# Let’s Ask Again: Refine Network for Automatic Question Generation


Preksha Nema*[1], Akash Kumar Mohankumar*[1], Mitesh M. Khapra[1], Balaji Vasan Srinivasan[2], Balaraman Ravindran[1]

[1]Robert Bosch Center for Data Science and Artificial Intelligence, IIT Madras, India

[2]Adobe Research India

{preksha,miteshk,ravi}@cse.iitm.ac.in, {makashkumar99}@gmail.com , {balsrini}@adobe.com


## 1. Motivation

- Existing Approaches focus on generating questions in *one single pass*. However, we observe that the generated questions look like an incomplete draft of the desired question with a clear scope for refinement.
- We propose a method which tries to mimic the human process of generating questions by first creating an initial draft and then refining it

**Passage 1**

Liberated by **Napoleon's** army in 1806, Warsaw was made the capital of the newly created Duchy of Warsaw.

**Generated Questions**

| | |
|---|---|
| ***Baseline*** | What was **the capital** of the newly duchy of Warsaw? |
| ***RefNet*** | **Who liberated** Warsaw in 1806? |
| ***Reward-RefNet*** | **Whose army** liberated Warsaw in 1806? |

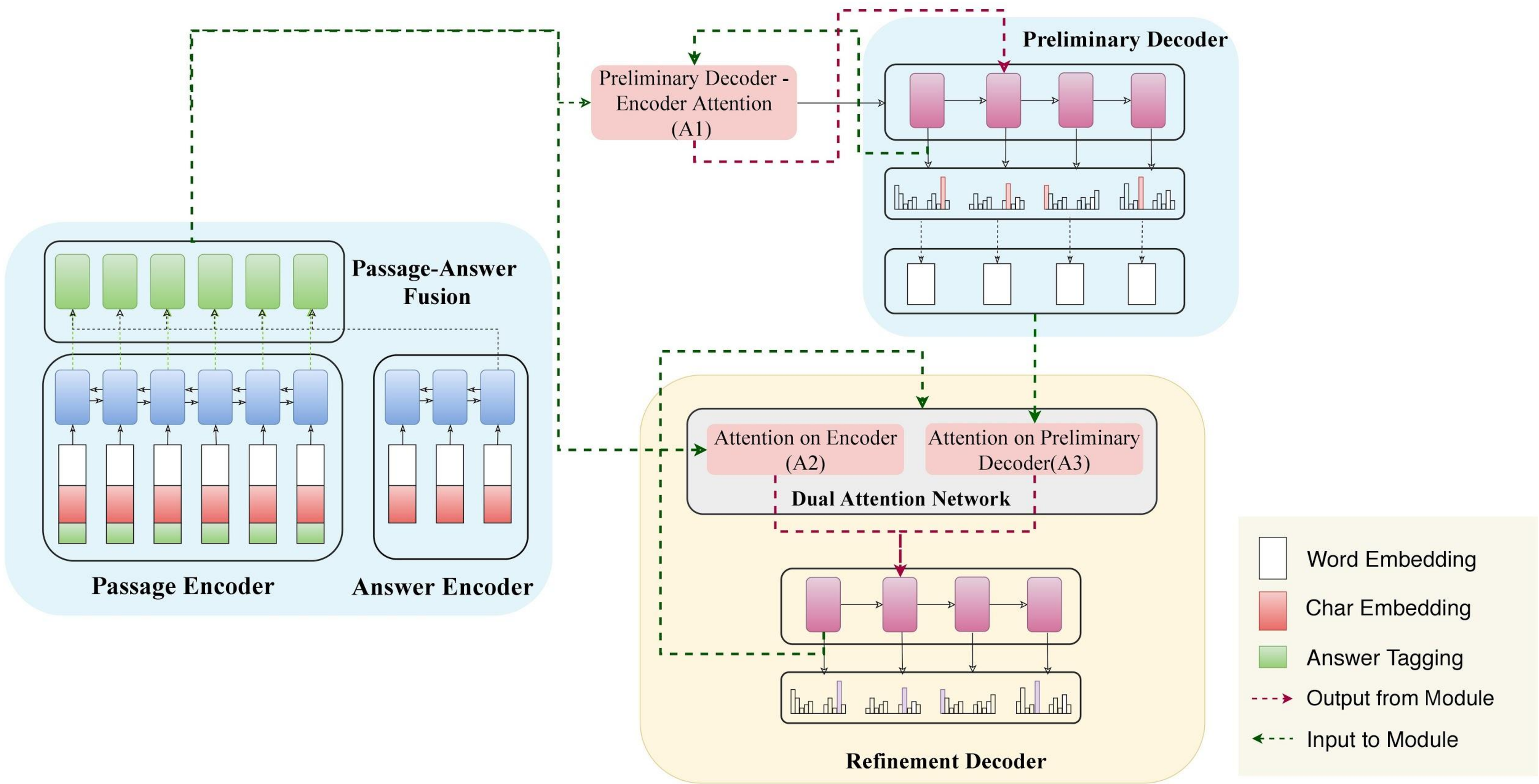
**Passage 2**

To fix carbon dioxide into sugar molecules in the process of photosynthesis, chloroplasts use **an enzyme called rubisco**

