Predicting Political Orientation in News with Latent Discourse Structure to Improve Bias Understanding

Nicolas Devatine[1], **Philippe Muller**[1,3], **Chloé Braud**[2,3]
[1]IRIT, University of Toulouse
[2]IRIT, CNRS
[3]Artificial and Natural Intelligence Toulouse Institute (ANITI)

firstname.lastname@irit.fr

Abstract

With the growing number of information sources, the problem of media bias becomes worrying for a democratic society. This paper explores the task of predicting the political orientation of news articles, with a goal of analyzing how bias is expressed. We demonstrate that integrating rhetorical dimensions via latent structures over sub-sentential discourse units allows for large improvements, with a +7.4 points difference between the base LSTM model and its discourse-based version, and +3 points improvement over the previous BERT-based state-of-the-art model. We also argue that this gives a new relevant handle for analyzing political bias in news articles.

1 Introduction

Misinformation is a major threat on modern democracy, influencing political agendas in an arguably unfair way, through multiple sources that are more or less transparent in their orientations. Biased media can influence public opinion by selecting reported facts and angles, oriented presentation of events, with a proven impact, e.g. on electoral behaviours (DellaVigna and Kaplan, 2007) or public health (Simonov et al., 2020). The automatic identification of such biases can thus help more transparent and democratic sharing of information, and the understanding of its typical expression.

The study of bias has generated a lot of interest in political sciences with some emphasis on its linguistics aspects (Lee and Solomon, 1990; Levasseur, 2008), which also gave rise to numerous studies on automating bias detection (Hamborg et al., 2019). NLP approaches mostly rely on lexical information (Recasens et al., 2013), or syntax (Iyyer et al., 2014), with, recently, the use of pretrained language models (Baly et al., 2020) or document-level bias distribution (Chen et al., 2020).

Bias can be expressed in more subtle ways however. In the excerpts below (Figure 1), discussing the 2019 Virginia Beach mass shooting[1], we can clearly identify the difference in coverage, with specific lexical choices ("epidemic", "refuse to cover") but also different ways of presenting the event: the style is either descriptive (BBC) or more emotional (WP, Townhall) ; the writer insists on particular topics or angles (*use of silencers*, *weapon prohibition*). The choice of topics is indeed an important aspect of information manipulation (Scheufele and Tewksbury, 2007), and has also generated NLP work, still lexically focused (Card et al., 2015; Baumer et al., 2015; Field et al., 2018; Morstatter et al., 2018)

"The Virginia Beach shooter put a sound suppressor (...) so that the death shots were muffled, perhaps denying others the warning that would have allowed them to escape. It is long past time to remove the silencer that seems to suppress action on gun-control legislation, to treat mass shooting as the epidemic it is, and do everything possible to save lives.'" **(Washington Post, left-leaning)**

"The attack began shortly after 16:00 (20:00 GMT), at Virginia Beach Municipal Center, in an area which is home to a number of city government buildings. The area was put into lockdown by police and employees were evacuated. 'We just heard people yelling and screaming at people to get down,' Megan Banton, an administrative assistant in the building, told local television news station WAVY." **(BBC, center)**

"The chilling fact is that mass public killers are attracted to targets where people can't defend themselves. (...) Ninety-eight percent of US mass public shootings since 1950 have occurred in places where people weren't allowed to defend themselves. But the news media refuses to cover this fact, which illustrates the need for self-defense, not for more gun control that doesn't work." **(Townhall, right-leaning)**

Figure 1: Excerpts from articles on the 2019 Virgina Beach mass shooting from media with different political tendencies.

In contrast, we investigate the task of predicting political orientation of news articles, while trying to consider global argumentative aspects instead of local, lexical ones. This classification task consists in predicting the political leaning of an article by considering, in our case, 3 political classes (left, center, right). Since text-level discourse analysis

[1]https://www.allsides.com/blog/virginia-beach-shooting-reinvigorates-gun-debate

Proceedings of 3rd Workshop on Computational Approaches to Discourse, pages 77–85
October 16 – 17, 2022. ©2022 International Conference on Computational Linguistics

is still a difficult problem (Zhang et al., 2020), our architecture encodes a document while automatically inducing latent structural dependencies as in Liu and Lapata (2018), with a focus on elementary discourse units instead of sentences. We hypothesize that structural information can help identify political sides and give insights into aspects related to the argumentative nature of different media.

