

The University of Edinburgh's English-German and English-Hausa Submissions to the WMT21 News Translation Task

Pinzhen Chen Jindřich Helcl Ulrich Germann Laurie Burchell Nikolay Bogoychev
 Antonio Valerio Miceli Barone Jonas Waldendorf Alexandra Birch Kenneth Heafield
 School of Informatics, University of Edinburgh

Overview

- **UEdin@WMT21**
 - **English↔German**: high-resource techniques focusing on filtering, back-translation, and fine-tuning.
 - **English↔Hausa**: low-resource transfer learning, iterative back-translation, fine-tuning, and vocabulary mapping.
- Only used provided resources, and created systems that are competitive with human and other unconstrained submissions.

English-German

- **Models**
 - Transformer-big, 32k SentencePiece, ensembling
 - different encoder-decoder layers: 6-6 and 8-4.
- **Parallel sentence filtering**
 - handcrafted rules: length ratio, language id, etc
 - dual conditional cross-entropy filtering

Data kept	De→En	En→De
top 50%	41.64	43.68
top 75%	42.15	43.40
all	42.02	42.70

Table: BLEU of filtering experiments on WMT19 test.

- **Back-translation**
 - for De→En: tagged BT
 - for En→De: a mix of sampling, greedy and beam search
 - BT slightly harms BLEU, so continue training on parallel

Configuration	De→En	En→De
Baseline	42.2	43.4
+ BT	41.8	43.0
+ continued training	42.5	43.6

Table: Average BLEU of BT experiments on WMT19 test.

- **Fine-tuning**
 - fine-tuning on past year test sets
 - test sentence splitting
 - tuning length normalization

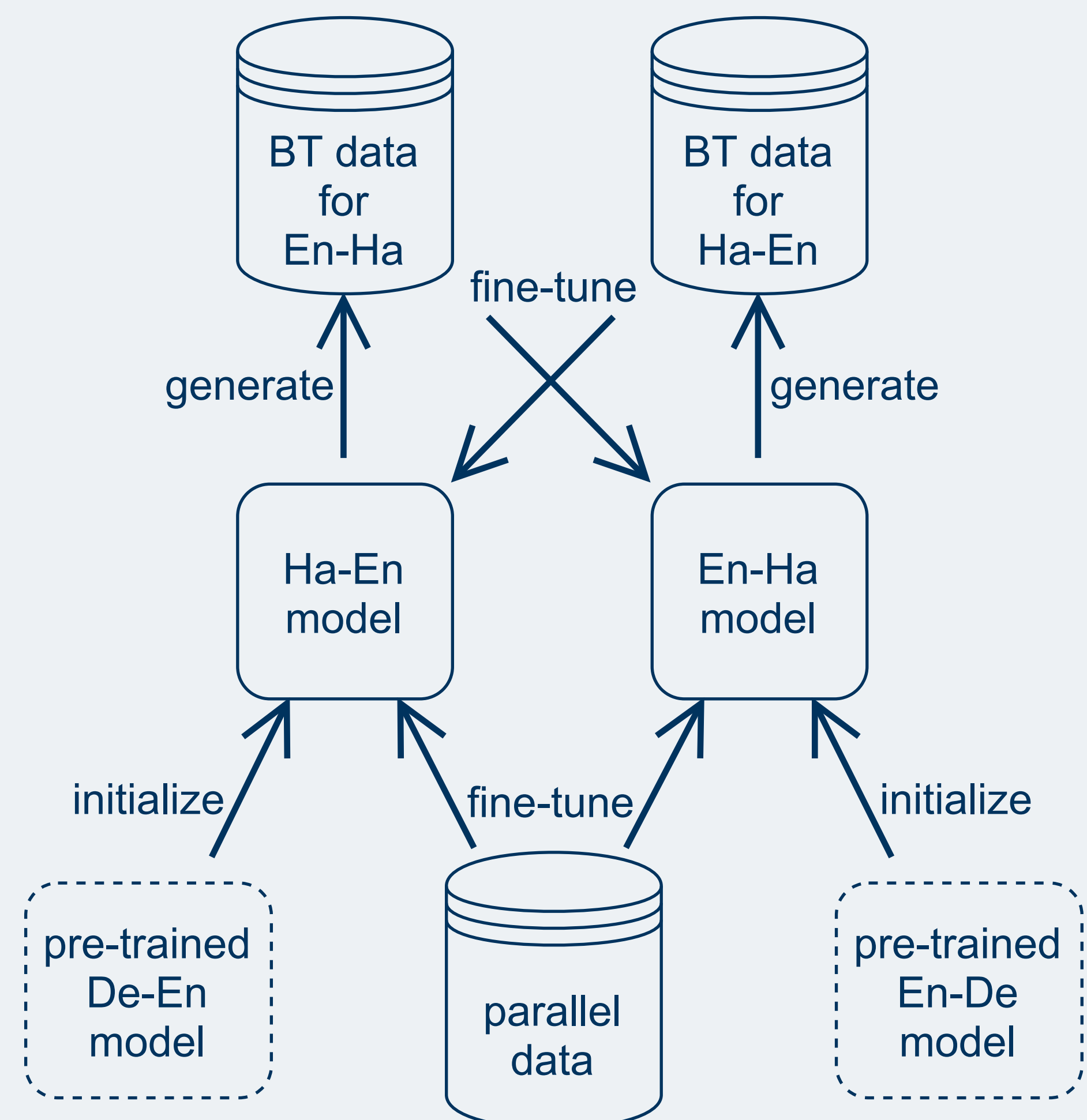
Fine-tuning	Test split	De→En	En→De
×	×	30.8	31.9
×	✓	41.7	35.2
✓	✓	41.9	36.7
✓, src only ¹	✓	42.5	36.9

¹ on data that originate from the source language only

Table: BLEU of fine-tuning and test-splitting experiments on a baseline on WMT20 test

English-Hausa

- **Iterative back-translation and fine-tuning**
 - initialize En↔Ha models with En↔De
 - iteratively back-translate and fine-tune the other direction



- **Vocabulary embedding mapping**
 - train a model with En, Ha, De joint vocabulary
 - map Ha to De embeddings during En↔Ha initialization
 - faster convergence but not better quality

Human Evaluation Results

- The organizers provided **human evaluation** of to-English systems
 - All German→English submissions, together with online engines and human translators, are **tied**.
 - Our Hausa→English system ties with many systems at the **2nd place**, despite that our model was accidentally overfit.

Ave.	Ave. z	System	Ave.	Ave. z	System
71.9	0.126	Participant 1	74.4	0.248	Participant 1
73.5	0.124	Online-A	68.8	0.118	Online-B
78.6	0.122	Online-W	66.6	0.062	Participant 2
...	66.5	0.059	Participant 3
73.1	-0.008	UEdin
...	59.6	-0.112	Participant 10
71.8	-0.061	Participant 13	60.4	-0.173	Participant 11
66.8	-0.081	<i>Human</i>	58.2	-0.205	UEdin
66.0	-0.120	Participant 14	56.9	-0.267	Participant 13

(a) De→En

(b) Ha→En

□ constrained

■ unconstrained

Table: To-English human evaluation results. Systems within a cluster are considered tied.