Simulated Multiple Reference Training Improves Low-Resource Machine Translation

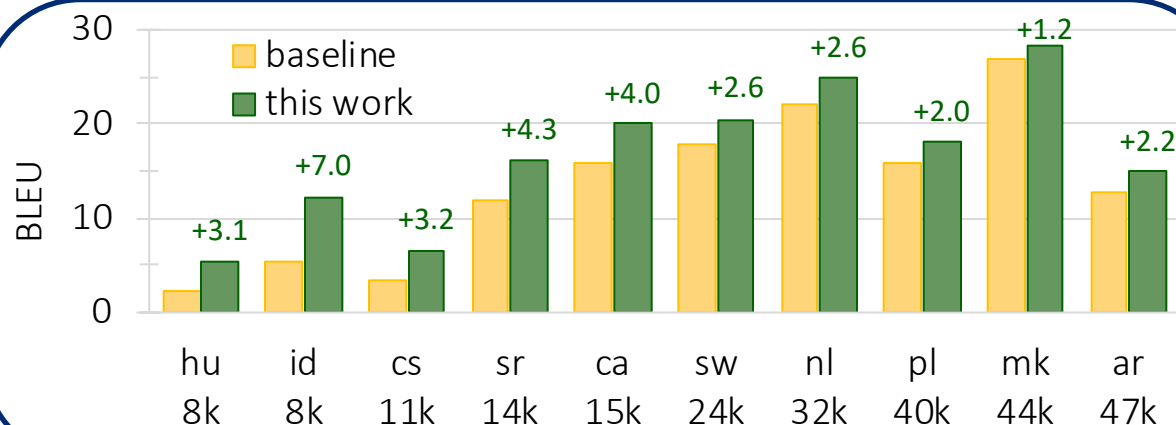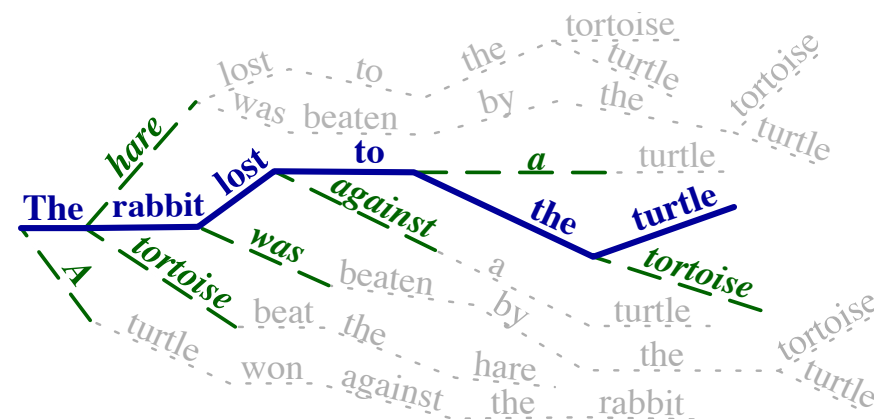**Huda Khayrallah**, Brian Thompson, Matt Post & Philipp Koehn | Johns Hopkins University

use a target language *paraphraser* in the training objective to overcome data sparsity in *low-resource* settings

Questions? Hiring? huda@jhu.edu

src: La tortuga ganó contra la liebre
ref: The turtle beat the hare

$$\text{NLL} -\sum_{v \in \mathcal{V}} \left[\underbrace{\mathbb{1}\{y_i = v\}}_{\text{Gold Target}} \times \log \underbrace{p_{\text{MT}}(y_i = v \mid x; y_{j<i})}_{\text{MT Model output distribution}} \right]$$

$$\text{SMRT} -\sum_{v \in \mathcal{V}} \left[\underbrace{p_{\text{para}}(y'_i = v \mid y; y'_{j<i})}_{\text{Paraphraser output distribution}} \times \log \underbrace{p_{\text{MT}}(y'_i = v \mid x; y'_{j<i})}_{\text{MT Model output distribution}} \right]$$

- Transformer models in fairseq (Ott et al., 2019)
- Global Voices corpora (Tiedemann, 2012)
- Translating to English
- English Paraphraser trained on ParaBank2 (Hu et al., 19)
- Use SMRT w/ 50% probability, NLL otherwise
- 4k SentencePiece vocab (Kudo & Richardson 2018)
- Code, Paper & more: data.statmt.org/smrt

BLEU chart comparing baseline vs this work:

language	hu 8k	id 8k	cs 11k	sr 14k	ca 15k	sw 24k	nl 32k	pl 40k	mk 44k	ar 47k
improvement	+3.1	+7.0	+3.2	+4.3	+4.0	+2.6	+2.6	+2.0	+1.2	+2.2