Cambridge ALTA
Institute for Automated Language Teaching and Assessment

UNIVERSITY OF CAMBRIDGE

DOCUMENT-LEVEL GRAMMATICAL ERROR CORRECTION

ZHENG YUAN AND CHRISTOPHER BRYANT

INTRODUCTION

- Grammatical error correction (GEC) attempts to automatically detect and correct grammatical errors in text.
- Document-level context can provide valuable information, which is crucial for correcting certain errors and resolving inconsistencies.
- Sentence-level systems may fail to correct document-level errors.

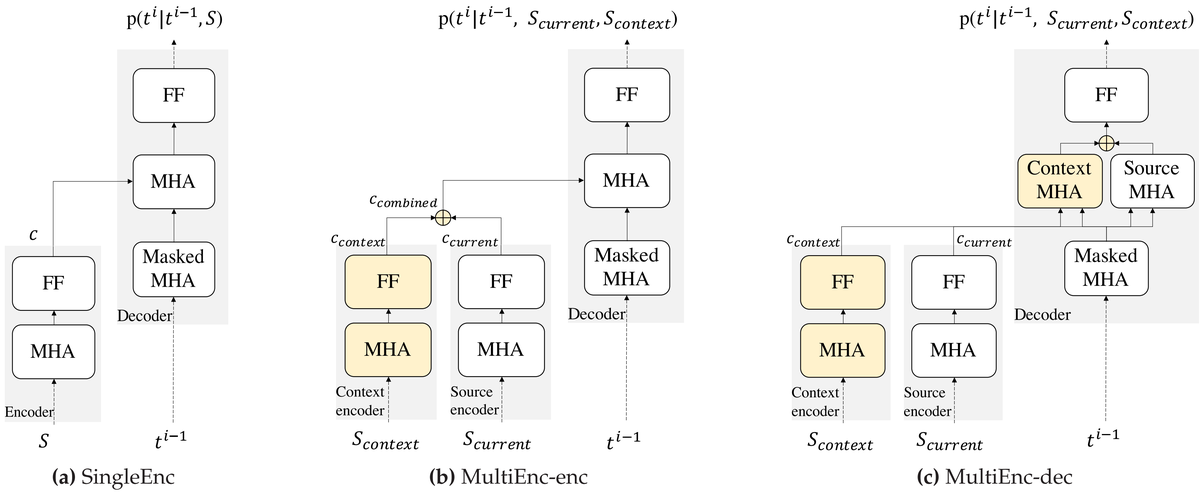
Example (a)

*In the chat room, she created a close relationship with eight people. She talks (**talked**) to them every night, trust (**trusted** / ~~**trusts**~~) them and share (**shared** / ~~**shares**~~) her life with them. Then eventually, she discovered that the eight people were one as the other person was using eight different identities to chat with her all the time.*

Example (b)

*I would like to recommend walking. Because there are a lot of beautiful trees. → I would like to recommend walking **because** there are a lot of beautiful trees.*

DOCUMENT-LEVEL GEC MODELS



(a) SingleEnc **(b)** MultiEnc-enc **(c)** MultiEnc-dec

Figure 1: Document-level GEC models (FF: Feed Forward, MHA: Multi-Head Attention). The newly introduced components are highlighted in yellow.

DOCUMENT-LEVEL EVALUATION

- We perform the first document-level GEC evaluation with the ERRANT Scorer.
- We produce new reference files at the document level to retain edits that cross sentence boundaries.
- For datasets with multiple references (i.e. CoNLL-2014), scores are computed against all the document-level edits of a single annotator simultaneously rather than mixed-and-matched from different annotators for each sentence.

RESULTS

Document-level evaluation:

Model	BEA-dev			FCE-test			CoNLL-2014		
	P	R	$F_{0.5}$	P	R	$F_{0.5}$	P	R	$F_{0.5}$
Baseline	58.49	38.29	52.91	63.65	42.27	57.80	59.96	27.08	48.25
SingleEnc	56.94	**43.16**	53.52	61.63	**44.95**	57.37	59.78	27.27	48.27
MultiEnc-enc	62.06	41.71	56.54	**65.55**	42.68	59.20	63.23	27.96	50.49
MultiEnc-dec	**62.64**	40.72	**56.55**	65.36	44.17	**59.64**	**64.57**	**28.65**	**51.62**

Sentence-level evaluation (comparison with NMT-based GEC systems):

System	FCE-test			CoNLL-2014		
	P	R	$F_{0.5}$	P	R	$F_{0.5}$
MultiEnc-dec	**69.9**	44.2	**62.6**	**74.3**	39.0	**62.9**
Chollampatt et al. (2019)	52.2	28.3	44.6	65.6	30.1	53.1
Kaneko et al. (2020)	65.0	49.6	61.2	69.2	45.6	62.6
Lichtarge et al. (2020)	-	-	-	69.4	43.9	62.1

ERROR ANALYSIS

- The biggest gains are observed for subject-verb agreement, preposition, noun number, determiner and pronoun errors.
- This confirms our hypothesis that correction of errors involving agreement, coreference or tense is more likely to rely on information outside the current sentence.
- Our system is good at handling errors that cross sentence boundaries.
- Manual inspection reveals that improvements also come from topic-aware lexical choice.

Example (a)

Context	*Then we went to Taxco.*
Source	*We stay in a very luxurious hotel.*
Baseline	*We stay in a very luxurious hotel.*
Our model	*We **stayed** in a very luxurious hotel.*

Example (b)

Context	*The motorcycle is the most dangerous transport ...*
Source	*... some riders still keep breaking the rule.*
Baseline	*... some cyclists still keep breaking the rule.*
Our model	*... some riders still keep breaking the rule.*

CONCLUSION

- Context is useful in GEC but very long context is not necessary for improved performance.
- Our best system outperforms all NMT-based single-model GEC systems and achieves state of the art on FCE-test.
- By drawing attention to this understudied area in GEC, we hope to motivate future efforts to build better context-aware GEC systems.