Can out-of-the-box NMT Beat a Domain-trained Moses on Technical Data?

Anne Beyer
Vivien Macketanz
Aljoscha Burchard
Philip Williams

EAMT Prague, 29.05.2017
Background

- German LSP interested in MT technology
- Moses-based system in-house
- Gathered some experience over the last years working together with translators (some even positive...)

Is NMT something for us?

- NMT is gaining more and more attention
- Open-source toolkits are available
- Is it worth the effort yet?
How?

- No extensive user experience
- Setting up a system requires additional resources
- How to evaluate?
How?

- No extensive user experience
  → We have to start somewhere
- Setting up a system requires additional resources
- How to evaluate?
How?

- No extensive user experience
  → We have to start somewhere
- Setting up a system requires additional resources
  → Let’s start with an existing model (Uni Edinburgh)
- How to evaluate?
How?

• No extensive user experience
  → We have to start somewhere

• Setting up a system requires additional resources
  → Let’s start with an existing model (Uni Edinburgh)

• How to evaluate?
  → Phenomenon-driven test-suite (DFKI)
What do we want to achieve?

Preliminary study on how out-of-the-box NMT can perform compared to a customised SMT system
The Method

- Phenomenon-driven approach
  - Re-introducing Linguistics into MT evaluation
- Analyse customer data with respect to linguistic phenomena
- Select the most frequent ones
- Compare NMT/SMT system’s performance on those categories
The Data

- Selection of “real world” customer data collected over a three month period
- Catalogue of technical tools
- German → English
- ~ 5,000 Segments
The SMT System

- Based on Moses toolkit
- Trained on customer data (TM + terminology)
- 337,600 segments
- Specific tag-handling using m4loc
The NMT System

- Provided by the University of Edinburg
- Based on Nematus toolkit
- Trained on WMT data
- Best performing system on WMT‘16 news translation task
## Manual Evaluation Results

<table>
<thead>
<tr>
<th>Feature</th>
<th>#</th>
<th>NMT</th>
<th>Moses</th>
</tr>
</thead>
<tbody>
<tr>
<td>formal address</td>
<td>138</td>
<td>90%</td>
<td>86%</td>
</tr>
<tr>
<td>genitive</td>
<td>114</td>
<td>92%</td>
<td>68%</td>
</tr>
<tr>
<td>modal construction</td>
<td>290</td>
<td>94%</td>
<td>75%</td>
</tr>
<tr>
<td>negation</td>
<td>101</td>
<td>93%</td>
<td>86%</td>
</tr>
<tr>
<td>passive voice</td>
<td>109</td>
<td>83%</td>
<td>40%</td>
</tr>
<tr>
<td>predicate adjective</td>
<td>122</td>
<td>81%</td>
<td>75%</td>
</tr>
<tr>
<td>prepositional phrase</td>
<td>104</td>
<td>81%</td>
<td>75%</td>
</tr>
<tr>
<td>terminology</td>
<td>330</td>
<td>35%</td>
<td>68%</td>
</tr>
<tr>
<td>tagging</td>
<td>145</td>
<td>83%</td>
<td>100%</td>
</tr>
<tr>
<td><strong>sum</strong></td>
<td>1453</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>average</strong></td>
<td></td>
<td><strong>89%</strong></td>
<td><strong>73%</strong></td>
</tr>
</tbody>
</table>
Example 1

Source: Schweißbänder erhöhen wesentlich den Tragekomfort eines Helmes.
Ref.: Sweatbands significantly increase the wearing comfort of a helmet.
NMT: Welding tapes significantly increase the comfort of a helmet.
Moses: Welding belts significantly increase the wearing comfort of a Helmes.
Example 2

Source: Dazu kann das Board werkzeuglos gedreht und wieder eingehängt werden.

Ref.: The board can be turned and re-attached without using tools.

NMT: The board can be rotated and re-mounted.

Moses: To do this, the board can be rotated and back.
Example 3

Source: Neben den Bedingungen zur Aufstellung und Inbetriebnahme wird eine Vielzahl von technischen und gesetzlichen Anforderungen an das Lager selbst gestellt, um z. B. wassergefährdende Flüssigkeiten, Säuren und Laugen oder auch entzündbare Flüssigkeiten gesetzeskonform aufzubewahren und zu lagern.

NMT: In addition to the conditions for installation and commissioning, a wide range of technical and legal requirements will be placed on the warehouse itself in order to maintain and store, for example, water-hazardous liquids, acids and foliage, or even flammable liquids.
Conclusion

- In this analysis NMT outperformed SMT, even though SMT was at an advantage
  - BUT: Tags and terminology are among the most important categories in commercial translation
- NMT development is only getting started

→ We should start looking into this, especially together with translators!
What’s next?

- Confirm linguistic findings with translators
  - Study usefulness of NMT pre-translation in their working environment
    - MT selection task with post-editing
    - Productivity tests

- Look into OpenNMT
  - Find means for tag-handling with NMT
  - Use/Add customer data to training corpus
Thank you!

anne.beyer@beo-doc.de
**Automatic Evaluation Results**

<table>
<thead>
<tr>
<th></th>
<th>NMT</th>
<th>Moses</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLEU</td>
<td>23.68</td>
<td>47.98</td>
</tr>
<tr>
<td>METEOR</td>
<td>28.46</td>
<td>38.26</td>
</tr>
</tbody>
</table>

Table 1: BLEU and METEOR scores.
Example 4

Source: Die Panoramaspiegelfläche mit integriertem Seitenschutz sorgt für eine <g id="1004">optimale Augenraumabdeckung</g>.

Ref.: The panoramic lens with integral side protection ensures <g id="1004">optimum coverage of the eye area</g>.

NMT: The panorama disc with integrated side protection ensures a <g id="1004">optimal eye room cover</g>.

Moses: The panoramic lens with integral side protection ensures <g id="1004">optimum Augenraumabdeckung</g>.