

The Automatic Generation of Contextual Questions and Answers for English Learners

Minji Kang¹, Sungjin Ahn², Hoyeol Kim³, Donghwan Kim⁴, and Sung Won Han¹
 Korea University¹ Sungkyunkwan University² Texas A&M University³ National University of Singapore⁴

Introduction

- We deployed the BERT, T5, and GPT-2 models for MC and Wh- question generation tasks.
- GPT-2 is a unidirectional language model, but it creates excellent results for text generation tasks.
- BERT is a bidirectional language model, the strength of which is predicting fill-in-the-blanks in context.
- T5 (Text-to-Text Transfer Transformer) is a pre-trained model with the C4 dataset. A T5 transformer can be fine-tuned for down-stream tasks.

Data

Description

- CoNLL 2014 and BEA 2019 are datasets containing essays written by ELs and native English speakers.
- Both datasets were originally designed for grammatical error correction tasks, so we fixed the errors on 170 of the essays to tailor them for our tasks.

Distribution Details

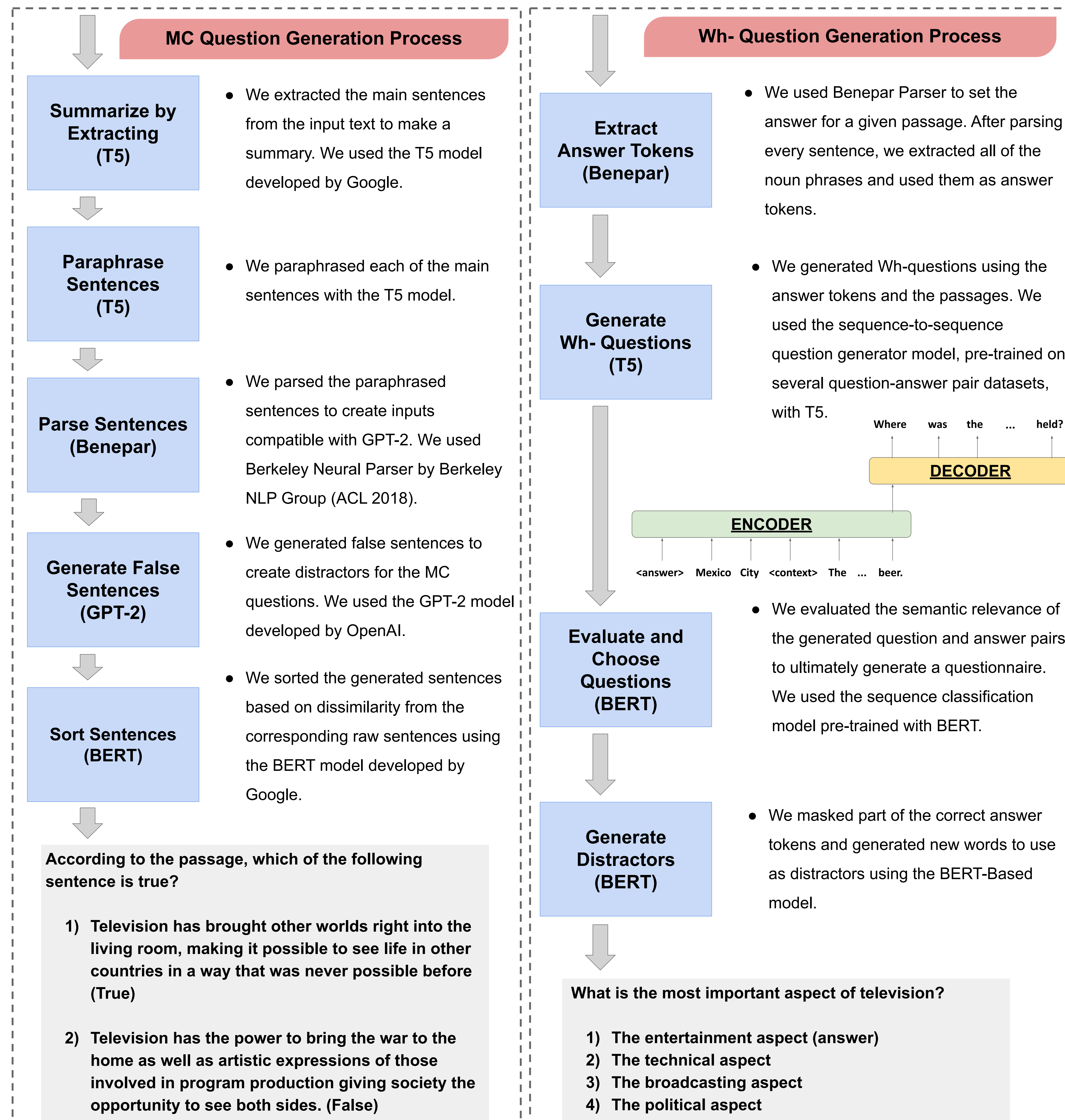
- To evaluate our generator, we randomly sampled 20 essays from the 170 corrected essays.

Datasets	Topics	Number of Essays
CoNLL 2014	Genetic risk	1
	Social media	3
BEA 2019	Write & Improve	10
	LOCNESS	6

- We created 20 MC and 20 Wh- questions, each with 4 answer choices.

Methodology

..... This instrument has brought other worlds right into the living room, making it possible to see life in other countries in a way that was never before possible. The entertainment aspect of television has offered society an easy escape from its problems and difficulties. Therefore, since the invention of television, society has become more informed but too lazy to think about this new information, and has distanced itself from family life.



Results

- Four experts rated 40 pairs of generated question and answer pairs using the given criteria and the results are shown in the table below.

- Q1: MC distractors acceptability (1~5)
- Q2: MC distractors grammar correctness (1~5)
- Q3: Wh-distractors acceptability (1~5)
- Q4: Wh-distractors grammar correctness (1~5)
- Q5: Wh-question acceptability (1~5)
- Q6: Wh-question grammar correctness (0, 1)*

	Q1	Q2	Q3	Q4	Q5	Q6
Evaluator 1	3.85	4.45	2.95	4.45	3.40	0.95
Evaluator 2	4.15	3.80	2.30	3.95	3.90	0.80
Evaluator 3	4.10	4.40	2.60	4.25	3.60	0.95
Evaluator 4	3.85	4.70	2.80	4.90	4.00	1.00
Average	3.99	4.34	2.66	4.39	3.73	0.93

Conclusion

- T5, BERT, and GPT-2 generate grammatically correct questions and distractors.
- Although the datasets were created by ESL students, they need to be tailored for MC and Wh- question generation tasks.
- There is room for improvement for MC and Wh- questions generation tasks. For instance, the model sometimes creates out-of-context Wh- distractors. Fine-tuning the model, curating the datasets, or creating tailored datasets can improve the results.

References

- Bryant, C., Felice, et al. (2019). The BEA-2019 Shared Task on Grammatical Error Correction. *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, 52-75. <https://doi.org/10.18653/v1/w19-4406>
- Devlin, Jacob, et al. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding.
- Ng, H. T., et al. (2014). The CoNLL-2014 Shared Task on Grammatical Error Correction. *Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task*, 1-14. <https://doi.org/10.3115/v1/W14-1701>
- Radford, Alec, et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- Raffel, Colin, et al. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer.
- Tutorials: <https://pseudo-lab.github.io/Tutorial-Book/chapters/NLP/intro.html>
- Homepage: <https://pseudo-lab.com/>