DEPARTMENT OF COMPUTER SCIENCE

Microsoft

Improving Multilingual Neural Machine Translation with Auxiliary Source Languages

Weijia Xu, Yuwei Yin, Shuming Ma, Dongdong Zhang, and Haoyang Huang

University of Maryland

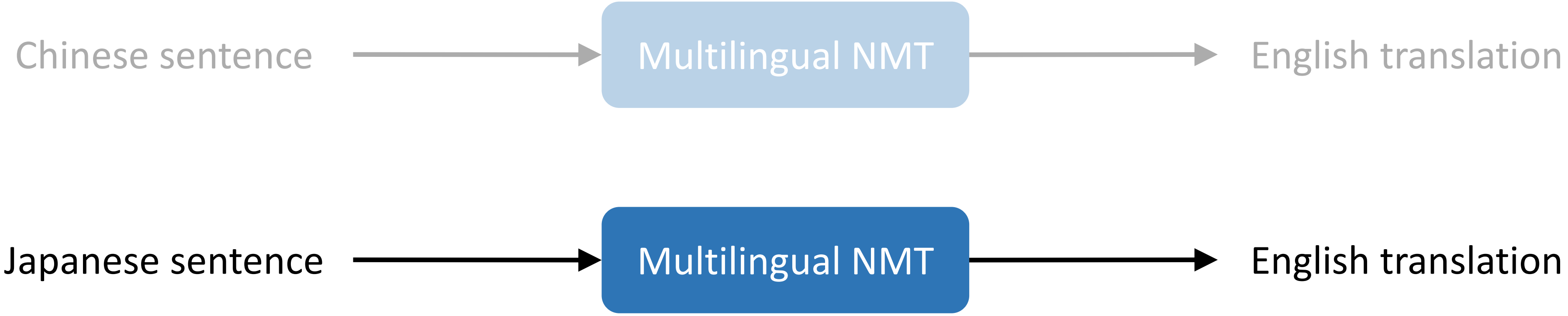
Microsoft Research

Limitations in Multilingual NMT

Multilingual NMT models can translate from multiple source languages, but typically handle one source sentence per time

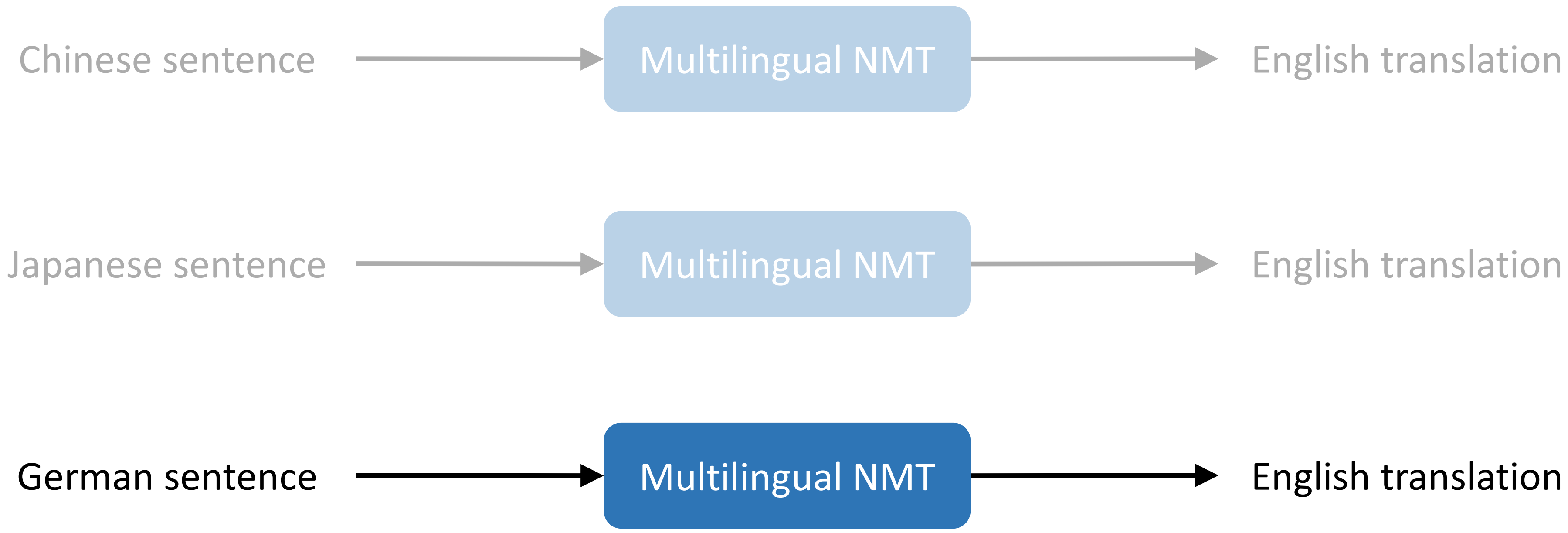
Limitations in Multilingual NMT

Multilingual NMT models can translate from multiple source languages, but typically handle one source sentence per time