We evaluate our approach on news articles (Baly et al., 2020) and also perform a preliminary interpretability study. Our contributions are: (i) a model predicting political orientation of news articles, inducing a latent structure over discourse segmented texts, with state-of-the-art results ; (ii) a preliminary analysis of the impact of lexical and structural information for bias detection. Our code is available at: `https://github.com/neops9/news_political_bias`.

2 Related work

There are multiple ways to consider the task of classifying political ideologies, especially by varying the number and type of classes, and the level of analysis. For example, one SemEval 2019 shared task focused on identifying hyperpartisan articles (Kiesel et al., 2019). Political bias can also be characterized by locating "propaganda techniques" in texts, as in the SemEval 2020 shared task (Da San Martino et al., 2020). Here, we consider the task proposed by Baly et al. (2020) based on 3 classes (left, center, right). A similar task was also considered in Li and Goldwasser (2021), but their dataset is not available for comparison. In addition, Baly et al. (2020) explore methods that prevent the model from using media-related information while remaining based on other lexical and syntactic ones (see section 3). They report at best 51.41% in accuracy.

Contrary to previous studies based solely on lexico-syntactic information, we hypothesize that document-level organization is crucial. Rather than relying on low-performing discourse parsers, we test Liu and Lapata (2018)'s approach: structural dependencies over sentences are induced while encoding the document. Their results indicate that the learned representations, without ever exposing the model to linguistic annotations or an external parser, achieve competitive performance on a range of tasks while arguably being meaningful. This approach is effective for summarization with the learned structures, while less complex than classical ones, capturing consistent information (Liu et al., 2019; Isonuma et al., 2019; Balachandran et al., 2021). A similar approach was shown to be effective for detecting fake/real news articles (Karimi and Tang, 2019). While focused on discourse-level phenomena, previous studies use sentences as basic units. We experiment with a fine-grained level, discourse segments, provided by a state-of-the-art segmenter.

3 Model

In Liu and Lapata (2018), the sentences in each document are composed of sequences of static word embeddings that are fed to a bi-LSTM to obtain hidden representations used to compute the sentence representations, that are then passed through another bi-LSTM to compute the document representation. At both levels, representations are built using the structured attention mechanism allowing for learning sentence dependencies, constrained to form a non-projective dependency tree. Finally a 2-layer perceptron predicts the distribution over class labels.

We modify the model to include the improvements proposed by Ferracane et al. (2019). In particular: (i) we remove the document-level bi-LSTM, (ii) for the pooling operation, we aggregate over units using a weighted sum based on root scores, instead of a max pooling, (iii) we perform several additional levels of percolation to embed information from the children's children of the tree, and not only direct children.

On top of that, we skip the sentence-level structure attention as it adds an unnecessary level of composition that was found to have a negative empirical impact on the results.

Segmentation The learning of a latent structure is supposed to let the model leverage rhetorical and argumentative processes that can reflect the author's political orientation. We change the relevant textual units from sentences to more discourse-oriented ones, as given by a discourse segmenter (Muller et al., 2019). Discourse segmentation is the first stage of discourse parsing, identifying text spans called Elementary Discourse Units (EDU) that will be linked by discourse relations.

Adversarial Adaptation Some specific cues (e.g. media name, common patterns) can reveal the media source. Since most articles from a media share the same political label, the classifier decisions are biased towards the source and models easily overfit the training set. But removing these cues is a costly, hard to generalize preprocessing step. Baly et al. (2020) suggest two approaches: adversarial adaptation, or AA (Ganin et al., 2016), and triplet loss pre-training (Schroff et al., 2015), and chose the latter based on preliminary results. On the contrary we found AA more promising: it works by adding a media classifier within the architecture whose loss will be maximized using a gradient reversal layer. The model thus learns to be discriminative for the main task while being media independent.

As the training set contains many media sources, with a long tail distribution, we only consider the 10 most frequent sources (74% of the data) for the adversarial part of the model.

