

Controllable Text Generation: Style & Content

Cansen CAGLAYAN

Department of Computer Engineering, Atılım University

Murat KARAKAYA

Department of Computer Engineering, Atılım University

To Cite This Chapter:

Caglayan, C., & Karakaya, M. (2023). Controllable text generation: Style & Content. In F. R. Dar & N. Kucukgencay (Eds.), *Current Studies in Social Sciences 2023* (pp. 55-62). ISRES Publishing.

Introduction

The generation of texts from human language by machines is called Automatic Text Generation (ATG). ATG is one of the most important research topics in Natural Language Processing (NLP). Automatic text generation is seen as a technological development that is expected to lead to significant changes in the world. It is a key component of language translation, question-answering systems, summarization, and several other applications that people interact with every day (Celikyilmaz et al., 2020). Chatbots like ChatGPT, which are very popular today, are one of the most important examples of this, which give very good results in generating text according to the prompts given. ChatGPT is a text generation structure that offers a very advanced service in question and answer developed by OpenAI. Thanks to its Transformer-based GPT large language models (LLM) (Radford et al., 2018) structure, it provides various benefits to users in every field by generating meaningful texts.

The model used in ATG is called the Language Model (LM). Simply, an LM learns the probability of occurrences of the token, i.e. the desired split smallest structure (symbol, character, or word) of a piece of text, based on given examples with considering the features of natural languages such as grammar, syntax, and semantics. This way, meaningful texts can be generated automatically (Caglayan, 2022). Although ATG has been tried with different methods before, the most significant results begin with the use of Deep Neural Network (DNN) models. New solutions are tried over time for problems such as the analysis of language structure, the length of words that can be associated, etc. encountered in text generation. With deep learning coming into the picture, recurrent neural networks (RNN) e.g., long short-term memory networks (LSTM), gated recurrent units (GRUs) methods are frequently used. Later sequence-to-sequence learning opens up a new chapter featured by the wide application of the encoder-decoder architecture. Although the encoder-decoder structure was strengthened with attention mechanisms and started to be used, it was still weak in capturing the relationship of words with each other since the language was examined sequentially. The success of all the deep learning-based methods relies heavily on large-scale datasets, posing a challenge for supervised and cross-domain text generation tasks (Zhang et al., 2022). As a solution the transformer structure (Vaswani et al., 2017), which adds the self-attention to the encoder-decoder structure

and examines it in parallel with the relationship of the tokens, came out. Self-attention is an attention mechanism relating different positions of a single sequence in order to compute a representation of the sequence (Vaswani et al., 2017). The transformer models showed big success in text generation as well as in all NLP fields. Later, transformer-based and pre-trained language models with large text data entered our lives. Pre-trained Large Language Models like GPT (Radford et al., 2018), BERT (Devlin et al., 2018), etc. are the most important examples of today. Pre-trained Large Language Models can learn to extensively model the distribution of natural language from their many different materials, so they can generate texts of unprecedented quality. In time while these developments in text generation continue, has also been focused on controllable text generation.

Controllable text generation (CTG), is an important part of ATG. CTG aims to ensure that the automatically generated texts have the desired qualities. The qualities to be controlled can be stylistic features such as politeness, sentiment, and formality; or maybe the characteristics of the hypothetical person writing the text, such as age, gender, and character; or maybe the topic, keywords, and information in the content of the text (Prabhumoye et al., 2020). As shown in Figure 1, basically CTG system has four components: an input text (I), the control element as (C) e.g. topic, the language model as process (P), and the generated text satisfying the control condition, as output (O).