Limitations in Multilingual NMT

Multilingual NMT models can translate from multiple source languages, but typically handle one source sentence per time

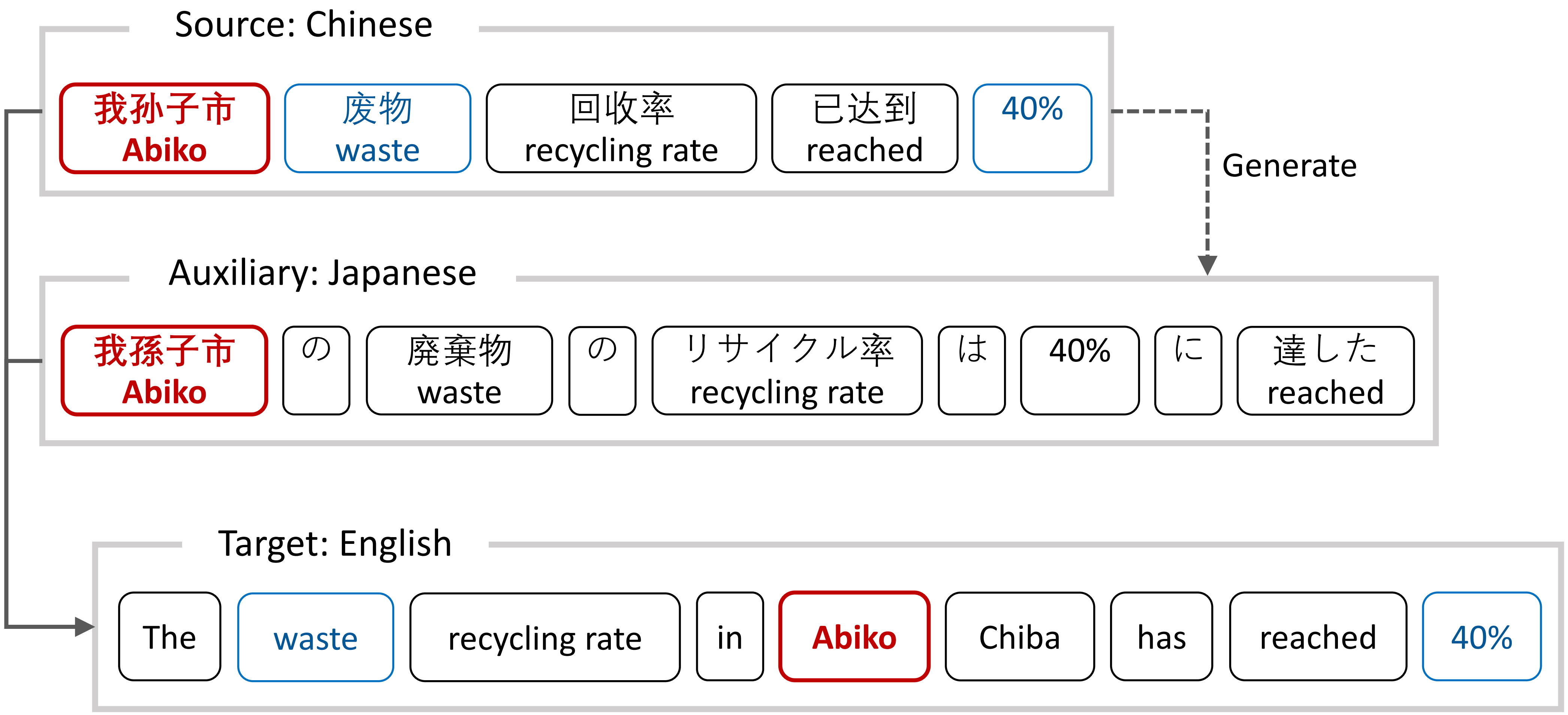


Translating from multiple source sentences brings improvements

Multi-source NMT [Och and Ney, 2001; Zoph and Knight, 2016]

- Better translation quality than single-source NMT
- Requires source sentence manually translated into all other source languages during inference

Can multilingual NMT benefit from synthetic sentences from auxiliary language?



Can multilingual NMT benefit from synthetic sentences from an auxiliary language?

