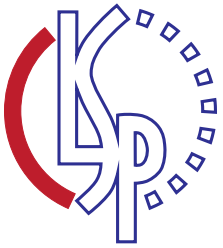
Simulated Multiple Reference Training (SMRT) Improves Low-Resource Machine Translation

Huda Khayrallah, Brian Thompson,

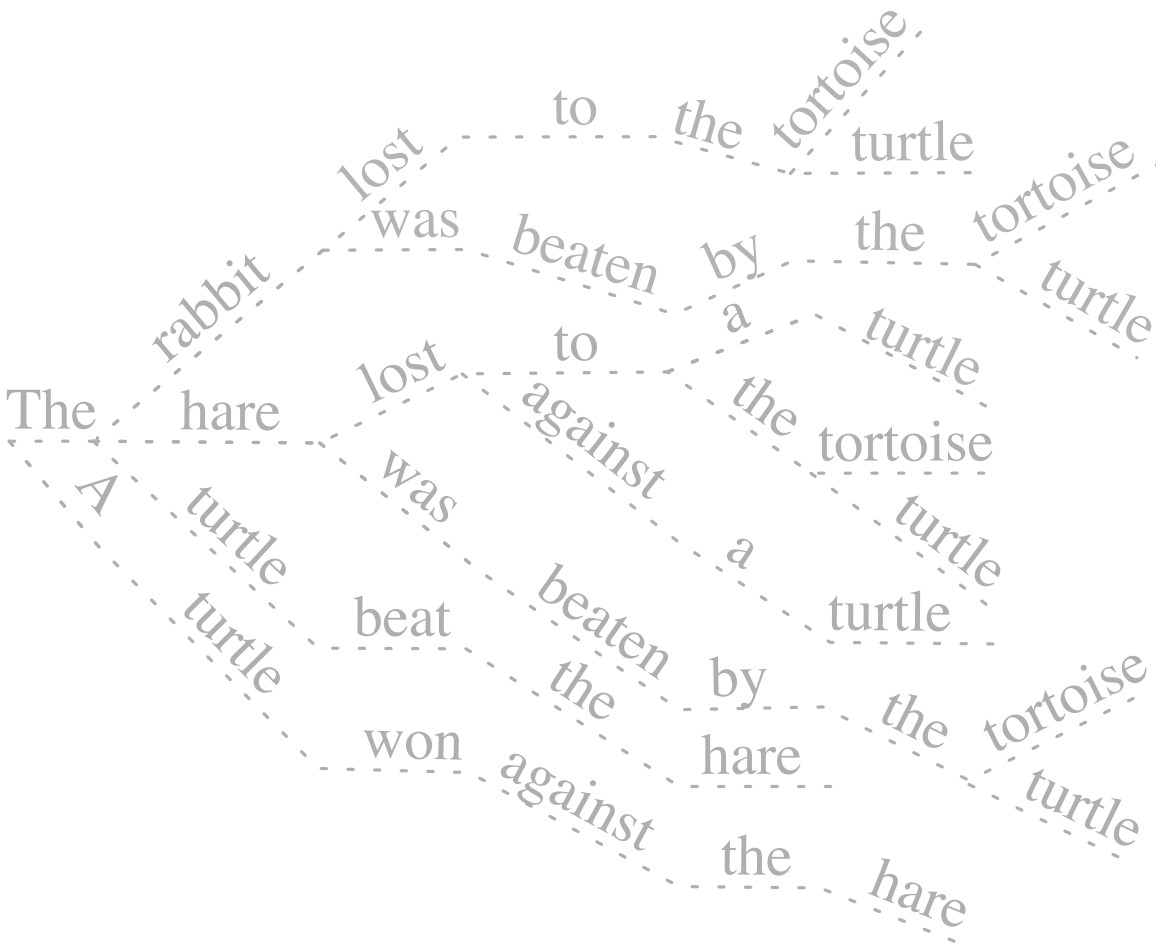
Matt Post & Philipp Koehn

JOHNS HOPKINS
UNIVERSITY

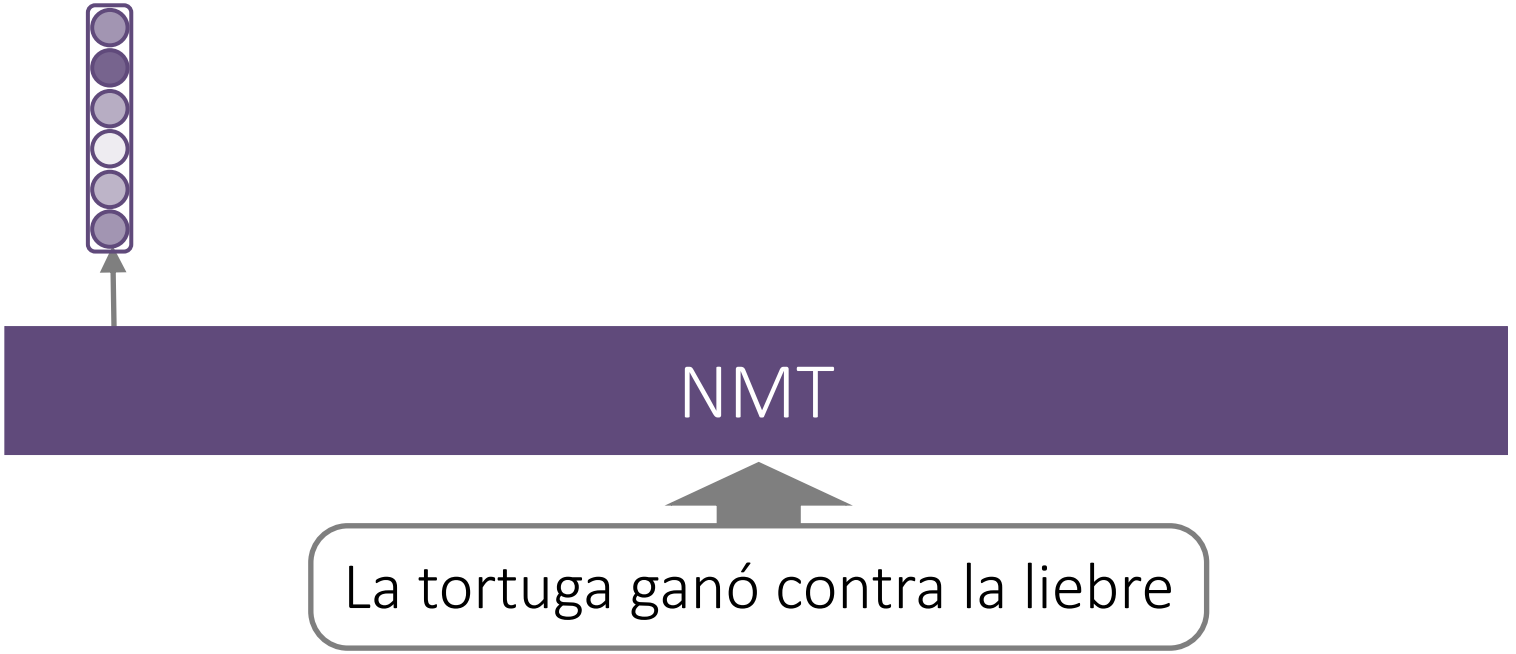


use target side ***paraphrasing***
to overcome data sparsity
in ***low-resource*** settings