**Generated Questions**

| | |
|---|---|
| ***Baseline*** | What does chloroplasts use? |
| ***RefNet*** | What does chloroplasts use **to fix carbon dioxide into sugar molecules?** |
| ***Reward-RefNet*** | What **do** chloroplasts use to fix carbon dioxide into sugar molecules? |

## 2. Proposed Model

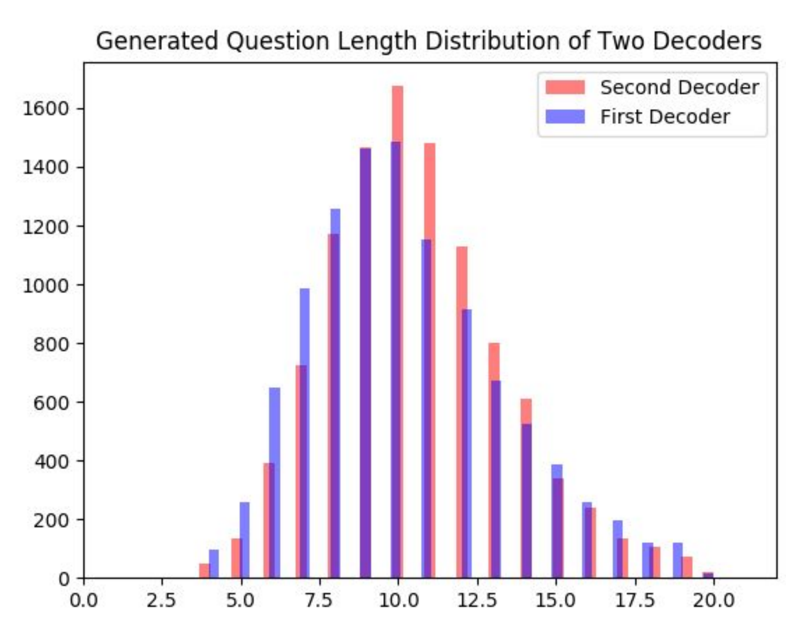



## 3. Results

| Dataset | Model | n-gram BLEU-1 | BLEU-2 | BLEU-3 | BLEU-4 | ROUGE-L | METEOR | QBLEU4 |
|---|---|---|---|---|---|---|---|---|
| SQuAD (Sentence Level) | (Sun et al., 2018) | 43.02 | 28.14 | 20.51 | 15.64 | - | - | - |
| | (Zhao et al., 2018) | 44.51 | 29.07 | 21.06 | 15.82 | 44.24 | 19.67 | - |
| | (Kim et al., 2019) | - | - | - | 16.17 | - | - | - |
| | EAD | 44.74 | 29.79 | 22.00 | 16.84 | 44.78 | 20.60 | 24.70 |
| | RefNet | **47.27** | **31.88** | **23.65** | **18.16** | **47.14** | **23.40** | **27.40** |
| SQuAD (Passage Level) | (Zhao et al., 2018)* | 45.07 | 29.58 | 21.60 | 16.38 | 44.48 | 20.25 | - |
| | EAD | 44.61 | 29.37 | 21.50 | 16.36 | 43.95 | 20.11 | 24.20 |
| | RefNet | **46.41** | **30.66** | **22.42** | **16.99** | **45.03** | **21.10** | **26.60** |
| Hotpot QA | (Zhao et al., 2018)* | 45.29 | 32.06 | 24.43 | 19.29 | 40.40 | 19.29 | 25.70 |
| | EAD | **46.00** | 32.47 | 24.82 | 19.68 | 41.52 | 23.27 | 26.20 |
| | RefNet | 45.45 | **33.13** | **26.05** | **21.17** | **43.12** | **25.81** | **28.70** |
| Drop Dataset | (Zhao et al., 2018)* | 39.56 | 29.19 | 22.53 | 18.07 | 45.01 | 19.68 | 31.40 |
| | EAD | 39.21 | 29.10 | 22.65 | 18.42 | 45.07 | 19.56 | 31.80 |
| | RefNet | **42.81** | **32.63** | **25.78** | **21.23** | **47.49** | **22.25** | **33.60** |

