Rapid Adaptation of Machine Translation to New Languages

Graham Neubig, Junjie Hu

@ EMNLP 11/2/2018

Carnegie Mellon University
Language Technologies Institute

NEULAB

Inspiration: Rapid Disaster Response

Disaster in Sri Lanka

Photo Credit: Wikimedia Commons

→

නාවලපිටිය යටකළ වැසි
වතුර – #HiruNews
#StandBy

ගංවතුර හා නායයාම් තත්වයන්ගෙන්
විපතට පත් වූ පිරිස්වලට සහන
සැපයූ ස්වේච්ඡා කණ්ඩායම්
දරුවන්ට කිරිපිටි ඉල්ලා පළ කළ
සමාජ ජාල පණිවුඩ පසුගිය දිනවල
දකින්නට ලැබිණි .

How can we **effectively** and **rapidly** adapt MT to new languages?

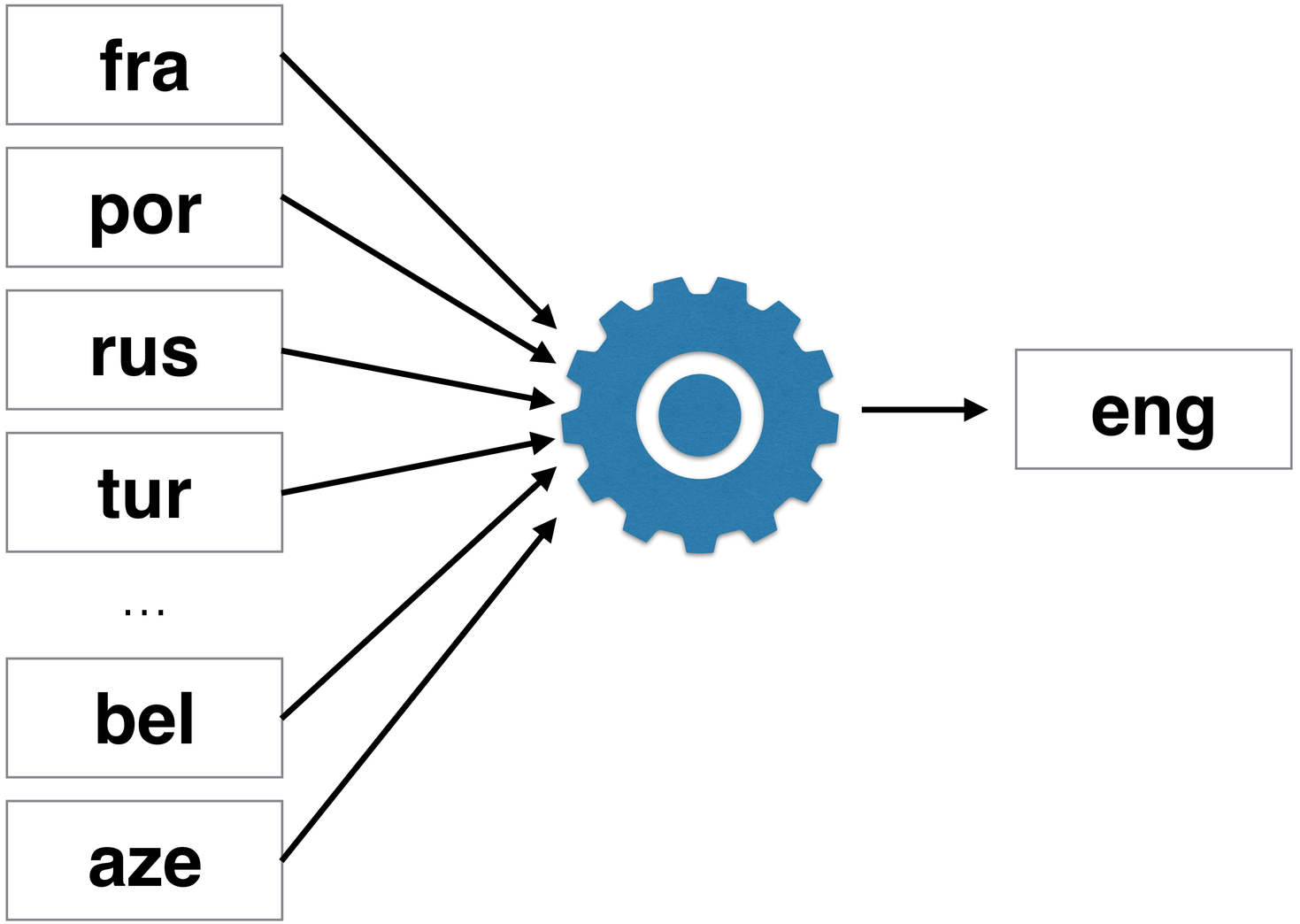
Some Crazy Ideas

- **Cross-lingual transfer:** can we create a machine translation system by transferring across language boundaries? [Zoph+16]
