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

JFLEG as a Benchmark Corpus
- 750 JFLEG sentences corrected by 4 leading GEC systems
  - AMU: Phrase-based MT [Junczys-Dowmunt 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)

Johns Hopkins Fluency Extended GUG Corpus (JFLEG)
- A new fluency data set for tuning and testing GEC
- JFLEG addresses limitations of CoNLL-14:
  - Changes: Minimal edits
  - Source: NUCLE [Dahlmeier et al., 2013]
  - Writers: University students in Singapore
  - Topics: 2
  - Proficiency: High
  - Size: 1300 sentences
  - Annotators per sentence: 2 English instructors

Tools for JFLEG
- CAMB16
- AMU
- NUS
- CAMB14

Error types in system output (JFLEG sentences)

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

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

* From manual analysis of 100 sentences

Types of edits made in annotations
- Has fluency edit: 12%
- Only minimal edits: 28%
- No changes: 60%

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

JFLEG as a Benchmark Corpus

<table>
<thead>
<tr>
<th>JFLEG</th>
<th>CoNLL-14</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAMB16</td>
<td>AMU</td>
</tr>
<tr>
<td>NUS</td>
<td>NUS</td>
</tr>
<tr>
<td>AMU</td>
<td>CAMB16</td>
</tr>
<tr>
<td>CAMB14</td>
<td>CAMB14</td>
</tr>
</tbody>
</table>

Error types in system output (JFLEG sentences)

- Awkward
- Orthographic error
- Ungrammatical

* From manual analysis of 100 sentences

JFLEG Example: Fluency Edits Are Better

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

with fluency edits:

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