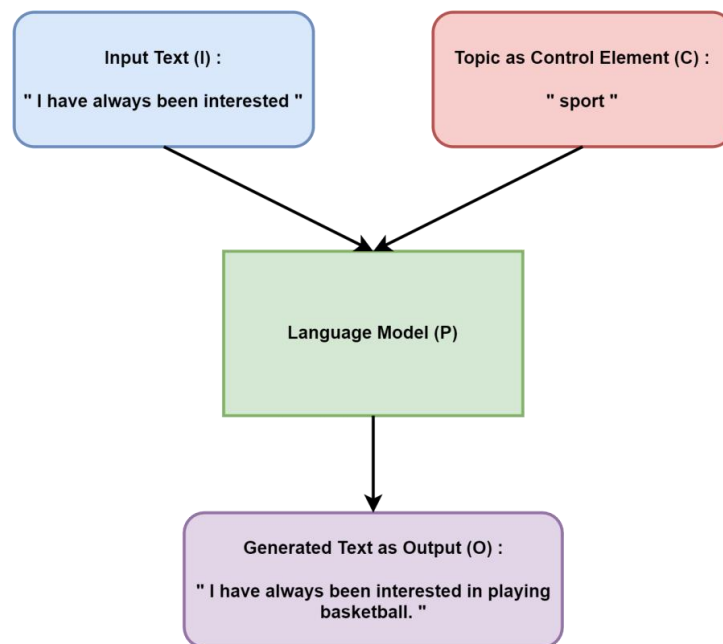


Figure 1. Example of topic-controlled text generation

In this case, the source sentence "I have always been interested" is completed based on the specified topic condition. As an example, if the topic is determined as sports in a topic-controlled text generation, the sentence can be completed with "in playing basketball" in accordance with this condition.

Given a vocabulary $\mathcal{V}$, the goal of CTG is to generate a target text $Y = \{y_1, y_2, \dots, y_n\}$, where $y_n \in \mathcal{V}$, with respect to a control element denoted as C . Then CTG can be formally described as:

$$P(Y|C) = (y_1, y_2, \dots, y_n|C)$$

The specific expression of C may vary according to different tasks. As for the sentence Y generated by the model, it is also expected to satisfy the constraint conditions while conforming to the general natural language characteristics such as fluency, rationality, and readability, to the greatest extent (Zhang et al., 2022).

Today, controllable text generation is made using a wide variety of models and techniques, but in this study, the results from the literature review are grouped as controllable text generation for style or content purposes. The aim of the study is to examine the academic studies on controllable text generation using different models and techniques and to collect them under the headings of style and content control. Thus, it may be useful for researchers working on this subject.

Style-Controlled Text Generation

Style is the way people do anything. Linguistic style is expressed in the choice of words or phrases as well as syntactic structures used to convey a piece of information (Prabhumoye, 2021). In this way, people can convey the information they want to give in a style they want, depending on their word or sentence choice. Creating a piece of information as text in a certain style is a fairly easy process for people. For example, people can convey a text with the same content in a ruder or more polite style in a matter of seconds. Gender, age, education, and personal styles, formality, politeness can be used as styles for text. In addition sentiment, humor, and romance etc. can be seen as different styles in texts and a text can include more than one style. However, it is very difficult for a machine that generates automatic texts in the desired style. Therefore, style-controlled text generation is a challenging and important study topic in the field of CTG. Figure 2 shows an example of how to control the style of two texts with the same content.

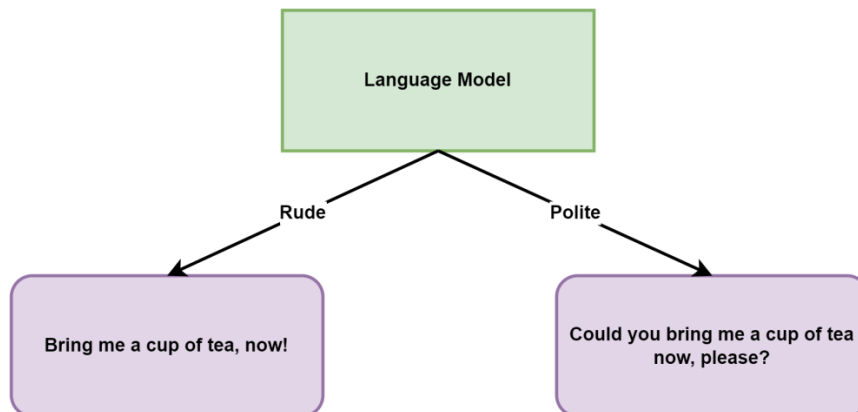


Figure 2. *Politeness-controlled text generation*

Style-controlled text generation task is also called style transfer. Style transfer is the task of rephrasing the text to contain specific stylistic properties without changing the

intent or effect within the context (Prabhumoye, 2021). The expression of the same emotion or thought by different personas may reveal a different style for each text.

There are many different types of tasks that we can use style transfer while generating text in real life. One example is to generate style-controlled responses for dialogue systems/chatbots. We can control the politeness, authority, persona of the chatbots. Also style transfer is widely used for generating texts for emails and stories.

