Generative Language Models for Paragraph-Level Question Generation

Asahi Ushio, *Fernando Alva-Manchego and Jose Camacho-Collados*

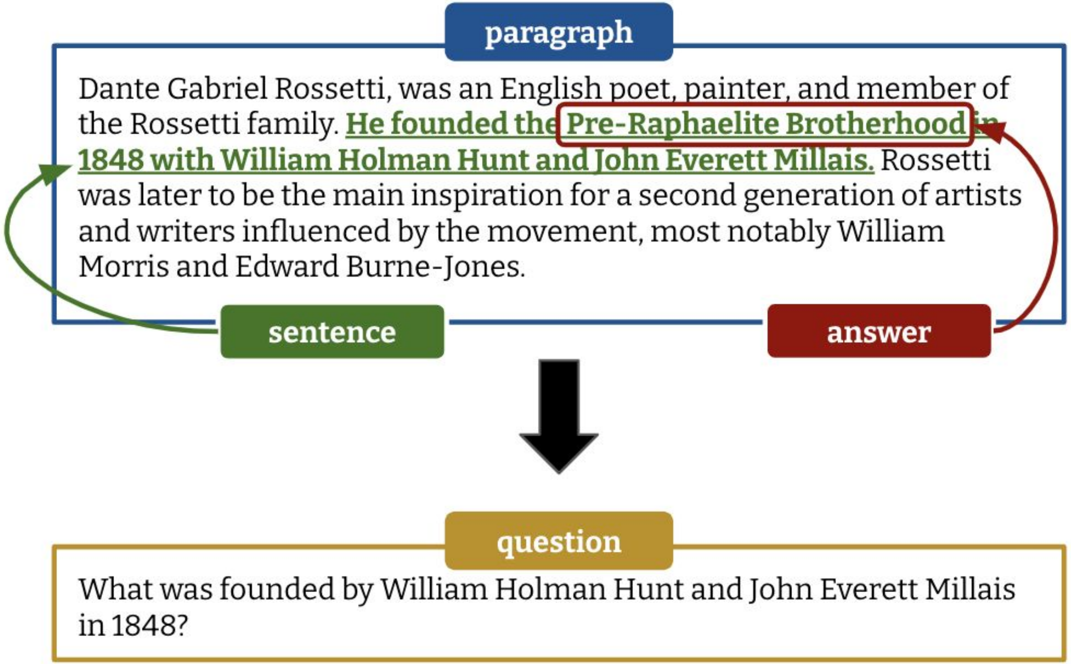
Computer Science & Informatics, Cardiff University, Cardiff NLP

https://github.com/asahi417/lm-question-generation

Outline

- What is question generation?
- QG-Bench: Unified Benchmark
  - Experimental Result
  - Manual Evaluation
- Resources

Paragraph-Level Question Generation



paragraph

Dante Gabriel Rossetti, was an English poet, painter, and member of the Rossetti family. **He founded the Pre-Raphaelite Brotherhood in 1848 with William Holman Hunt and John Everett Millais.** Rossetti was later to be the main inspiration for a second generation of artists and writers influenced by the movement, most notably William Morris and Edward Burne-Jones.

sentence

answer

question

What was founded by William Holman Hunt and John Everett Millais in 1848?

Applications of Question Generation (QG)

- Educational Service: [Heilman 2010], [Lindberg 2013]
- Domain Adaptation of QA Models: [Shakeri 2020]
- Adversarial Data Augmentation: [Paranjape 2021], [Bartolo 2021]
- LM pre-training: [Jia 2021]
- Unsupervised QA Models: [Lewis 2019], [Puri 2020]
- Nearest Neighbour QA: [Lewis 2021]
- Question Re-writing: [Lee 2020]
- Semantic Role Labeling: [Pyatkin 2021]
- Multihop Question Decomposition: [Perez 2020]
- Visual QA: [Krishna 2018]

Question generation remains understudied…🤔

- Model Selection
  - BART, T5, or ERNIE-GEN?
- Dataset
  - Domains
  - Languages
- Evaluation
  - BLEU4…?
- Effect of Input Type
  - With/without answer or paragraph/sentence

QG-Bench

← Available on HuggingFace 🤗

Multilingual & multidomain QG Benchmark Dataset.