- **Zero-shot transfer:** can we do it with *no data* in the low-resource language?

Multi-lingual Training

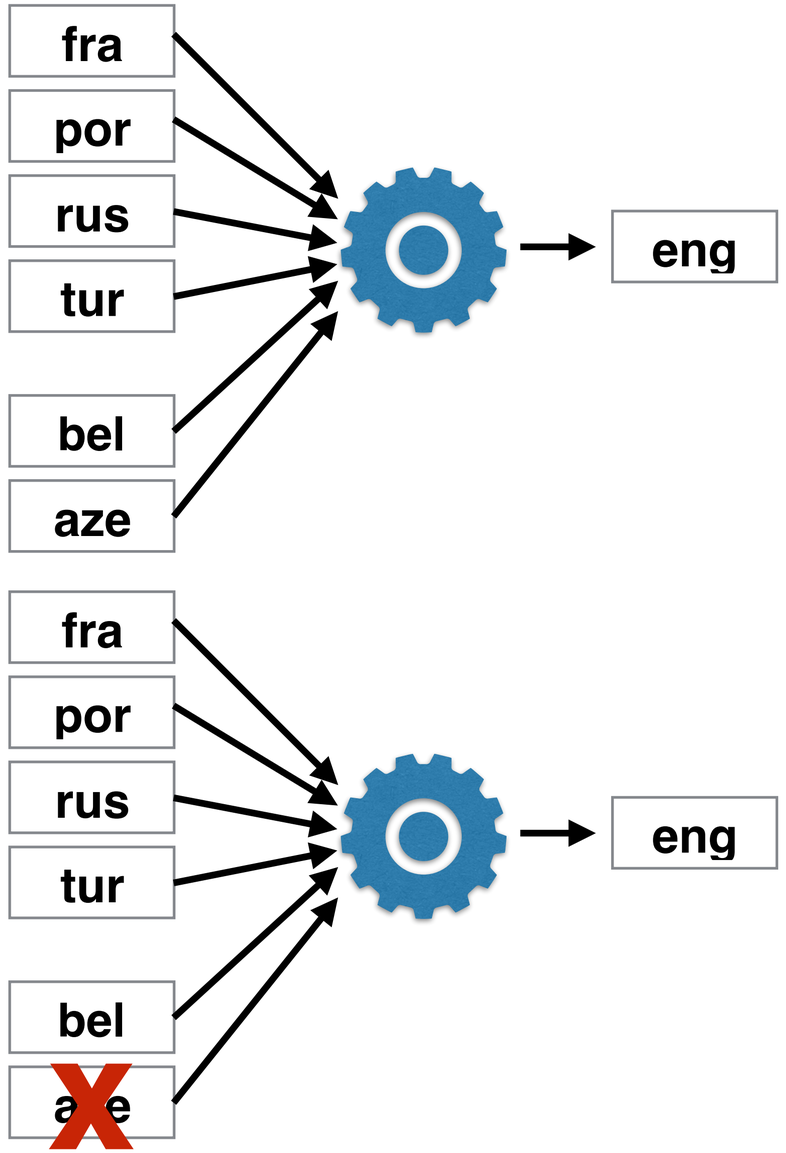
[Firat+16, Johnson+17, Ha+17]

- Train a large multi-lingual MT system, and apply it to a low-resource language



Two Multilingual Training Paradigms

- **Warm-start training:** (indicated w/ "+")
  - We already have *some* data in the test language
  - Train a model starting with that data
- **Cold-start training:** (indicated w/ "-")
  - We initially have no data in the test language
  - Possibilities for **completely unsupervised** transfer
  - Suitable for **rapid adaptation** to new languages



