LINGCONV: An Interactive Toolkit for Controlled Paraphrase Generation with Linguistic Feature Control

Mohamed Elgaar and **Hadi Amiri**
University of Massachusetts Lowell
{melgaar,hadi}@cs.uml.edu

Abstract

We introduce a paraphrase generation system that can control and adjust several linguistic attributes of input text. Given a sentence and a desired level of linguistic complexity, the model can generate multiple paraphrases that preserve the meaning while satisfying the specified level of linguistic complexity. To ensure high-quality outputs, our approach interpolates the embedding of linguistic indices to locate the closest and most achievable set of indices suited to the target complexity level. Applications of linguistic conversion are diverse and impactful. It can make text accessible to everyone, including those with low literacy skills, cognitive or learning disabilities, aphasia, or dementia; enables personalized communication; serves as a research tool for linguists and NLP researchers; and facilitates second language learning by adjusting text complexity based on learner's proficiency level. The system is available at https://mohdelgaar-lingconv.hf.space, and a two-minute video demo is available at https://youtu.be/Evs5dwpT4Jc

1 Introduction

Controllable generation is the task of generating text that conforms to a given set of attributes, such as sentiment, formality, negation, and toxicity (Jin et al., 2022). Applications of controllable generation include text simplification (Lee and Lee, 2023b,a; Vajjala and Lučić, 2018; Zhang and Lapata, 2017; Xu et al., 2015), toxicity control (Zheng et al., 2023; Zhang and Song, 2022; Liu et al., 2021), emotion and topic control (Yang et al., 2023), and personalized dialog generation (Huang et al., 2023b; Niu and Bansal, 2018).

Existing methods have been developed and tested to control up to three attributes. However, increasing the number of control attributes poses a major challenge to approaches that use prefix-tuning of large language models (Bandel et al., 2022) or employ separate discriminators for each attribute (Keskar et al., 2019; Lyu et al., 2021; Dathathri et al., 2019; Krause et al., 2021; Yang and Klein, 2021; Liu et al., 2022, 2023).

We develop a system for compositional control of a set of 40 linguistic attributes, covering lexical, syntactic, and discourse attributes. This system is capable of generating paraphrases of a given source text with varying levels of linguistic complexity. Moreover, it can generate text that exhibits specific user-specified linguistic characteristics, useful for personalized communication, and language learning and development.

Our system, named LINGCONV, extends a T5 encoder-decoder model (Raffel et al., 2020) with an attribute control approach that integrates linguistic complexity into the decoding process. The model's training leverages two key training objectives to guide the generation process: translation cross-entropy loss, and linguistic discriminator loss. This enables the model to generate text that adheres to specified linguistic complexity levels while maintaining semantic consistency.

From a technical perspective, LINGCONV benefits from a quality control mechanism in the generation process to ensure high-quality outputs. Since not all combinations of linguistic indices are feasible, it interpolates the embeddings of linguistic indices to identify the closest and most achievable set of indices for generating a target text. This interpolation process iteratively adjusts the target linguistic indices until the generation is in proximity to the desired linguistic attributes. This mechanism enforces the generation to closely align with the target indices, even for incoherent or complex specifications of linguistic indices.

LINGCONV offers a range of advanced features and functionalities to provide users with greater control and flexibility over the generation process. Figure 1 illustrates the layout of the system. Users can choose from three distinct paraphrase genera-