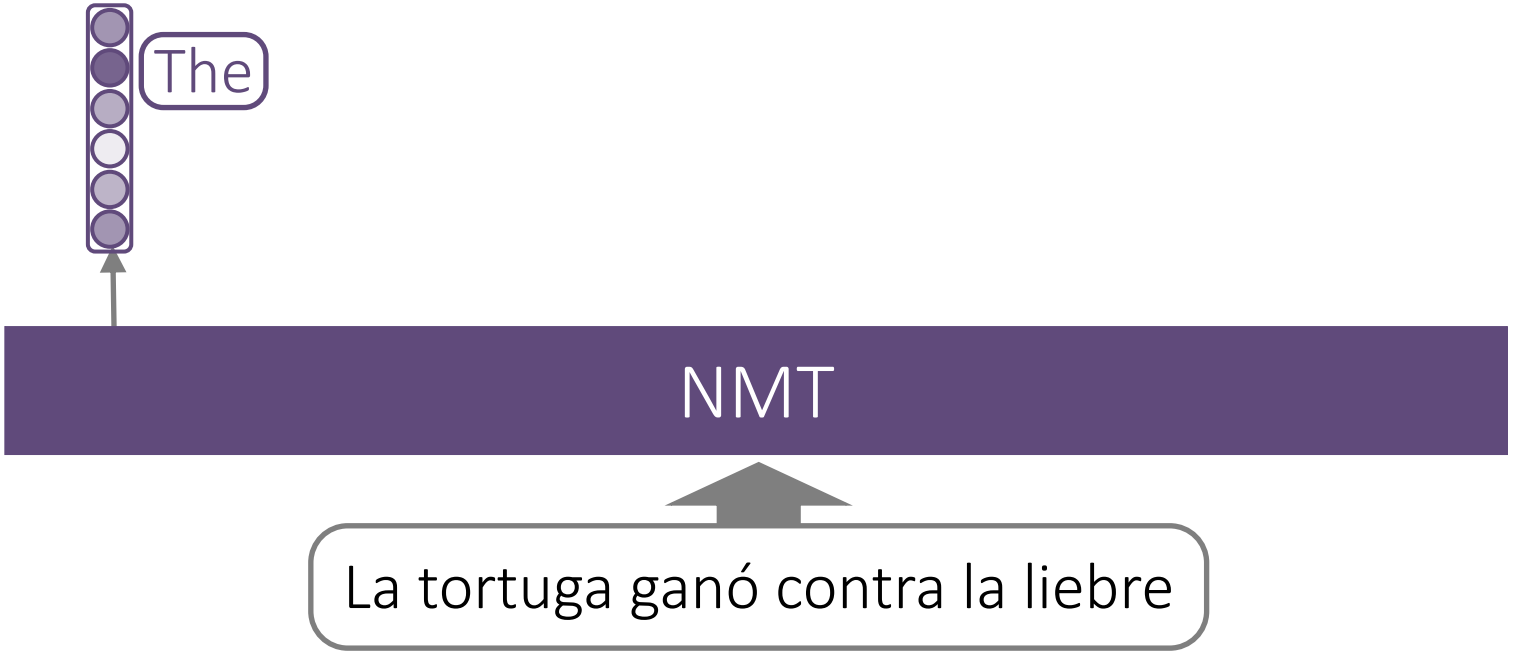
La tortuga ganó contra la liebre | The turtle beat the hare



