

THE MLLP-UPV SPANISH-PORTUGUESE AND PORTUGUESE-SPANISH MACHINE TRANSLATION SYSTEMS FOR WMT19 SIMILAR LANGUAGE TRANSLATION TASK

Pau Baquero-Arnal Javier Iranzo-Sánchez Jorge Civera Alfons Juan

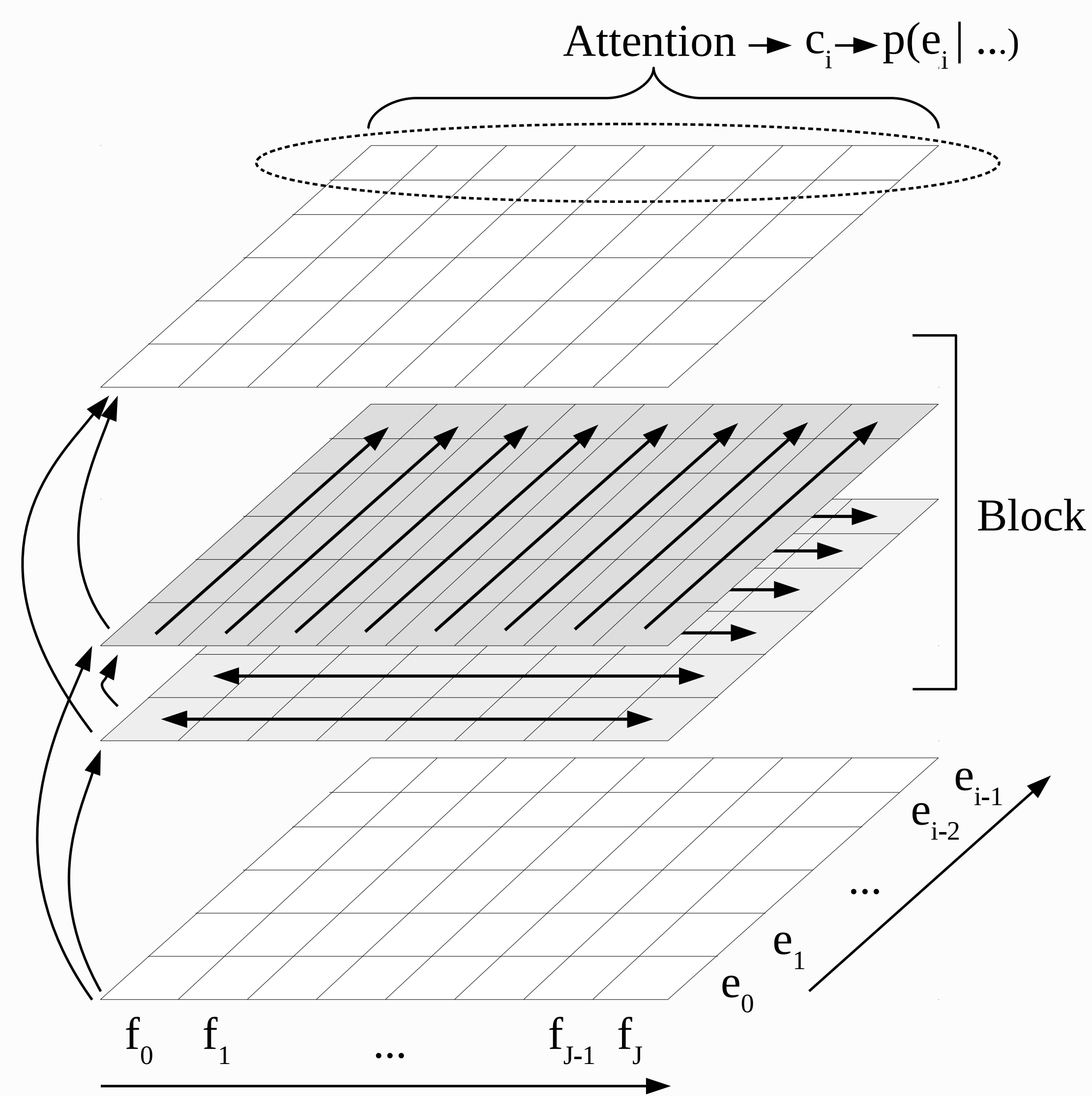
www.mllp.upv.es



INTRODUCTION

- Neural Machine Translation (NMT) system created for the WMT19 Similar Language Translation shared task (ES→PT)
- Standard NMT approach to similar language translation
- 2 NMT architectures explored:
 - Multi-headed self-attention (Transformer)
 - 2D Alternating RNN
- Domain adaptation carried out using fine-tuning

2D ALTERNATING RNN



- First grid: concatenate combinations of all source/target positions
- Block: two recurrent layers, one for each direction
- Output: concatenation of each layer
- Context vector from a row of vectors with attention mechanism