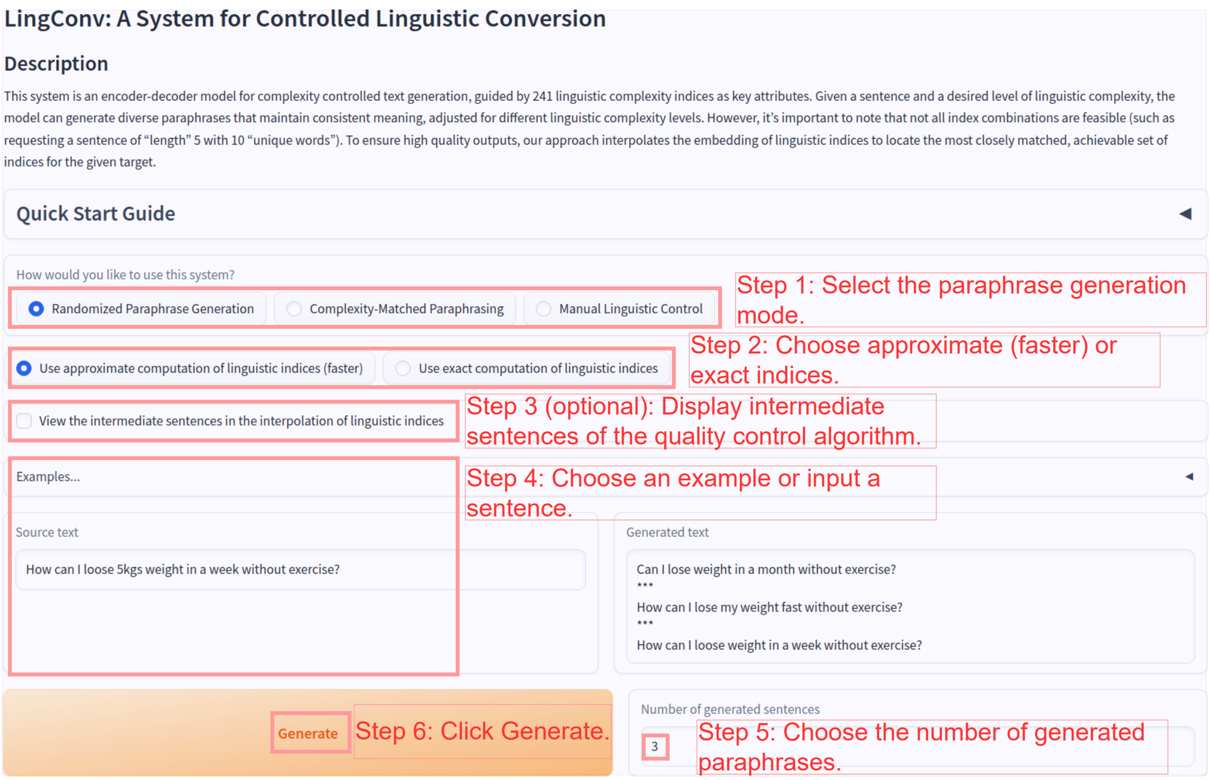


Figure 1: Layout of LingConv. The system offers three modes of generation: (a): Randomized paraphrase generation, (b): Paraphrase copying the attribute of an additional input sentence, (c): Controlled paraphrase generation with fine-grained control of linguistic attributes. The figure displays the functionality of mode (a).

Source Text	Paraphrases
Do people still believe in demonic possession?	Is there any real faith in demonic possession? Why do some people still believe in demonic possession? What's the best way to believe in demonics?
What do I do to increase my height?	What should I do when I want to increase my height? How do I increase my height? "What are the ways to increase "height"?"
When is the iPhone 7 coming out?	What is the new year iphone 7 release? What is the next posible release for iPhone 7? Will there be an iPhone 7 or something?

Table 1: Examples of three source sentences, and three paraphrases generated by LingConv for each source.

tion strategies: *Randomized Paraphrase Generation*, *Complexity-Matched Paraphrasing*, and *Manual Linguistic Control*. Additionally, the system allows users to select between exact or approximate linguistic index computation, which is used in interpolation and the manual setting of linguistic attributes. The system provides the option to show the intermediate sentences generated during the quality control process, enhancing transparency and interpretability.

As linguistic complexity indices, we consider lexical, syntactic, as well as a diverse set of semantic, discourse, and traditional linguistic indices. We extract these indices using multiple sources developed by Lu (2010, 2012) and Lee and Lee (2023a). Their extensive coverage of linguistic attributes enables our system to produce highly nuanced paraphrases of given source texts.