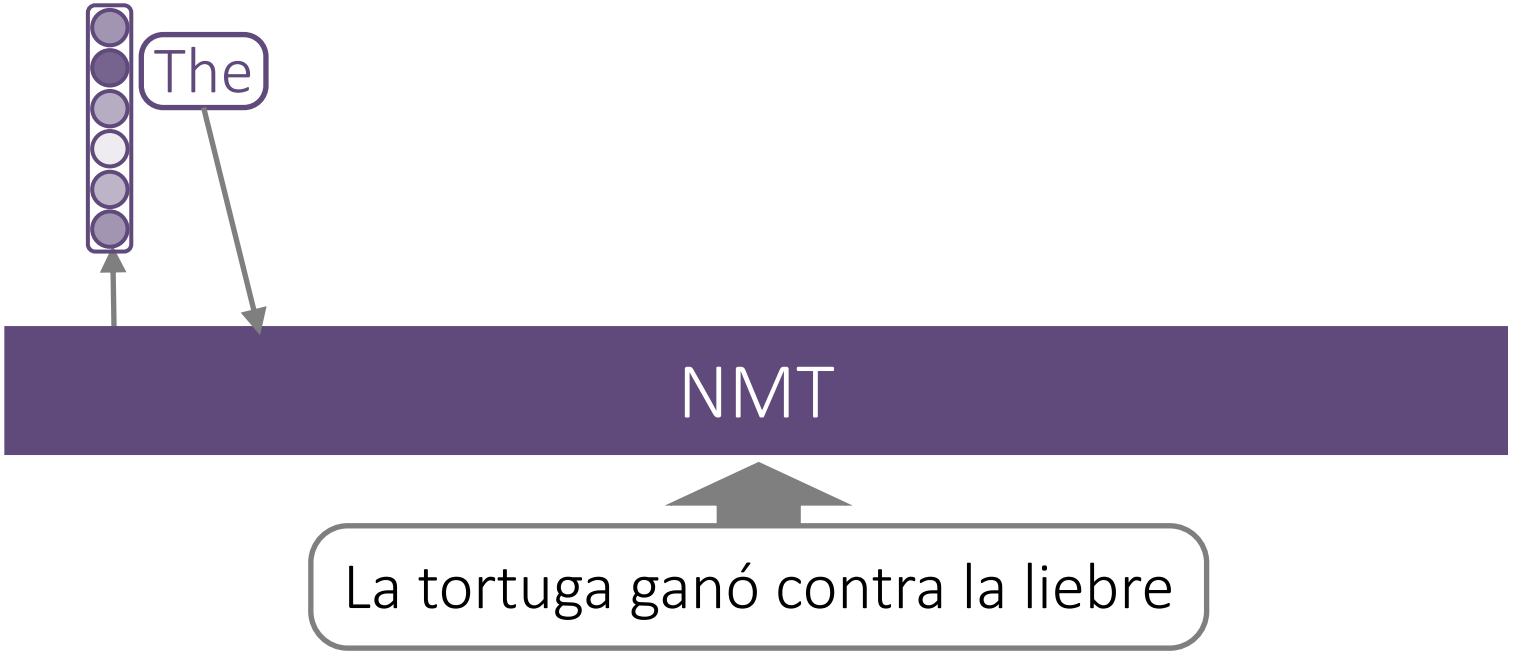
Machine Translation



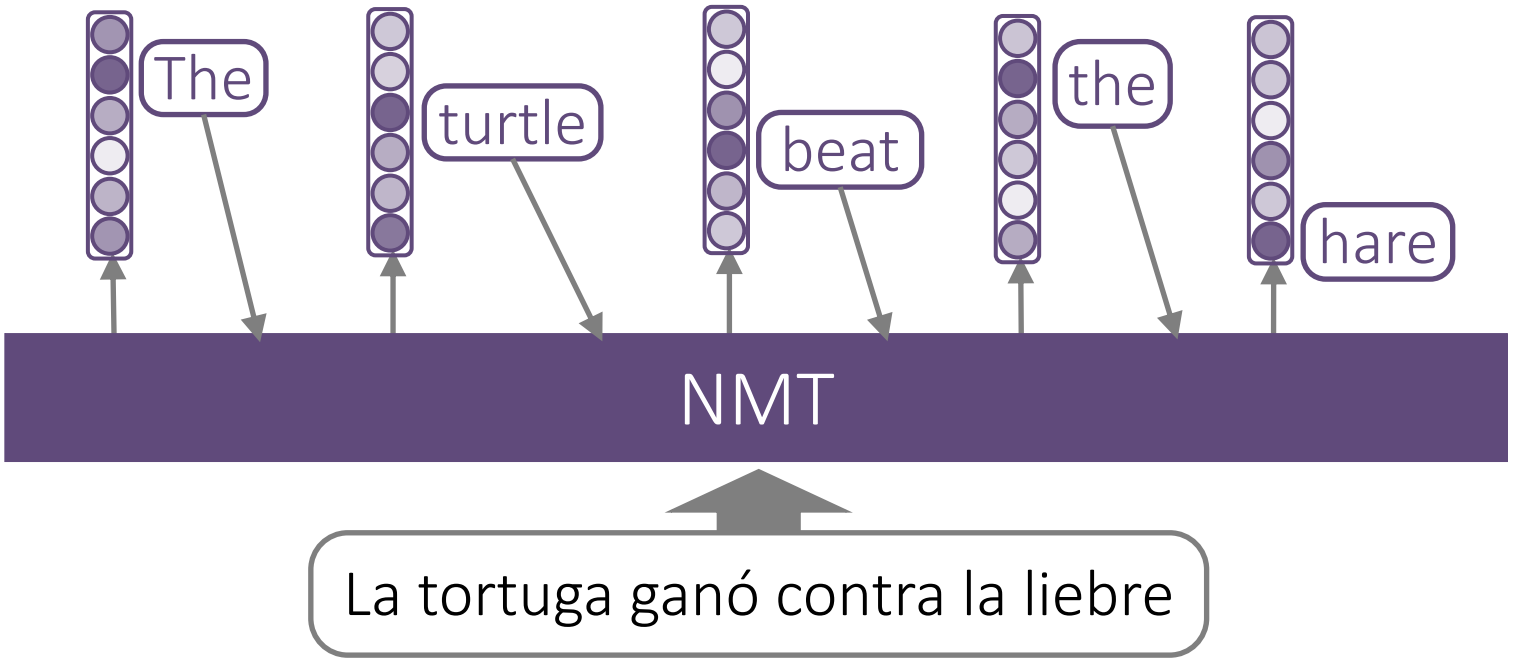
Machine Translation



Machine Translation

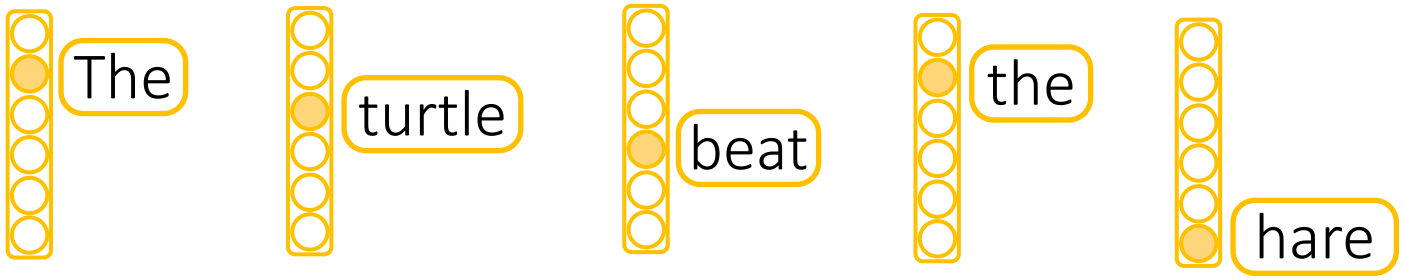


Machine Translation



La tortuga ganó contra la liebre | The turtle beat the hare

La tortuga ganó contra la liebre | The turtle beat the hare

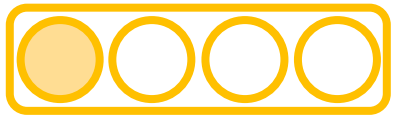
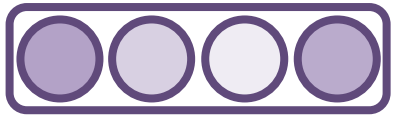


NLL Objective

$$-\sum_{v \in \mathcal{V}} \left[\mathbb{1}\{y_i = v\} \times \log p_{\mathrm{MT}}(y_i = v \mid x; y_{j<i}) \right]$$

Gold Target

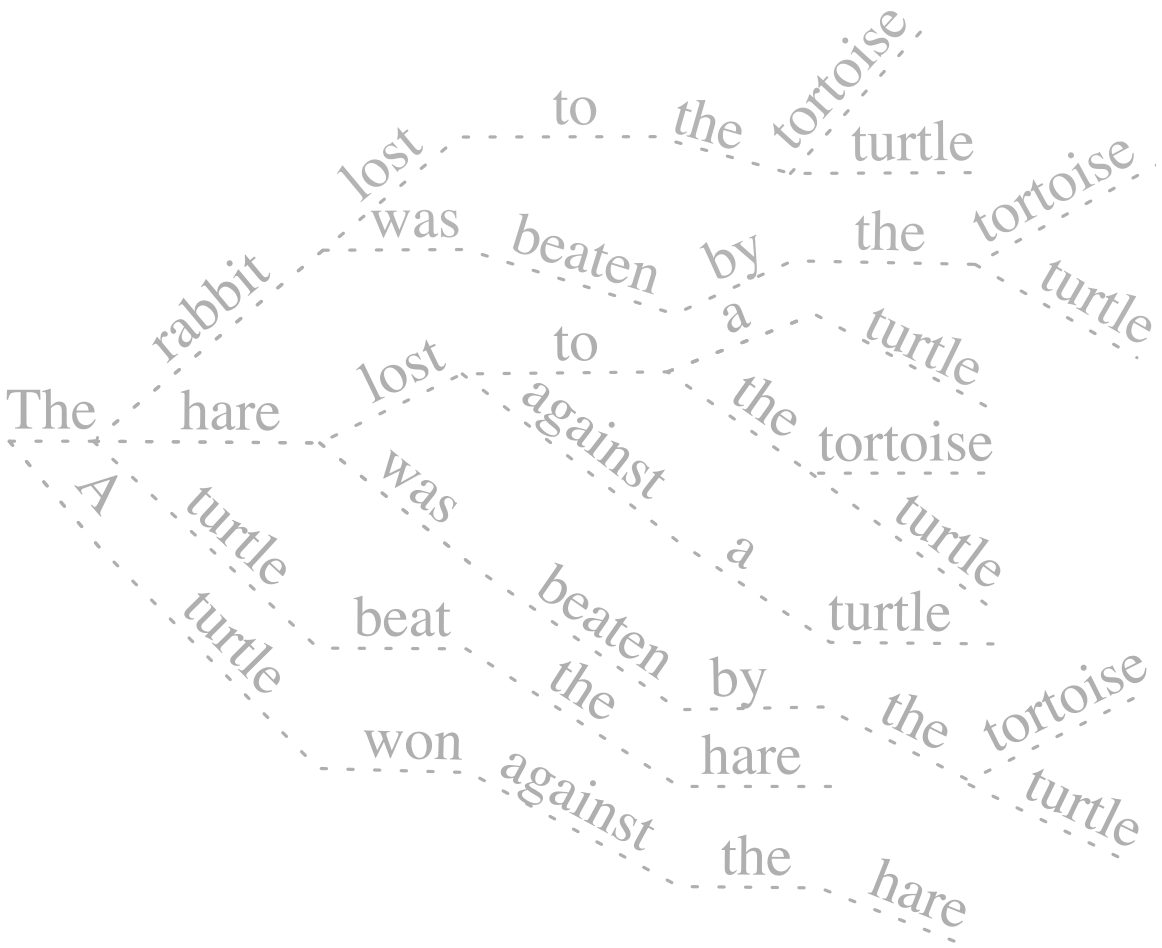
MT Model output

Cross Entropy( , )

