Background

**Modern QA approaches require context**

- A table show highlight or high language policy at story.

Why not translate?

- Machine and human translation can introduce artifacts
- Preserved word order when not required
- Constrained language
- Might not reflect all question types
- Might not reflect language specific topics
- Translatability

These observations are among the ones driving QA dataset development.

Datasets

- Active development of datasets for English
- Wikidata (Yang et al., 2015)
- Natural Questions (Kiatkawalwiri et al., 2016)
- Typologically diverse QA (TyDi QA, Clark et al., 2020)

Other languages and formats

- See QA dataset explosion (Bojanowski et al., 2021)
- Question Answer pair datasets for Icelandic (without context)

Typological diversity of Icelandic

<table>
<thead>
<tr>
<th>Language</th>
<th>Latin script</th>
<th>White space</th>
<th>Sentences</th>
<th>Word formation</th>
<th>Gender</th>
<th>Pro-drop</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>+</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Arabic</td>
<td>-</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Bengali</td>
<td>-</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Finnish</td>
<td>+</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Japanese</td>
<td>-</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Kannada</td>
<td>+</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Kasah</td>
<td>+</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Korean</td>
<td>+</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Russian</td>
<td>+</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Telugu</td>
<td>-</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Thai</td>
<td>+</td>
<td>e</td>
<td>e</td>
<td>e</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Icelandic</td>
<td>+</td>
<td>e</td>
<td>+</td>
<td>e</td>
<td>+</td>
<td>*</td>
</tr>
</tbody>
</table>

Reading Comprehension

We follow the methodology used to create TyDi QA (Clark et al., 2020).

**Dataset**

<table>
<thead>
<tr>
<th>Model</th>
<th>Dataset</th>
<th>F1</th>
<th>Exact match</th>
</tr>
</thead>
<tbody>
<tr>
<td>mBERT</td>
<td>NGI</td>
<td>78.0</td>
<td>58.4</td>
</tr>
<tr>
<td>XLM-R-base</td>
<td>NGI</td>
<td>72.1</td>
<td>96.1</td>
</tr>
<tr>
<td>XLM-R-base</td>
<td>TyDi English</td>
<td>67.7</td>
<td>96.6</td>
</tr>
<tr>
<td>XLM-R-base</td>
<td>TyDi Finnish</td>
<td>70.3</td>
<td>44.4</td>
</tr>
</tbody>
</table>

**Question type**

- What: 182: 65.4: 37.4
- How: 25: 78.8: 80.0
- When: 60: 64.0: 41.7
- Where: 40: 70.6: 50.0
- In: 18: 79.6: 72.2
- Who: 110: 86.3: 47.3
- Which: 43: 63.0: 46.5
- Why: 32: 56.8: 43.8
- Other: 16: 97.9: 87.5

**Discussion**

Wikipedia was almost too small for our setting:
- Use alternative sources
- Build a larger Wikipedia
- Crowdsource to improve diversity
- Clearer instructions to reduce annotator disagreement
- Quality checks

Make the answer labelling task easier and more effective:
- Rank passages by likelihood of containing answer
- Mark short and uncertain passages for labeling (active learning)

**References**

- Berga Tómass Tómass Einarsson, Héldur Víking Gunnarsson, Hildur Bjarnadóttir, Ingjörg Kára Andréassdóttir, and Urra Ingjó Saurausdóttir

**Bergur Tareq Tamimi Einarsson**

**Vésteinn Snæbjarnarson & Hafstein Einarsson**