4 Dataset and Settings

Allsides Dataset The articles are crawled from the Allsides website,[2] with 192 news sources covering 109 topics. Allsides is a platform that offers an analysis of the political leaning of various English-language media at the article level by annotating them with 5 political classes that cover the whole political spectrum from the Left to the Right. The published version of the dataset[3] used in Baly et al. (2020) does not match their paper as it includes resp. $2,817$ and 119 additional articles and media. Although it complicates results comparison, we kept the published dataset which is large and seems well designed.[4] This dataset comes with two organizations: article-based or media-based. We chose the latter ($30,246$ articles) where media present at training time are excluded from evaluation, which avoids evaluating the model on articles that come from media already seen during training. For complexity reasons, we removed from the training set the longest articles in terms of number and size of segments, using a threshold of 100. The final dataset contains $27,146$ articles, see Table 1. Note that the original Allsides data are divided into 5 classes, but Baly et al. (2020) merged the two Left (resp. Right) classes.

Segmentation We kept the pre-processed data as in Baly et al. (2020) but we experimented with both sentence- and EDU-segmented texts (see Section 3). We rely on the DISRPT2019 shared task winner (Muller et al., 2019) that only needs plain text as input.[5] The model is based on the BERT pretrained transformer language model, fine-tuned for sequence tagging on plain documents from the GUM corpus (Zeldes, 2016), the English dataset which has the most varied document types. We end up with an average of 49 EDUs per article, and an average of 19 words per EDU.

	Left	Center	Right	Total
Train	$9,618$ (41%)	$6,683$ (28%)	$7,189$ (31%)	$23,490$
Valid.	98 (4%)	618 (26%)	$1,640$ (70%)	$2,356$
Test	599 (46%)	299 (23%)	402 (31%)	$1,300$

Table 1: Statistics about the dataset (media-based split).

Settings We built on Ferracane et al. (2019)'s implementation,[6] itself based on Liu and Lapata (2018)'s. We adapted the code according to the modifications and additions proposed in our approach as detailed in Section 3. Hyper-parameters were set using grid search: 200 for the hidden size of bi-LSTM and 2-layer perceptron, 0.01 for learning rate, 0.5 for dropout and 8 for batch size. We used pretrained 300D GloVe vectors. For Adversarial Adaptation, best results used a weighting factor $\lambda = 0.7$ for the adversarial part of the loss. Training is done with Adagrad optimizer, on a Nvidia GeForce GTX 1080 Ti GPU card.

Evaluation We evaluate four versions: (i) keeping only the bi-LSTM (Ours Base), (ii) full architecture with structural attention and sentence segmentation (Ours+SA/Sent), (iii) full architecture but with EDU segmentation (Ours+SA/EDU), and (iv) full architecture but keeping only the first 512 tokens of each text as in Baly et al. (2020) (Ours: 512t, +SA/EDU). We report standard measures but also the mean absolute error (MAE) as this is an ordinal problem. We compare to scores reported in Baly et al. (2020) on the same split for their LSTM and BERT versions (limited to 512 tokens).

5 Results and Analysis

Results obtained by the different models are given in Table 2. We also report scores per class in Table 3 (best model) to control that the model does

[2] http://allsides.com/
[3] https://github.com/ramybaly/Article-Bias-Prediction
[4] Note that the original version is not available.
[5] Recent approaches reported improvements (Zeldes et al., 2021), but require more preprocessing, e.g. syntactic parses.
[6] https://github.com/elisaF/structured/

Model	Acc.	Macro F_1	MAE
Ours Base	46.97	44.41	0.69
Ours+SA/Sent	48.76	45.84	0.67
Liu&Lapata+SA/EDU	51.01	48.61	0.72
Ours+SA/EDU	54.39	51.36	0.57
Ours: 512t, +SA/EDU	50.04	45.23	0.70
Baly 20: 512t, LSTM	46.42	45.44	0.62
Baly 20: 512t, BERT	51.41	48.26	0.51

Table 2: Accuracy%, macro-F_1%, Mean Absolute Error (MAE, lower is better) on the test set for different versions of the model. "Baly 20" refers to the results reported in Baly et al. (2020), we did not replicate their experiments. "512t" means that only the first 512 tokens of the inputs were used to train the model. "SA" = for Structured Attention, and "Sent"/"EDU" is for inputs segmented in sentence or discourse units. We also evaluate on the original model proposed by Liu and Lapata (2018) without the improvements added in our version of the model. The 95% binomial proportion confidence interval for the best model classification accuracy is 2.9%

not overpredict most represented classes. We observe significant differences in performance between models that use structured attention ("+SA") gaining about 7.4 points in accuracy and 6 in macro F_1 for the best version (+SA/EDU). Our full model, using GloVe, obtains higher scores that those reported in Baly et al. (2020) (LSTM version), +8 points acc. and +6 in F_1, and also a +3 improvement in both over their best BERT-based system.