Experiments: Training Setting

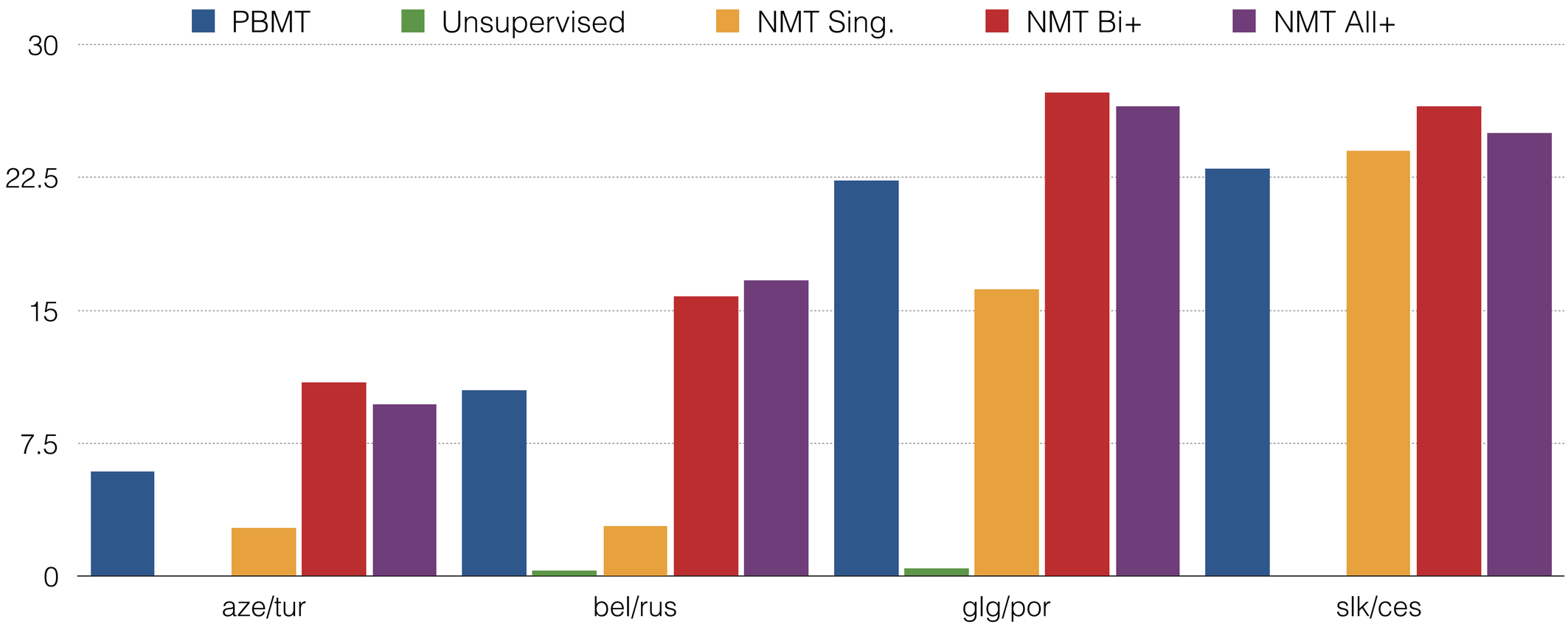
- TED multi-lingual corpus (Qi et al. 2018)
  https://github.com/neulab/word-embeddings-for-nmt
- 57 source languages, plus English
- Testbed languages: Azerbaijani (aze), Belarusian (bel), Galician (glg), Slovak (slk)
- Related languages: Turkish (tur), Russian (rus), Portuguese (por), Czech (ces)

LRL	train	dev	test	HRL	train
aze	5.94k	671	903	tur	182k
bel	4.51k	248	664	rus	208k
glg	10.0k	682	1,007	por	185k
slk	61.5k	2,271	2,445	ces	103k

