In other words, target texts were obtained by translating the source texts into a second language, and then translated back to Portuguese, hence constituting an artificial 'back-translated' text style distinct from the source text with presumably minimal meaning alteration.

UstanceBR consists of 47,470 tweets representing favourable and unfavourable attitudes towards six target topics (Lula, Bolsonaro, Sinovac vaccine, Hydroxychloroquine, the church, and Globo TV), and it has been created for the development of stance detection models in Portuguese (e.g., dos Santos and Paraboni (2019); Pavan et al. (2020); Flores et al. (2022); Pavan et al. (2023)). These texts were submitted to back translation in order to create a second version (or a rewrite in a second style) to be used as a target, hereby called *UstanceBrback* corpus. Despite the lexical changes that the method incurs, a number of studies have suggested that back translation is generally capable of preserving meanings across multiple NLP tasks (Wieting et al., 2017; Edunov et al., 2018).

Back translation was performed using the public Google API, which has been shown to obtain satisfactory results for a number of practical purposes (Johnson et al., 2017). Table 2 illustrates the linguistic variation obtained by back-translating the

UstanceBR corpus with the aid of three intermediate languages (Japanese, English and Czech).

Language	Bleu	Edit dist.
Japanese	65.18	66.36
English	81.31	33.06
Czech	72.33	51.39

Table 2: Original and back-translated corpora

Since Japanese provided both the greatest perturbation in the text (as represented by edit distances), and also the best lexical and semantic preservation (as represented by Bleu scores), we chose Japanese as the language for back translation.

As in the case of social media text in general, *UstanceBR* texts are naturally prone to noise. For that reason, we chose to perform a data cleaning step to remove non-standard expressions; symbols and punctuation were normalised, and sentences containing fewer than three words were removed. Table 3 presents descriptive statistics of the original and back-translated corpora.

Corpus	Sent.	Words	Sent. len.	Vocab.
UstanceBR	22,194	551,247	24,28	53,000
UstanceBrback	21,215	468,284	22.08	56,490

Table 3: Corpus descriptive statistics

Results from Table 3 show a 6% variation in number of sentences, and a 9% variation in number of words. This arguably represents a moderate degree of modification in the global corpus features from original to back-translated version.

Finally, we carried out additional post-processing after back translation to remove sentences that did not pass the confidence criteria of the text classifier in Shuyo (2010), which determines whether a piece of text is actually Portuguese. Empty or otherwise ill-formed sentences were also removed. After post-processing, we randomly selected 90% of the aligned corpus ($38,120$ sentence pairs) for training, from which 5% ($2,007$ pairs) were taken as the validation set. The remainder 10% ($4,458$ sentence pairs) makes our test set.

4 Generative models

We implemented 9 generative models for our experiments, divided into two main categories: 7 sequence-to-sequence (hereby s2s) models, an architecture that has been shown to be simple and effective solutions for a range of text generation tasks (Goldberg, 2016; Goodfellow et al., 2016), and 2 transformer-based models that rely on self-attention (Vaswani et al., 2017), and which may be considered closer to the current state-of-the-art in the field. Table 4 summarises these alternatives, and further details are discussed below.

#	Model	Size	Neurons	Layers	Pre-train?
i	s2s+GeA	100	100	2	N
ii	s2s+GeA	200	200	2	N
iii	s2s+GeA	300	300	2	N
iv	s2s+GeA	400	400	2	N
v	s2s+GeA	400	400	4	N
vi	s2s+GeA	300	400	4	y
vii	s2s+GlA	300	400	4	y
viii	tr+MhA	512	400	6	N
ix	PTT5finne	768	3072	12	y

Table 4: Model configurations

Models (i) to (vii) implement the sequence-to-sequence approach with either general – GeA, in (i) to (vi) – or global attention mechanism – GlA, in (vii) – (Bahdanau et al., 2014; Cho et al., 2014), and varying model sizes. In all these cases, we used the architecture described in Bahdanau et al. (2014) with one LSTM network for encoding, and a second network for decoding, varying the embedding size and number of layers.

Model (viii) (tr+MhA) follows the architecture proposed in Vaswani et al. (2017), whereas model (ix) (PTT5finne) fine-tunes the PTT5-base model in Carmo et al. (2020), a Portuguese version of T5 (Raffel et al., 2020) that has been pre-trained on the BrWac corpus (Filho et al., 2018).

