

A User Study on Machine Translation Fine-tuning with Custom Translation Memories in 3 Language Pairs

Gökhan Doğru

Postdoctoral Researcher @ Universitat Autònoma de Barcelona
Visiting Researcher @ Dublin City University
24.11.2022



Outline

- Introduction
- Desktop MT Fine-tuning
 - ◆ OpusCAT & OpusMT
- Methodology
- Corpus Statistics
- Results
 - ◆ Automatic evaluation
 - ◆ Human evaluation
- Conclusion & Future Work

PART I

Introduction

- Research on translator interaction with NMT has tended to focus on the productivity or quality rather than its “usefulness.. as a tool for professionals” (Ragni & Nunes Vieira 2022, 153).
- This study aims to address the usefulness of free, customizable, desktop NMT: OpusCAT MT (Nieminen, 2021).

OPUSCAT and OPUSMT

OPUSCAT (Nieminen 2021)

- Free
- Desktop (Secure and Private)
- Pre-trained NMT engines
- Fine-tuning with private translation memories
- Integration to CAT tools

OPUSMT (Tiedemann & Thottingal 2020)

- A repository of over 1,000 pre-trained NMT models
- Based on Marian-NMT
- Trained on OPUS data
- Bilingual & multilingual engines
- Free MT resources & tools

The Study

- We assess the effect of data augmentation in OPUSCAT fine-tuning in 3 language pairs:
 - ◆ English → Turkish
 - ◆ English → Spanish
 - ◆ English → Catalan
- Domain: Localization
- Objective: To evaluate the usefulness of OPUSCAT for professional translators

PART II
Methodology

Methodology

- Pre-trained NMT engines are fine-tuned with small and big localization corpora in 3 language pairs in OPUSCAT
- Automatic evaluation
 - ◆ BLUE, chrF3 and TER
- Human evaluation
 - ◆ Ranking, adequacy and fluency

MT Engines and Corpus Statistics

<i>Lang</i>	<i>Baseline</i>	<i>Small Corpus</i>	<i>Large Corpus</i>	<i>Source</i>
<i>EN-TR</i>	No fine-tuning (en-tr-1)	501,371 (en-tr-2)	2,253,304 (en-tr-3)	Microsoft
<i>EN-ES</i>	No fine-tuning (en-es-1)	501,979 (en-es-2)	2,745,645 (en-es-3)	Microsoft
<i>EN-CA</i>	No fine-tuning (en-ca-1)	503,188 (en-ca-2)	2,269,533 (en-ca-3)	Softcatala

Test Corpora

- 210 strings from Microsoft Corpus for 3 language pairs
 - ◆ With human reference translation
 - ◆ Not used for fine-tuning
 - ◆ From the same file
 - ◆ Repetitions removed
 - ◆ Used for automatic evaluation and human evaluation

Human Evaluators

- 3 evaluators per language pairs
 - ◆ Native professional translators and reviewers
- KantanLQR
- Scale of 3 for ranking
- Scale of 5 for adequacy and fluency

PART III

Results

Automatic Evaluation Results: English - Turkish

<i>MT Name</i>	<i>BLEU</i> ↑	<i>chrF3</i> ↑	<i>TER</i> ↓
<i>en-tr-1</i>	23	53.5	67.52
<i>en-tr-2</i>	49.6	67.86	49.13
<i>en-tr-3</i>	51.7	67.51	47.88

Automatic Evaluation Results: English - Spanish

<i>MT Name</i>	<i>BLEU</i> ↑	<i>chrF3</i> ↑	<i>TER</i> ↓
<i>en-es-1</i>	37.3	66.36	46.64
<i>en-es-2</i>	38.5	69.99	43.1
<i>en-es-3</i>	48.1	74.41	38.12

Automatic Evaluation Results: English - Catalan

<i>MT Name</i>	<i>BLEU</i> ↑	<i>chrF3</i> ↑	<i>TER</i> ↓
<i>en-ca-1</i>	38	63.11	53.33
<i>en-ca-2</i>	42.5	63.11	49.56
<i>en-ca-3</i>	47.2	67.84	42.4

Human Evaluation Results: Ranking

Ranking (%)	Baseline	Small Corpus	Big Corpus
EN → TR	67.57	77.94	80.63
EN → ES	75.03	77.25	80.21
EN → CA	67.72	72.38	79.05

Human Evaluation Results: Adequacy

Adequacy (%)	Baseline	Small Corpus	Big Corpus
EN - TR	72.35	81.17	82.73
EN - ES	80.92	83.49	84.22
EN - CA	76.35	76.38	80.51

Human Evaluation Results: Fluency

Fluency (%)	Baseline	Small Corpus	Big Corpus
EN - TR	75.94	83.37	84.22
EN - ES	81.97	83.02	84.38
EN - CA	73.90	76.79	80.25

Conclusions and Future Work

- Automatic and human evaluation suggest that data augmentation improves overall engine quality
- Automatic evaluation scores imply a significant quality increase after fine-tuning with a small corpus but when 2M+ big corpus is added, quality only improves slightly
- Human adequacy and fluency results suggest a gradual quality increase with data augmentation

Conclusions and Future Work

- Overall high scores in human evaluation suggest that professional translators can use OPUSCAT for their translation project
 - ◆ Secure and private, customizable MT
- Inter-annotator agreement measurement is needed
- Best engines will be compared to commercial systems to compare usefulness

References

Nieminen, T. (2021). OPUS-CAT: Desktop NMT with CAT integration and local fine-tuning. *EACL16: System Demonstrations*, 288–294.

OPUS Website: <https://helsinki-nlp.github.io/OPUS-CAT/>

Ragni, V., & Nunes Vieira, L. (2022). What has changed with neural machine translation? A critical review of human factors. *Perspectives*, 30(1), 137–158.

Tiedemann, J., & Thottingal, S. (2020). OPUS-MT – Building open translation services for the World. *EAMT22*, 479–480.