To the best of our knowledge, LingConv is the first system designed to generate paraphrases with specific linguistic complexity levels while providing fine-grained control over a wide range of linguistic indices within the generated text. Thus, it not only enhances paraphrase generation but also reshapes the way users can manipulate and understand linguistic complexity. We note that there are several approaches that can generate syntactically diverse paraphrases of a given sentence by altering its syntactic form. For example, the work by Huang et al. (2023a) rotates the abstract meaning representation (AMR) tree, Goyal and Durrett (2020) reorders the segments in a sentence parse tree, and Iyyer et al. (2018) uses templates of constituency parses to generate the target sentence. LINGCONV accounts for various aspects of controlled linguistic generation, including lexical, syntactic, semantic, and discourse aspects.

2 System Architecture

2.1 Model

We employ a T5-Base (Raffel et al., 2020) encoder-decoder model augmented with a linguistic complexity control approach to perform complexity-controlled paraphrase generation. Given a dataset $\mathcal{D} = \{s, t\}$ of paraphrase source and target pairs s and t, we compute the linguistic indices of the source l^s and target l^t. We create a linguistic embedding layer $h(l) = \mathbb{R}^k \to \mathbb{R}^{d_{model}}$, where k is the number of linguistic indices, and d_{model} is the input embedding dimension of T5. These linguistic embeddings enhance the decoding process through element-wise addition to the decoder inputs.

The model is trained using two objectives. First, the translation loss from source to target paraphrases using cross-entropy. Second, a linguistic discriminator loss. Using a differentiable pre-trained classifier of linguistic indices, we minimize the mean squared error (MSE) between the indices of the generated text and the target indices. The two losses work together to achieve generations that adhere to the target linguistic indices while conserving the semantics as much as possible.

2.2 Generation Quality Control

Since not all user-specified linguistic index combinations are feasible (such as requesting a sentence of "length" 5 with 10 "unique words"), LINGCONV implements a quality control approach for the generation process. To ensure high-quality outputs, it interpolates the embedding of linguistic indices to locate the closest and most achievable set of indices for the given target.

The system generates a candidate output and compares the embedding of its linguistic indices against the embedding of the target linguistic indices. If the two embeddings are not close enough, the target linguistic indices are replaced with the midpoint between the indices of the generation and the original indices. This process is iteratively repeated until the linguistic indices of the generated text are close to the target linguistic indices. As mentioned before, users have the option to manually specify desired linguistic complexity by providing target indices. If these indices are incoherent, more interpolation steps will be needed until convergence. This process further enforces the generation to be close to the given target indices.

3 Features and Functionalities

LINGCONV offers three generation strategies: Randomized Paraphrase Generation, Complexity-matched Paraphrasing, and Manual Linguistic Control. Moreover, it offers the choice to use exact or approximate linguistic indices computation, used in interpolation and in the manual setting of linguistic indices, and the option to show the intermediate sentences generated during quality control.

The linguistic indices we use are extracted from three sources: lexical indices developed by Lu (2012), syntactic indices by Lu (2010), and a diverse set of semantic, lexical, discourse, and traditional indices by Lee and Lee (2023a).

3.1 Randomized Paraphrase Generation

The Paraphrase Generation feature provides a straightforward yet powerful way to generate multiple paraphrases of a given source sentence. Users begin by inputting the source sentence and indicating the desired number of paraphrases to be generated. The system then employs its linguistic indices sampling and text generation algorithm to generate a set of distinct paraphrases, each adhering to a unique set of target linguistic indices.

LINGCONV employs a large-scale repository of precomputed linguistic index sets. These sets, extracted from the training dataset, encompass a wide spectrum of linguistic attributes. By randomly selecting index sets from this collection, the system ensures that the generated paraphrases carry diverse linguistic characteristics. Table 1 shows examples of paraphrases generated by LINGCONV using the randomized paraphrase generation option.

