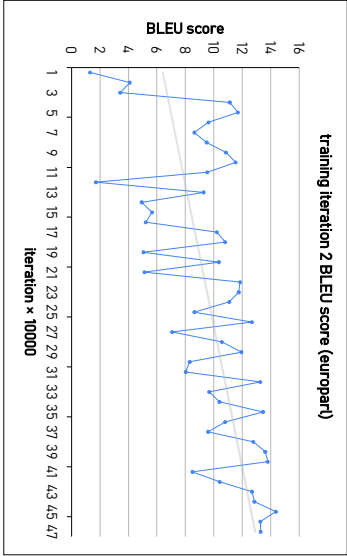


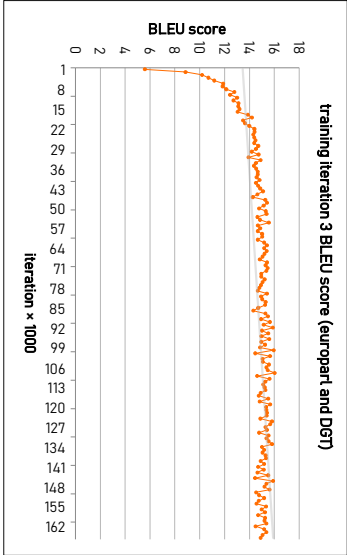
Iteration 1 = First test training with small subset to evaluate software
Iteration 2 = Actual training with only europarl v10 data
Iteration 3 = Actual training with europarl v10 and DGT 2019 data



Best BLEU: 14.354700
Least BLEU: 1.72
STD: 3.28721446
n 47

Checkpoints to analyze and why

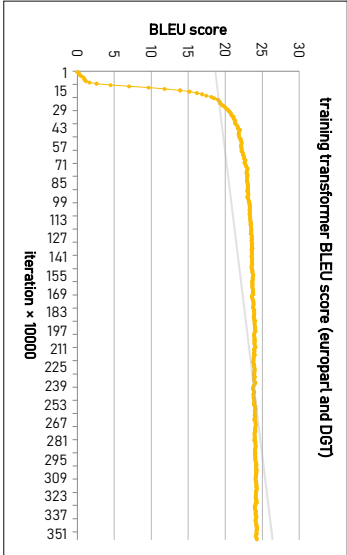
- 1. epoch start
- 4. epoch BLEU > 10
- 12. epoch sudden drop
- 26. epoch middle high
- 45. epoch highest BLEU



Best BLEU: 16.0208
Least BLEU: 5.58434
STD: 1.25665258
n 166

Checkpoints to analyze and why

- 1. epoch start
- 3. epoch BLEU > 10
- 18. epoch BLEU > 14
- 44. epoch BLEU > 15
- 108. epoch highest BLEU



Best BLEU: 24.2892
Least BLEU: 0
STD: 4.245557556
n 356

Checkpoints to analyze and why

- Epoch 2 (iteration 15): BLEU score 10
- Epoch 4 (iteration 39): BLEU score > 20
- Epoch 21 (iteration 255): BLEU score 24 (no more changes)
- Epoch 28 (iteration 355): Final epoch, BLEU score 24

Category	Ambiguity	Coordination & ellipsis	False friends	Function word	Long distance dependency & interrogative	MWE	Named entity & terminology	Negation	Non-verbal agreement	Punctuation	Sub-ordination	Verb tense/ aspect/ mood	Verb valency
s2s #2	20%	30%	50%	60%	40%	22%	20%	60%	20%	60%	70%	20%	30%
s2s #3	10%	30%	60%	44%	50%	0%	40%	50%	30%	20%	30%	10%	20%
Transformer	20%	10%	50%	60%	70%	10%	80%	60%	50%	50%	80%	40%	50%

The figures show the BLEU score progress over all recorded iterations. Please note that you have to multiply the stated iteration number with 10,000. All systems are tested against newstest2015 and newstest2016