The politeness plays a very important role in the style of the text generated. (Madaan et al., 2020) converts non-polite sentences to polite sentences for emails while preserving the meaning. An example they use in their work is to replace "send me the data" with a more polite version of "could you please send me the data". (Prabhakaran et al., 2014) works on emails according to control of politeness based on gender. (Peterson et al., 2011) examines the controlling formality of emails. (Sennrich et al., 2016) uses the English to German translation task to provide a model that can create formal or informal sequences based on verbs and pronouns to control the politeness. (Niu and Bansal, 2018) generates polite dialogues, their work focuses on contextual dialogue response generation. In the task of dialogue response generation, not only politeness but also to control emotion (Li et al., 2020), persona (Zhang et al., 2018) (Li et al., 2016), formality, authority etc. is also required.

Gender-related style differences are also an important issue in text styles. According to the researches, the word choices of women and men, the frequency of using pronouns and exclamations differ (Bamman et al., 2014).

Style transfer is used to generate persona based dialog responses and generating stylistic sentences for different type of personalities such as romantic or humorous (Shuster et al., 2018) (Li et al., 2018), politically slanted (Prabhumoye et al., 2018) or sentiment based like being more positive or negative (Shen et al., 2017) (Fu et al., 2018). Also to generate stories with different emotional endings (Peng et al., 2018) and creating them according to the different persona types (Chandu et al., 2019).

Another example is we can transfer the specific person's style to generated text. (Xu et al., 2017) use a phrase-based statistical model, and (Jhamtani et al., 2017) use a standard encoder-decoder model to convert modern language to Shakespeare-style language by treating style transfer as a translation task (Niu et al., 2018).

Style-controlled text generation is very popular subject and research is being done to improve it day by day. The biggest challenge of style-controlled text generation is the lack of parallel data. Also, this makes it very difficult to measure whether the generated text has the desired properties. Since pre-trained large language models trained with a very large corpus have a lot of information, they show improvement in reflecting the desired style on the given prompt. However, the issue of the fact that the generated text should be accurate and fluent as well as having all the desired features and the methods for measuring the results needs improvement.

Content-Controlled Text Generation

Controlling what content and information the text will have is a very important subject, as much as controlling the style. Having the desired content of an automatically generated text is called content-controlled text generation or content transfer task. Unless trained language models are specifically conditioned, the contents of texts may be random rather than limited. One of the most important sub-tasks is the topic-controlled text generation. As shown in Figure 1, the topic of the text to be generated can be limited. When the desired topic is "sport", the prompt given is completed in accordance with this topic. For example in our previously published work on this subject, we add the topic information to the input during the language model training (Caglayan et al., 2021). In this way, while the language model is expected to generate user reviews according to the characteristics of the language, the text should also be about the desired topic. (Tang et al., 2019) and (Prabhumoye et al., 2020b) controls the topic sequence for dialogue response generation task. Accordingly, it is possible to limit the answers generated to the desired topic. Also (Huang et al., 2019), (Wang et al., 2019), (Yang et al., 2019), (Chang et al., 2021), (Lin and Riedl, 2021) controls the topic sequence for story generation task.

Another important sub-task of content transfer is document grounded. In document-grounded controlled text production, the content of the text produced is determined by an external source. A lot of information resides in an unstructured format in the form of books, Encyclopedias, news articles, Wikipedia articles (Prabhumoye, 2021) and this information is used to control the content of the text to be generated. Figure 3 shows that if the continuation of the text given as input to the language model is generated by using an external source, for example, a Wikipedia article, the content of the generated text will be determined according to the given document.