Models (vi) and (vii) use pre-trained GloVe embeddings (Pennington et al., 2014) available from Hartmann et al. (2017). For models that do not use word embeddings, we used Xavier initialisation (Glorot and Bengio, 2010). Out-of-vocabulary words were modelled as *UNKNOWN*.

5 Evaluation

Models (i) to (ix) described in the previous section were trained using back-propagation, and were subject to a two-step evaluation procedure. Each evaluation step used a different set of evaluation metrics as discussed below. In both steps, we used Adam optimiser with an initial learning rate of 0.001 for 600 epochs, and a mini batch size of 256.

In the first step, we identified the three top-performing models according to their validation results by measuring accuracy and perplexity. This choice of evaluation metrics was motivated by the need to identify those models that are most capable of preserving both lexicality and semantics whilst maximising vocabulary variation. Thus, accuracy is intended to measure the mean lexical assertiveness of the text, and perplexity is intended to capture the correlation between generated text and input vocabulary. Table 5 shows perplexity and accuracy results over the validation dataset.

#	Model	Perplexity	Accuracy
i	s2s+GeA	115.4	37.01
ii	s2s+GeA	115.4	37.01
iii	s2s+GeA	30.46	43.59
iv	s2s+GeA	42.09	43.91
v	s2s+GeA	43.39	43.91
vi	s2s+GeA	40.92	44.63
vii	s2s+GlA	31.03	42.23
viii	tr+MhA	9.42	53.12
ix	PTT5finne	**4.00**	**70.12**

Table 5: Validation results

Results from Table 5 suggest that the transformer-based models (viii) (tr+MhA) and (ix) (PTT5finne) outperform the alternatives, and that the latter was the best of all.

The three models with lowest perplexity (iii, viii, and ix), were further assessed for their ability to generalise over the test data. To this end, we measured edit-distance, Bleu, BERT score (BERT.sc) (Zhang et al., 2020) and cosBERT. The choice for edit-distance is intended to measure lexical similarity. The choice for Bleu (Papineni et al., 2002) is motivated by the need to capture both lexical and syntactical similarities by measuring the degree of n-gram overlap. BERT.sc is intended to represent semantic similarity, and cosBERT is intended to represent word-level semantic (cosine) similarity using BERTimbau (Souza et al., 2020). Table 6 summarises the test results for the three selected models.

#	Model	Edit d.	Bleu	BERT.sc	cosBert
iii	s2s+GeA	68.93	53.66	0.38	0.72
viii	tr+MhA	93.33	21.48	0.08	0.37
ix	PTT5-finne	58.83	68.59	0.56	0.84

Table 6: Best-performing models.

As expected, results from Table 6 show once again that model (ix) (PTT5finne) generally outperforms the alternatives. As a means to further assess PTT5finne, Table 7 provides more fine-grained Bleu results according to target topic (e.g., Lula, Bolsonaro, etc.) and stance polarity (for or against).

Target	For	Against	Overall
Lula	73.18	71.74	72.38
Bolsonaro	53.89	47.89	50.51
Sinovac	73.42	73.09	73.27
Hydrox.	74.23	72.35	73.42
Church	71.74	71.74	71.74
Globo TV	67.75	67.39	67.39

Table 7: PTT5finne Bleu score results per class

Generally speaking, PTT5finne displays uniform results across target topics and polarity. As a means to illustrate the kinds of output text produced by PTT5finne, we randomly selected three test samples representing low, moderate and high generation error levels according to their closeness to the corresponding target text. These samples are presented below using only their original Portuguese format as translating them would obscure the kinds of error made by the generative model, and therefore rendering the analysis unhelpful.

(low error level)
target: *vou te levar para a igreja*
generated: *eu vou te levar para a igreja*

(moderate error level)
target: *a avó do meu irmão está morrendo de vontade de me levar à igreja ela ficará surpresa quando descobrir que sou ateu*
generated: *a avó do meu irmão está com vontade de me levar para a igreja ela fica surpreso quando eu descobrir que sou ateu*

(high error level)
target: *concordo é um deputado é um médico e se opõe a bloqueios ele é a favor da cloroquina ajudou no combate ao hn tem todos os requisitos para o cargo melhor nome que temos atualmente outros nomes faltam experiência política e precisam estar alinhados com o presidente*
generated: *aceito ele era ajudante médico se opôs ao bloqueio a favor da cloroquina e ajudou a combater o hn existem todos os requisitos para uma posição o melhor nome que temos agora outros nomes não têm experiência política e devem ser iguais ao do presidente*

We notice that some errors stem from originally ill-formed texts, as in the high error level example. Other issues seem to be related to sentence length, which makes generation increasingly complex and more prone to hallucination.

