

Cross-Corpora Evaluation and Analysis of Grammatical Error Correction Models

Is Single-Corpus Evaluation Enough?

Masato Mita^{1,2} (masato.mita@riken.jp), Tomoya Mizumto¹, Masahiro Kaneko^{3,1}, Ryo Nagata^{4,1}, Kentaro Inui^{2,1}

¹RIKEN Center for Advanced Intelligence Project (AIP), ²Tohoku University, ³Tokyo Metropolitan University, ⁴Konan University

Ans: No. it's not enough. GEC systems should be tested over multiple diverse datasets!!

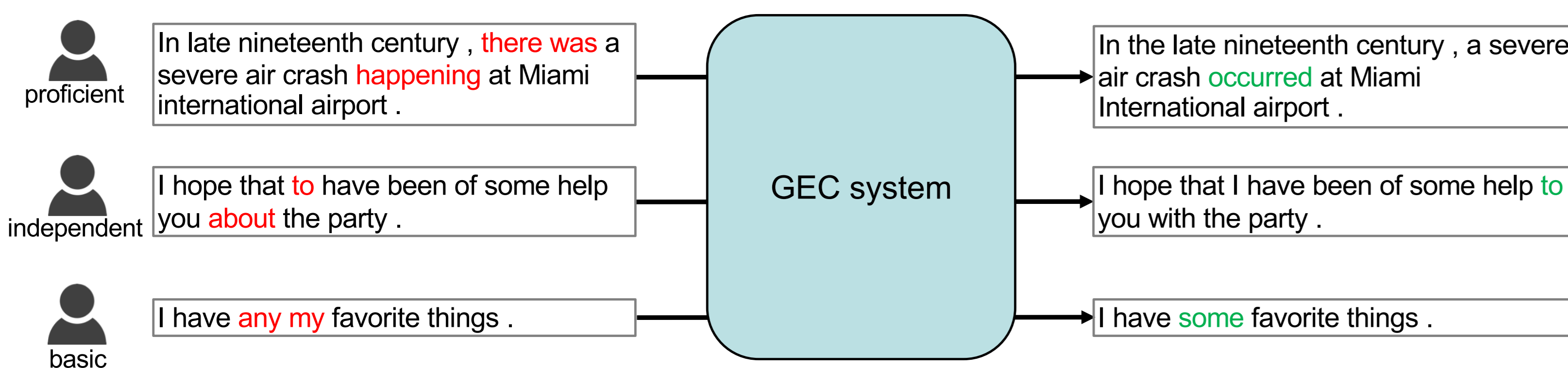
CoNLL-2014	CoNLL-2013	FCE	JFLEG	KJ	ICNALE
Transformer 48.62	Transformer 36.20	LSTM 41.41	Transformer 60.06	LSTM 45.64	LSTM 43.02
LSTM 48.57	LSTM 33.76	CNN 40.06	CNN 57.49	CNN 45.40	CNN 40.78
SMT 46.80	CNN 33.67	Transformer 39.57	LSTM 57.47	Transformer 42.80	Transformer 37.72
CNN 46.16	SMT 32.30	SMT 26.79	SMT 42.35	SMT 32.04	SMT 32.91

https://github.com/tomo-wb/GEC_CCE

Background

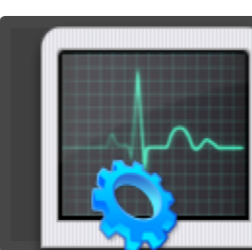
Task:

- Input: **ungrammatical** sentence
- Output: **grammatical** sentence
- GEC input is expected to have many variations due to the influence of writer's proficiency



Issue:

- GEC system performance is still primarily benchmarked against the narrow domain data set ---CoNLL-2014 test set



Experiments on Cross-Corpora

Systems:

- LSTM
- CNN [Chollampatt et al., 2017]
- Transformer
- SMT [Junczyes-Dowmunt et al., 2017]