3.2 Complexity-matched Paraphrasing

Give a source sentence and a reference sentence. The model extracts the linguistic indices from the reference. Utilizing these extracted indices as a guide, the system generates a paraphrase of the source sentence that mirrors the linguistic attributes of the reference. This form of textual style transfer enables users to seamlessly adapt their content to match a specific style or level of complexity. It is

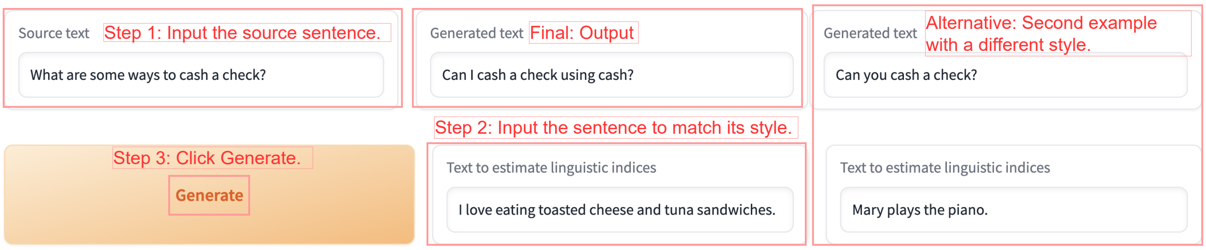
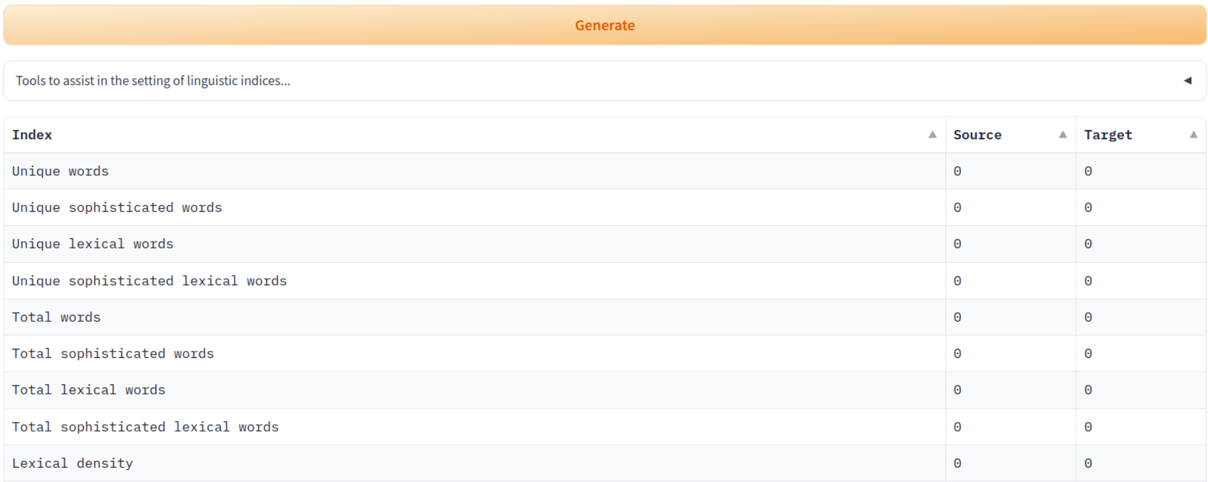


Figure 2: Style-matched paraphrase generation. This mode allows the generation of a paraphrase that copies the attributes of an additionally given sentence. In the example, the source sentence is "What are some ways to cash a check?", and two example generations are shown, where an alternative sentence is supplied to copy the target attributes from. The model generates paraphrases that hold the attributes of the supplied sentence.



Index	Source	Target
Unique words	0	0
Unique sophisticated words	0	0
Unique lexical words	0	0
Unique sophisticated lexical words	0	0
Total words	0	0
Total sophisticated words	0	0
Total lexical words	0	0
Total sophisticated lexical words	0	0
Lexical density	0	0

Figure 3: The third mode of using the system is by manually inputting the desired target attributes, and generating a paraphrase that holds those attributes. This allows for fine-grained control over the linguistic attributes of the generation.

a valuable tool for authors, marketers, communicators, and clinicians looking to tailor their text to distinct audiences or contexts while maintaining semantic coherence. Figure 2 displays an example of using this mode.

