17 February 2022
Virtual Meeting

CUNI Submission for the MT Task

Round 2

Covid-19 MLIA @ Eval

Ivana Kvapilíková, Ondřej Bojar

kvapilikova@ufal.mff,cuni.cz

Charles University
Institute of Formal and Applied Linguistics

Introduction

- We participated in all language pairs (EN → AR, DE, EL, ES, IT, FR, SV)
- All our systems are constrained
- Training data for Round 2, no filtering
- Different settings:
  1. Multilingual
  2. Transfer learning

System Description

- Same architecture
  - 6-layer Transformer with 8 heads
  - Vocabulary size 50k - shared among all languages

- Multilingual
  - One system, seven languages
  - Joint training
  - Target language embeddings

- Transfer learning
  - Pretraining on one parent language pair
  - Pretraining on a sequence of two language pairs
  - Pretraining on a mix of all language pairs

Transfer learning

Train until convergence

1. 

Unidirectional MT system
EN → DE

Legend:

One batch of parallel sentences

Training

Transfer learning

Swap training corpus and fine-tune

2. 

Unidirectional MT system
EN → DE

Legend:

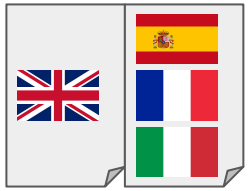
One batch of parallel sentences

Training

Transfer learning with multilingual pretraining

Train until convergence

1.



Unidirectional MT system
EN → DE

Legend:

One batch of parallel sentences

Training

Transfer learning with multilingual pretraining

Swap training corpus and fine-tune

2. 

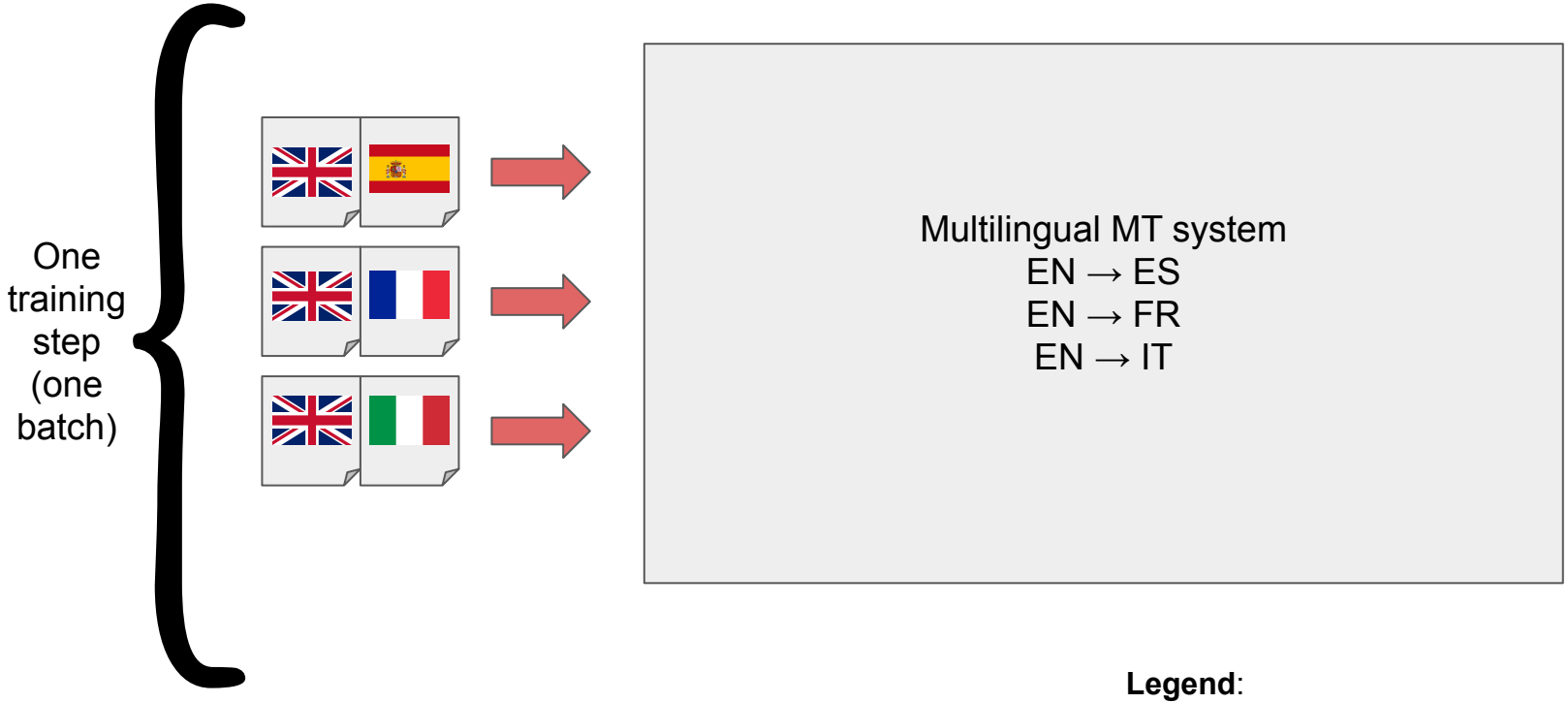
Unidirectional MT system
EN → DE

Legend:

One batch of parallel sentences

Training

Multilingual training



Results: Transfer learning

		ar	de	el	es	fr	it	sv
	ar		38.50	41.18	56.09	57.12	45.06	16.90
	de	23.67		43.09	56.31	57.56	46.15	18.74
	el	21.51	39.12		56.11	57.58	45.8	18.08
transfer from	**es**	23.27	39.89	43.58		**57.87**	**47.38**	**18.75**
	fr	**24.82**	**40.21**	**43.86**	**56.58**		47.35	18.71
	it	24.49	39.18	41.93	56.1	57.45		18.38
	sv	23.87	38.38	43.11	55.77	57.38	45.28	

Table 1: Transfer learning for various language pairs (child - horizontal axis, parent - vertical axis). BLEU scores on dev set.

Results: Transfer learning

		ar	de	el	es	fr	it	sv
	ar		38.50	41.18	56.09	57.12	45.06	16.90
	de	23.67		43.09	56.31	57.56	46.15	18.74
	el	21.51	39.12		56.11	57.58	45.8	18.08
transfer from	**es**	23.27	39.89	43.58		**57.87**	**47.38**	**18.75**
	fr	**24.82**	**40.21**	**43.86**	**56.58**		47.35	18.71
	it	24.49	39.18	41.93	56.1	57.45		18.38
	sv	23.87	38.38	43.11	55.77	57.38	45.28	
	all	**24.95**	39.83	**44.05**	54.76	55.43	47.39	**19.4**

Results (BLEU scores on test set)

EN → AR	
Multilingual	17
Transfer	**19.1**

EN → DE	
Multilingual	27
Transfer	**27.5**

EN → EL	
Multilingual	34.9
Transfer	**32.4**

EN → ES	
Multilingual	42.1
Transfer	**48.4**

EN → FR	
Multilingual	44.1
Transfer	na

EN → IT	
Multilingual	35.2
Transfer	**38.6**

EN → SV	
Multilingual	**14.7**
Transfer	13.9

Conclusion

- All language pairs significantly benefit from transfer learning
- In our setup, transfer learning is more effective than multilingual training
- The larger the parent dataset, the better
- Multilingual pretraining is beneficial for low-resource language pairs
- Possible improvements:
  - data preprocessing
  - individual vocabularies
  - back-translation

Thank you for your attention!

Any questions?