EXPERIMENTAL SETUP

Transformer

- “Base” config: model size: 512; feed-forward size: 2048; 6 layers
- Single-GPU with gradient accumulation each 4 batches
- Fairseq NMT toolkit

2D Alternating RNN

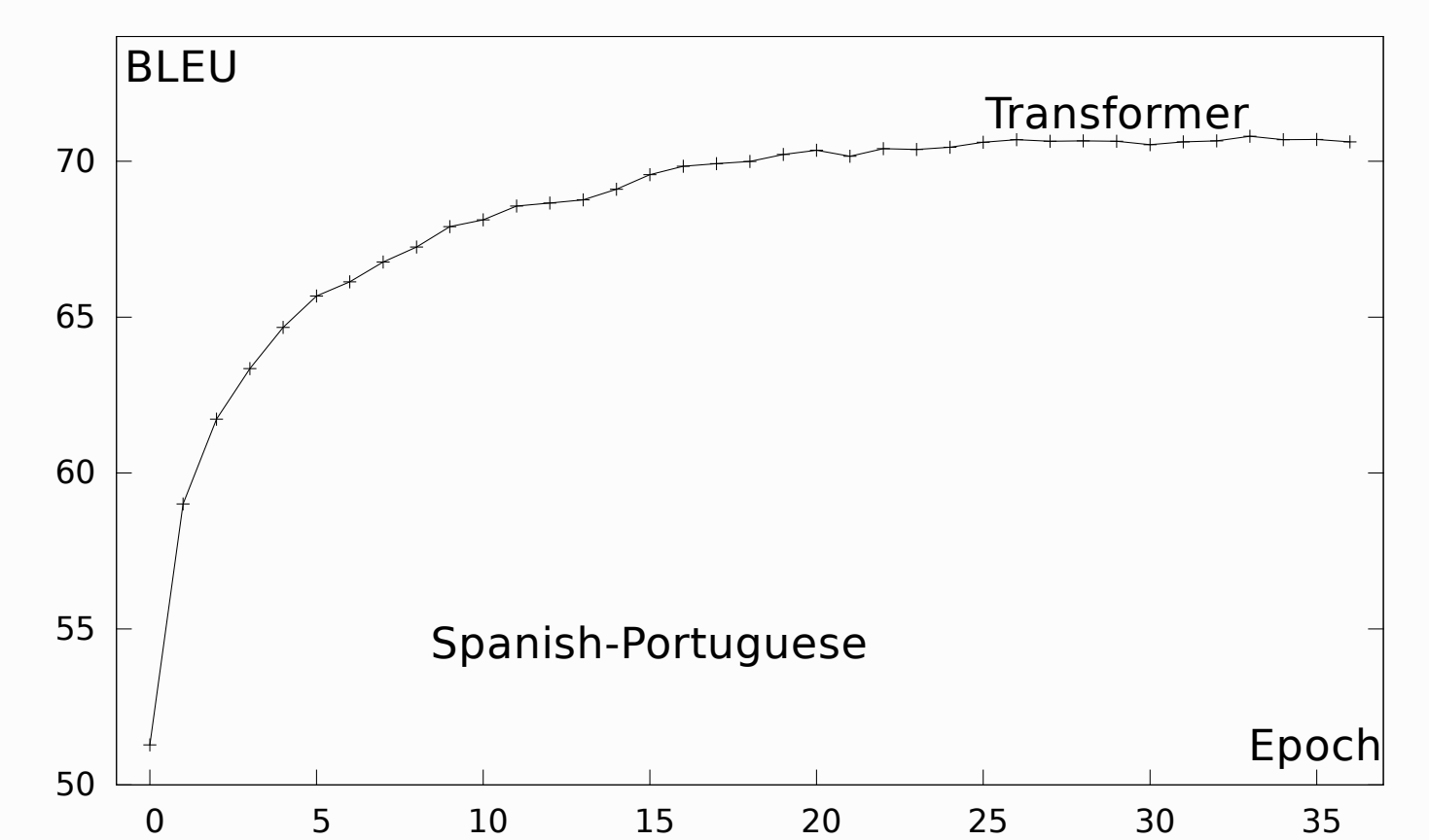
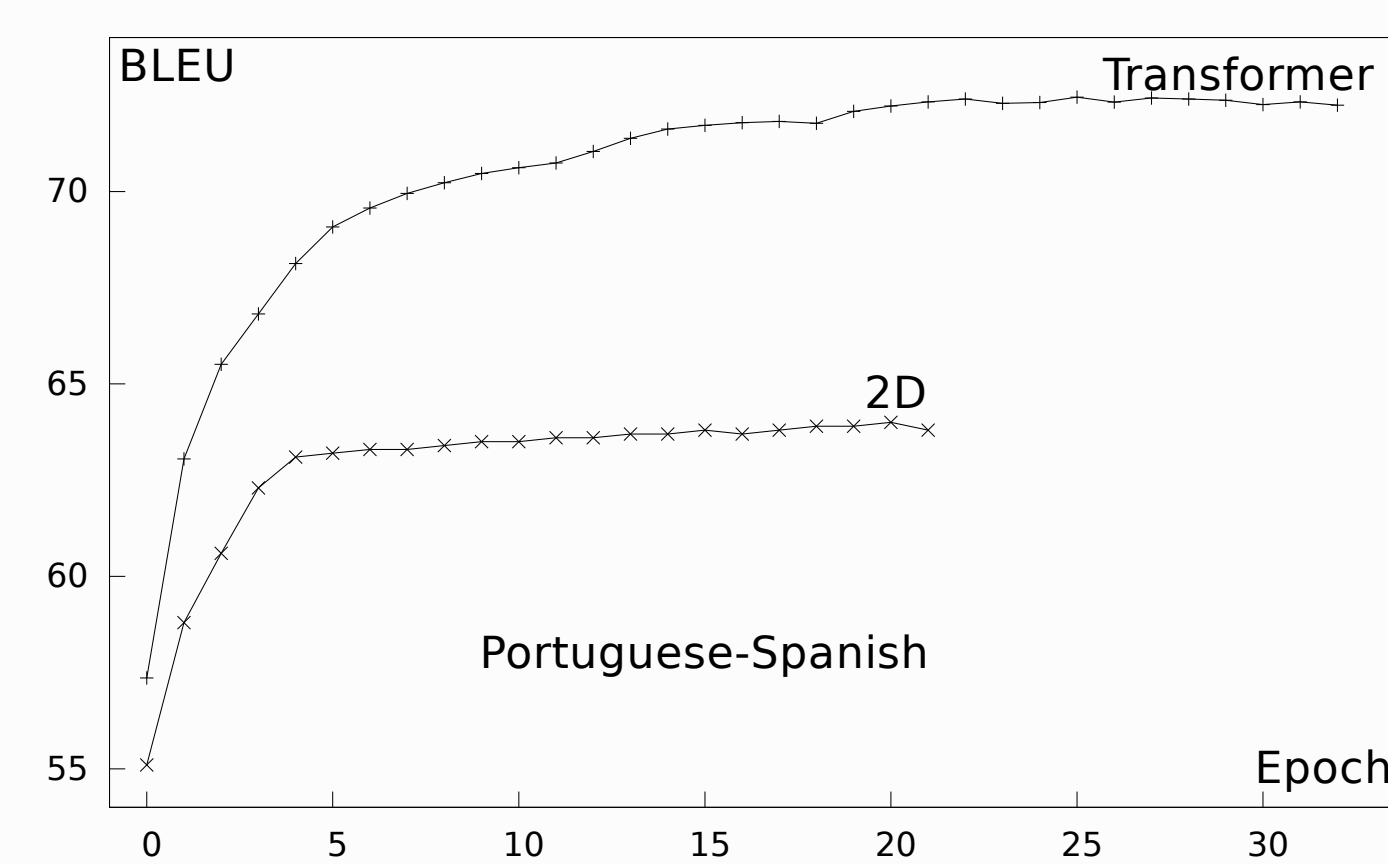
- Embedding size: 258; Recurrent state size: 128; 1 block
- Self-developed implementation on TensorFlow

