

Unsupervised German to Czech: 15.3 BLEU

Shuffling synthetic data to avoid German word order in Czech

Initial translations of named entities were a mess
→ special treatment during fine-tuning

CUNI Systems for the Unsupervised News Translation Task in WMT 2019



Ivana Kvapilíková, Dominik Macháček and Ondřej Bojar

Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Czech Republic

Introduction

Our system competes in the news shared task of unsupervised machine translation from German to Czech using monolingual training data only.

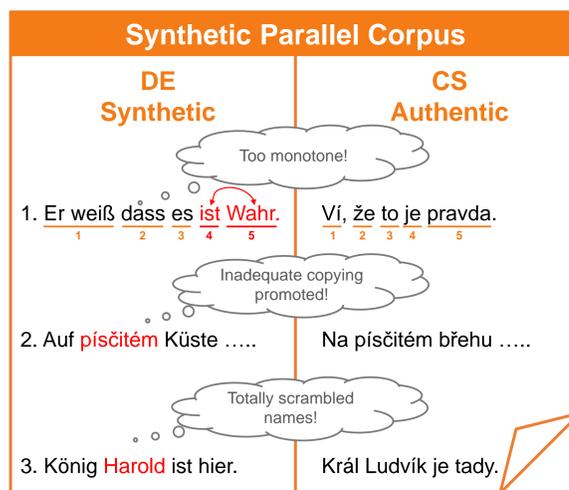
System Description

Our pipeline combines a PBMT and a NMT system in the following steps:

- Create an **unsupervised PBMT system** (Monoses by Artetxe et al., 2018);
- Use the PBMT system to produce a synthetic parallel corpus;
- Train a **Transformer NMT system** on the synthetic data (seg-level Marian as in the poster “English-Czech Systems in WMT19: Document-Level Transformer”);
- **Enhance the synthetic corpus** and fine-tune the system.

Challenges

1. Wrong word order;
2. Non-translated words;
3. Randomly mistranslated **named entities**.



Solutions

Fine-tuning of the NMT model on new synthetic corpora enhanced by:

1. Random reordering of words within a 5-word window;
2. Replacing non-translated Czech words on the German side by <unk>;
3. Removing mistranslated named entities or replacing them by the original names.

Post-processing to correct final translations of named entities.

Final Automatic Evaluation

System Name	BLEU uncased
CUNI-Unsup-base	13.6
CUNI-Unsup	15.3
CUNI-Unsup-NER	14.6
CUNI-Unsup-NER-post	14.4
Benchmark-Supervised	19.3
Benchmark-TransferEN	13.6

System performance on newstest2019.

Manual Comparison

Better Output By	Sentences	
	with NEs	no NEs
CUNI-Unsup	28%	26%
CUNI-Unsup-NER	52%	28%
Both equal	20%	46%
Total	100%	100%

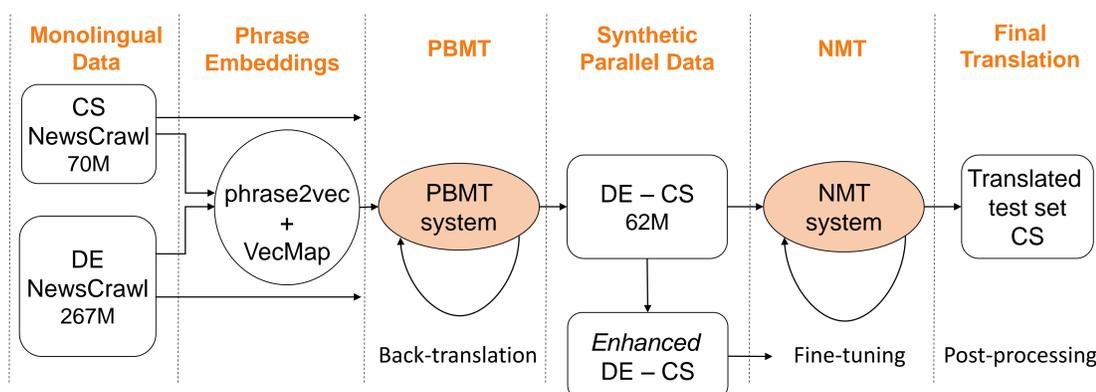
Better Output By	Sentences	
	with NEs	no NEs
CUNI-Unsup-NER	14%	0%
CUNI-Unsup-NER-post	18%	0%
Both equal	68%	100%
Total	100%	100%

Results of manual evaluation on a stratified subset of the validation data set created by randomly selecting 100 sentences with NEs and 100 sentences without NEs.

Legend

CUNI-Unsup-base	no fine-tuning
CUNI-Unsup*	fine-tuned on 1 & 2
CUNI-Unsup-NER*	fine-tuned on 1, 2 & 3
CUNI-Unsup-NER-post**	post-processed
Benchmark-Supervised	trained on 8.8M parallel sentences
Benchmark-TransferEN	trained for the en→cs task and fine-tuned on synthetic de→cs data

* submitted; ** submitted as primary



The full paper



This study was supported in parts by the grants SVV 260 453, 1050119 of the Charles University Grant Agency, 18-24210S of the Czech Science Foundation and the EU grant H2020-ICT-2018-2-825460 (ELITR).