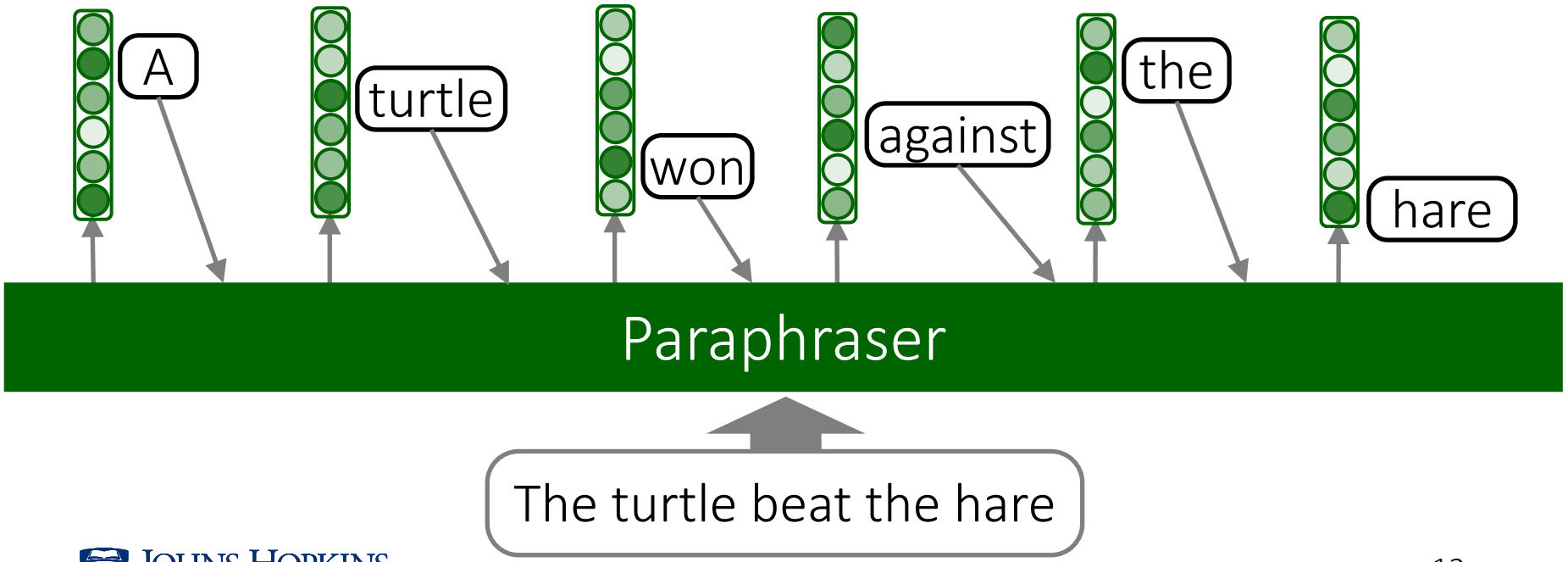
Gold Target

MT Model output

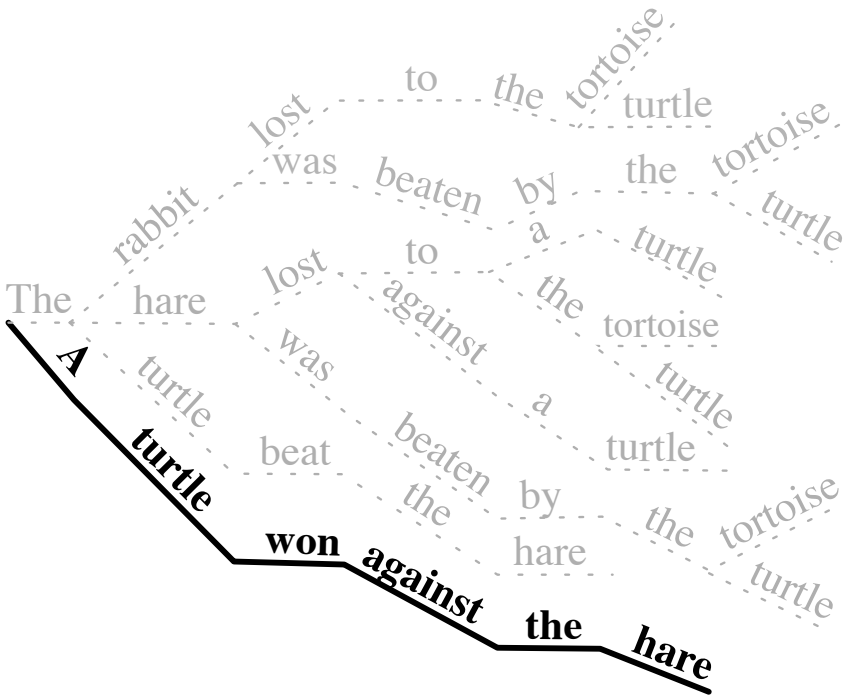
La tortuga ganó contra la liebre | The turtle beat the rabbit