FINE-TUNING

- Significant domain mismatch between train and test data
- Fine-tuning (after training converges) on a subset of dev data

System	Portuguese → Spanish		Spanish → Portuguese	
	test	test-hidden	test	test-hidden
Transformer	57.4	51.9	51.3	45.5
+ fine-tuned	72.4	66.6	70.7	64.7
2D altern. RNN	55.1	49.7	-	-
+ fine-tuned	64.0	-	-	-

- First fine-tuning epochs are the most useful. After that, we get diminishing returns until the BLEU curve flattens.



SYSTEM EVALUATION

Portuguese → Spanish			Spanish → Portuguese		
Team	BLEU	TER	Team	BLEU	TER
MLLP	66.6	19.7	MLLP	64.7	20.8
NICT	59.9	25.3	UPC-TALP	62.1	23.0
U. Helsinki	58.4	25.3	NICT	53.3	29.1
Kyoto U.	56.9	26.9	U. Helsinki	52.0	29.4
BSC	54.8	29.8	UBC-NLP	46.1	36.0
UBC-NLP	52.3	32.9	BSC	44.0	37.5
MLLP-2D	49.7	32.1			

CONCLUSIONS

- Generalist approach to similar language translation
- For this task, fine-tuning with in-domain data was critical
- We introduced a novel NMT architecture still in development to test against other participants

Acknowledgments

The research leading to these results has received funding from the **European Union's Horizon 2020** research and innovation programme under grant agreement no. 761758 (X5gon); the **Government of Spain's** research project Multisub, ref. RTI2018-094879-B-I00 (MCIU/AEI/FEDER, EU); and the **Universitat Politècnica de València's** PAID-01-17 R&D support programme.