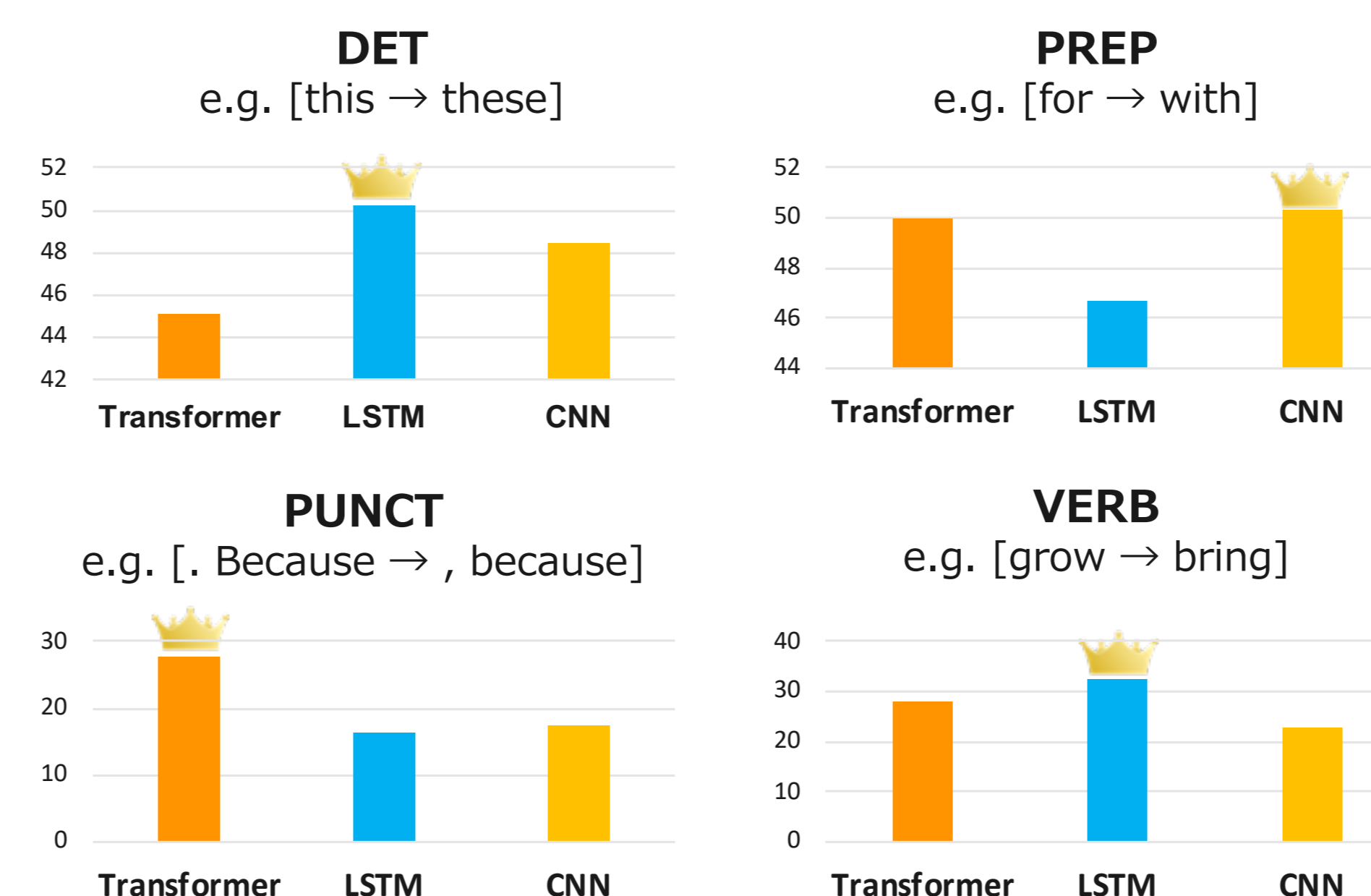
Data set:

- Train: Lang-8 + NUCLE (1.3 M)
- Dev: NUCLE (5.4K)

Corpus	# topics	Multiple L1	Multiple proficiency
CoNLL-2014 [Ng et al., 2014]	2		
CoNLL-2013 [Ng et al., 2013]	2		
FCE [Yannakoudakis et al., 2011]	10	✓	✓
JFLEG [Napoles et al., 2017]	Many	✓	✓
KJ [Nagata et al., 2011]	10		
ICNALE [Ishikawa, 2013]	2	✓	✓

Analysis1: System Performance by Error Type (CoNLL-2014)

- Each system has different strengths and weaknesses



Analysis2: Top System by Error Type (Cross-Corpora)

- Each corpus has different tendency errors

	CoNLL-2014	CoNLL-2013	FCE	JFLEG	KJ	ICNALE
DET	LSTM	LSTM	LSTM	LSTM	CNN	CNN
PREP	CNN	Transformer	Transformer	Transformer	LSTM	LSTM
PUNCT	Transformer	Transformer	Transformer	Transformer	LSTM	Transformer
VERB	Transformer	CNN	Transformer	LSTM	LSTM	LSTM