12 - 14 January 2021
Virtual Meeting

CUNI Submission for the MT Task

Covid-19 MLIA @ Eval

Ivana Kvapilíková, Ondřej Bojar

Charles University
Institute of Formal and Applied Linguistics

Introduction

- We participated in all language pairs (EN → DE, EL, ES, IT, FR, SV)
- All our systems are constrained
- Different settings:
  1. Standard MT (Baseline)
  2. Online back-translation
  3. Multilingual
  4. Transfer learning

System Description

- Same architecture
  - 6-layer Transformer
  - 1024 hidden dimension
  - GELU activations
  - Language embeddings
- Training
  - Dropout 0.2
  - Vocabulary size 30k
- Inference
  - Beam search with beam size 4

Baseline

One training step (one batch)

Unidirectional MT system
EN → DE

Legend:

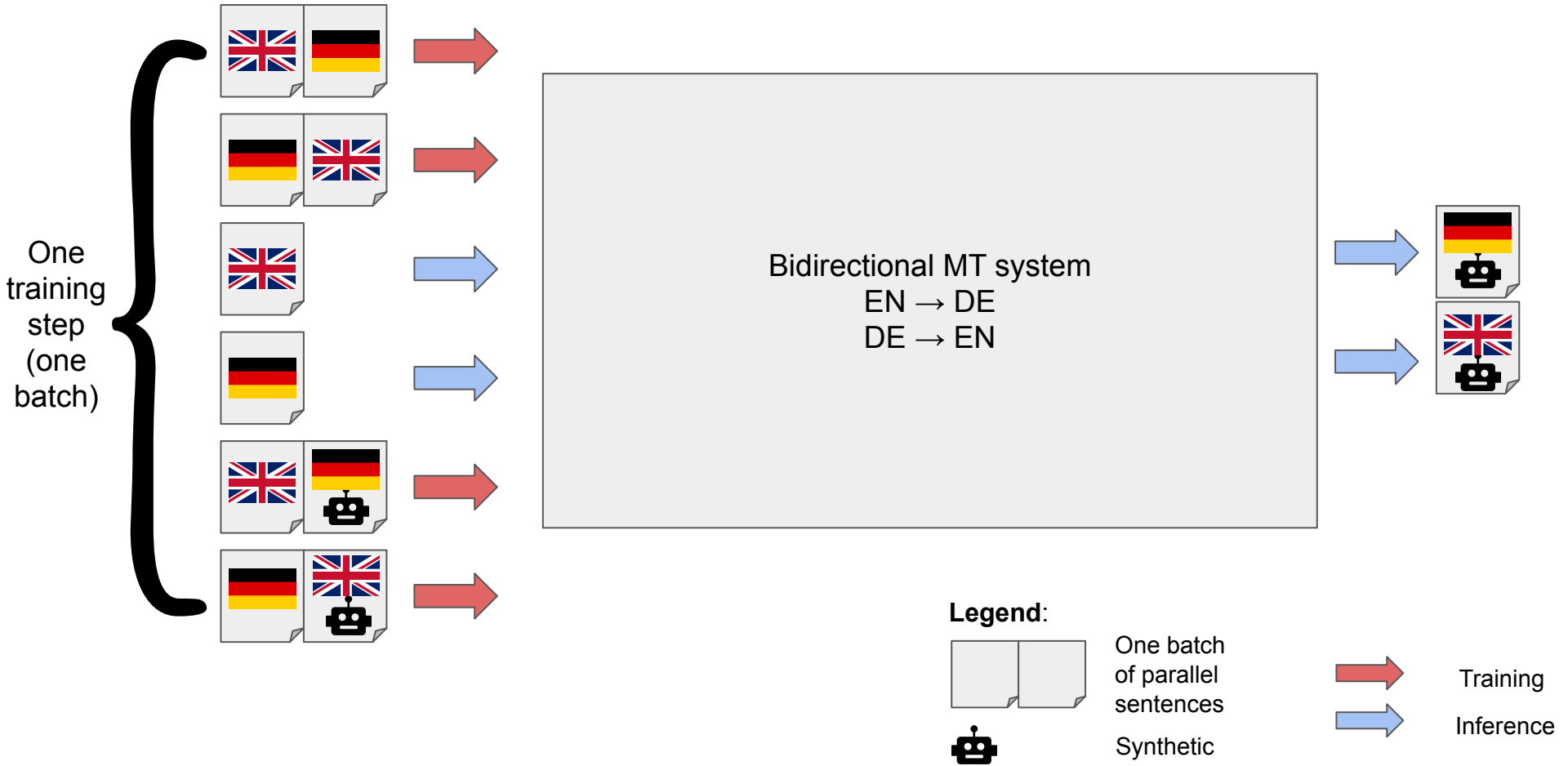
One batch of parallel sentences

Synthetic

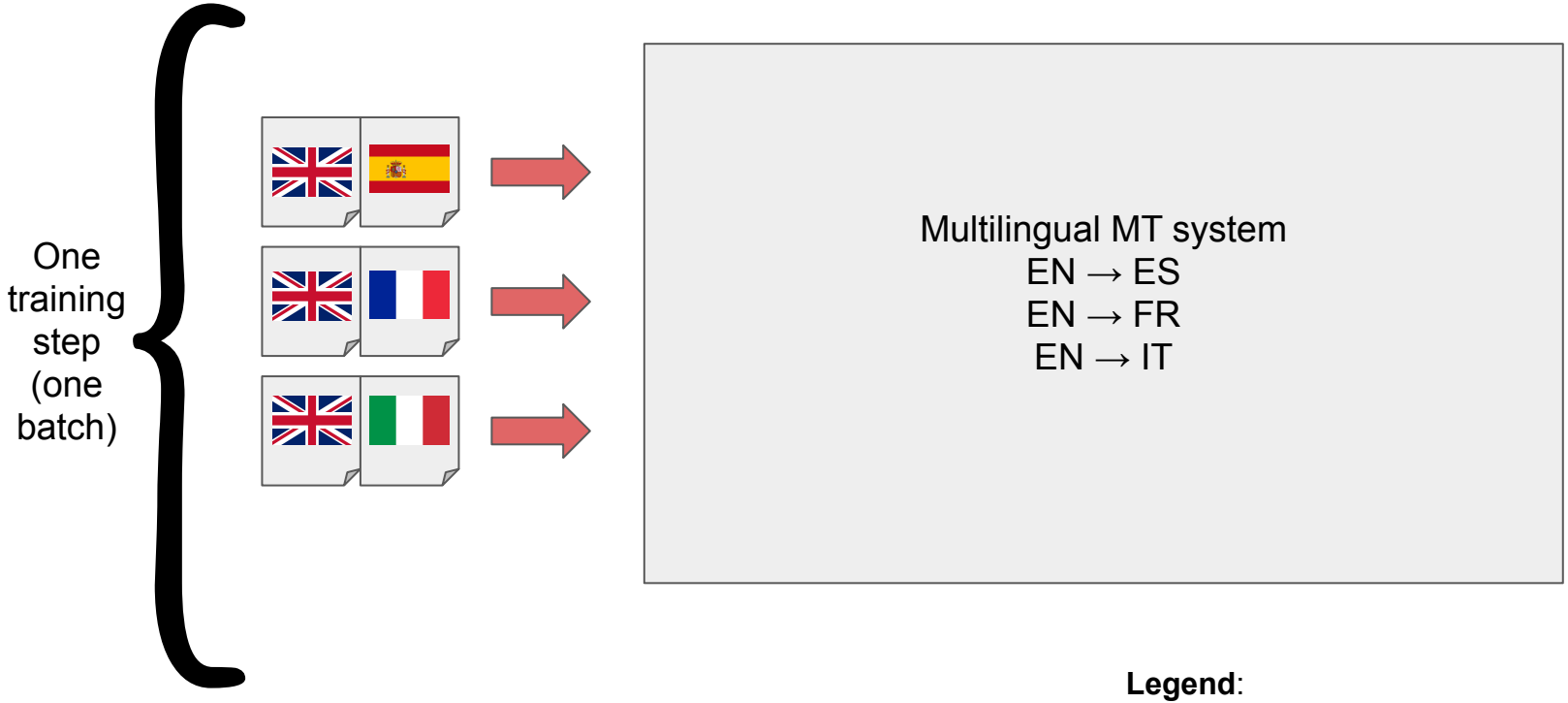
Training

Inference

Online back-translation



Multilingual training



Transfer learning

Train until convergence

1. 

Unidirectional MT system
EN → DE

Legend:

Corpus of parallel sentences

Synthetic

Training

Inference

Transfer learning

Swap training corpus and fine-tune

2. 

Unidirectional MT system
EN → DE

Legend:

Corpus of parallel sentences

Synthetic

Training

Inference

Transfer learning “with a detour”

Train until convergence

1. 

Unidirectional MT system
EN → DE

Legend:

Corpus of parallel sentences

Synthetic

Training

Inference

Transfer learning “with a detour”

Swap training corpus and fine-tune until converge nce

2. 

Unidirectional MT system
EN → DE

Legend:

Corpus of parallel sentences

Synthetic

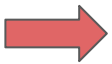
Training

Inference

Transfer learning “with a detour”

Swap training corpus and fine-tune

3.



