Acknowledgements This work was supported by the framework of the PON programme FSE REACT-EU, Ref. DOT1303118. We also acknowledge the work of Maria Fasoi from the Athens University of Economics and Business, Greece, who provided us with the GLC corrections.

Introduction

- Issue: scarcity of resources for GEC purposes in Greek.
- Approach: MT5 multilingual text-to-text transformer (Xue et al., 2020).
- Evaluation with the Greek version of the Error Annotation Toolkit scorer (ERRANT) (Bryant et al., 2017).

Data

- Greek Native Corpus (Korre et al., 2021)
- Greek Learner Corpus (Tantos and Papadopoulou, 2018) \(\rightarrow\) GLC2

Method

MT5 employs an Encoder Decoder Transformer (Vaswani et al., 2017) that is pre-trained with masked language modeling by masking consecutive spans of input tokens and then trying to reconstruct them.

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Dev</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences</td>
<td>GNC</td>
<td>322</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>GLC</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tokens</td>
<td>GNC</td>
<td>8440</td>
<td>451</td>
</tr>
<tr>
<td></td>
<td>GLC</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Error Analysis

- MT5 performs well: accent and spelling errors, under-performs when the sentences are more complex
- Reduces the length or modifies other parts of the sentence. E.g., Το φαινόμενο αυτό αποτελεί θέμα μεγάλης ανησυχίας στην εποχή μας [This phenomenon is a matter of great concern nowadays]. \(\rightarrow\) Τέλος αυτό αποτελεί θέμα μεγάλης ανησυχίας στην εποχή μας [Finally, this is a matter of great concern nowadays].

Conclusions

- Second highest compared to published results in F0.5 and only 16 percent units below the state of the art in English GEC.
- The performance drops when it is evaluated on a learners’ dataset, possibly due to high error frequency.

Limitations and future work

- Inter-annotator agreement was very low for the GLC2 annotation.
- Small dataset sizes \(\rightarrow\) Synthetic data.
- BASE version of MT5, due to constrained resources \(\rightarrow\) better results with larger available models.