- SQuAD style QG in 8 languages:
  - Language: en/es/de/ja/ko/fr/it/ru
  - Source: Wikipedia
- 10 domains in 2 styles (English only):
  - Objective: Amazon/Wiki/News/Reddit
  - Subjective: Book/Elec./Grocery/Movie/Restaurant/Trip

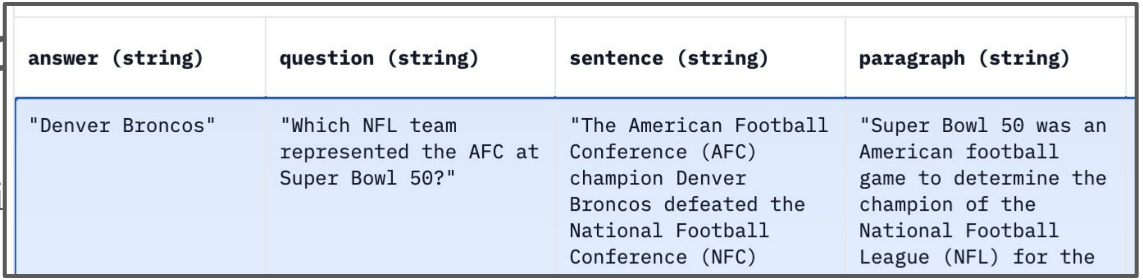
GitHub: *https://github.com/asahi417/lm-question-generation*

QG-Bench

←*Available on HuggingFace* 🤗

Multilingual & multidomain Q

- SQuAD style QG in 8 languages:
  - Language: en/es/de/ja/ko/fr/i
  - Source: Wikipedia
- 10 domains in 2 styles (English only):
  - Objective: Amazon/Wiki/News/Reddit
  - Subjective: Book/Elec./Grocery/Movie/Restaurant/Trip



answer (string)	question (string)	sentence (string)	paragraph (string)
"Denver Broncos"	"Which NFL team represented the AFC at Super Bowl 50?"	"The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC)	"Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the

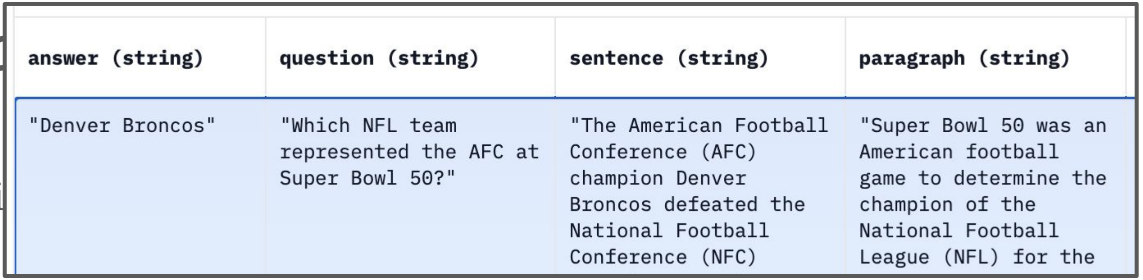
GitHub: [*https://github.com/asahi417/lm-question-generation*](https://github.com/asahi417/lm-question-generation)