We performed a control experiment on the size of the input as Baly et al. (2020) only consider the first 512 tokens of the articles, as this is a hard constraint on the BERT model. Reducing the input size (line 4 in Table 2) decreases model performance, showing the importance of considering the whole text and which represents an important limitation of BERT. The experiments with EDUs show the importance of having fine-grained level discourse phenomena: SA based on sentences only improves results by less than 2 points, while SA based on EDUs is much more efficient. In addition, we show the benefits of modifications made to the implementation of Liu and Lapata (2018) that include those proposed by Ferracane et al. (2019) with a +3 points improvement in accuracy. The detailed results by class show that our approach does not overspecialize, although the center class is harder to predict.

As said above, dataset differences and the lack of detailed results per class means the comparison with Baly et al. (2020) should be considered with caution. In particular, since they do not yet provide an implementation to replicate their experiments, we cannot control the overspecialization issues.

Side	Prec.%	Recall%	F_1%
Left	67.39	27.19	38.75
Center	39.59	72.76	51.28
Right	66.53	61.74	64.05

Table 3: Scores per class (best model): Ours+SA/EDU.

Regarding biases towards the topics covered, we rely on the analysis by Baly et al. (2020) for their dataset: they showed that topics covered are fairly represented in each class and thus that it should not significantly impact the model decisions.

We also want to give here some insights into the model by an analysis with interpretablity methods at the lexical level but also with respect to the induced structure.

Saliency Map A saliency map in NLP is a method for visualizing a deep learning model by computing relative importance of each token (word) in the input based on gradients (Ribeiro et al., 2016; Murdoch et al., 2018). It allows us to identify the lexical cues that provide partial understanding of the decisions made by the model. Here, we considered the vanilla gradient approach (Simonyan et al., 2014), focusing on the gradient of the loss with respect to each token embedding. From these, we can first clearly assess the positive impact of the AA method. Lexical cues used by the model without AA, such as the name of the media source, are no longer as relevant for the prediction, although still present. We notice that the model focuses on specific lexical fields depending on the political orientation of the article, such as health, numbers/statistics, economy, for left, center and right leaning articles respectively. We found that crucial information for the model are the mentions or quotes of political figures (e.g., Donald Trump, Hillary Clinton, @realDonaldTrump, Barack) by media sources of the same political side, but they also represent an important source of errors when it appears in articles of the opposite side as the model tends to use this information alone without considering its context.

It also confirms our intuition that there is relevant information in the middle and at the end of articles, even though the model usually focuses on small portions of text, and it explains why reducing the entry size results in a loss of performance. An example is provided in appendix A as heatmap.

Structured Attention Regarding structured attention, we extracted the maximum spanning trees from the attention scores using the Chu-Liu-Edmonds algorithm (Chu and Liu, 1965; Edmonds, 1967). An example of dependency tree is given in appendix A. For a first qualitative analysis, we looked at some statistics following Ferracane et al. (2019) methodology. In particular, we measure the average height of trees (10.68), the average proportion of leaf nodes (0.77), and the average normalized arc length (0.35). Statistics per class are equivalent. The learned trees have complex (non-flat) structures which show that relevant information to the model has been encoded in them in contrast to the results obtained by Ferracane et al. (2019). We observed that they have marked differences with "natural" structures, such as distant links and it could be interesting to add more constraints.

6 Conclusion

We proposed an original approach for predicting the political orientation of newspaper articles based on learning a latent structure showing the importance of considering elementary discourse units over sentences to include the argumentative dimension, allowing for large improvements over past approaches. We provide preliminary qualitative results on interpreting the predictions to characterize bias. Further work will focus on relying on contextual pretrained models while overcoming limitations on document size, and improving output structures and analyses.

7 Acknowledgments

Nicolas Devatine's work is supported by the SLANT project (ANR-19-CE23-0022). This work was partially supported by the ANR (ANR-19-PI3A-0004) through the AI Interdisciplinary Institute, ANITI, as a part of France's “Investing for the Future — PIA3" program. This work is also partially supported by the AnDiaMO project (ANR-21-CE23-0020). Chloé Braud and Philippe Muller are part of the programme DesCartes and are also supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

8 Ethical considerations

We used the same data as Baly et al. (2020) for comparison purposes. They consist in news articles referenced by the Allsides website, which also assigns political orientation to media sources based on their expertise and some polling.[7] While the exact method is undisclosed, they allow user feedback, which is a way of validating the labels. The fact remains that political labelling is potentially subjective, evolving, and labelling the source is not the same as labelling an article from the source. We train models on that approximate information nonetheless, and it can affect the prediction performance. Also, we merged all labels from the same "side" (left/right) to have only 3 classes instead of Allsides 5 categories. The dataset is not entirely balanced between left/center/right classes, but it's not possible to tell if the distribution is representative of the whole set of potential journalistic sources.

