Google’s Multilingual Neural Machine Translation System:
Enabling Zero-Shot Translation

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, Jeffrey Dean

Presented by: Emma Peng
State-of-the-art: Neural Machine Translation (NMT)
Multilingual NMT? Previously...

Multiple Encoders → Multiple Decoders [1]

- Er-Es Encoder
- Er-NL Encoder
- Er-Fr Encoder

→

- Er-Es Decoder
- Er-NL Decoder
- Er-Fr Decoder

Shared Encoder → Multiple Decoders [2]

- Shared Encoder

→

- En-Es Decoder
- En-NL Decoder
- En-Fr Decoder

Multiple Encoders → Shared Decoder [3]

- Es-Er Encoder
- NL-Er Encoder
- Fr-Er Encoder

→

- Shared Decoder
Google’s Multilingual NMT System Benefits

- **Simplicity**: single model
- **Low-resource language improvements**
- **Zero-shot translation**
Google’s Multilingual NMT System Architecture
Artificial token at the beginning of the input sentence to indicate the target language

Hello, how are you? -> ¿Hola como estás?

Add <2es> to indicate that Spanish is the target language

<2es> Hello, how are you? -> ¿Hola como estás?
Google’s Multilingual NMT System Experiments

- **WMT’14:**
  - Comparable performance: English $\rightarrow$ French
  - State-of-the-art: English $\rightarrow$ German, French $\rightarrow$ English

- **WMT’15:**
  - State-of-the-art: German $\rightarrow$ English
Google’s Multilingual NMT System Zero-Shot Translation

- **Train:**
  - Portuguese → English, English → Spanish (Model 1)
  - Or, English ↔ {Portuguese, Spanish} (Model 2)

- **Test:**
  - Portuguese → Spanish  **Zero-Shot!**

Table 5: Portuguese→Spanish BLEU scores using various models.

<table>
<thead>
<tr>
<th>Model</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) PBMT bridged</td>
<td>28.99</td>
</tr>
<tr>
<td>(b) NMT bridged</td>
<td>30.91</td>
</tr>
<tr>
<td>(c) NMT Pt→Es</td>
<td>31.50</td>
</tr>
<tr>
<td>(d) Model 1 (Pt→En, En→Es)</td>
<td>21.62</td>
</tr>
<tr>
<td>(e) Model 2 (En↔{Es, Pt})</td>
<td>24.75</td>
</tr>
<tr>
<td>(f) Model 2 + incremental training</td>
<td>31.77</td>
</tr>
</tbody>
</table>
Thank you!