QG-Bench

←*Available on HuggingFace* 🤗

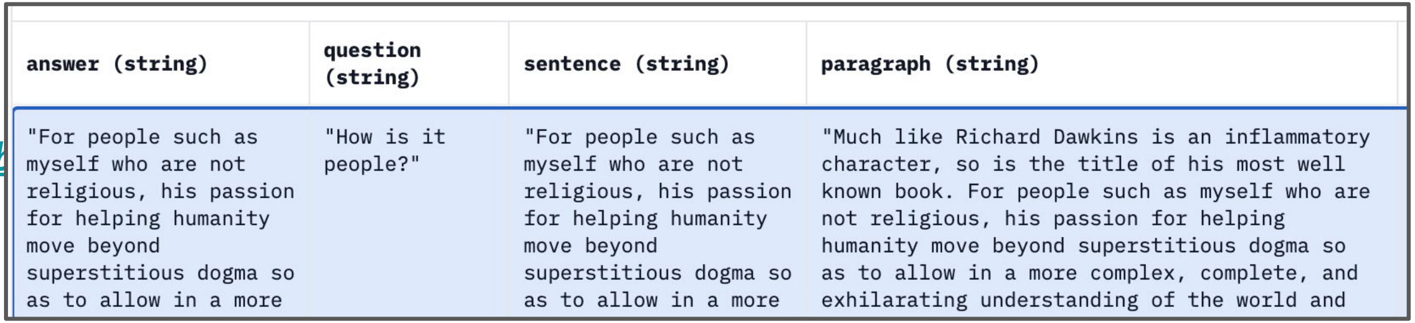
Multilingual & multidomain QC

- SQuAD style QG in 8 languages:
  - Language: en/es/de/ja/ko/fr/i
  - Source: Wikipedia
- 10 domains in 2 styles (English only):
  - Objective: Amazon/Wiki/News/Reddit
  - Subjective: Book/Elec./Grocery/Movie/Restaurant/Trip



answer (string)	question (string)	sentence (string)	paragraph (string)
"Denver Broncos"	"Which NFL team represented the AFC at Super Bowl 50?"	"The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC)	"Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the

GitHub: https://githul



answer (string)	question (string)	sentence (string)	paragraph (string)
"For people such as myself who are not religious, his passion for helping humanity move beyond superstitious dogma so as to allow in a more	"How is it people?"	"For people such as myself who are not religious, his passion for helping humanity move beyond superstitious dogma so as to allow in a more	"Much like Richard Dawkins is an inflammatory character, so is the title of his most well known book. For people such as myself who are not religious, his passion for helping humanity move beyond superstitious dogma so as to allow in a more complex, complete, and exhilarating understanding of the world and

Experiment: English QG-Bench (SQuAD)

- T5-Large achieves SoTA (ERNIE-GEN previously).
- T5-Base already outperforms previous SoTA with less parameters.
- Performance scales with the model size.

	Param (M)	BLEU4	METEOR	ROUGE-L
ERNIE-GEN	340	25.40	26.92	52.84
T5 Small	60	24.40	25.84	51.43
T5 Base	220	26.13	26.97	53.33
T5 Large	770	**27.21**	**27.70**	**54.13**