This study is not intended to provide an accurate tool for predicting the political orientation of a news article. The prediction model is a means to analyze differences in linguistic expressions of different biases, with post-hoc analysis of the model internal representations. While revealing orientation of media sources could be a legitimate goal in itself (and is the purpose of the Allsides website), note that current models do not make reliable predictions, and their results should not be taken as such without evidence supporting their decision. This is why part of our work is to analyze and look for linguistic regularities with respect to political orientation. As existing clues are currently either shallow (lexicon) or subject to further validation (structure analysis), it does not dispense of human judgement to decide if a text if showing a bias, openly or not, towards a position on the political spectrum.

[7]https://www.allsides.com/media-bias/media-bias-rating-methods

References

Vidhisha Balachandran, Artidoro Pagnoni, Jay Yoon Lee, Dheeraj Rajagopal, Jaime Carbonell, and Yulia Tsvetkov. 2021. StructSum: Summarization via structured representations. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2575–2585, Online. Association for Computational Linguistics.

Ramy Baly, Giovanni Da San Martino, James Glass, and Preslav Nakov. 2020. We can detect your bias: Predicting the political ideology of news articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4982–4991, Online. Association for Computational Linguistics.

Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1472–1482, Denver, Colorado. Association for Computational Linguistics.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Wei-Fan Chen, Khalid Al Khatib, Benno Stein, and Henning Wachsmuth. 2020. Detecting media bias in news articles using Gaussian bias distributions. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4290–4300, Online. Association for Computational Linguistics.

Yoeng-Jin Chu and Tseng-Hong Liu. 1965. On the shortest arborescence of a directed graph. *Scientia Sinica*, 14:1396–1400.

Giovanni Da San Martino, Alberto Barrón-Cedeño, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020. SemEval-2020 task 11: Detection of propaganda techniques in news articles. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1377–1414, Barcelona (online). International Committee for Computational Linguistics.

Stefano DellaVigna and Ethan Kaplan. 2007. The Fox News Effect: Media Bias and Voting*. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Jack Edmonds. 1967. Optimum branchings. *Journal of Research of the national Bureau of Standards*, 71:233–240.

Elisa Ferracane, Greg Durrett, Junyi Jessy Li, and Katrin Erk. 2019. Evaluating discourse in structured text representations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 646–653, Florence, Italy. Association for Computational Linguistics.

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in Russian news: a computational analysis of intricate political strategies. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3570–3580, Brussels, Belgium. Association for Computational Linguistics.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59):1–35.

Felix Hamborg, Karsten Donnay, and Bela Gipp. 2019. Automated identification of media bias in news articles: an interdisciplinary literature review. *Int. J. Digit. Libr.*, 20(4):391–415.

Masaru Isonuma, Junichiro Mori, and Ichiro Sakata. 2019. Unsupervised neural single-document summarization of reviews via learning latent discourse structure and its ranking. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2142–2152, Florence, Italy. Association for Computational Linguistics.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1113–1122, Baltimore, Maryland. Association for Computational Linguistics.

Hamid Karimi and Jiliang Tang. 2019. Learning hierarchical discourse-level structure for fake news detection. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3432–3442, Minneapolis, Minnesota. Association for Computational Linguistics.

Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. SemEval-2019 task 4: Hyperpartisan news detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 829–839, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

M. Lee and N. Solomon. 1990. *Unreliable Sources: A Guide to Detecting Bias in News Media*. Lyle Smart, New York.

D. G Levasseur. 2008. Media bias. In L. L. Kaid, editor, *Encyclopedia of political communication*. Thousand Oaks, CA: Sage Publications.

Chang Li and Dan Goldwasser. 2021. Mean: Multi-head entity aware attention networkfor political perspective detection in news media. In *Proceedings of the Fourth Workshop on NLP for Internet Freedom Censorship, Disinformation, and Propaganda (NLP4IF)*.

