

Simple Fusion: Return of the Language Model

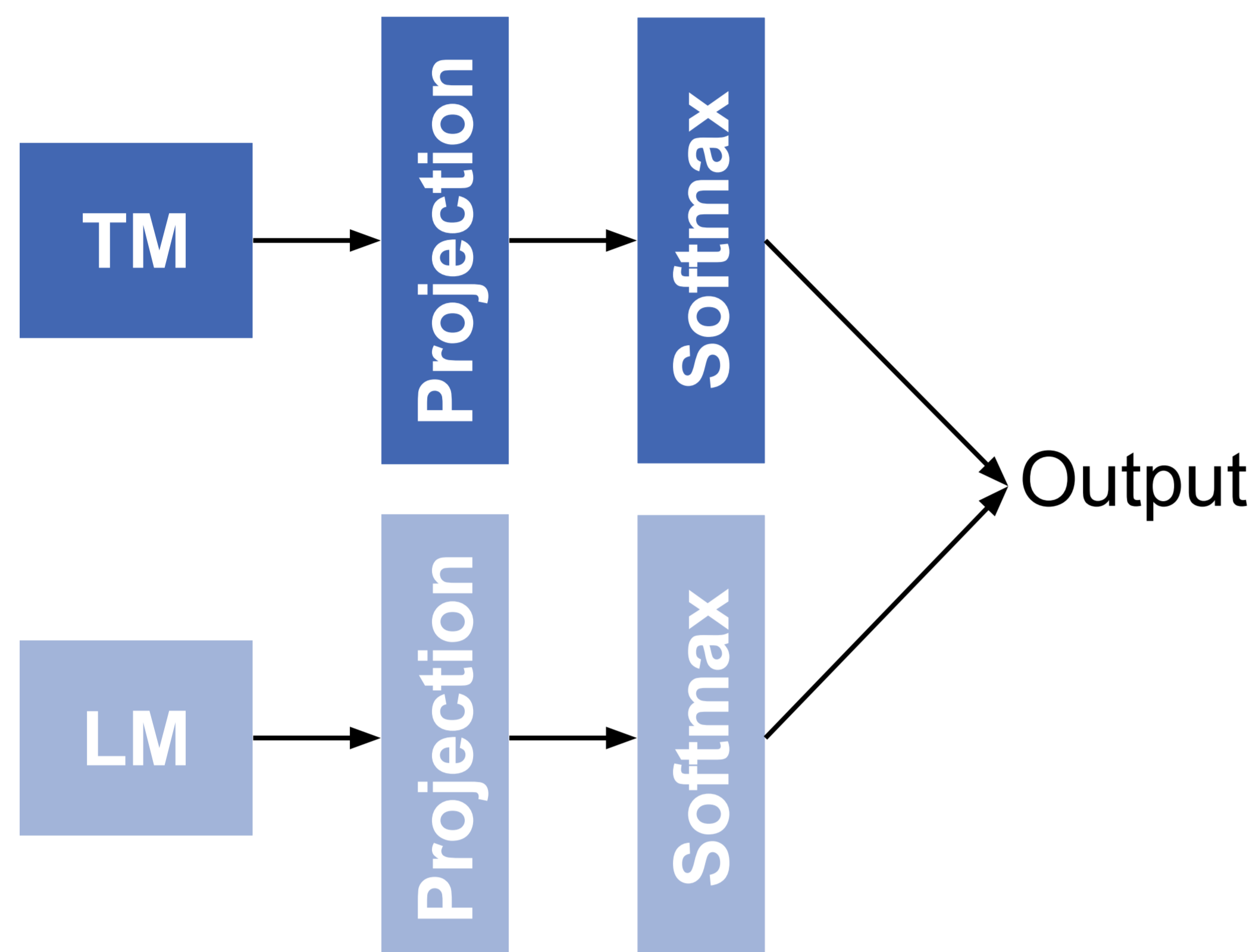
Felix Stahlberg, James Cross, Veselin Stoyanov