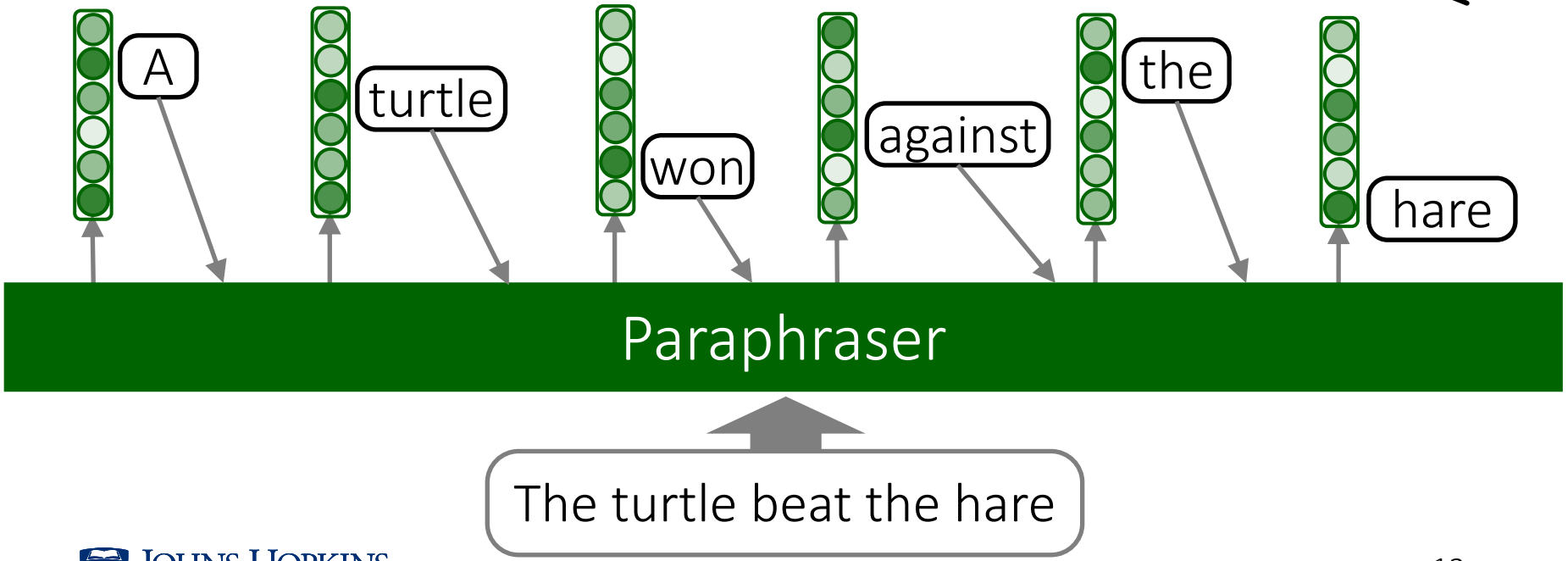


Paraphraser

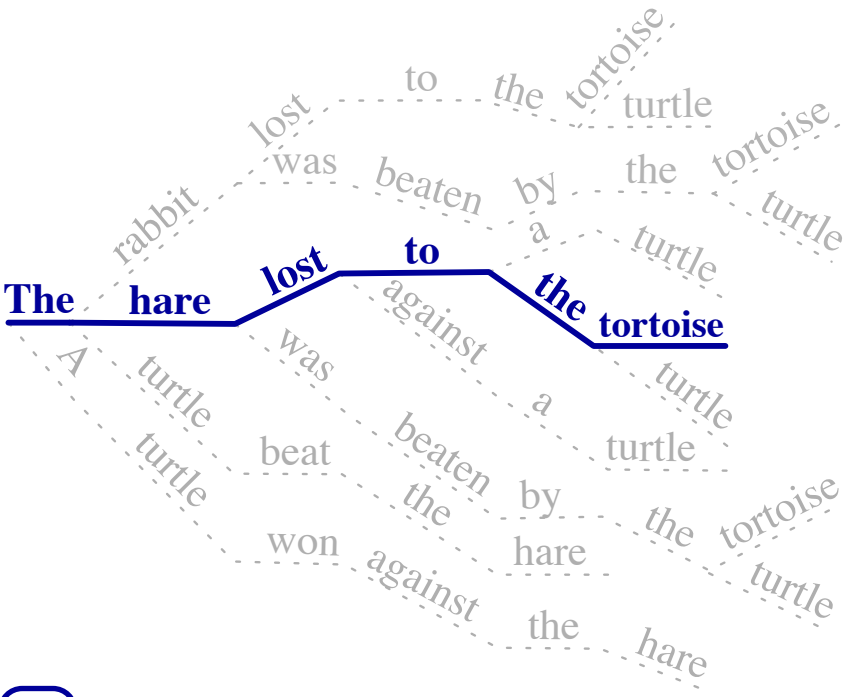


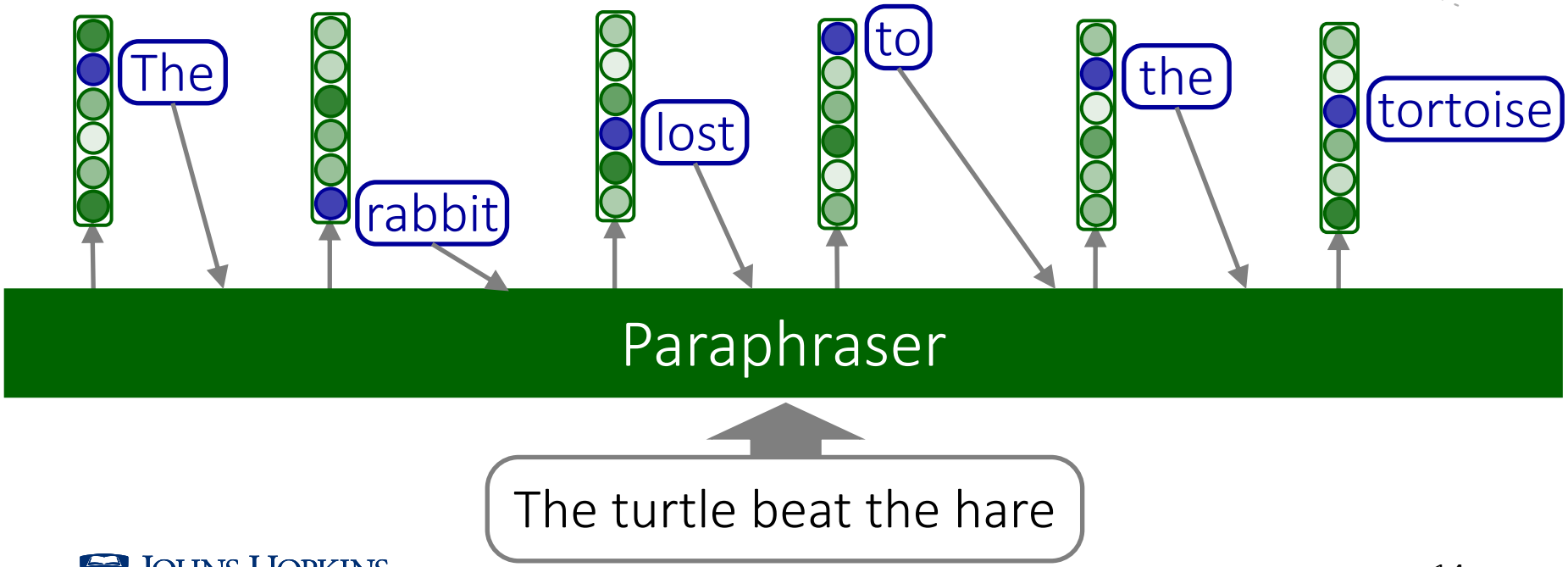
Paraphraser



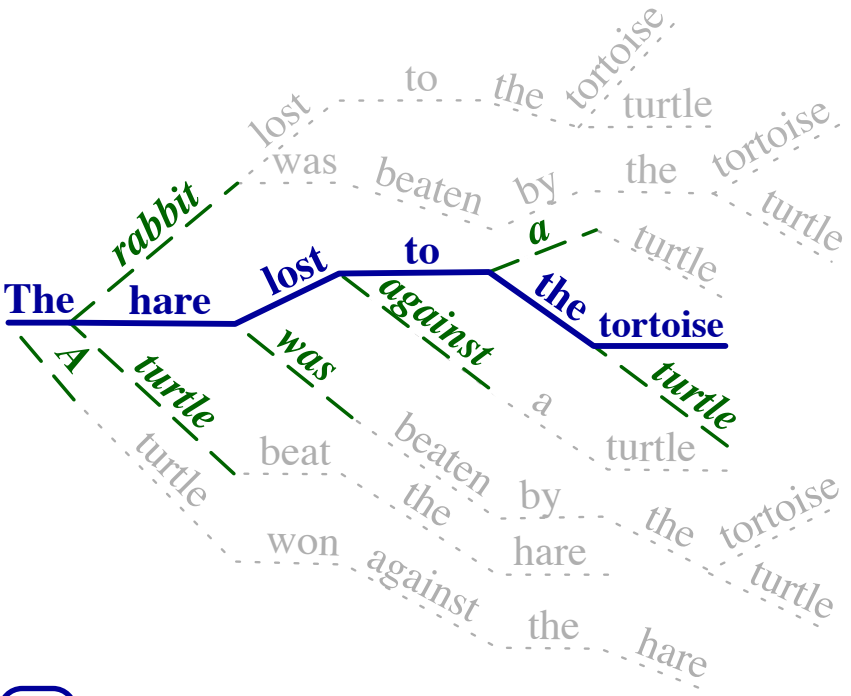


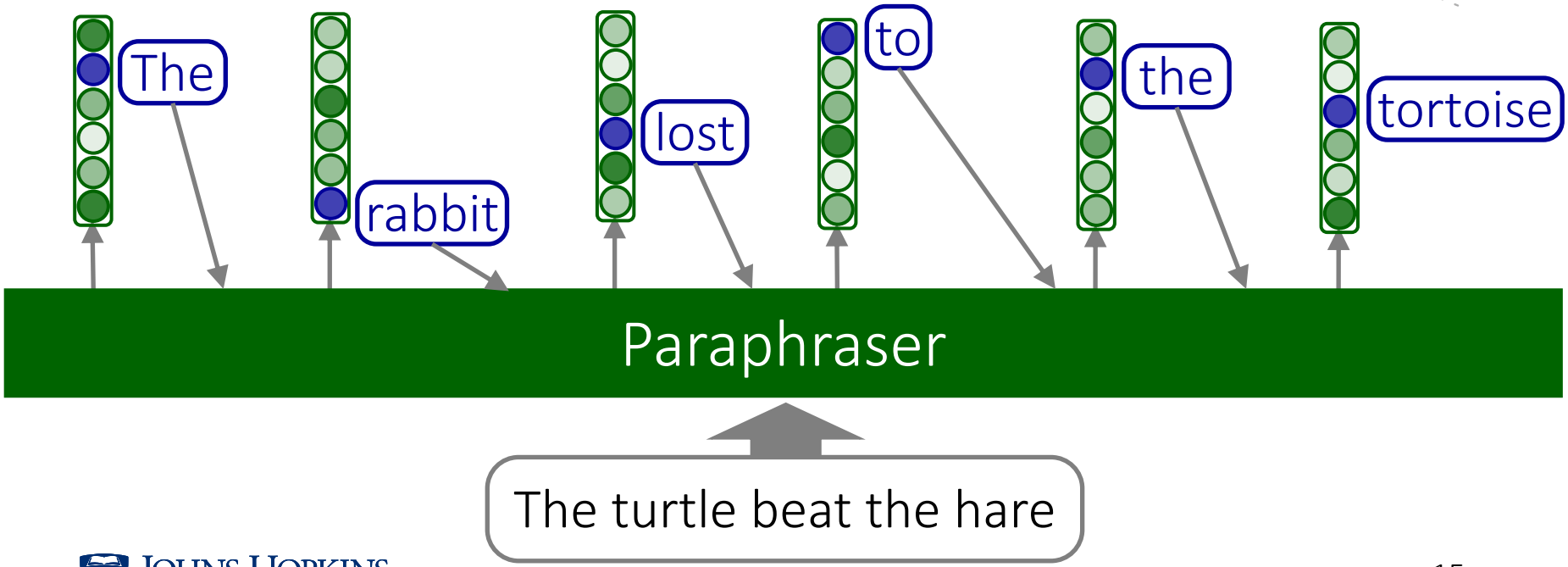
Sampling


The hare lost to the tortoise
rabbit was beaten by a turtle
lost to the tortoise turtle
to the tortoise turtle
A turtle was beaten by the hare
turtle beat the hare
won against the hare
against a turtle
the tortoise turtle