6 Final remarks

This paper reported a first experiment in text style transfer for Portuguese text generation using a back-translated aligned corpus as an hypothetical example of target style. Results suggest that a transformer-based model outperforms sequence-to-sequence alternatives according to several intrinsic evaluation metrics.

As future work, we intend to allow further linguistic variation by replacing the current method for a paraphrase-based strategy as in Krishna et al. (2020); Wieting et al. (2021), and substitute the current 'artificial' target style for an actual style obtained from aligned corpora of real language use. Moreover, we intended to use more robust LLMs as a means to reduce hallucination and improve grammaticality, and carry out a more detailed evaluation work with the aid of human judges.

Acknowledgements The present research has been supported by the São Paulo Research Foundation (FAPESP grant #2021/08213-0).

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Yu Bao, Hao Zhou, Shujian Huang, Lei Li, Lili Mou, Olga Vechtomova, Xinyu Dai, and Jiajun Chen. 2019. Generating sentences from disentangled syntactic and semantic spaces. *arXiv preprint arXiv:1907.05789*.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems*, pages 1877–1901.

Diedre Carmo, Marcos Piau, Israel Campiotti, Rodrigo Nogueira, and Roberto Lotufo. 2020. Ptt5: Pretraining and validating the t5 model on brazilian portuguese data. *arXiv preprint arXiv:2008.09144*.

Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder–decoder approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111, Doha, Qatar. Association for Computational Linguistics.

Wesley Ramos dos Santos and Ivandré Paraboni. 2019. Moral Stance Recognition and Polarity Classification from Twitter and Elicited Text. In *Recents Advances in Natural Language Processing (RANLP-2019)*, pages 1069–1075, Varna, Bulgaria. INCOMA Ltd.

Sergey Edunov, Myle Ott, Michael Auli, and David Grangier. 2018. Understanding back-translation at scale. *arXiv preprint arXiv:1808.09381*.

Jorge A. Wagner Filho, Rodrigo Wilkens, Marco Idiart, and Aline Villavicencio. 2018. The brWaC corpus: A new open resource for Brazilian Portuguese. In *11th International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Arthur Marçal Flores, Matheus Camasmie Pavan, and Ivandré Paraboni. 2022. User profiling and satisfaction inference in public information access services. *Journal of Intelligent Information Systems*, 58(1):67–89.

Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. 2018. Style transfer in text: Exploration and evaluation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32:1.

Xavier Garcia, Noah Constant, Mandy Guo, and Orhan Firat. 2021. Towards universality in multilingual text rewriting. *arXiv preprint arXiv:2107.14749*.

Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings.

Yoav Goldberg. 2016. A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57:345–420.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. 2016. *Deep learning*, volume 1:2. MIT press Cambridge.

Nathan Hartmann, Erick Fonseca, Christopher Shulby, Marcos Treviso, Jessica Rodrigues, and Sandra Aluisio. 2017. Portuguese word embeddings: Evaluating on word analogies and natural language tasks. *arXiv preprint arXiv:1708.06025*.

Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P Xing. 2017. Toward controlled generation of text. In *International conference on machine learning*, pages 1587–1596. PMLR.

Di Jin, Zhijing Jin, Zhiting Hu, Olga Vechtomova, and Rada Mihalcea. 2022. Deep learning for text style transfer: A survey. *Computational Linguistics*, 48(1):155–205.

Vineet John, Lili Mou, Hareesh Bahuleyan, and Olga Vechtomova. 2018. Disentangled representation learning for non-parallel text style transfer. *arXiv preprint arXiv:1808.04339*.

Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Kalpesh Krishna, Deepak Nathani, Xavier Garcia, Bidisha Samanta, and Partha Talukdar. 2022. Few-shot controllable style transfer for low-resource multilingual settings. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7439–7468, Dublin, Ireland. Association for Computational Linguistics.

Kalpesh Krishna, John Wieting, and Mohit Iyyer. 2020. Reformulating unsupervised style transfer as paraphrase generation. In *Empirical Methods in Natural Language Processing*.

