

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



Anne Beyer



Vivien Macketanz
Aljoscha Burchard



Philip Williams

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 Edinburgh
- Based on Nematus toolkit
- Trained on WMT data
- Best performing system on WMT'16 news translation task

Manual Evaluation Results

	#	NMT	Moses
formal address	138	90%	86%
genitive	114	92%	68%
modal construction	290	94%	75%
negation	101	93%	86%
passive voice	109	83%	40%
predicate adjective	122	81%	75%
prepositional phrase	104	81%	75%
terminology	330	35%	68%
tagging	145	83%	100%
sum	1453		
average		89%	73%

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

	NMT	Moses
BLEU	23.68	47.98
METEOR	28.46	38.26

Table 1: BLEU and METEOR scores.

Example 4

Source: Die Panoramascheibe 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>.