Question Generation for Reading Comprehension Assessment by Modeling How and What to Ask

[1]Bilal Ghanem, Lauren Lutz Coleman, Julia Rivard Dexter,
Spencer McIntosh von der Ohe, Alona Fyshe

[1]bghanem@ualberta.ca

Problem Setting

Reading is integral to everyday life, and yet learning to read is a struggle for many **children**.

Teachers can use **comprehension questions** to increase engagement, test reading skills, and improve retention. However, writing questions time consuming.

Here we focus on **inferential** questions, which are a better test reading comprehension skills that the literal questions in previous work.

We ask:

- Can generative models generate inferential questions?
- Can we control the type of the generated questions?

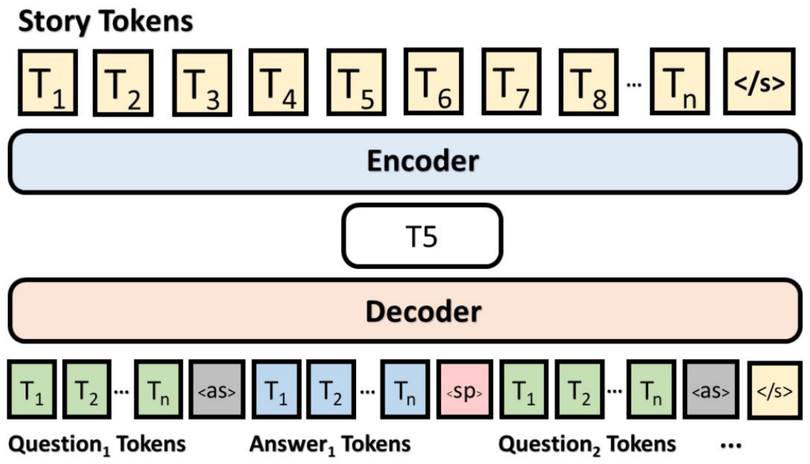


Figure 1: Input and output format of the **How to Ask** (**HTA**) model.

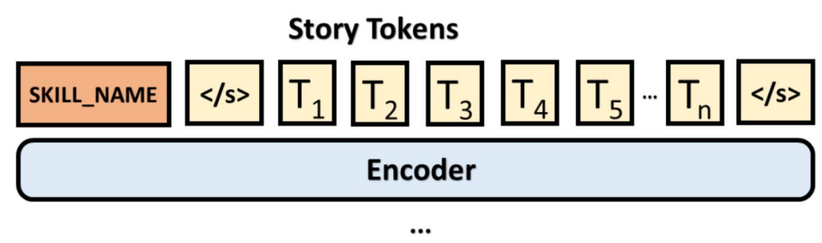


Figure 2: Input format of the **What to Ask** (**WTA**) model. The output format is the same as in HTA model (see Figure 1).

Examples

Story: "The map says left", said Bri. "*But my heart says right!*" cried Rob. "Is your heart full of hidden treasure?" asked Bri. "Yes." Rob replied. "At least, that's what my mom says."

Generated Inferring question: Why do you think Bri's heart says "But my heart says right!"?

Generated answer (correct): Because she thinks she has found something.

Story: The Line 1 Toronto train was a subway like many others you've seen. *He rocketed down Yonge Street, around the Union loop, and rattled off towards Vaughn*. At Vaughn he'd let out a loud, hissing sigh and a clanking sort of grunt, then reverse and do the whole thing backwards all over again. He liked his transit union job well enough, but he couldn't help thinking about the lights at the end of his tunnels. No matter how long he'd been running, or how much he wished for anything else, that little hopeful point of light always turned out to be just one more dirty subway platform.

Generated Figurative Language question: Reread this sentence: "He rocketed down Yonge Street, around the Union loop, and rattled off towards Vaughn." Which figurative language technique is being used here?

Generated answer (wrong): Alliteration.

Data

We built questions for the following skill types:

- **Basic Story Elements**
- **Character Traits**
- **Close Reading**
- **Figurative Language**
- **Inferring**
- **Predicting**
- **Summarizing**
- **Visualizing**
- **Vocabulary**

Our dataset has 726 children's stories and 4K question-answer pairs, with an average of 5.5 pairs per story. 25 professionals (18 teachers, 7 graduate students) worked together to create the dataset.

Approach

We propose a two-steps approach to train T5 transformer models:

- In the first step, we train the model on external data to teach it How to Ask (**HTA**), without controlling the skill of the generated questions (Figure 1).
- Secondly, we used our new dataset to to teach the model What to Ask (**WTA,** Figure 2).

Results

Models' performances

Model	BLEU-4	BLEURT
Vanilla Seq2seq	2.37	08.42
NQG-Seq	2.49	11.13
NQG-Max	2.77	31.78
CGC-QG	5.68	29.28
AnswerQuest	4.71	29.15
One-Step	2.94	29.45
T5-WTA	3.92	32.96
HTA-WTA	**7.67**	**34.82**

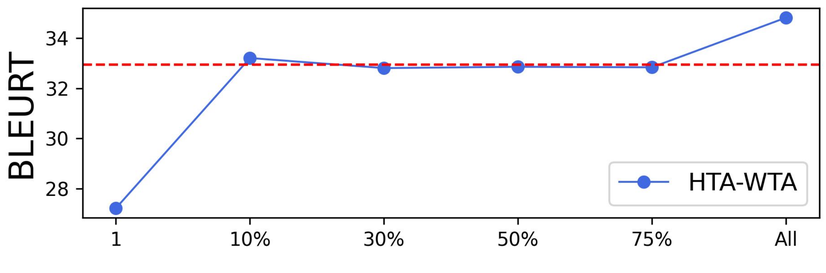
Inferential vs Literal BLEURT scores

Model	Inferential	Literal
One-Step	28.44	30.63
T5-WTA	33.13	32.78
HTA-WTA	35.45	34.08

Human Evaluation: Answerability (Ay), Fluency (Fy), and Grammaticality (Gy)

Model	Ay	Fy	Gy	Skills Accuracy
One-Step	3.82	4.28	4.37	0.16
HTA-WTA	3.89	4.29	4.45	0.8

Effect of dataset size on BLEURT



This work has been done in collaboration with:

DREAMSCAPE

UNIVERSITY OF ALBERTA
CAMPUS DESIGN & PRINT SOLUTIONS