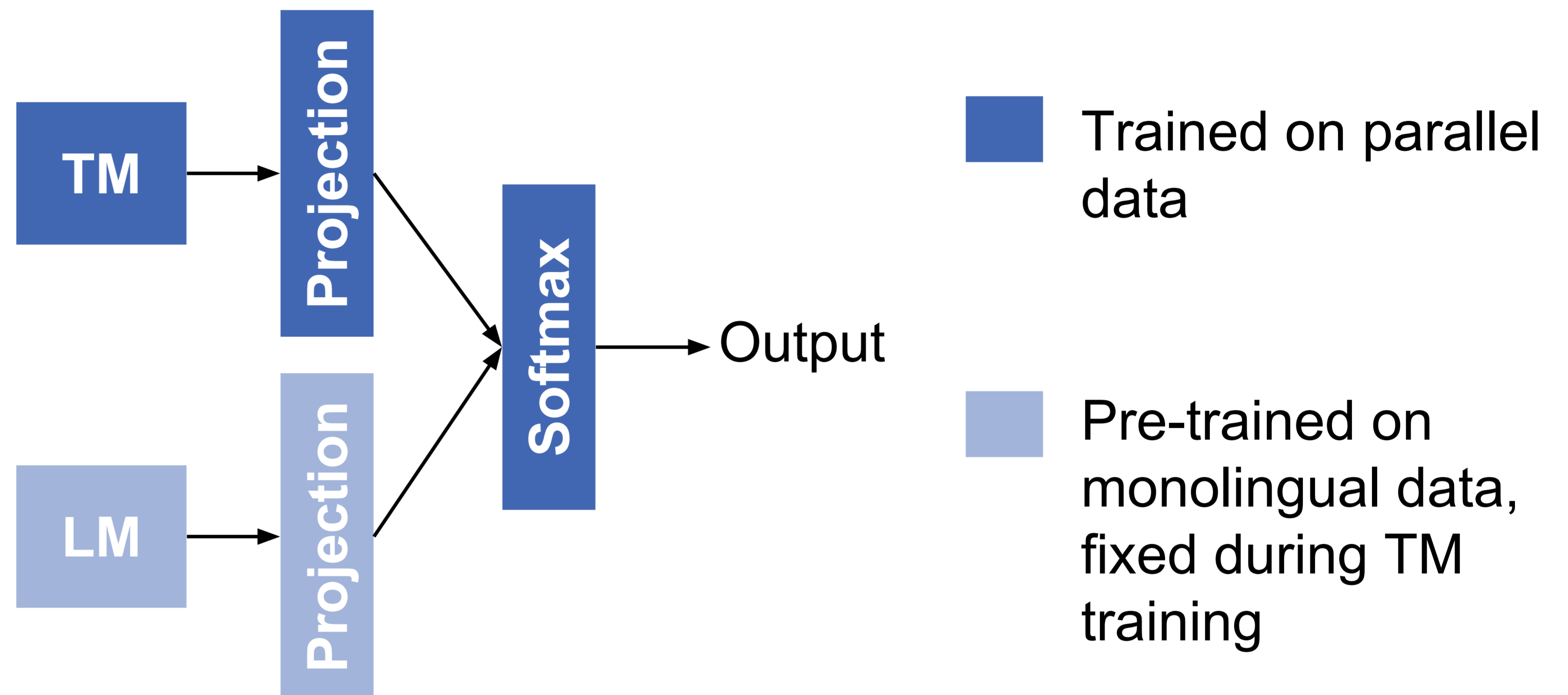
Simple Fusion Combination Strategies

- Combine a pre-trained and fixed language model (LM) with the translation model (TM) while the TM is trained from scratch.
- The TM predicts the residual probability added to the prediction of the LM and thus can focus its capacity on modeling the source sentence.