## 4. Discussions

### Preliminary and Refinement Decoder




| Model | Decoder | BLEU-4 | QBLEU-4 |
|---|---|---|---|
| ***without $A_3$*** | Refinement Decoder | 17.16 | 25.80 |
| | Preliminary Decoder | 17.59 | 26.00 |
| ***with $A_3$*** | Refinement Decoder | 18.37 | 27.40 |
| | Preliminary Decoder | 17.89 | 26.00 |

### Reward-RefNet

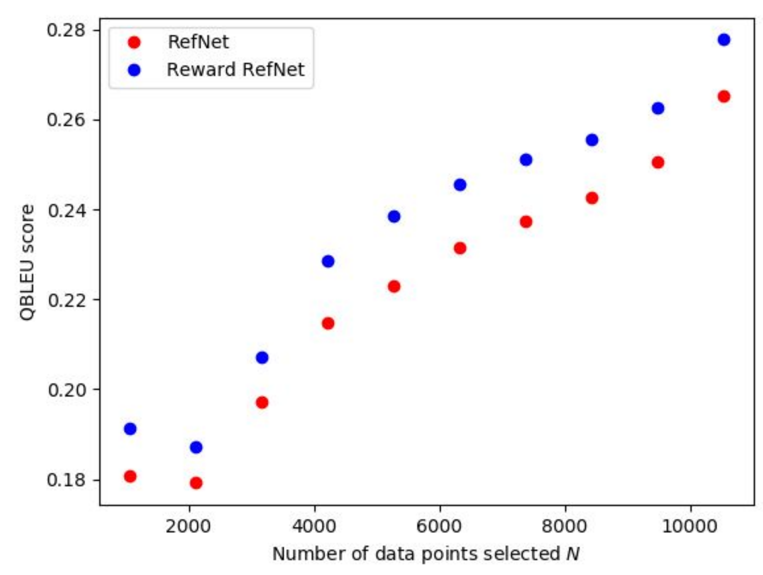
Given Preliminary & Refinement Decoder’s Generated word sequence $\tilde{\mathbf{Q}} = \{\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_T\}$ and $\mathbf{Q} = \{q_1, q_2, \ldots, q_T\}$ respectively, the training loss is defined as follows:

$$L(\mathbf{Q}) = (r(\mathbf{Q}) - r(\tilde{\mathbf{Q}})) \cdot \sum_{t=1}^{T} \log p(q_t | q_{t-1}, \ldots, q_1, \tilde{\mathbf{Q}}, \mathbf{U}, \mathbf{h}^a)$$

where r(**Q**) and r($\tilde{\mathbf{Q}}$) are the rewards obtained by comparing with the reference question **Q*** .

| Model | BLEU Reward Signal | | Answerability Reward Signal | |
|---|---|---|---|---|
| | *BLEU-4* | *%preference* | *QBLEU-4* | *%preference* |
| ***RefNet*** | 18.37 | 32.9% | 36.9 | 30% |
| ***Reward-RefNet*** | 18.52 | 67.1% | 37.5 | 70% |

### Analysis on Originality




**Passage:** McLetchie was elected on the Lothian regional list with Conservatives suffered a net loss of five seats, with leader **Annabel Goldie** **claiming that their support had held firm,** nevertheless, she too announced she would step down as leader of the party

| | |
|---|---|
| ***Baseline*** | Who announced she would step down as leader of the Conservatives ? |
| ***RefNet*** | Who **claiming that their support had held firm ?** |
| ***Reward-RefNet*** | Who was the leader of the conservatives ? |

## 5. Conclusion

- RefNet outperforms existing state-of-the-art models on the SQuAD, HOTPOT-QA and DROP datasets.
- We validate our empirical results using human evaluations.
- We show that Reward-RefNets improves the initial draft on specific aspects like Answerability, Fluency and Originality.

## Acknowledgements


We thank Amazon Web Services for providing free GPU compute and Google for supporting Preksha Nema’s contribution in this work through Google Ph.D. Fellowship programme. We would like to acknowledge Department of Computer Science and Engineering, IIT Madras and Robert Bosch Center for Data Sciences and Artificial Intelligence, IIT Madras (RBC-DSAI) for providing us sufficient resources. We would also like to thank Patanjali SLPSK, Sahana Ramnath, Rahul Ramesh, Anirban Laha, Nikita Moghe and the anonymous reviewers for their valuable and constructive suggestions.


## References


•Xingwu Sun, Jing Liu, Yajuan Lyu, Wei He, Yan-jun Ma, and Shi Wang. 2018. Answer-focused and position-aware neural question generation.

•Yao Zhao, Xiaochuan Ni, Yuanyuan Ding, and QifaKe. 2018. Paragraph-level neural question generation with maxout pointer and gated self-attention networks.

•Yanghoon Kim, Hwanhee Lee, Joongbo Shin, and Kyomin Jung. 2019. Improving neural question generation using answer separation


*Code Available at :*
*https://github.com/PrekshaNema25/RefNet-QG*