Systems

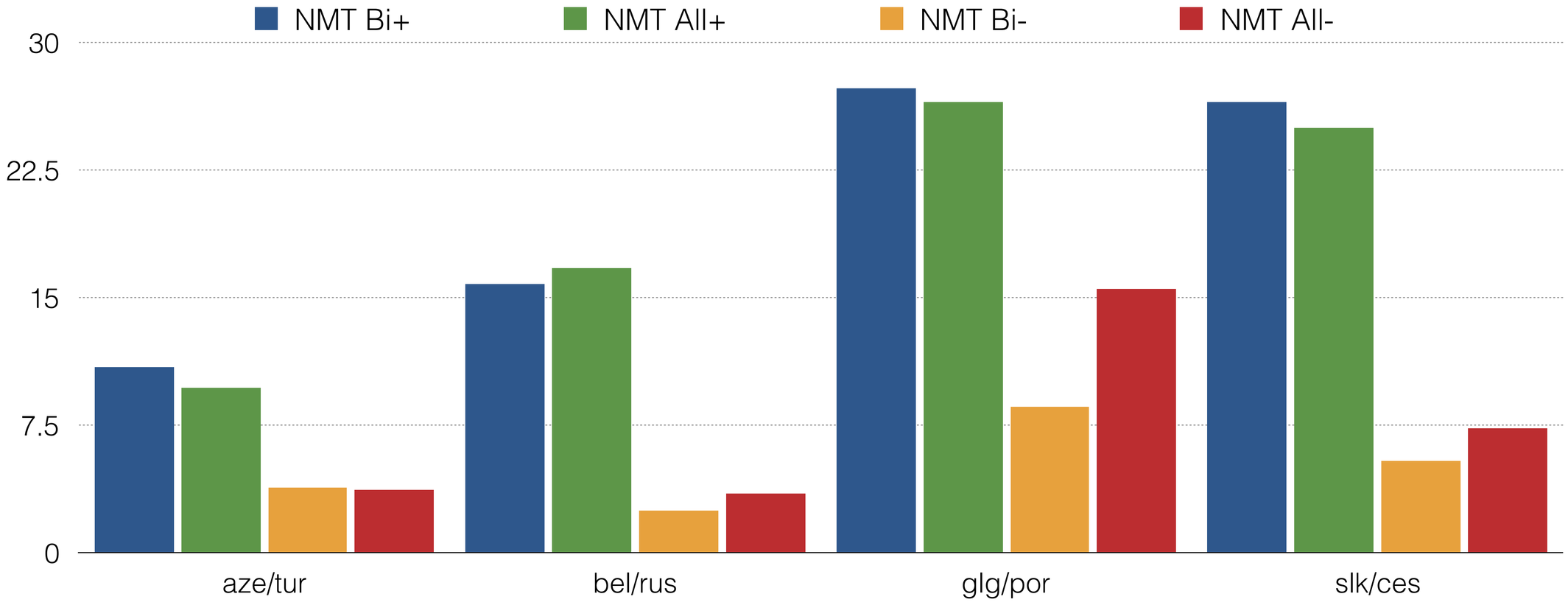
- Test Systems:
  - **Single-source Neural MT (Sing.):** Test source language only
  - **Bi-source Neural MT (Bi.):** Test source language and related source
  - **All-source Neural MT (All):** All source languages
- Other Baselines:
  - **Phrase-based MT:** Shown to be strong in low-resource settings
  - **Unsupervised MT [Artetxe+17]:** Learn system using only monolingual data in source/target languages (cited as effective in low-resource settings)

How does Cross-lingual Transfer Help?



- Unsupervised translation not competitive
- Without transfer, NMT worse than PBMT
- With transfer NMT significantly better (transfer barely helped PBMT)

How Does Cold-start Compare?

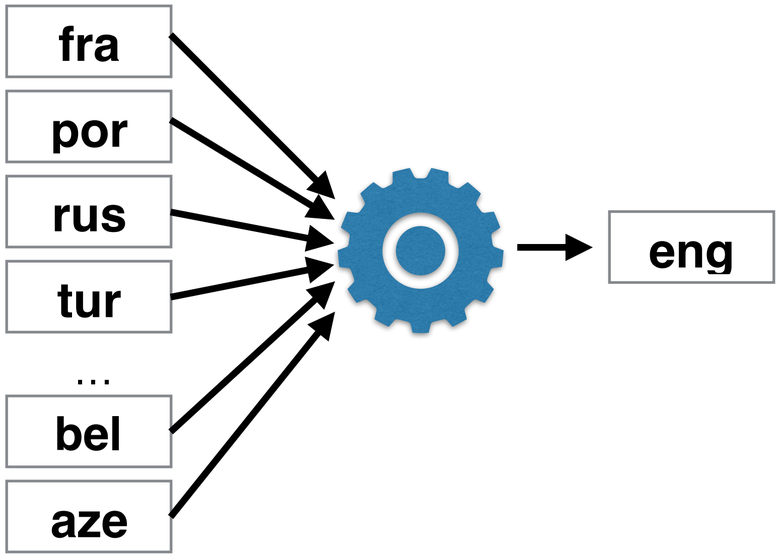


- Large drop, but still much better than nothing
- Up to 15 BLEU with *no* training data in test language