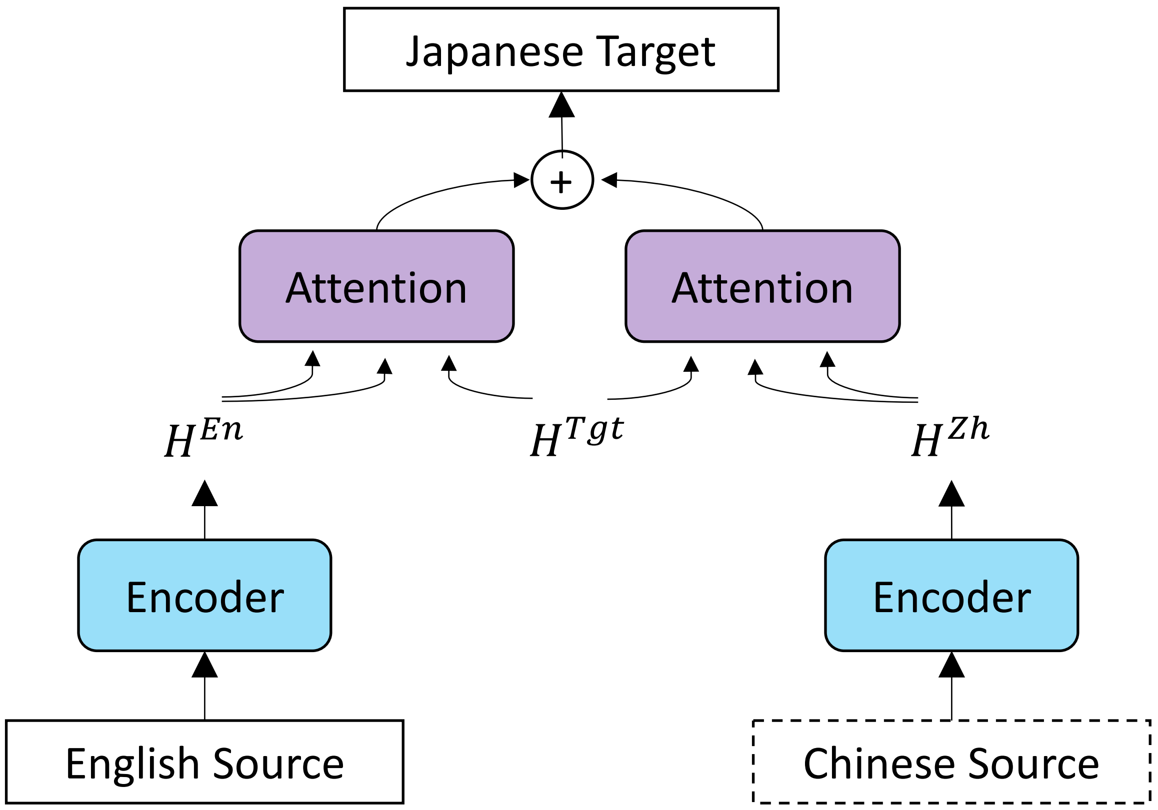
Improve multilingual NMT by incorporating synthetic sentences from an auxiliary language

How?

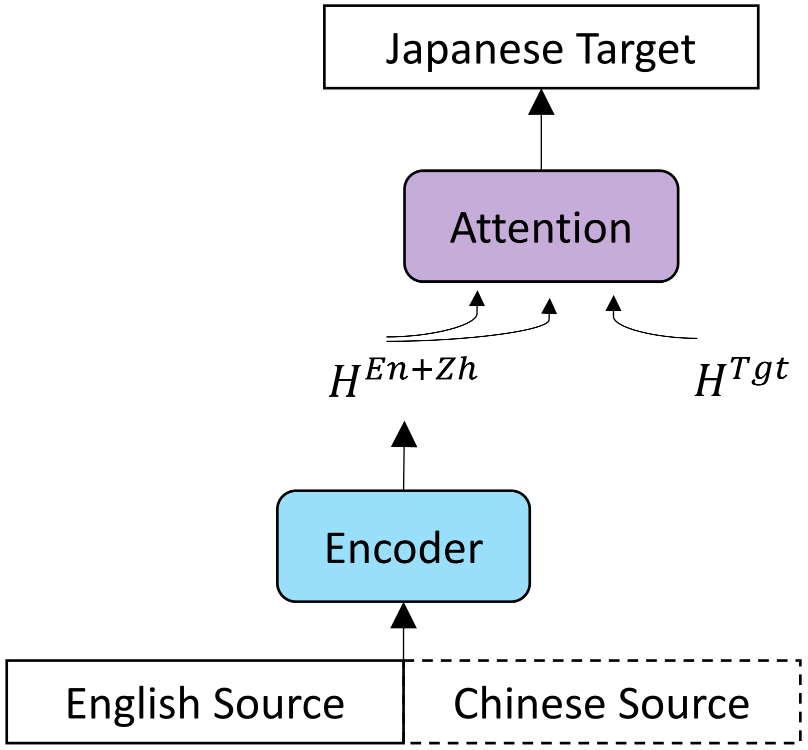
- Train a bi-source NMT model to leverage synthetic sentences from an auxiliary language
- Enable single-source and bi-source modes for flexible reference

