Generating Discourse Connectives with Pre-trained Language Models: Do Discourse Relations Help? *

Symon Jory Stevens-Guille† Aleksandre Maskharashvili† Xintong Li‡ and Michael White†
†The Ohio State University ‡Baidu Research

1 Motivation and Setup

Traditional approaches to discourse have shown the essential importance of discourse (rhetorical) relations in providing coherence to a text [1, 2, 3]. Current approaches to natural language generation (NLG) employing pre-trained models have been shown to excel in generating well-formed text [4], but their ability to produce coherent texts structured with the help of discourse connectives is understudied [5]. Therefore, the study of how well pre-trained models realize discourse relations is of significant interest in the NLG community.

We report results of our experiments using BART [6] and the Penn Discourse Tree Bank [7] (PDTB) to generate texts with correctly realized discourse relations. We address a question left open by previous research [8, 9] concerning whether conditioning the model on the intended discourse relation—which corresponds to adding explicit discourse relation information into the input to the model—improves its performance.

BART, being a transformer [10] based language model, is trained on purposefully corrupted data so that the model learns to 'denoise' the corrupted input in the process of reconstructing the original data. Fine-tuning BART on different versions of input and output lets us probe whether the underlying language model needs or benefits from explicit cues to consistently reconstruct an adequate discourse connective. The PDTB is one of the few corpora developed to identify discourse dependencies between texts. It provides a well-developed ontology of discourse relations; these discourse relations are used to annotate the Wall Street Journal corpus. We consider versions of the corpus differing in (i) whether the order of the arguments in the output is explicitly encoded in the input, (ii) whether the output is the connective or the connective embedded in the corresponding WSJ text, (iii) whether a discourse relation is included in the input and how specific it is. The third is the most important difference since it corresponds to whether the model is conditioned on discourse relation information. We refer to models conditioned on discourse relations by BART$_{D+}$ and models not conditioned on discourse relations by BART$_{D-}$.

In order to determine how well the models perform in realizing discourse relations, we employ standard metrics, e.g. precision, recall, F-1, and devise some new metrics inspired by psycholinguistic and corpus studies to determine the degree to which the models' preferences for realizing different discourse relations correspond to reported human preferences for realizing those relations [11, 12, 8]. While space precludes reporting of the results on these new metrics in the abstract, we intend to report them subsequently.

2 Results and Discussion

Our results show that fine-tuning BART on the different versions of PDTB inputs and outputs mentioned in the foregoing consistently produces discourse connectives which match those used in the text. The BART$_{D+}$ models nonetheless outperform the BART$_{D-}$ models. It's noteworthy that the best BART$_{D+}$ model matched (recall = 79%) on hundregs of additional data points compared to the best BART$_{D-}$ (recall = 71.3%) model (McNemar's Test Statistic 313; p $<$ .000). With respect to matching on explicit connectives, the BART$_{D+}$ (69.8%) model matched significantly more than the BART$_{D-}$ (54.3%) model (McNemar's Test Statistic 157; p $<$ .000). With respect to matching on implicits the BART$_{D+}$ (89.2%) model is slightly but significantly worse than the BART$_{D-}$ (90.6%) model (McNemar's Test Statistic 118; p $<$ .025), though this seems to reflect the overprediction of implicits by the BART$_{D-}$ model.

*This research was supported by a collaborative open science research agreement between Facebook and The Ohio State University. E-mail: stevensguille.1@buckeyemail.osu.edu

The results reported above are in line with the view that information concerning discourse relations should be present in the inputs of neural approaches to NLG [13, 14, 5], which has not typically been the case. When the metrics are extended to include whether non-matching connectives chosen by the model fit the intended discourse relation, the best BART$_{D+}$ model continues to outperform the best BART$_{D-}$ model. When producing non-matching connectives, we find that the chosen connectives of the BART$_{D+}$ models correspond to the intended discourse relations more frequently than those produced by the BART$_{D-}$ models.

The main conclusion one can draw from our results is that discourse relation information is essential for consistently generating matching discourse connectives beyond the sentence level. While large-scale human judgement experiments on our model's predictions are the most obvious next step, the improvement of the BART$_{D+}$ models over the BART$_{D-}$ models with respect to exact matching is encouraging, especially in light of recent results showing that humans don't uniformly accept substitution of discourse connectives which express the same discourse relation [8]. With respect to whether mere arguments suffice to predict the discourse connective holding between them, our results indicate that the purely distributional meaning of texts induced by the models under-determines the meaning of explicit discourse connectives. Directly conditioning on explicit discourse relations significantly improves the match between discourse connective produced and discourse relation intended to be expressed.

To sum up, our results suggest that the intended discourse relation cannot always be inferred from the arguments using pre-trained models. Inclusion of the discourse relation in the input provides an immediate boost to control over output coherence.

References

[1] W. C. Mann and S. A. Thompson, *Rhetorical structure theory: A theory of text organization.* University of Southern California, Information Sciences Institute Los Angeles, 1987.

[2] A. Lascarides and N. Asher, "Segmented discourse representation theory: Dynamic semantics with discourse structure," in *Computing meaning*, pp. 87–124, Springer, 2008.

[3] A. Kehler and A. Kehler, *Coherence, reference, and the theory of grammar.* CSLI publications Stanford, CA, 2002.

[4] M. Kale and A. Rastogi, "Text-to-text pre-training for data-to-text tasks," in *Proceedings of the 13th International Conference on Natural Language Generation*, pp. 97–102, 2020.

[5] A. Maskharashvili, S. Stevens-Guille, X. Li, and M. White, "Neural methodius revisited: Do discourse relations help with pre-trained models too?," in *Proceedings of the 14th International Conference on Natural Language Generation*, (Aberdeen, Scotland, UK), pp. 12–23, Association for Computational Linguistics, Aug. 2021.

[6] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *arXiv preprint arXiv:1910.13461*, 2019.

[7] B. Webber, R. Prasad, A. Lee, and A. Joshi, "The penn discourse treebank 3.0 annotation manual," *Philadelphia, University of Pennsylvania*, 2019.

[8] F. Yung, M. Scholman, and V. Demberg, "A practical perspective on connective generation," in *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, pp. 72–83, 2021.

[9] W.-J. Ko and J. J. Li, "Assessing discourse relations in language generation from gpt-2," in *Proceedings of the 13th International Conference on Natural Language Generation*, pp. 52–59, 2020.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *CoRR*, vol. abs/1706.03762, 2017.

[11] J. D. Murray, "Connectives and narrative text: The role of continuity," *Memory & Cognition*, vol. 25, no. 2, pp. 227–236, 1997.

[12] T. Sanders, "Coherence, causality and cognitive complexity in discourse," in *Proceedings/Actes SEM-05, First International Symposium on the exploration and modelling of meaning*, pp. 105–114, University of Toulouse-le-Mirail Toulouse, 2005.

[13] A. Balakrishnan, V. Demberg, C. Khatri, A. Rastogi, D. Scott, M. Walker, and M. White, "Proceedings of the 1st workshop on discourse structure in neural nlg," in *Proceedings of the 1st Workshop on Discourse Structure in Neural NLG*, 2019.

[14] S. Stevens-Guille, A. Maskharashvili, A. Isard, X. Li, and M. White, "Neural NLG for methodius: From RST meaning representations to texts," in *Proceedings of the 13th International Conference on Natural Language Generation*, (Dublin, Ireland), pp. 306–315, Association for Computational Linguistics, Dec. 2020.