Yang Liu and Mirella Lapata. 2018. Learning structured text representations. *Transactions of the Association for Computational Linguistics*, 6:63–75.

Yang Liu, Ivan Titov, and Mirella Lapata. 2019. Single document summarization as tree induction. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1745–1755, Minneapolis, Minnesota. Association for Computational Linguistics.

Fred Morstatter, Liang Wu, Uraz Yavanoglu, Stephen R. Corman, and Huan Liu. 2018. Identifying framing bias in online news. *Trans. Soc. Comput.*, 1(2).

Philippe Muller, Chloé Braud, and Mathieu Morey. 2019. ToNy: Contextual embeddings for accurate multilingual discourse segmentation of full documents. In *Proceedings of the Workshop on Discourse Relation Parsing and Treebanking 2019*, pages 115–124, Minneapolis, MN. Association for Computational Linguistics.

W. James Murdoch, Peter J. Liu, and Bin Yu. 2018. Beyond word importance: Contextual decomposition to extract interactions from lstms. *CoRR*, abs/1801.05453.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1650–1659. Association for Computational Linguistics.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "why should i trust you?": Explaining the predictions of any classifier.

Dietram A. Scheufele and David Tewksbury. 2007. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*, 57(1):9–20.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. *CoRR*, abs/1503.03832.

Andrey Simonov, Szymon K Sacher, Jean-Pierre H Dubé, and Shirsho Biswas. 2020. The persuasive effect of fox news: non-compliance with social distancing during the covid-19 pandemic. Technical report, National Bureau of Economic Research.

Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. 2014. Deep inside convolutional networks: Visualising image classification models and saliency maps.

Amir Zeldes. 2016. The GUM corpus: Creating multilayer resources in the classroom. In *Proceedings of LREC*.

Amir Zeldes, Yang Janet Liu, Mikel Iruskieta, Philippe Muller, Chloé Braud, and Sonia Badene. 2021. The DISRPT 2021 shared task on elementary discourse unit segmentation, connective detection, and relation classification. In *Proceedings of the 2nd Shared Task on Discourse Relation Parsing and Treebanking (DISRPT 2021)*, pages 1–12, Punta Cana, Dominican Republic. Association for Computational Linguistics.

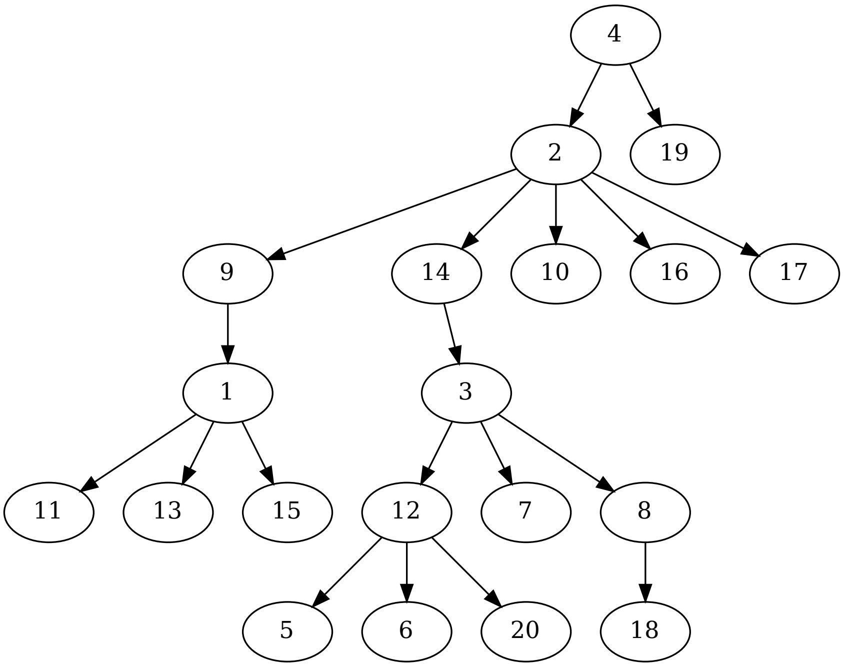
Longyin Zhang, Yuqing Xing, Fang Kong, Peifeng Li, and Guodong Zhou. 2020. A top-down neural architecture towards text-level parsing of discourse rhetorical structure. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6386–6395, Online. Association for Computational Linguistics.