Bi-Source Multilingual NMT: Model

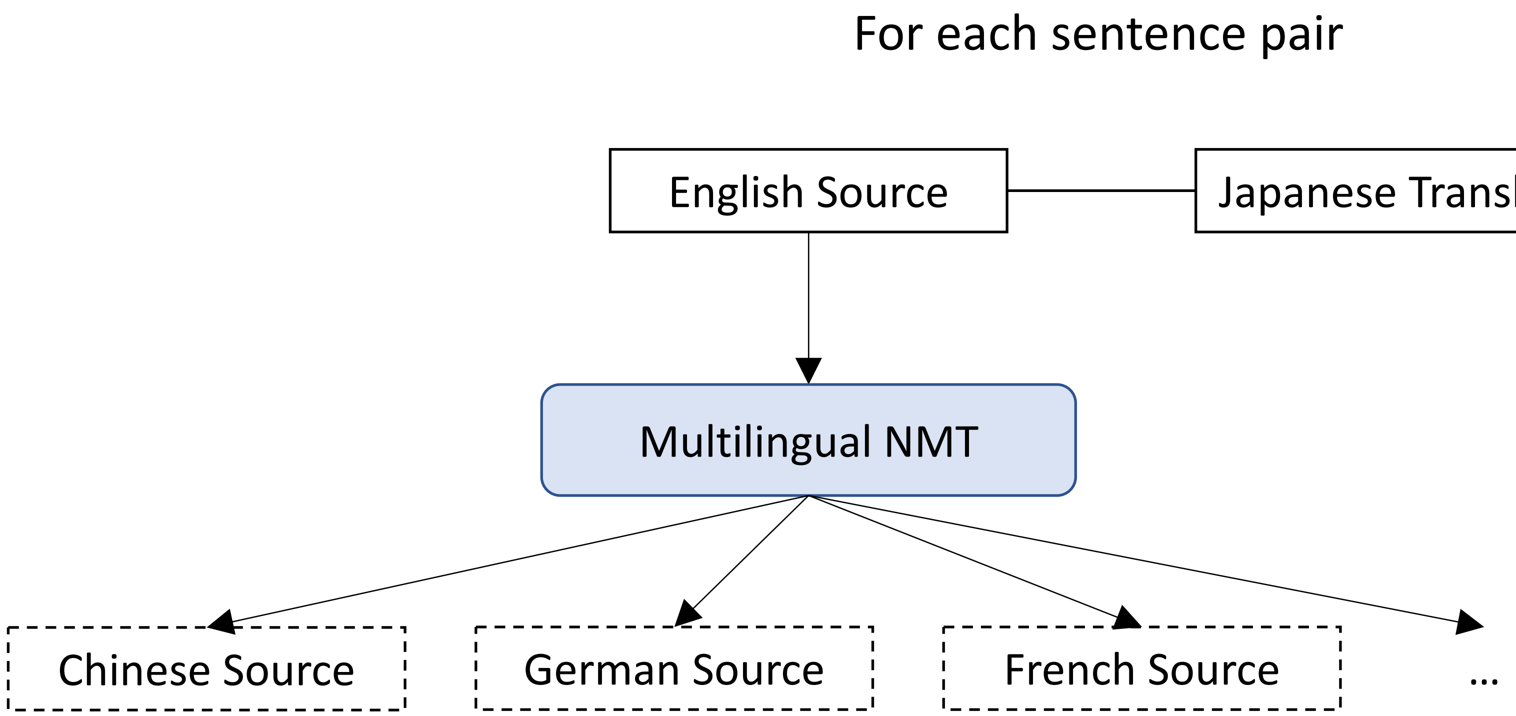
Single-Encoder



Multi-Encoder

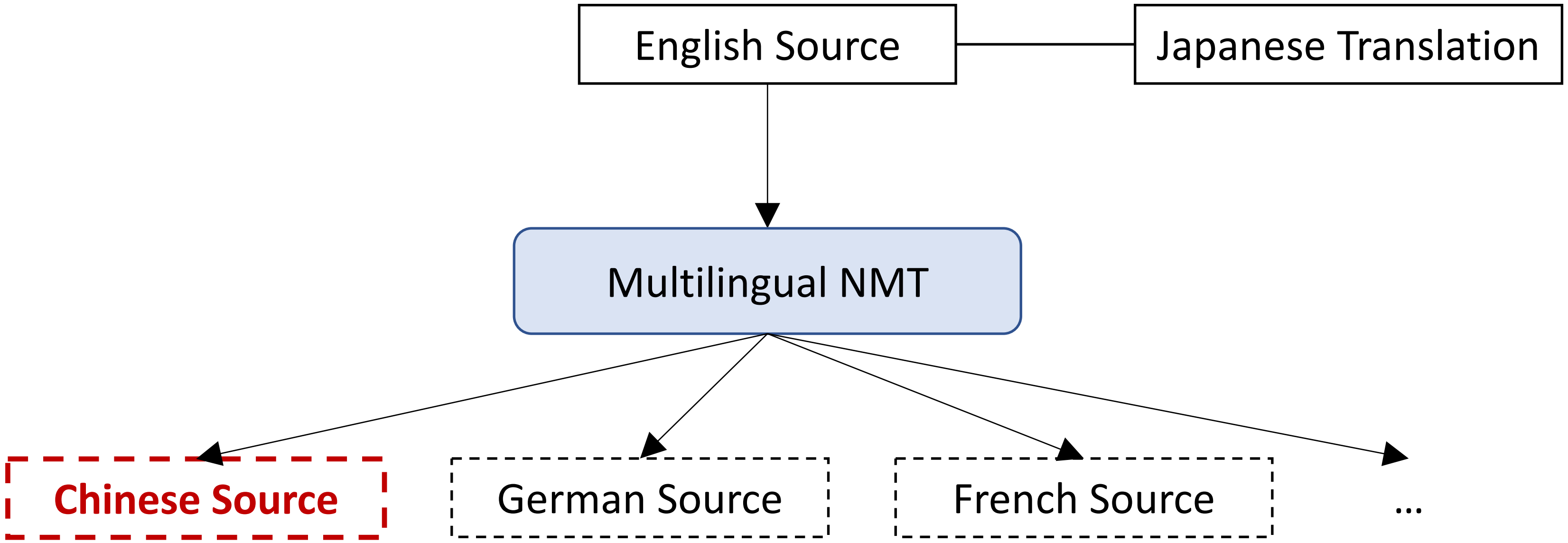


Bi-Source Multilingual NMT: Synthetic Data Generation



Bi-Source Multilingual NMT: Training

Randomly pick a triplet of source, target and auxiliary languages for each training batch



Bi-Source Multilingual NMT: Training

- Randomly pick the source, target and auxiliary languages for each training batch
- Train the bi-source model to translate from the source and auxiliary languages to the target

Bi-Source Multilingual NMT: Training

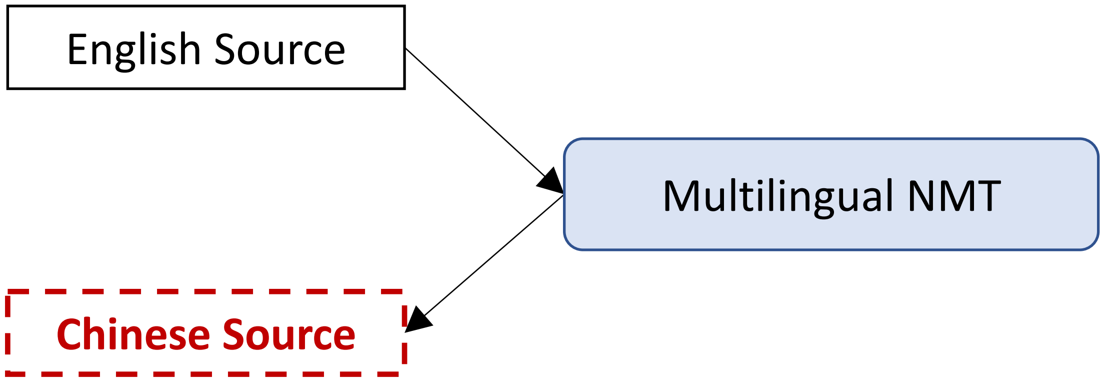
- Randomly pick the source, target and auxiliary languages for each training batch
- Train the bi-source model to translate from the source and auxiliary languages to the target
- Mask out the auxiliary sentence with probability p_{mask} to enable flexible inference

Bi-Source Multilingual NMT: Inference

- Single-source inference
- Bi-source inference

Bi-Source Multilingual NMT: Inference

- Single-source inference
- Bi-source inference
  - Translate the source to an auxiliary language



Bi-Source Multilingual NMT: Inference

- Single-source inference
- Bi-source inference
  - Translate the source to an auxiliary language
  - Translate the source and auxiliary sentences into the target language

Can multilingual NMT benefit from synthetic sentences from an auxiliary language?

