

# JFLEG: A Fluency Corpus and Benchmark for Grammatical Error Correction

Courtney Napoles



Keisuke Sakaguchi

Joel Tetreault



## Introduction

- Before now the only public GEC test set was from the CoNLL 2014 Shared Task (CoNLL-14)
- CoNLL-14 and other GEC corpora only have minimal edit corrections and not holistic **fluency** edits
- We worry that GEC is optimizing to CoNLL-14 and minimal edits

## Johns Hopkins Fluency Extended GUG Corpus (JFLEG)

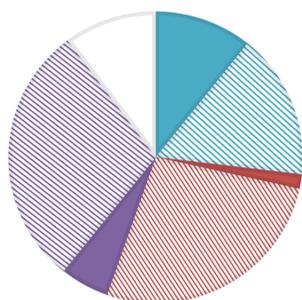
- A new *fluency* data set for tuning and testing GEC
- JFLEG addresses limitations of CoNLL-14:

	CoNLL-14	JFLEG
<b>Changes:</b>	Minimal edits	Fluency + minimal edits as needed
<b>Source:</b>	NUCLE [Dahlmeier et al. 2013]	GUG [Heilman et al. 2014]
<b>Writers:</b>	University students in Singapore	Non-native speakers from many countries
<b># Topics:</b>	2	Many
<b>Proficiency:</b>	High	Low–High
<b>Size:</b>	1300 sentences	1500 sentences
<b>Annotators per sentence:</b>	2 English instructors	4 Turkers in USA

- What kinds of errors and corrections are in JFLEG?

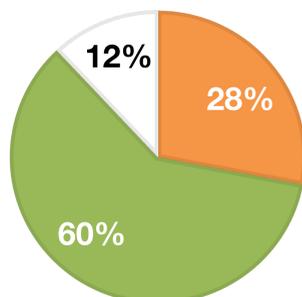
### Error types in original sentence

- Awkward
- Orthographic error
- Ungrammatical
- Error fixed in annotation
- No error



### Types of edits made in annotations

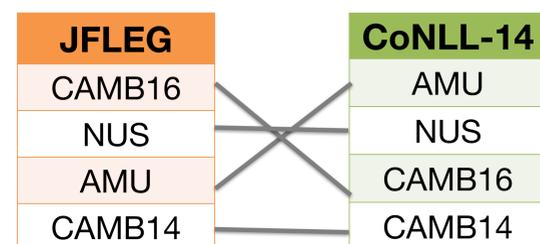
- Has fluency edit
- Only minimal edits
- No changes



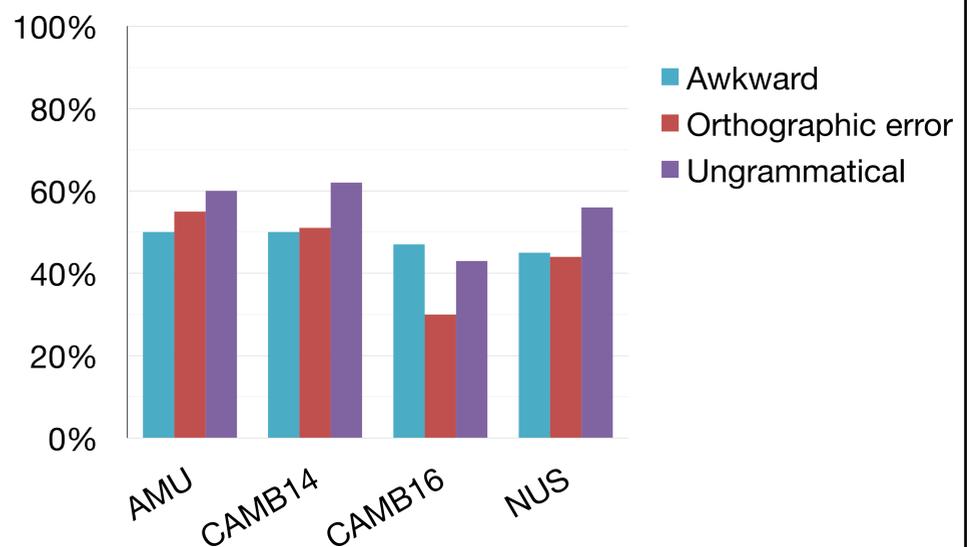
\* From manual analysis of 100 sentences

## JFLEG as a Benchmark Corpus

- 750 JFLEG sentences corrected by 4 leading GEC systems
  - AMU: Phrase-based MT [Junczys-Dowmnut and Grundkiewicz, 2016]
  - CAMB14: Hybrid rule-based MT with LM reranking [Felice et al., 2014]
  - CAMB16: Neural MT [Yuan and Briscoe, 2016]
  - NUS: Hybrid PBMT with NN [Chollampatt et al., 2016]
- Neural approaches do better on fluency corpus (JFLEG) while PBMT alone is best on minimal edit corpus (CoNLL-14)



### Error types in system output (JFLEG sentences)



\* From manual analysis of 100 sentences

## Conclusion

- JFLEG is the first and only fluency test set for GEC
- It enables GEC advancement beyond minimal edits
- Systems that do well on minimal edits are not necessarily the best on fluency edits

Download from <https://github.com/keisks/jfleg>

## JFLEG Example: Fluency Edits Are Better

only minimal edits:

One person, if ~~they~~ don't have good health, ~~that~~ means so many things ~~they~~ could ~~be~~ lost.

with fluency edits:

if <sup>a</sup> One person (if) ~~don't~~ <sup>doesn't</sup> have good health, ~~that means~~ so many things ~~they~~ could ~~be~~ lost.

JFLEG