3.3 Manual Linguistic Control

For precise control over generated text, users have the option to manually input specific values for target linguistic indices. Upon entering these values into the provided table, see Figure 3, the system generates a paraphrase in accordance with the defined indices. To facilitate the manual setting of linguistic values, the system offers a set of useful tools. These tools can be accessed by clicking on the "Tools to assist in setting linguistic indices" prompt within the user interface. As illustrated in Figure 4, The tools provide the following features:

- **Computing/Estimating Linguistic Indices**: This tool aids users in calculating or estimating the linguistic indices of a given source or reference sentence.
- **Copying Indices**: Users can copy indices from the source column to the target column. This streamlined process is particularly useful for making incremental adjustments to the source indices.
- **Adding and Subtracting Noise**: This feature serves a dual purpose: altering the source text through incremental noise addition and generating slight variations of the existing output by introducing minor noise perturbations.

By offering both automated and manual mechanisms for complexity-controlled text generation, LINGCONV enhances paraphrase generation and reshapes the way users can control, manipulate and understand linguistic complexity.

4 Data and Evaluation

We train and evaluate our model using the combination of three datasets that consist of semanti-

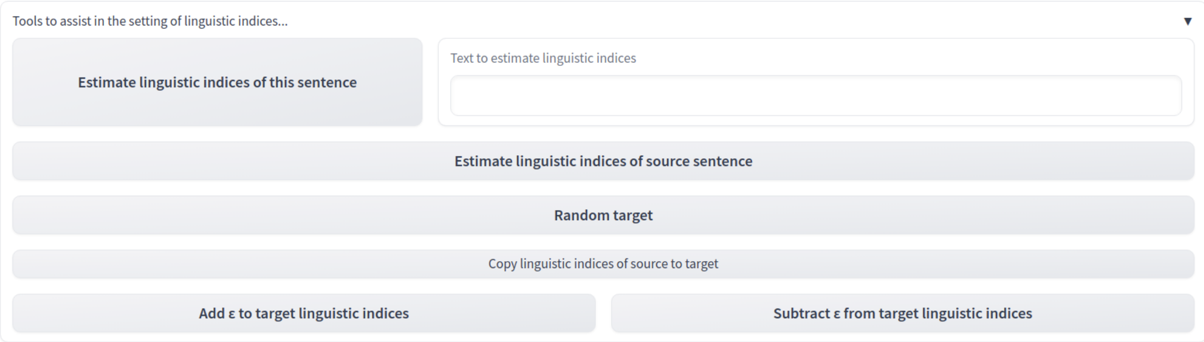


Figure 4: In the manual setting of target attributes, the system provides a set of tools to aid the user. The user may load a pre-existing set of attributes randomly, or extract them from a given sentence, or copy them from the source sentence. These attributes may be used as a starting baseline as particular linguistic indices are modified.

Source Text	Intermediate Texts for each Paraphrase
Why do some people still believe in demonic possession?	1. Why do people still believe in demonic possession? 2. Why do some people still believe in democratic possession in 2016? → Why do some people still believe in demonic possession?
What do I do to increase my height?	1. What do I do to increase my height? If yes, then what should I do that will help me to increase my height? 2. What do I do to increase my height after I have already crossed her? 3. What should I do if I want to increase my height after I already have ended? 4. What should I do if I have gained some height but I don't think so? 5. What should I do if I want to increase his height? 6. What should I do if I want to increase myself? → What should I do when I want to increase my height?
What do I do to increase my height?	1. What is the best way to "make" your height? 2. What are the best ways to increase "height"? 3. What are the ways to get "attractive"? → What are the ways to increase "height"?

Table 2: Example intermediate texts iteratively generated at inference time using the quality control approach for text generation. Arrows (→) indicate the final step in the iterative generation process.

cally equivalent pairs of sentences: The Microsoft Research Paraphrase Corpus (MRPC) (Dolan and Brockett, 2005), The Semantic Textual Similarity Benchmark (STS-B) (Cer et al., 2017), and The Quora Question Pairs[1].