Unidirectional MT system
EN → DE

Legend:

Corpus of parallel sentences

Synthetic

Training

Inference

Results (BLEU scores on dev set)

EN → DE	
Baseline	20.76
Online BT	21.52
Multilingual	na
Transfer	**22.60**

EN → EL	
Baseline	22.7
Online BT	22.3
Multilingual	na
Transfer	**23.29**

EN → ES	
Baseline	40.46
Online BT	40.94
Multilingual	40.15
Transfer	**41.34**

EN → FR	
Baseline	35.57
Online BT	**38.46**
Multilingual	36.07
Transfer	35.10

EN → IT	
Baseline	30.97
Online BT	**33.17**
Multilingual	32.76
Transfer	33.07

EN → SV	
Baseline	19.13
Online BT	20.61
Multilingual	na
Transfer	**22.60**

Results (BLEU scores on dev set)

EN → DE	
Baseline	20.76
Online BT	21.52
Multilingual	na
Transfer	**22.60**

Round 1 winner
31.6 BLEU
on test set

N → EL	
	22.7
Online BT	22.3
Multilingual	na
Transfer	**23.29**

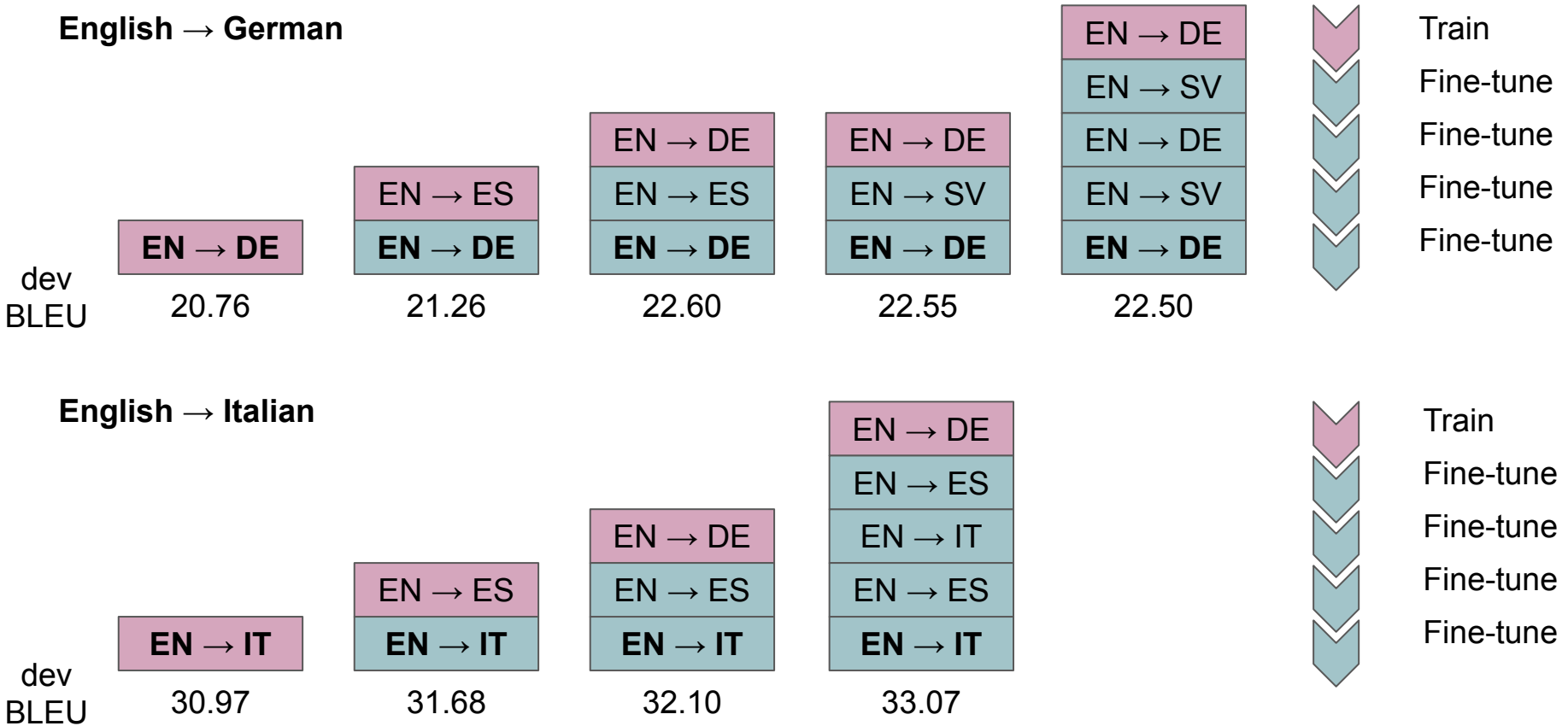
EN → ES	
Baseline	40.46
Online BT	40.94
Multilingual	40.15
Transfer	**41.34**

EN → FR	
Baseline	35.57
Online BT	**38.46**
Multilingual	36.07
Transfer	35.10

EN → IT	
Baseline	30.97
Online BT	**33.17**
Multilingual	32.76
Transfer	33.07

EN → SV	
Baseline	19.13
Online BT	20.61
Multilingual	na
Transfer	**22.60**

Analysis: Transfer learning



Conclusion

- No single winning approach
- All language pairs except for English-French significantly benefit from transfer learning
- Transfer learning more effective than multilingual training
- "Detours" during transfer learning help the final model (e.g. transferring via two languages)
- Next steps
  - Further investigate transfer learning combinations
  - Use a pretrained model from general domain (unconstrained)

Thank you for your attention!

Any questions?