

Full Results – correlations excluding Transformer

BLEU vs. other metrics:

-0.57 ± 0.31 (en → cs)

-0.54 ± 0.27 (en → de)

Pairwise average
(except BLEU):

0.78 ± 0.32 (en → cs)

0.62 ± 0.23 (en → de)