We evaluate the quality of generation using two quantitative metrics of text generation, BLEU score (Papineni et al., 2002) and BERTScore (Zhang et al., 2020), one metric of controllable text generation, the mean squared error (MSE) between normalized linguistic indices of the generated text and target, and a qualtitative human evaluation of generated paraphrase validity; if the generation is a true paraphrase of the source. We compare our model with a Vanilla T5 (Raffel et al., 2020) that we fine-tune on our datasets, and Quality Controlled Paraphrase Generation (QCPG) (Bandel et al., 2022), a model for attribute-controlled paraphrase generation. Qualitative evaluation is performed only on LingConv and QCPG. Results in Table 3 shows that the attribute control of LINGCONV is 40% more accurate than the baseline, while being comparable in semantic equivalence. We attribute this higher performance to the effective use of linguistic complexity indices in the decoding phase of LingConv.

Table 2 presents examples of intermediate text outputs generated at inference time during the interpolation process for quality control of the generated texts. These results provide a clear illustration of the step-by-step generation process that progressively moves towards generating target sentences that meet desired levels of linguistic complexity.

5 Conclusion

We developed a new text conversion system, LINGCONV, which offers a range of features and functionalities for complexity-controlled text generation. Through the careful integration of linguistic indices and model architecture, LINGCONV provides users with the tools to generate text that adheres to both specific and diverse linguistic com-

[1] https://quoradata.quora.com/First-Quora-Dataset-Release-Question-Pairs

Model	BLEU ↑	BERTScore ↑	Indices MSE ↓	Paraphrase Validity ↑
Vanilla T5	38.6	94.2	6.86	-
QCPG	48.1	94.9	4.92	**81.3%**
LingConv	**56.6**	**95.6**	**2.96**	77.3%

Table 3: LingConV performance against Vanilla T5 and QCPG baselines.

plexity levels.

The system's paraphrase generation feature allows users to create a variety of paraphrases that align with distinct linguistic characteristics. By drawing from a vast collection of precomputed linguistic index sets, LINGCONV can generate of linguistically diverse paraphrases. Furthermore, the complexity-matched paraphrasing feature allows users to seamlessly transform the linguistic style of their texts. This functionality is particularly useful for tailoring content to specific contexts or audiences while preserving the underlying meaning.

Comparative evaluations against a recent controlled generation baseline and the vanilla T5 model, across several evaluation metrics, and the accuracy of generated linguistic indices verify the reliability and effectiveness of LINGCONV.

References

Elron Bandel, Ranit Aharonov, Michal Shmueli-Scheuer, Ilya Shnayderman, Noam Slonim, and Liat Ein-Dor. 2022. Quality controlled paraphrase generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 596–609, Dublin, Ireland. Association for Computational Linguistics.

Daniel Cer, Mona Diab, Eneko Agirre, Iñigo Lopez-Gazpio, and Lucia Specia. 2017. SemEval-2017 task 1: Semantic textual similarity multilingual and crosslingual focused evaluation. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 1–14, Vancouver, Canada. Association for Computational Linguistics.

Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2019. Plug and play language models: A simple approach to controlled text generation. In *International Conference on Learning Representations*.

William B. Dolan and Chris Brockett. 2005. Automatically constructing a corpus of sentential paraphrases. In *Proceedings of the Third International Workshop on Paraphrasing (IWP2005)*.

Tanya Goyal and Greg Durrett. 2020. Neural syntactic preordering for controlled paraphrase generation. *ArXiv*, abs/2005.02013.

Kuan-Hao Huang, Varun Iyer, I-Hung Hsu, Anoop Kumar, Kai-Wei Chang, and Aram Galstyan. 2023a. ParaAMR: A large-scale syntactically diverse paraphrase dataset by AMR back-translation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8047–8061, Toronto, Canada. Association for Computational Linguistics.