A Example Appendix

Analysis As mentioned in Section 5, it is clear from this example that there is relevant information in the middle and at the end of articles, even though the model focuses on small portions of text, which confirms the value of keeping the whole text. Political figures play an important role for the model, with entities such as "Trump" or "Mattis" (from the Right) having high scores. Furthermore, the model focuses on words or, more generally, on lexical fields that relate to the main subject of the article and that seem to be particularly sensitive for the political side considered here.

President Trump ' s decision late Friday to ban transgender Americans from serving in the U.S. military was blasted by House Minority Leader Nancy Pelosi , who called the move " cowardly " and " disgusting . " The Trump administration issued a memorandum that bars people with a history of " gender dysphoria , " which would require medical treatment , from being admitted to the U.S. military " except under certain limited circumstances . " Pelosi , a San Francisco Democrat , immediately released a statement slamming the memorandum and condemning the Trump administration . " The President ' s ban is a cruel and arbitrary decision designed to humiliate transgender Americans who have stepped forward to serve our country , " she said in a statement . " This bigoted ban weakens our military readiness and our country , and shows this president ' s stunning lack of loyalty to those who risk all to defend our freedoms . " We will continue to fight this discriminatory action , which has no place in our country . House Democrats will never allow hate and prejudice to dictate our national security . " The current policy was based on recommendations made by Defense Secretary James Mattis , who said the Pentagon found that exempting transgender people from military standards could undermine its readiness for combat . " Exempting such persons from well - established mental health , physical health , and sex - based standards , which apply to all Service members , including transgender Service members without gender dysphoria , could undermine readiness , disrupt unit cohesion , and impose an unreasonable burden on the military that is not conducive to military effectiveness and lethality , " read the recommendation that was included in a court filing . DOJ ASKS SUPREME COURT TO TAKE UP CASE OF MILITARY TRANSGENDER BAN The Trump administration asked the Supreme Court to issue an unusually quick ruling on the Pentagon 's policy of restricting military service by transgender people in a bid to bypass lower courts that previously ruled against the administration and its policy barring transgender recruits . The Pentagon initially lifted its ban on transgender troops serving openly in the military in 2016 under the orders of the Obama administration . Trump reversed the policy , prompting outrage and lawsuits , which were ruled against the Trump administration .

Figure 2: Article from "Fox News" (right-leaning) correctly predicted: "Pelosi blasts Trump's move to bar transgender troops, calls it 'disgusting' and 'cowardly'". The darker it is, the higher the relevance of the word to the model.



1 - President-elect Donald Trump has chosen Republican National Committee chairman Reince Priebus his new chief of staff.
2 - He also named conservative media executive Stephen K. Bannon as his senior counselor.
3 - "I am thrilled to have my very successful team continue with me in leading our country", Trump said in a statement.
4 - Trump's transition team made the announcement, Sunday, in the first steps toward solidifying the President-elect's administration.
5 - Priebus, is a Washington veteran with deep ties to Republican leadership, particularly House Speaker Paul Ryan, The Associated Press reports.
6 - "It is truly an honor to join President-elect Trump in the White House as his Chief of Staff", Priebus said in the statement.
7 - "I am very grateful to the President-elect for this opportunity to serve him and this nation
8 - as we work to create an economy that works for everyone, secure our borders, repeal and replace Obamacare and destroy radical Islamic terrorism.
9 - He will be a great President for all Americans."
10 - Bannon is believed to have been in the running for the position, but will now serve as chief strategist and senior counselor.
11 - He ran the conservative website Breitbart News before joining the presidential campaign during the general election.
12 - "Steve and Reince are highly qualified leaders who worked well together on our campaign and led us to a historic victory.
13 - Now I will have them both with me in the White House
14 - as we work to make America great again",
15 - Trump said.
16 - The campaign's statement described Bannon and Priebus as "equal partners".
17 - "Bannon and Priebus will continue the effective leadership team they formed during the campaign, working as equal partners to transform the federal government,
18 - making it much more efficient, effective and productive", it said.
19 - According to CNN, Trump's picks signal that he will look to build bridges in Washington and keep continuity with the Republican party's agenda.
20 - "We will have that same partnership in working to help President-elect Trump achieve his agenda", Bannon said.

Figure 3: Example of a tree induced by the structured attention mechanism. Article from "CBN" (leaning-right) correctly predicted: "Donald Trump Names Reince Priebus as Chief of Staff"