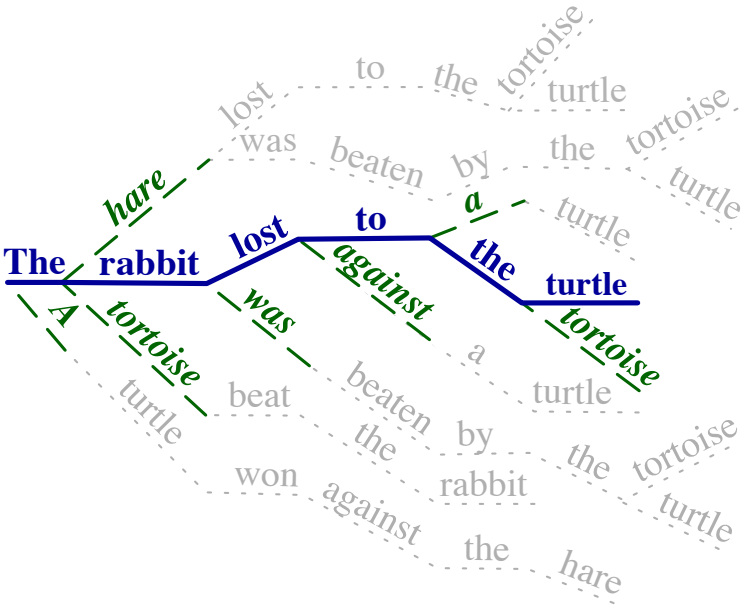

The
rabbit
lost
to
the
tortoise
Paraphraser
The turtle beat the hare

Distribution





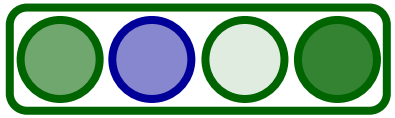
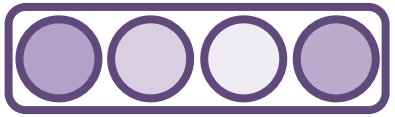
SMRT Objective



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

Paraphraser Output

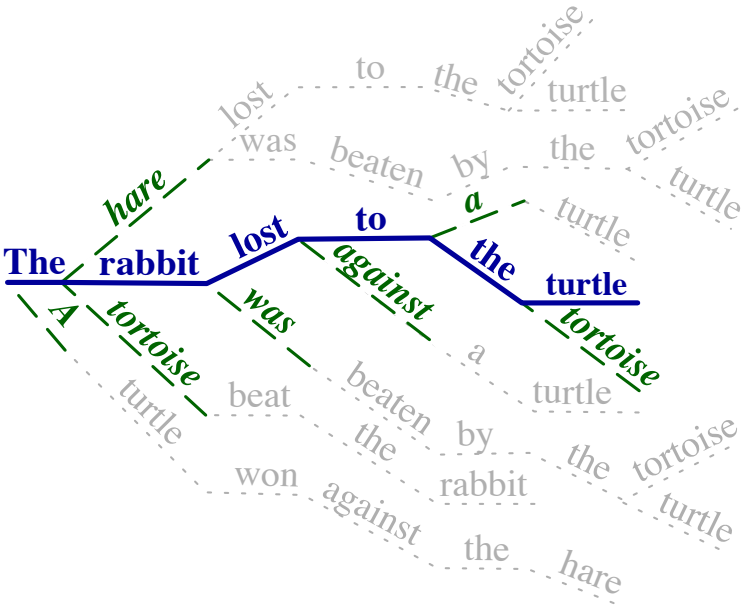
MT Model output

Cross Entropy( , )

Paraphraser Output

MT Model output

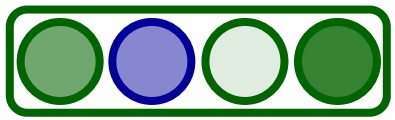
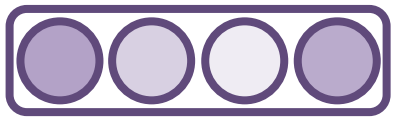
SMRT Objective



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

Paraphraser Output (teacher)

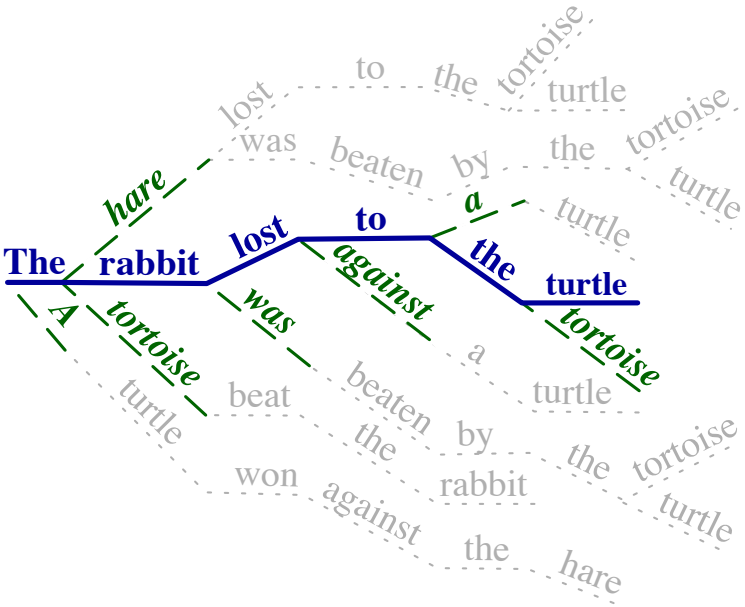
MT Model output (student)

Cross Entropy( , )

Paraphraser Output (teacher)

MT Model output (student)

SMRT Objective



$$-\sum_{v \in \mathcal{V}} \Big[p_{\text{para}}(y'_i = v \mid y; y'_{j<i}) \times \log p_{\text{MT}}(y'_i = v \mid x; y'_{j<i}) \Big]$$

Paraphraser Output

MT Model output

NLL $$-\sum_{v \in \mathcal{V}} \Big[\mathbb{1}\{y_i = v\} \times \log p_{\text{MT}}(y_i = v \mid x; y_{j<i}) \Big]$$