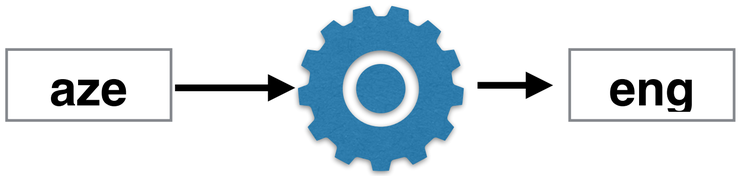
Adaptation to New Languages

- Training on all languages can be less effective, esp. in cold-start case
- Can we further adapt to new languages?
- Problem: overfitting

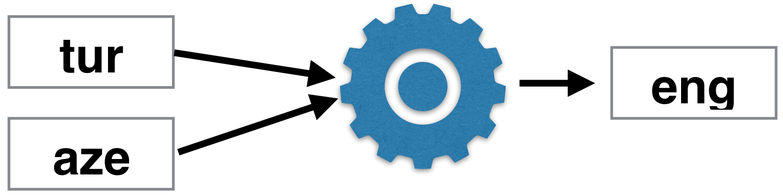
Pre-training



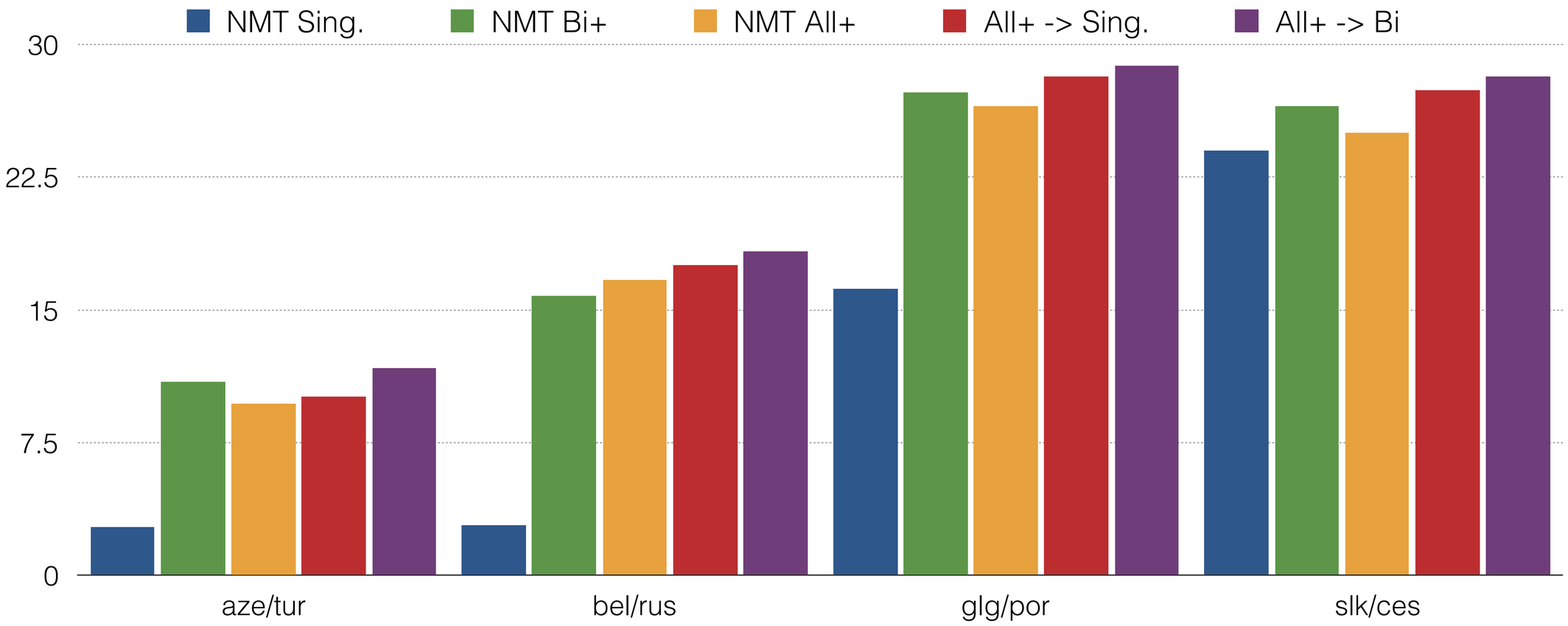
Adaptation
(All→Sing.)