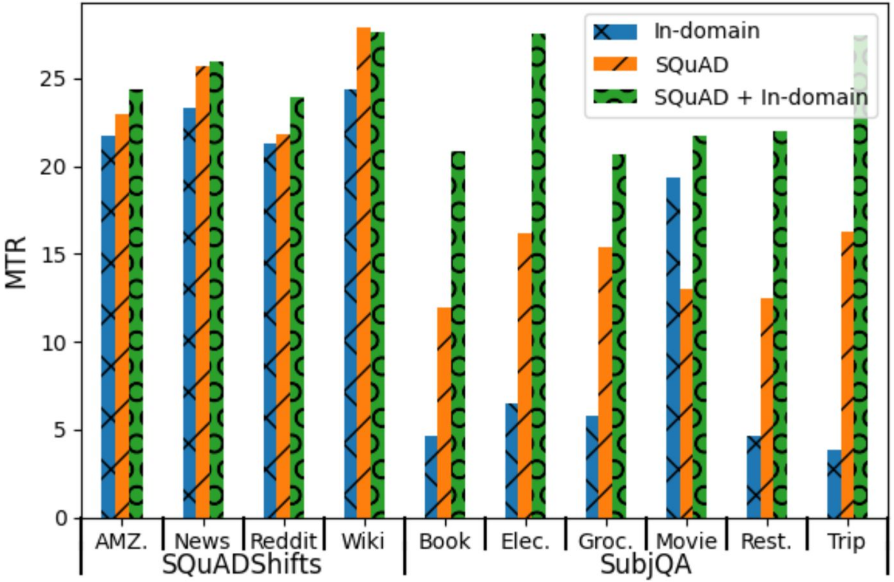
Experiment: Non-English QG-Bench

- Fine-tuning with MT5 Base
- Spanish > Italian > Korean > Russian > Japanese > French > German
- The size of the training instance matters.
    - Spanish (77,025) and Italian (46,550) vs French (17,543) and German (9,314)

Language	BLEU4	METEOR	ROUGE-L
EN	23.03	25.18	50.67
RU	17.63	28.48	33.02
JA	32.40	30.58	52.67
IT	7.70	18.00	22.51
KO	12.18	29.62	28.57
ES	10.15	23.43	25.45
DE	0.87	13.65	11.10
FR	6.14	15.55	25.88

Experiment: Domain Adaptability

- Fine-tuning with T5 Large on:
  - Domain's training set
  - SQuAD training set
  - SQuAD -> domain's training set
- In-domain training set is too small for fine-tuning (~3k).
- SQuAD fine-tuned models cannot generalize well, especially on SubjQA.
- Continuous fine-tuning (SQuAD -> domain) works the best.



Manual Evaluation

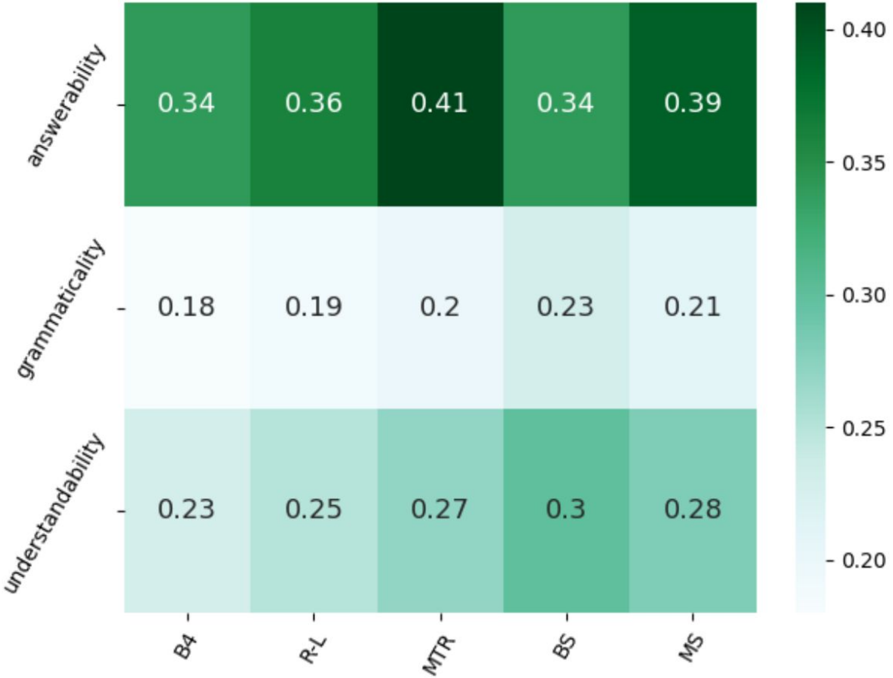
Criteria:

- **Answerability**: whether the question can be answered by the given input answer.
- **Grammaticality**: grammatical correctness.
- **Understandability**: whether the question is easy to be understood by readers.