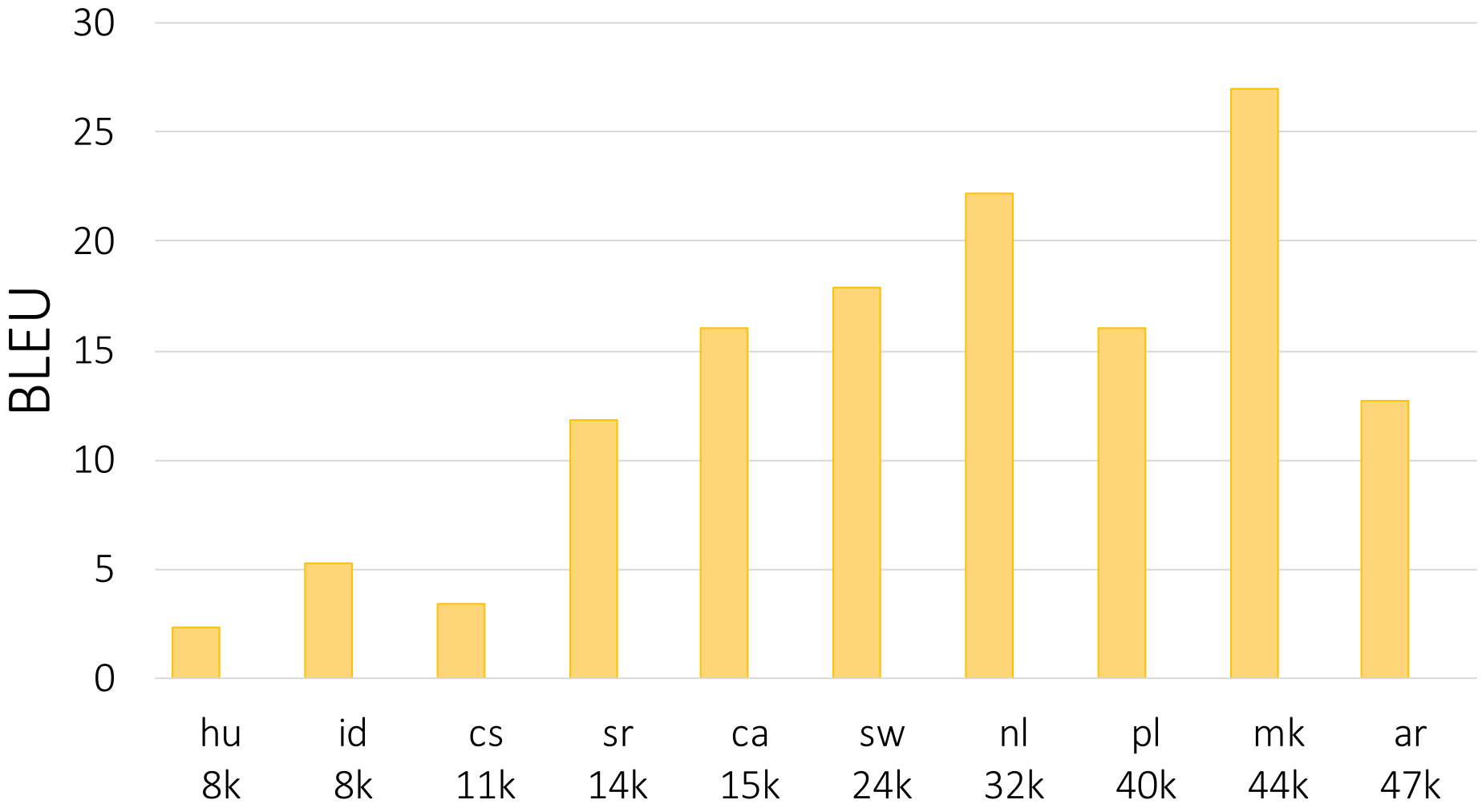
Gold Target

MT Model output

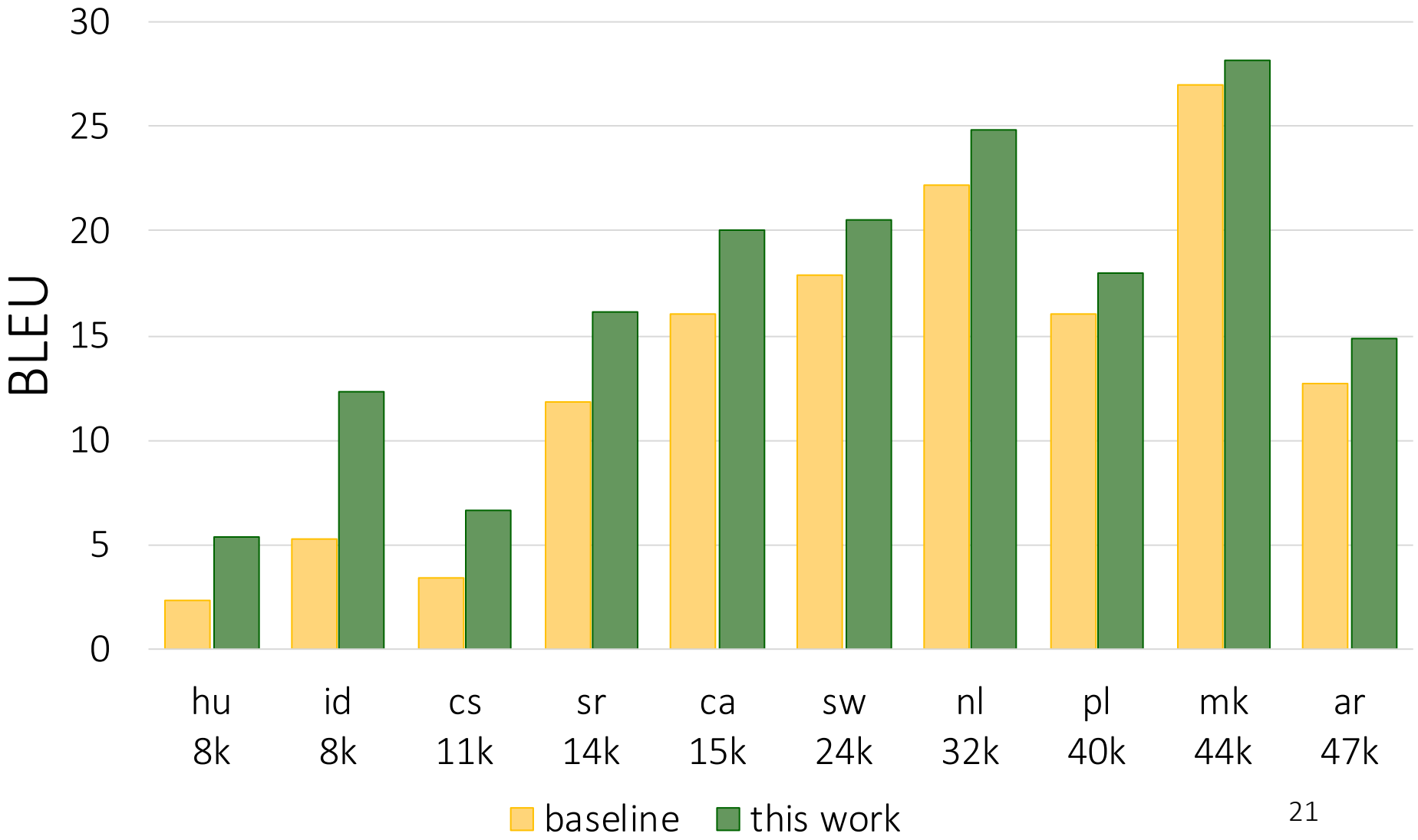
Experimental Details

- transformer model in fairseq (Ott et al., 2019)
- Global Voices corpora (Tiedemann, 2012)
  - (+ MATERIAL corpora in paper)
- Use SMRT w/ 50% probability, NLL otherwise
- English Paraphraser trained on ParaBank2 (Hu et al., 2019)
- 4k SentencePiece vocab (Kudo & Richardson 2018)
- Code, Global Voices Data splits & paraphraser released: data.statmt.org/SMRT