Improve multilingual NMT by incorporating synthetic sentences from an auxiliary language

How?

- Train a bi-source NMT model to leverage synthetic sentences from an auxiliary language
- Enable single-source and bi-source modes for flexible reference

Experimental Settings

- Datasets
  - Chinese/English -> Japanese (in/out-of-domain performance)
  - English -> X (10 languages)
- Baselines
  - Bilingual baseline
  - Multilingual baseline

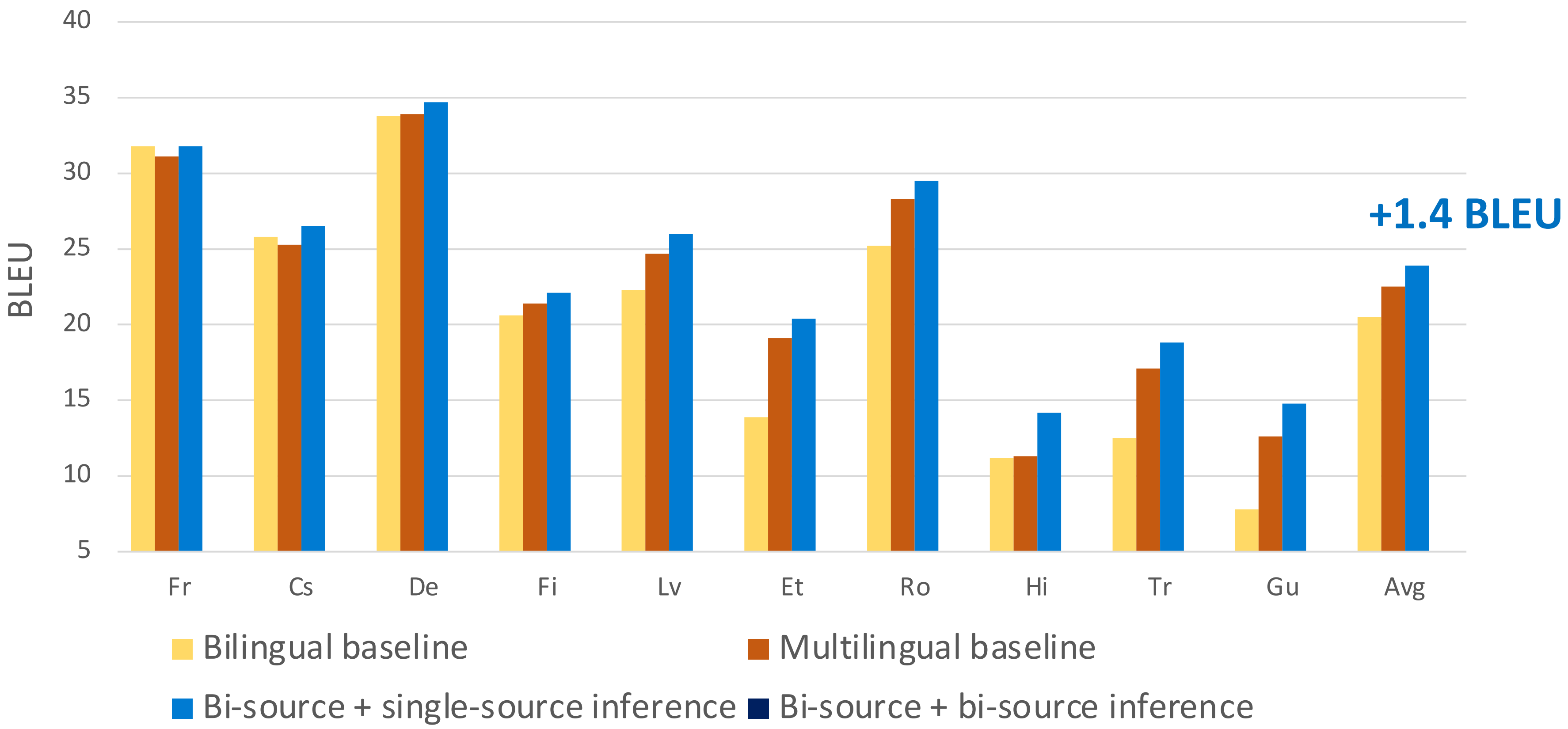
English -> X

	Train Size	Test
Fr-En	10.00M	newstest13
Cs-En	10.00M	newstest16
De-En	4.60M	newstest16
Fi-En	4.80M	newstest16
Lv-En	1.40M	newsdev17
Et-En	0.70M	newsdev18
Ro-En	0.50M	newsdev16
Hi-En	0.26M	newsdev14
Tr-En	0.18M	newstest16
Gu-En	0.08M	newsdev19

