

# Construction of a Quality Estimation Dataset for Automatic Evaluation of Japanese Grammatical Error Correction

Daisuke Suzuki ♠, Yujin Takahashi ♠, Ikumi Yamashita ♠, Taichi Aida ♠, Toshio Hirasawa ♠,  
 Michitaka Nakatsuji ♠, Masato Mita ♣, ♠, Mamoru Komachi ♠

<https://github.com/tmu-nlp/jqe4gec/>

♠ Tokyo Metropolitan University; ♣ RIKEN

## Introduction

- Quality estimation (QE) models can evaluate grammatical error correction (GEC) systems without relying on reference sentences
- Yoshimura+ (2020) showed English QE model has high correlation with human, but Japanese QE models are not available owing to the absence of datasets

## Contribution

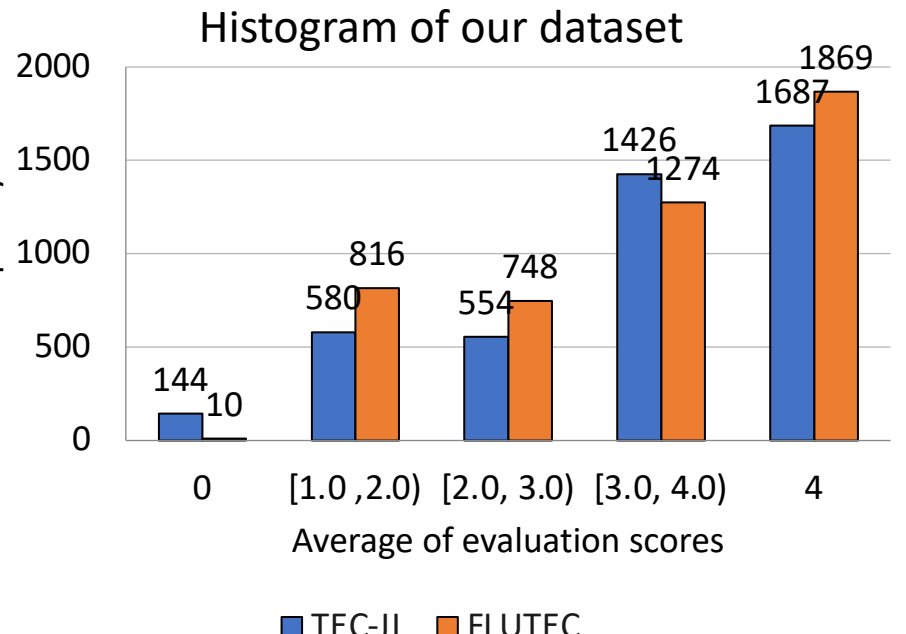
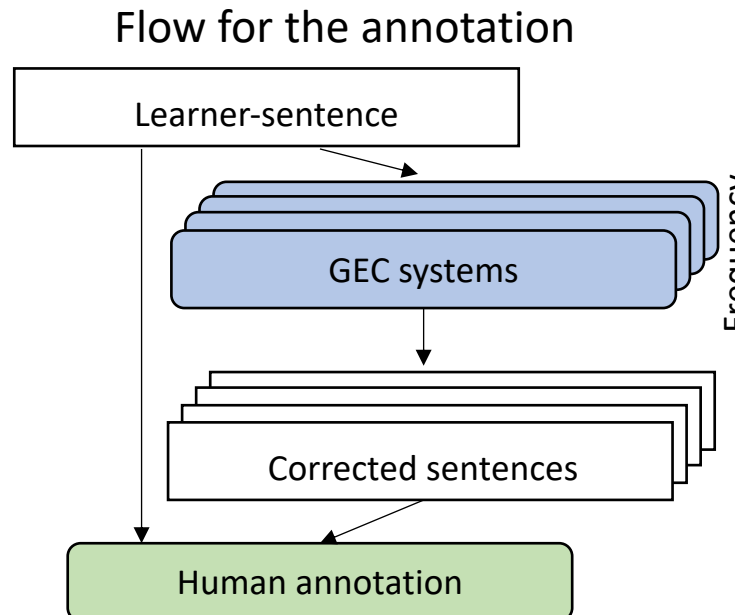
- Created a Japanese QE dataset for GEC via manual evaluation based on the NAIST Lang-8 Learner Corpora (Mizumoto+, 2011)
- Constructed a QE model using the dataset and demonstrated the usefulness of a QE model for Japanese GEC

## Annotation

- Get learner-sentences from Lang-8 derived corpora
  - TEC-JL (Koyama+, 2020): Evaluation dataset for Japanese GEC (2,042 sents)
  - FLUTEC (Kiyama+, 2022): Fluency-edited evaluation dataset (2,100 sents)
- Use a variety of GEC systems to generate system outputs
  - SMT (moses), RNN (fairseq), CNN (fairseq), Transformer (fairseq)
- Retain unique pairs of learner-sentences and system-outputs
- Assign QE scores (0-4) to the pairs by three Japanese-speaking university students

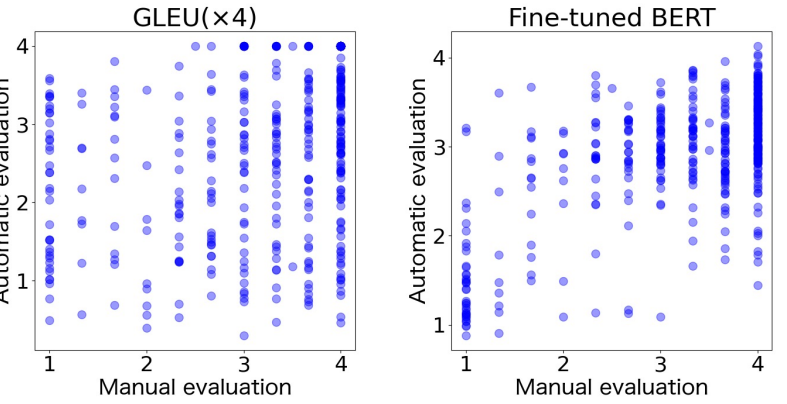
## Experiments

- Built a BERT-based QE model using our dataset (4,391 sents)
- Baseline: GLEU (Napoles+, 2016)
- Compare Pearson's and Spearman's coefficients against human scores



Results of the QE model

Method	Pearson	Spearman
GLEU	0.320	0.362
Fine-tuned BERT	<b>0.580</b>	<b>0.413</b>



### Successful examples (S = src, C = correction, R = reference, E = evaluation)

Example 1

S ソンあさんが好きがっているのは推理小説です。  
 (Son A wants to like reading mysteries.)

C ソンあさんが好きになっているのは推理小説です。  
 (Son A is becoming fond of mystery novels.)

R ソンアさんが好んでいるのは推理小説です。  
 (Sonna likes to read mysteries.)  
 ソンアさんが好きなのは推理小説です。  
 (Sonna likes to read mysteries.)

E GLEU, BERT, Human / 0.59, 3.30, 3.00