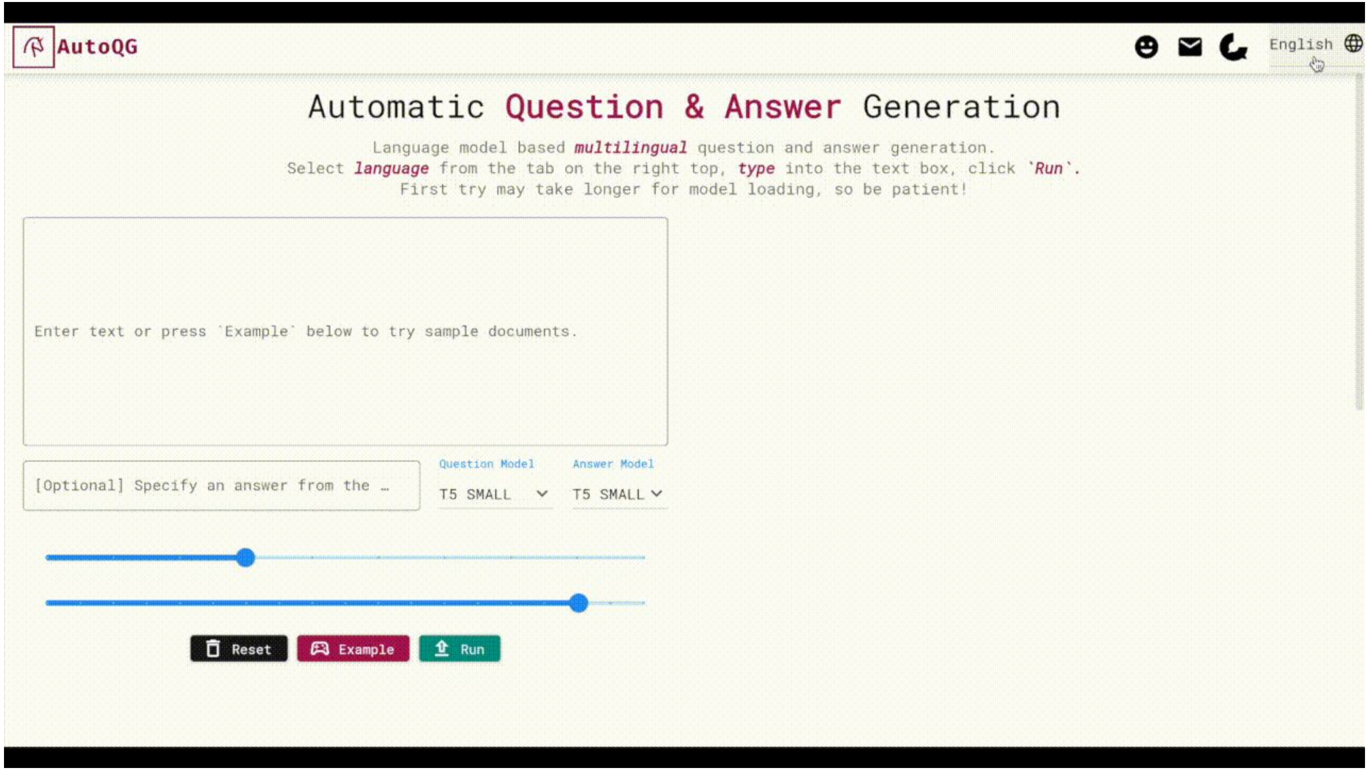
Models: BART, T5, LSTM models.

BLEU4 (B4) does not correlate well with human judgements.

METEOR (MTR) & BERTScore (BS) are more robust.



AutoQG https://autoqg.net/



AutoQG

English

Automatic Question & Answer Generation

Language model based *multilingual* question and answer generation.
Select *language* from the tab on the right top, *type* into the text box, click `Run`.
First try may take longer for model loading, so be patient!