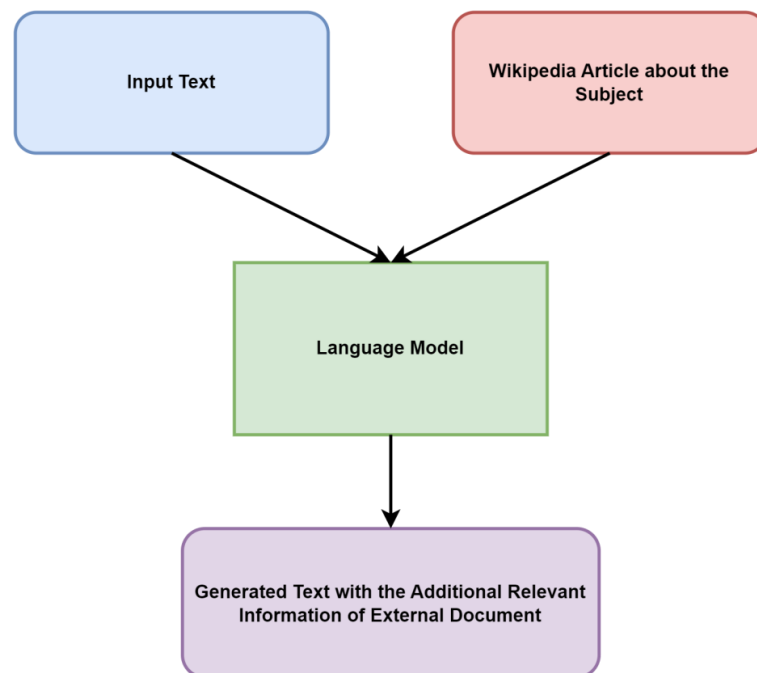


Figure 3. Document grounded controlled text generation

(Prabhumoye, 2021) introduces a new dataset that addresses the concerns of grounding in conversation responses, context, and coherence for dialogue responses, presents a dataset that has real human conversations with a grounding in a document with Wikipedia articles about movies and generates responses. (Dinan et al., 2018) also introduce a dataset of human-human conversations that are grounded in Wikipedia articles. (Zhou et al., 2018) and (Ghazvininejad et al., 2018) generates grounding the responses in external source of information for dialogue response generation task.

The most critical difficulties in document-grounded controlled text generation are that the documents used are generally unstructured, the extraction of information from the document requires extra operations such as summarization, which is open to development, and of course, the difficulty level of these operations varies according to the length of the document. It is also very difficult to generate precisely distinguishable texts for topic-controlled text generation. Because of all these, content-controlled text generation is a very important research area.

Conclusion

As a result, controllable text generation is an important and challenging issue today in terms of both style and content control. In addition to providing fluent and accurate text generation, it is necessary to generate texts in the desired style and content. This study was written to give general information about controllable text generation also to explain style, and content-controlled text generation, and gather the research under these two important headings. Nowadays, by considering these various factors, controllable text generation studies are carried out quite a lot and the importance of this field will increase gradually.

References

- Bamman D., Eisenstein J., & Schnoebelen T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135–160.
- Caglayan C. (2022). *A Novel Deep Learning Approach for Controlled Multi-topic Text Generation*. Atılım University Graduate School of Natural and Applied Sciences, Ankara, Turkey.
- Caglayan C., & Karakaya M. (2021). Topic-Controlled Text Generation. 6th International Conference on Computer Science and Engineering (UBMK), 533-536.
- Celikyilmaz, A., Clark, E., & Gao, J. (2020). Evaluation of Text Generation: A Survey. *ArXiv, abs/2006.14799*.
- Chandu K., Prabhumoye S., Salakhutdinov R., & Black A. W. (2019). “my way of telling a story”: Persona based grounded story generation. In *Proceedings of the Second Workshop on Storytelling*, pages 11–21.
- Chang H., Yuan J., Iyyer M., & McCallum A. (2021). Changing the Mind of Transformers for Topically-Controllable Language Generation. *arXiv preprint arXiv:2103.15335*.
- Devlin J., Chang M., Lee K., & Toutanova K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*, pp. 4171-4186.
- Dinan E., Roller S., Shuster K., Fan A., Auli M., & Weston J. (2018). Wizard of wikipedia: Knowledge-powered conversational agents. In *International Conference on Learning Representations*.
- Fu Z., Tan X., Peng N., Zhao D., & Yan R. (2018). Style transfer in text: Exploration and evaluation. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