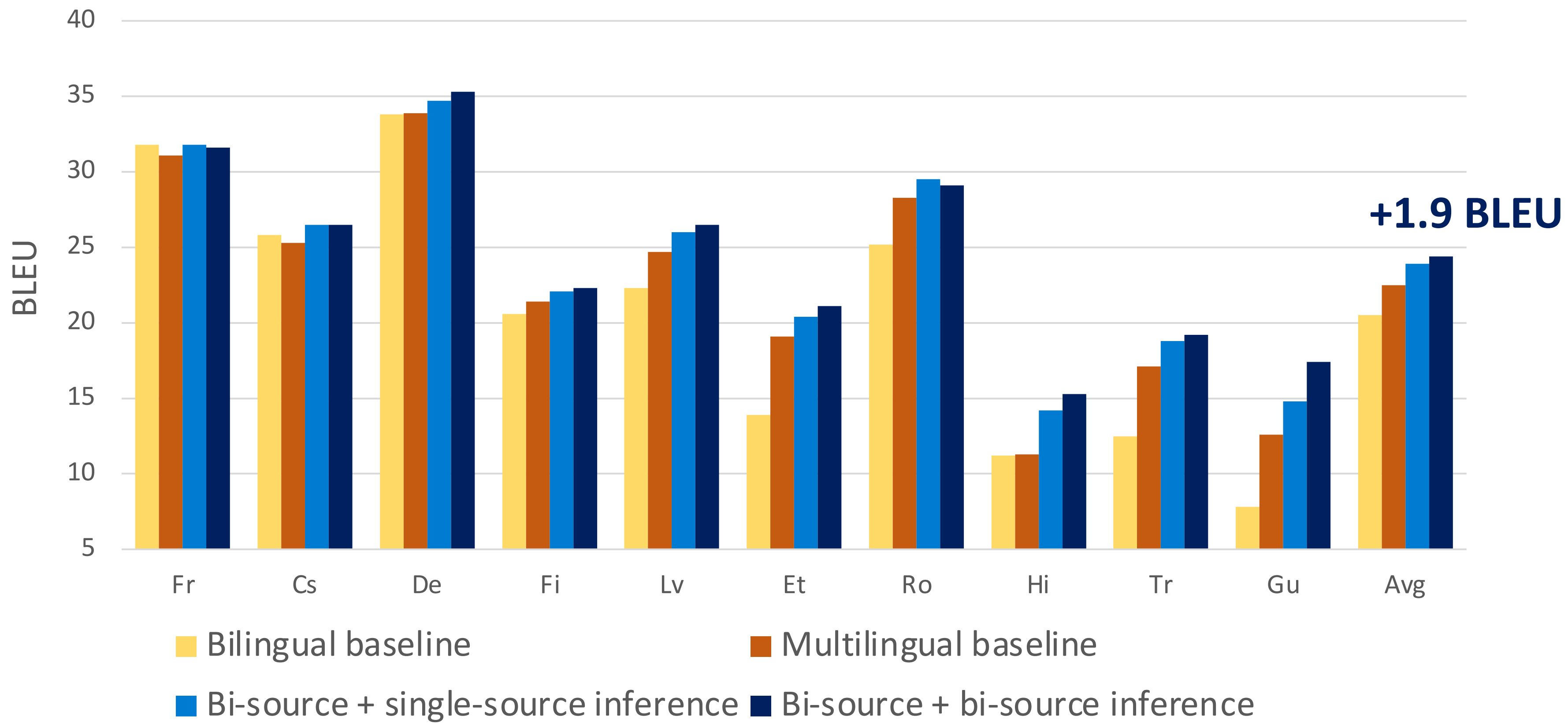
high-resource

low-resource

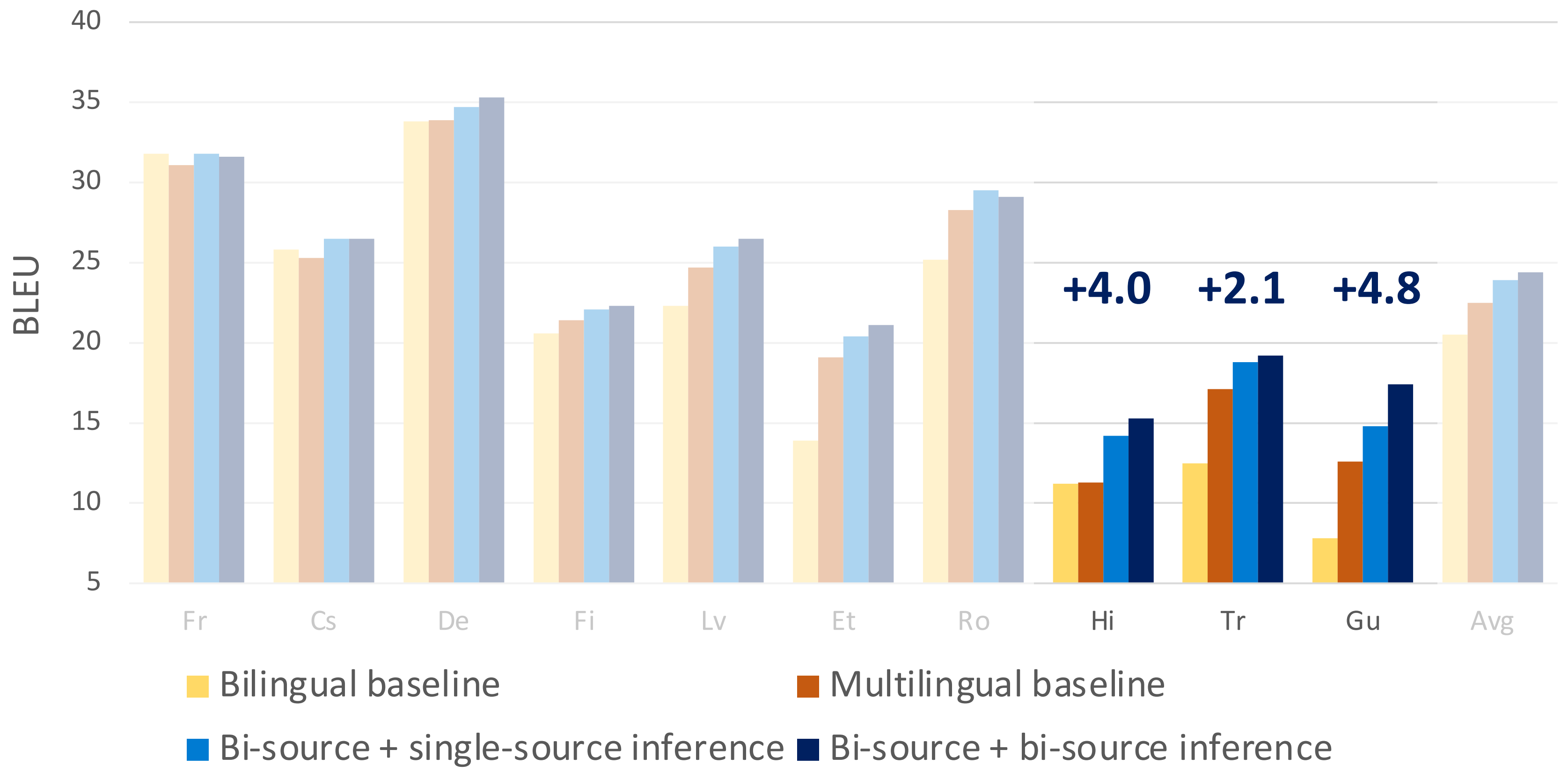
Bi-source model + single-source inference improves over baselines



Bi-source inference brings further improvements



Bi-source inference improves the most on low-resource languages



Bi-source inference helps disambiguate word senses

- Word Sense Disambiguation Test Suite (MuCoW) [Raganato et al.,2019]
- Bi-source model + single-source inference outperforms multilingual baseline
- Bi-source model + bi-source inference achieves the highest coverage score

	Model	Coverage↑
En→Cs	Multilingual baseline	56.05
	Ours (single-source)	56.27
	Ours (bi-source)	**56.96**
En→De	Multilingual baseline	57.21
	Ours (single-source)	60.44
	Ours (bi-source)	**60.61**

Improving Multilingual Neural Machine Translation with Auxiliary Source Languages

Contributions

- Bi-source multilingual NMT model that leverages a synthetic source sentence from an auxiliary language
- A novel training algorithm to enable flexible inference in single-source or bi-source mode
- Bi-source model improves over multilingual baselines especially on low-resource languages

More results and analysis in the paper

Weijia Xu, Yuwei Yin, Shuming Ma, Dongdong Zhang, and Haoyang Huang - EMNLP 2021