PostNorm Strategy



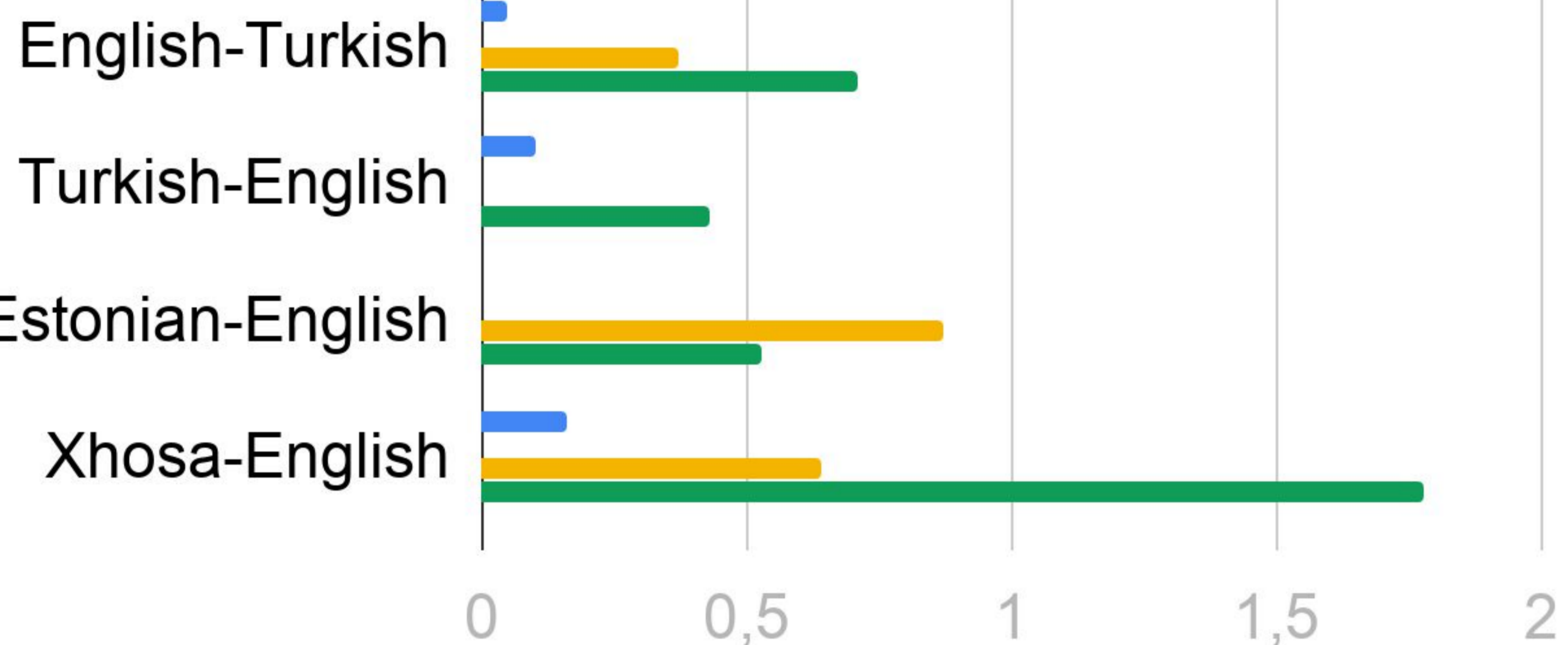
PreNorm Strategy



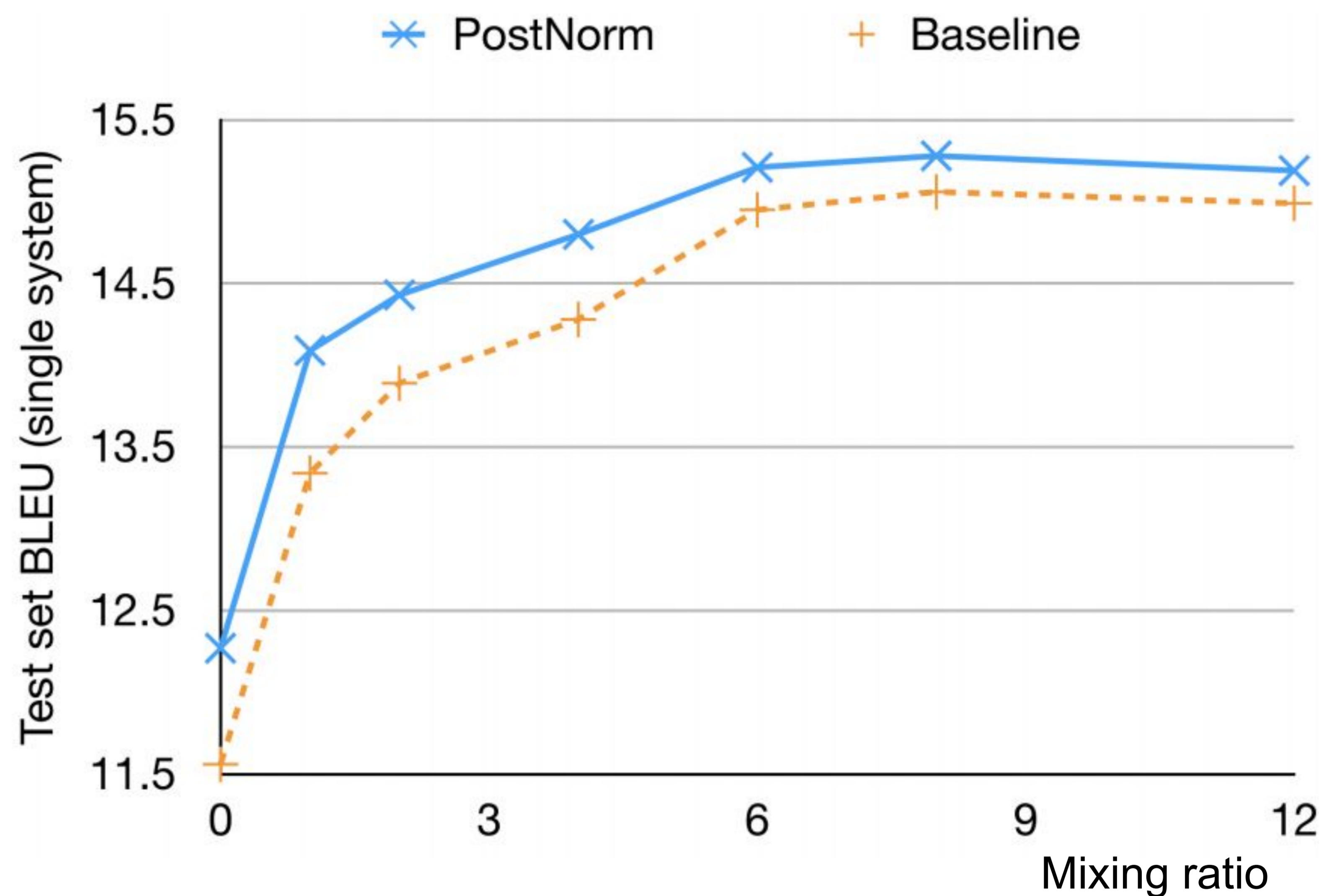
■ Trained on parallel data
■ Pre-trained on monolingual data, fixed during TM training