Qiushi Huang, Yu Zhang, Tom Ko, Xubo Liu, Bo Wu, Wenwu Wang, and H Tang. 2023b. Personalized dialogue generation with persona-adaptive attention. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 12916–12923.

Mohit Iyyer, John Wieting, Kevin Gimpel, and Luke Zettlemoyer. 2018. Adversarial example generation with syntactically controlled paraphrase networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1875–1885.

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.

Nitish Shirish Keskar, Bryan McCann, Lav R Varshney, Caiming Xiong, and Richard Socher. 2019. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*.

Ben Krause, Akhilesh Deepak Gotmare, Bryan McCann, Nitish Shirish Keskar, Shafiq Joty, Richard Socher, and Nazneen Fatema Rajani. 2021. GeDi: Generative discriminator guided sequence generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4929–4952, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bruce W. Lee and Jason Lee. 2023a. LFTK: Handcrafted features in computational linguistics. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 1–19, Toronto, Canada. Association for Computational Linguistics.

Bruce W Lee and Jason Lee. 2023b. Prompt-based learning for text readability assessment. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1774–1779.

Alisa Liu, Maarten Sap, Ximing Lu, Swabha Swayamdipta, Chandra Bhagavatula, Noah A. Smith, and Yejin Choi. 2021. Dexperts: Decoding-time controlled text generation with experts and anti-experts.

In *Annual Meeting of the Association for Computational Linguistics*.

Guangyi Liu, Zeyu Feng, Yuan Gao, Zichao Yang, Xiaodan Liang, Junwei Bao, Xiaodong He, Shuguang Cui, Zhen Li, and Zhiting Hu. 2022. Composable text control operations in latent space with ordinary differential equations. *arXiv preprint arXiv:2208.00638*.

Xin Liu, Muhammad Khalifa, and Lu Wang. 2023. BOLT: Fast energy-based controlled text generation with tunable biases. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 186–200, Toronto, Canada. Association for Computational Linguistics.

Xiaofei Lu. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15:474–496. Citation Key: Lu2010.

Xiaofei Lu. 2012. The relationship of lexical richness to the quality of esl learners' oral narratives. *Source: The Modern Language Journal*, 96(2):190–208. Citation Key: Lu2012.

Yiwei Lyu, Paul Pu Liang, Hai Pham, Eduard Hovy, Barnabás Póczos, Ruslan Salakhutdinov, and Louis-Philippe Morency. 2021. StylePTB: A compositional benchmark for fine-grained controllable text style transfer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2116–2138, Online. Association for Computational Linguistics.

Tong Niu and Mohit Bansal. 2018. Polite dialogue generation without parallel data. *Transactions of the Association for Computational Linguistics*, 6:373–389.

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 of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

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.

Sowmya Vajjala and Ivana Lučić. 2018. OneStopEnglish corpus: A new corpus for automatic readability assessment and text simplification. In *Proceedings of the Thirteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 297–304, New Orleans, Louisiana. Association for Computational Linguistics.

Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3:283–297.

Kevin Yang and Dan Klein. 2021. FUDGE: Controlled text generation with future discriminators. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3511–3535, Online. Association for Computational Linguistics.

Kexin Yang, Dayiheng Liu, Wenqiang Lei, Baosong Yang, Mingfeng Xue, Boxing Chen, and Jun Xie. 2023. Tailor: A soft-prompt-based approach to attribute-based controlled text generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 410–427, Toronto, Canada. Association for Computational Linguistics.

Hanqing Zhang and Dawei Song. 2022. Discup: Discriminator cooperative unlikelihood prompt-tuning for controllable text generation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3392–3406.

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*.

Xingxing Zhang and Mirella Lapata. 2017. Sentence simplification with deep reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594, Copenhagen, Denmark. Association for Computational Linguistics.

Carolina Zheng, Claudia Shi, Keyon Vafa, Amir Feder, and David Blei. 2023. An invariant learning characterization of controlled text generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3186–3206, Toronto, Canada. Association for Computational Linguistics.