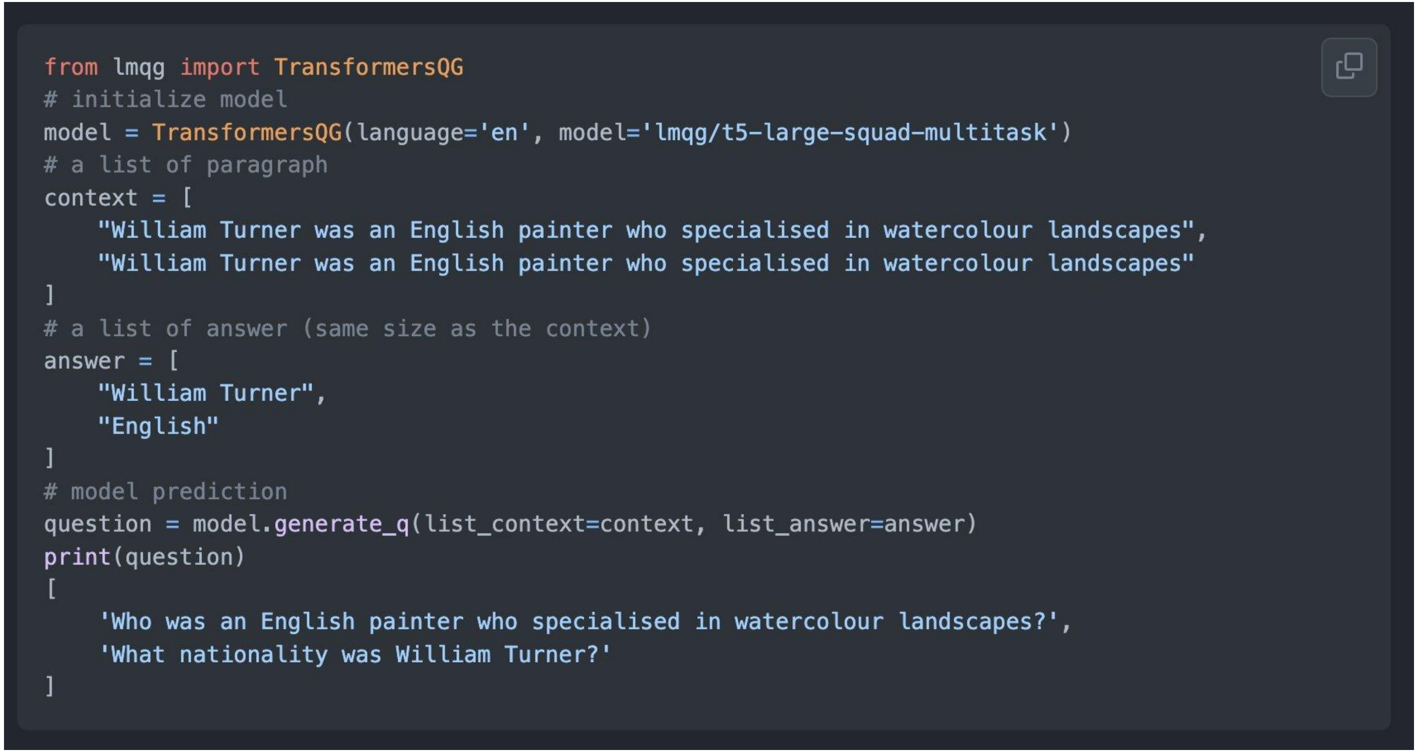
Enter text or press `Example` below to try sample documents.

[Optional] Specify an answer from the …

Question Model: T5 SMALL

Answer Model: T5 SMALL

QG Models Available via `pip install lmqg`



```
from lmqg import TransformersQG
# initialize model
model = TransformersQG(language='en', model='lmqg/t5-large-squad-multitask')
# a list of paragraph
context = [
    "William Turner was an English painter who specialised in watercolour landscapes",
    "William Turner was an English painter who specialised in watercolour landscapes"
]
# a list of answer (same size as the context)
answer = [
    "William Turner",
    "English"
]
# model prediction
question = model.generate_q(list_context=context, list_answer=answer)
print(question)
[
    'Who was an English painter who specialised in watercolour landscapes?',
    'What nationality was William Turner?'
]
```

Thank you!!