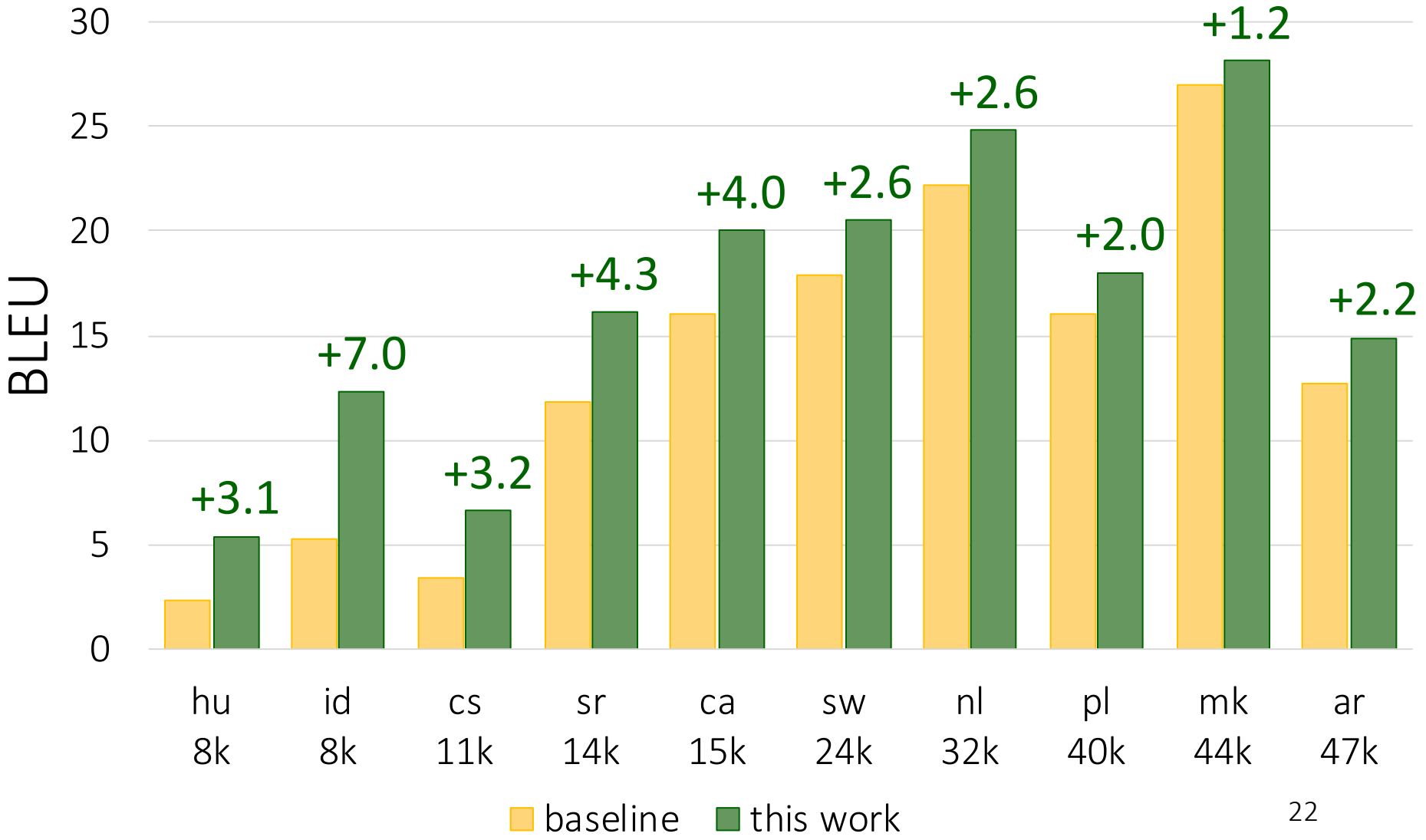
Results


BLEU
30
25
20
15
10
5
0
hu 8k
id 8k
cs 11k
sr 14k
ca 15k
sw 24k
nl 32k
pl 40k
mk 44k
ar 47k

Results


30
25
20
15
10
5
0
BLEU
hu 8k
id 8k
cs 11k
sr 14k
ca 15k
sw 24k
nl 32k
pl 40k
mk 44k
ar 47k
baseline
this work

Results


BLEU
30
25
20
15
10
5
0
+3.1
+7.0
+3.2
+4.3
+4.0
+2.6
+2.6
+2.0
+1.2
+2.2
hu 8k
id 8k
cs 11k
sr 14k
ca 15k
sw 24k
nl 32k
pl 40k
mk 44k
ar 47k
baseline
this work

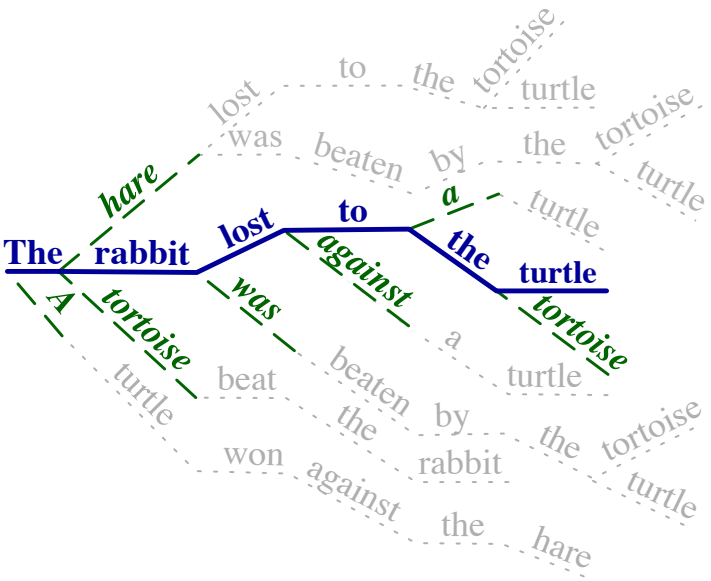
But wait, there's more! (in the paper)

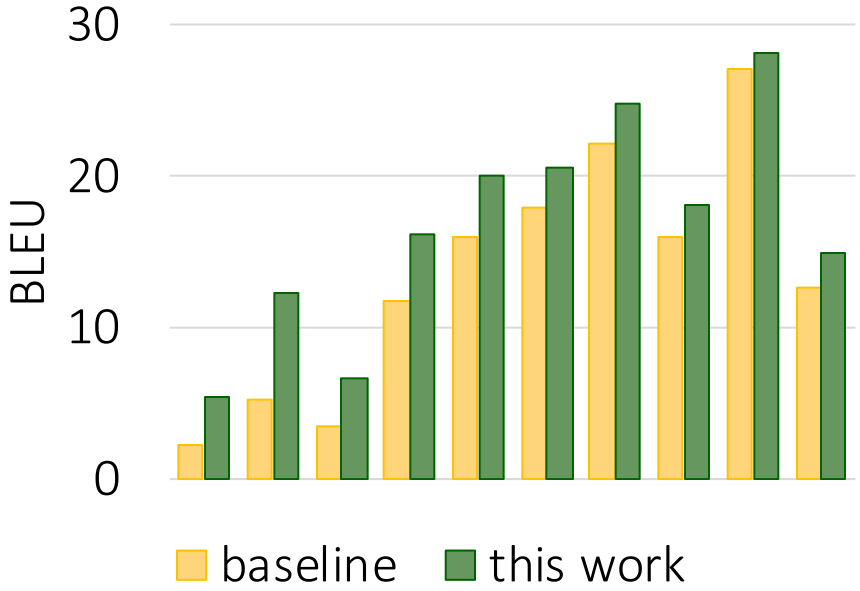
- Comparison to back-translation
  - Both work, SMRT is better than BT in very low resource
  - Can combine for larger improvement
- Data ablation
  - Larger improvements in lower resource settings
- Method ablation
  - Both sampling and the distribution in the loss are helpful
- Sequence-Level Paraphrastic Augmentation
  - It works; SMRT is better

But wait, there's more! (other papers)

- Thompson & Post have a new multilingual paraphraser (Prism) that works for 39 languages and is a great MT metric (@EMNLP)
- Khayrallah & Sedoc apply SMRT + Prism to Chatbots (@EMNLP findings)

Summary





Code, Global Voices Data splits & paraphraser released: data.statmt.org/SMRT

Questions? Hiring?
Huda Khayrallah
huda@jhu.edu

JOHNS HOPKINS UNIVERSITY

Khayrallah et al.