Adaptation w/ Similar Language Regularization
(All→Bi.)

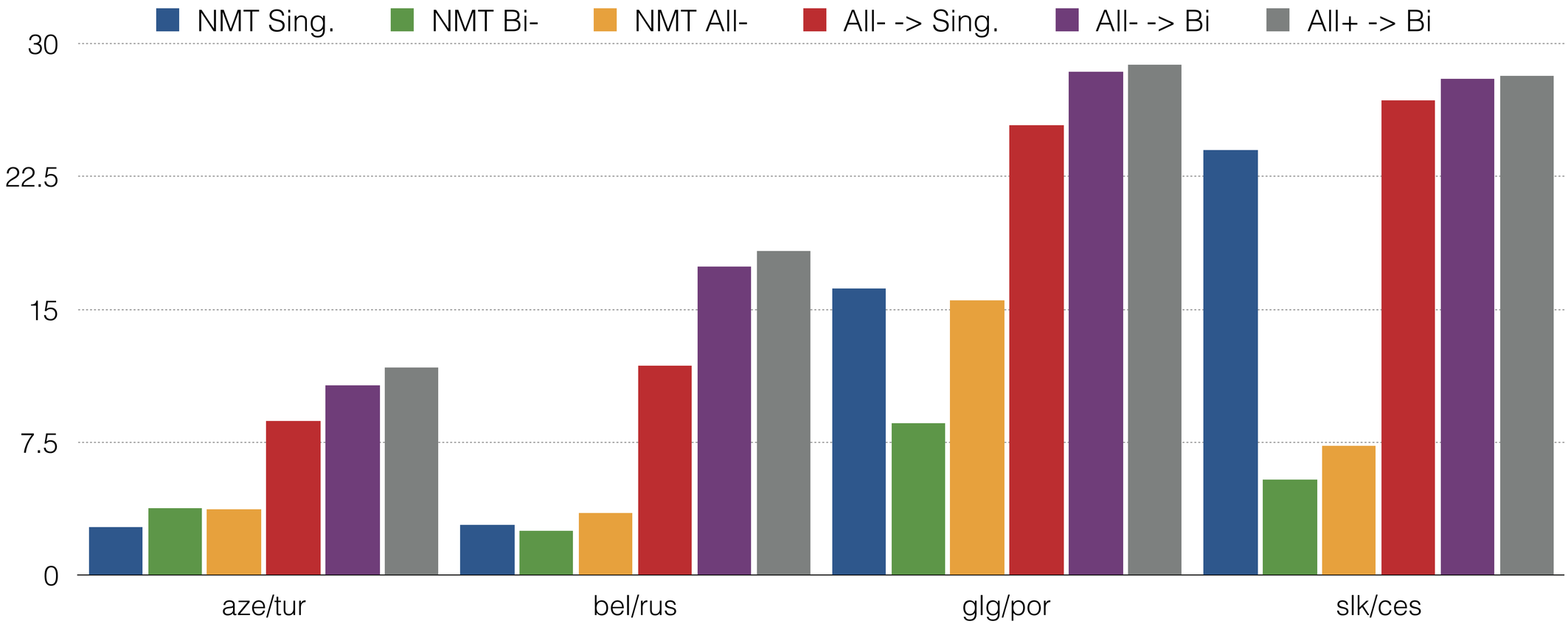


Warm-start + Adaptation



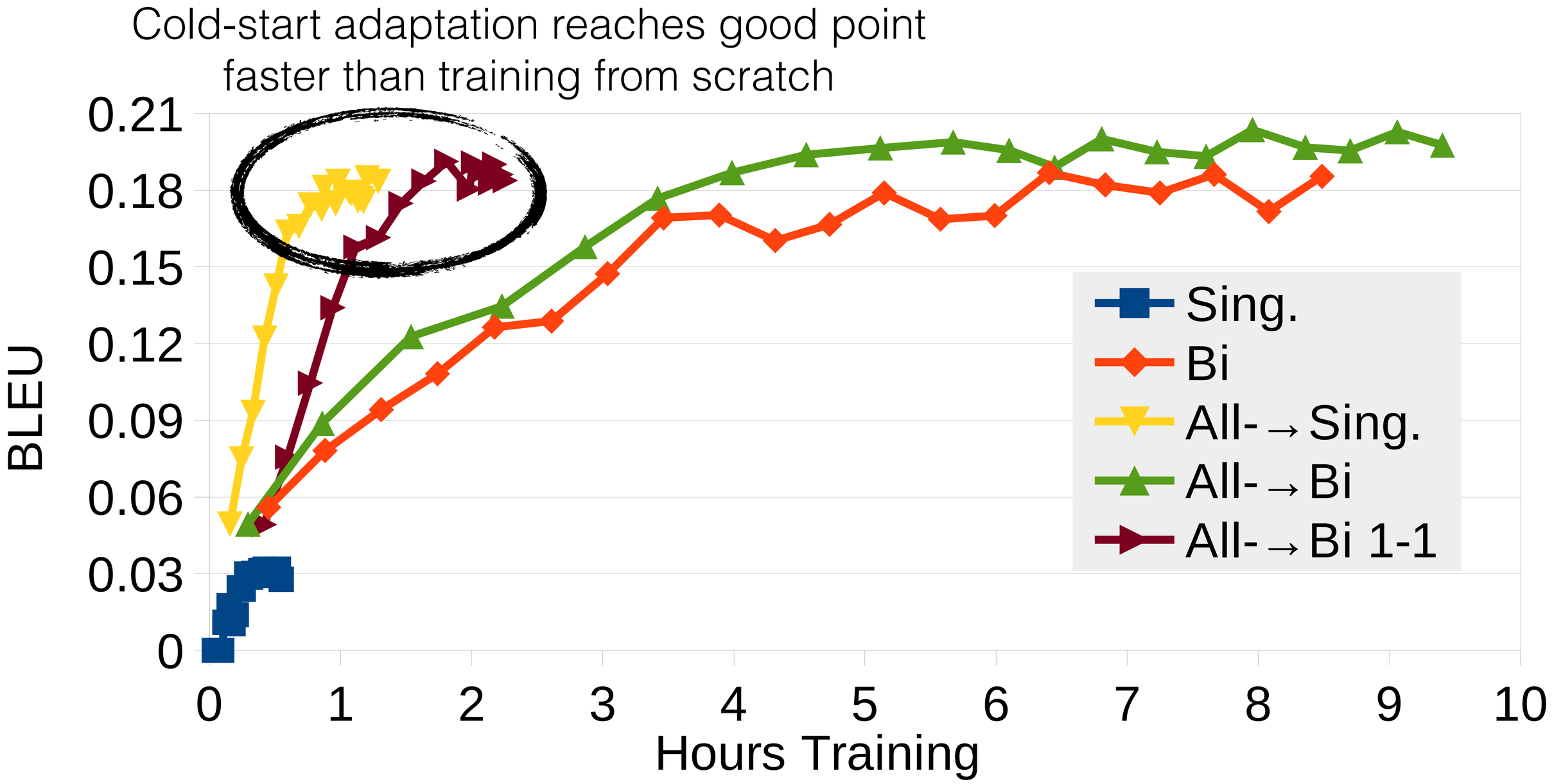
- Adaptation helps!
- Helps more w/ similar language regularization

Cold-start + Adaptation



- Adaptation w/ similar-language regularization gains more
- Approaches quality of warm-start; doesn't need data a-priori

How Fast can we Adapt?



Take-aways

- NMT with **massively multi-lingual cross-lingual transfer**: a stable recipe for low-resource translation
- **Better results than phrase-based, unsupervised MT** in real low-resource languages
- Adaptation w/ **similar language regularization**: simple and effective, even in cold-start scenarios

https://github.com/neubig/rapid-adaptation

Questions?