Single Systems

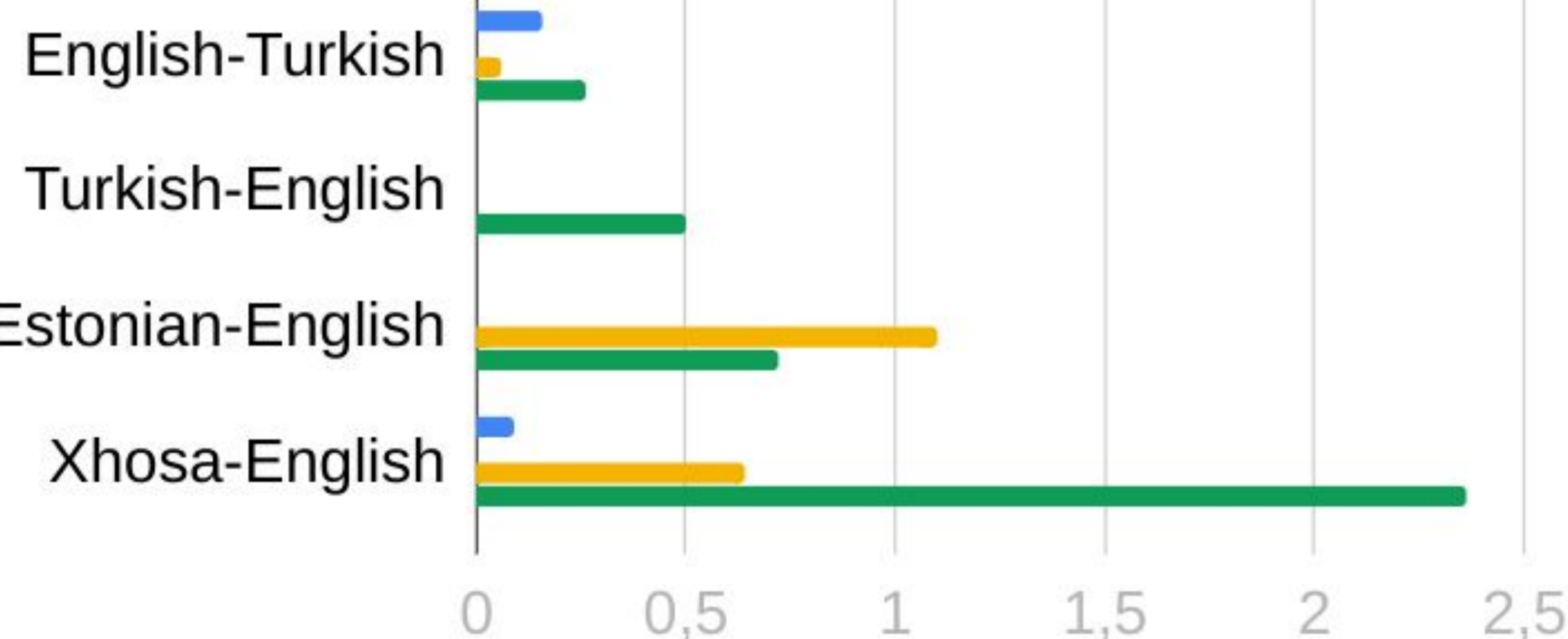
Abs. BLEU gain over baseline w/o LM



Backtranslation



4-Ensembles



■ Shallow Fusion (Gulcehre et al., 2015) ■ Cold Fusion (Siriam et al., 2017)
■ PreNorm (this work) ■ PostNorm (this work)

N-gram Precisions (Estonian-English)

Method	BLEU	Precisions			
		1-gram	2-gram	3-gram	4-gram
Baseline (no LM)	17.91	53.0	23.7	12.3	6.6
This work: PreNorm	19.01	54.0	24.9	13.4	7.4
Relative gain	+6.14%	+1.89%	+5.06%	+8.94%	+12.12%