Neural Machine Translation – Localization and beyond

Emanuele Dias
Principal Machine Learning Engineer
Autodesk – What do we do

Construction

Manufacturing

Media and Entertainment
Our Localization Challenges

Terminology
Consistency
Quality
Autodesk’s MT history

**Early 2000s**
Rule-based MT engines

**~2010**
First in-house SMT engines

**2017**
Initial NMT explorations

**Late 2017**
Google’s “Attention is all you need”

**2019**
SMT engines discontinued

**Today**
16 in-house NMT engines
Why in-house?

Confidentiality
Privacy

Quality

Know-how
Quality

- Better overall quality and a more consistent handling of “external entities”
  - Less post-editing (PE) required
  - Better PE rates
  - More raw MT content

- Ability to fix problems quickly and with increased precision

<table>
<thead>
<tr>
<th>Language</th>
<th>Difference (average rating points)</th>
<th>Autodesk</th>
<th>Best-Performing Competitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czech</td>
<td>2.46</td>
<td>78.56</td>
<td>76.10 (Google)</td>
</tr>
<tr>
<td>German</td>
<td>5.49</td>
<td>91.64</td>
<td>86.15 (Google)</td>
</tr>
<tr>
<td>Hungarian</td>
<td>9.17</td>
<td>73.80</td>
<td>64.63 (Google)</td>
</tr>
<tr>
<td>Chinese (Simplified)</td>
<td>5.71</td>
<td>87.68</td>
<td>81.97 (Google)</td>
</tr>
<tr>
<td>Japanese</td>
<td>1.72</td>
<td>89.21</td>
<td>87.60 (Google)</td>
</tr>
<tr>
<td>Portuguese (Brazilian)</td>
<td>2.32</td>
<td>89.30</td>
<td>86.96 (Google)</td>
</tr>
<tr>
<td>Spanish</td>
<td>5.16</td>
<td>90.12</td>
<td>84.96 (Google)</td>
</tr>
</tbody>
</table>
Know-how

- Work on NLP problems that apply horizontally
- Expand the scope to other NLP solutions that can benefit L10N
- Look into NMT to satisfy L10N’s needs
NMT Beyond Localization

Better quality increases adoption

NMT for L10N

NMT being used outside of L10N