Frustratingly Easy Multilingual Grapheme-to-Phoneme Conversion

Nikhil Prabhu and Katharina Kann
University of Colorado Boulder

\{nikhil.prabhu, katharina.kann\}@colorado.edu

**Overview:**
- We present two CU Boulder submissions for SIGMORPHON 2020 Task 1 of Multilingual Grapheme-to-Phoneme Conversion (G2P)
- We improve over the high performance of a standard transformer by adding two modifications: i) multi-task training ii) ensembling

**Multi-task Training:**
- We train our model jointly on two tasks:
  i. G2P: \( g(w) \rightarrow p(w) \), with dataset \( D_{g2p} \)
  ii. P2G: \( p(w) \rightarrow g(w) \), with dataset \( D_{p2g} \)
- \( \lambda_g \) and \( \lambda_p \) are task embeddings which indicate which task the input belongs to
- We then obtain model parameters \( \theta \) that maximize the joint log-likelihood of both datasets:
  \[
  \mathcal{L}(\theta) = \sum_{w \in D_{g2p}} \log p_{\theta}(g[w]|g[w], \lambda_g) + \sum_{w \in D_{p2g}} \log p_{\theta}(g[w]|p[w], \lambda_p)
  \]

**Ensembling:**
Our models are ensembles created via majority voting
- i) CU-1, is an ensemble of 5 standard G2P transformers and 5 multi-task transformers
- ii) CU-2, is an ensemble of 5 multi-task transformers

**Grapheme-to-Phoneme Conversion:**
- Let \( \Sigma_g \) and \( \Sigma_p \) be an alphabet of graphemes and phonemes, respectively
- Consider \( g(w) \in \Sigma_g^* \) and \( p(w) \in \Sigma_p^* \) as the grapheme and phoneme representations of \( w \), respectively
- G2P then refers to the mapping \( g(w) \rightarrow p(w) \)

**Results:**
![Language vs WER](chart.png)

<table>
<thead>
<tr>
<th>Grapheme</th>
<th>Phoneme</th>
<th>Phoneme</th>
<th>Grapheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>!andacht</td>
<td>anduxt</td>
<td>?anduxt</td>
<td>aandacht</td>
</tr>
<tr>
<td>!basson</td>
<td>basson</td>
<td>?basson</td>
<td>basson</td>
</tr>
<tr>
<td>!begin</td>
<td>begin</td>
<td>?begin</td>
<td>begin</td>
</tr>
<tr>
<td>!gerst</td>
<td>xirst</td>
<td>?xirst</td>
<td>gierst</td>
</tr>
</tbody>
</table>