- Ghazvininejad M., Brockett C., Chang M., Dolan B., Gao J., Yih W., & Galley M. (2018). A knowledge-grounded neural conversation model. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- Huang Q., Gan Z., Celikyilmaz A., Wu D., Wang J., & He X. (2019). Hierarchically structured reinforcement learning for topically coherent visual story generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8465–8472.
- Jhamtani H., Gangal V., Hovy E., & Nyberg E. (2017). Shakespearizing modern language using copy-enriched sequence-to-sequence models. In *Proceedings of the Workshop on Stylistic Variation*, pages 10–19.
- Li J., Galley M., Brockett C., Spithourakis G. P., Gao J., & Dolan B. (2016). A persona-based neural conversation model. In *Proc. ACL*.
- Li J., Jia R., He H., & Liang P. (2018). Delete, retrieve, generate: a simple approach to sentiment and style transfer. *Association for Computational Linguistics: Human Language Technologies*, Volume 1 pages 1865–1874.
- Li S., Feng S., Wang D., Song K., Zhang Y., & Wang W. (2020). EmoElicitor: An Open Domain Response Generation Model with User Emotional Reaction Awareness. In *IJCAI*. 3637–3643.
- Lin Z. & Riedl M. O. (2021). Plug-and-Blend: A Framework for Plug-and-Play Controllable Story Generation with Sketches. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 17. 58–65.
- Madaan A., Setlur A.R., Parekh T., Póczos B., Neubig G., Yang Y., Salakhutdinov R., Black A.W., & Prabhumoye S. (2020). Politeness Transfer: A Tag and Generate Approach. *Annual Meeting of the Association for Computational Linguistics*.
- Niu T., & Bansal M. (2018). Polite dialogue generation without parallel data. *Transactions of the Association for Computational Linguistics*, 6:373–389.
- Peng N., Ghazvininejad M., May J., & Knight K.. (2018). Towards controllable story generation. In *Proceedings of the First Workshop on Storytelling*, pages 43–49.
- Peterson K., Hohensee M., & Xia F. (2011). Email formality in the workplace: A case study on the Enron corpus. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, pages 86–95, Portland, Oregon. Association for Computational Linguistics.
- Prabhakaran V., Reid E. E., & Rambow O. (2014). Gender and Power: How Gender and Gender Environment Affect Manifestations of Power. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1965–1976, Doha, Qatar. Association for Computational Linguistics.
- Prabhumoye S. (2021). *Controllable Text Generation And Ethical Implications*. Language Technologies Institute School of Computer Science Carnegie Mellon University, Pittsburgh.
- Prabhumoye S., Black A. W., & Salakhutdinov R. (2020a). Exploring Controllable Text Generation Techniques. *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 1–14.
- Prabhumoye S., Li M., Urbanek J., Dinan E., Kiela D., Weston J., & Szlam A. (2020b). I love your chain mail! making knights smile in a fantasy game world: Open-domain goal-oriented dialogue agents. *arXiv preprint arXiv:2002.02878*.
- Prabhumoye S., Tsvetkov Y., Salakhutdinov R., & Black A. W. (2018). Style transfer through back-translation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 866–876.
- Radford A., Narasimhan K., Salimans T., & Sutskever I. (2018). Improving Language Understanding by Generative Pre-Training. *OpenAI*, pp. 1–10.
- Sennrich R., Haddow B., & Birch A. (2016). Controlling politeness in neural machine translation via side constraints. In *Proceedings of North American Chapter of the Association for Computational Linguistics*, pages 35–40.
- Shen T., Lei T., Barzilay R., & Jaakkola T. (2017). Style transfer from non-parallel text by cross-alignment. In *Advances in neural information processing systems*, pages 6830–6841.
- Shuster K., Humeau S., Bordes A., & Weston J. (2018). Engaging image chat: Modeling personality in grounded dialogue. *arXiv preprint arXiv:1811.00945*.
- Tang J., Zhao T., Xiong C., Liang X., Xing E., & Hu Z. (2019). Target-guided open-domain conversation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5624–5634.

- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., & Polosukhin I. (2017). Attention is all you need. NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 6000–6010.
- Wang R., Wei Z., Cheng Y., Li P., Shan H., Zhang J., Zhang Q., & Huang X. (2019). Keep it consistent: Topic-aware storytelling from an image stream via iterative multi-agent communication. arXiv preprint arXiv:1911.04192.
- Xu W., Ritter A., Dolan B., Grishman R., & Cherry C. (2012). Paraphrasing for style. In Proceedings of the 24th International Conference on Computational Linguistics, pages 2899–2914.
- Yang P., Li L., Luo F., Liu T., & Sun X. (2019). Enhancing topic-to-essay generation with external commonsense knowledge. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.
- Zhang H., Song H., LI S., Zhou M., & Song D. (2022). A Survey of Controllable Text Generation using Transformer-based Pre-trained Language Models. [CoRR abs/2201.05337](https://arxiv.org/abs/2201.05337).
- Zhang S., Dinan E., Urbanek J., Szlam A., Kiela D., & Weston J. (2018). Personalizing Dialogue Agents: I have a dog, do you have pets too?. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Melbourne, Australia, 2204–2213. <https://doi.org/10.18653/v1/P18-1205>.
- Zhou K., Prabhume S., & Black A. W. (2018). A dataset for document grounded conversations. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 708–713.