Juncen Li, Robin Jia, He He, and Percy Liang. 2018. Delete, retrieve, generate: a simple approach to sentiment and style transfer. *arXiv preprint arXiv:1804.06437*.

Fuli Luo, Peng Li, Jie Zhou, Pengcheng Yang, Baobao Chang, Zhifang Sui, and Xu Sun. 2019. A dual reinforcement learning framework for unsupervised text style transfer. *arXiv preprint arXiv:1905.10060*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

Matheus Camasmie Pavan, Vitor Garcia dos Santos, Alex Gwo Jen Lan, Jo ao Trevisan Martins, Wesley Ramos dos Santos, Caio Deutsch, Pablo Botton da Costa, Fernando Chiu Hsieh, and Ivandré Paraboni. 2023. Morality classification in natural language text. *IEEE transactions on Affective Computing*, 14(1):857–863.

Matheus Camasmie Pavan, Wesley Ramos dos Santos, and Ivandré Paraboni. 2020. Twitter Moral Stance Classification using Long Short-Term Memory Networks. In *BRACIS-2020 proceedings LNAI 12319*, pages 636–647. Springer.

Matheus Camasmie Pavan and Ivandré Paraboni. 2022. Cross-target stance classification as domain adaptation. In *Advances in Computational Intelligence - MICAI 2022 - Lecture Notes in Artificial Intelligence vol 13612*, pages 15–25, Cham. Springer Nature Switzerland.

J. Pennington, R. Socher, and C. D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of EMNLP-2014*, pages 1532–1543.

Camila Pereira, Matheus Pavan, Sungwon Yoon, Ricelli Ramos, Pablo Costa, Laís Cavalheiro, and Ivandré Paraboni. 2023. UstanceBR: a multimodal language resource for stance prediction. *arXiv:2312.06374*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Sudha Rao and Joel Tetreault. 2018. Dear sir or madam, may i introduce the gyafc dataset: Corpus, benchmarks and metrics for formality style transfer. *arXiv preprint arXiv:1803.06535*.

Emily Reif, Daphne Ippolito, Ann Yuan, Andy Coenen, Chris Callison-Burch, and Jason Wei. 2022. A recipe for arbitrary text style transfer with large language models. In *60th Annual Meeting of the Association for Computational Linguistics*, pages 837–848, Dublin, Ireland. Association for Computational Linguistics.

Parker Riley, Noah Constant, Mandy Guo, Girish Kumar, David Uthus, and Zarana Parekh. 2021. TextSETTR: Few-shot text style extraction and tunable targeted restyling. In *59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 3786–3800, Online. Association for Computational Linguistics.

Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2017. Style transfer from non-parallel text by cross-alignment. *Advances in neural information processing systems*, 30.

Nakatani Shuyo. 2010. Language detection library for java.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. BERTimbau: pretrained BERT models for Brazilian Portuguese. In *9th Brazilian Conference on Intelligent Systems, BRACIS*.

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. Curran Associates, Inc.

Yunli Wang, Yu Wu, Lili Mou, Zhoujun Li, and Wenhan Chao. 2019. Harnessing pre-trained neural networks with rules for formality style transfer. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3573–3578.

John Wieting, Kevin Gimpel, Graham Neubig, and Taylor Berg-Kirkpatrick. 2021. Paraphrastic representations at scale. *arXiv preprint arXiv:2104.15114*.

John Wieting, Jonathan Mallinson, and Kevin Gimpel. 2017. Learning paraphrastic sentence embeddings from back-translated bitext. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 274–285, Copenhagen, Denmark. Association for Computational Linguistics.

Chen Wu, Xuancheng Ren, Fuli Luo, and Xu Sun. 2019. A hierarchical reinforced sequence operation method for unsupervised text style transfer. In *57th Conference of the Association for Computational Linguistics, ACL 2019*, pages 4873–4883, Florence, Italy. Association for Computational Linguistics.

Jingjing Xu, Xu Sun, Qi Zeng, Xuancheng Ren, Xiaodong Zhang, Houfeng Wang, and Wenjie Li. 2018. Unpaired sentiment-to-sentiment translation: A cycled reinforcement learning approach. *arXiv preprint arXiv:1805.05181*.

Wei Xu, Alan Ritter, William B Dolan, Ralph Grishman, and Colin Cherry. 2012. Paraphrasing for style. In *Proceedings of COLING 2012*, pages 2899–2914.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.