

## 1. SUMMARY

### Motivation

- Improving translation from English to low resource Asian languages
- Focus on Languages from the Asian Language Treebank (ALT): Bengali, Filipino, Indonesian, Japanese, Khmer, Malay, Vietnamese

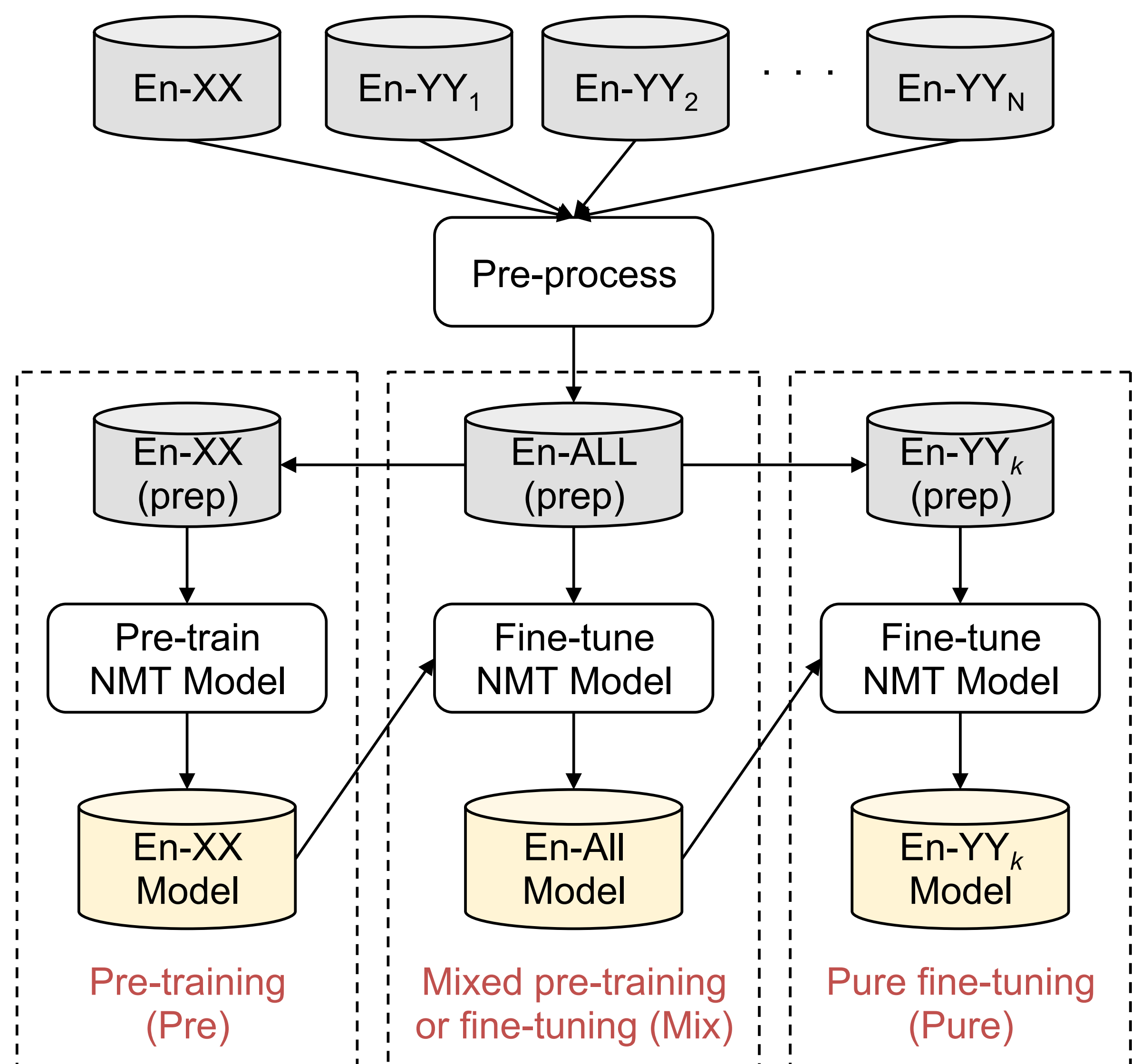
### Constraints

- Language pair specific training data is extremely small
- Extremely small data is N-way multilingual (content redundant)
- Availability of a helping corpus with an unrelated target language

### Contributions

- Multi-stage training involving strongly trained multilingual models
- Stronger pre-training via multilingualism through N-way corpora
- Robust transfer learning giving 3–9 BLEU improvements

## 3. MULTILINGUAL MULTISTAGE FINE-TUNING



### Expectation

- Multilingualism for strongly pre-trained intermediate models
- Multistage fine-tuning for efficient learning via division of labor

## 5. EXPERIMENTAL DETAILS

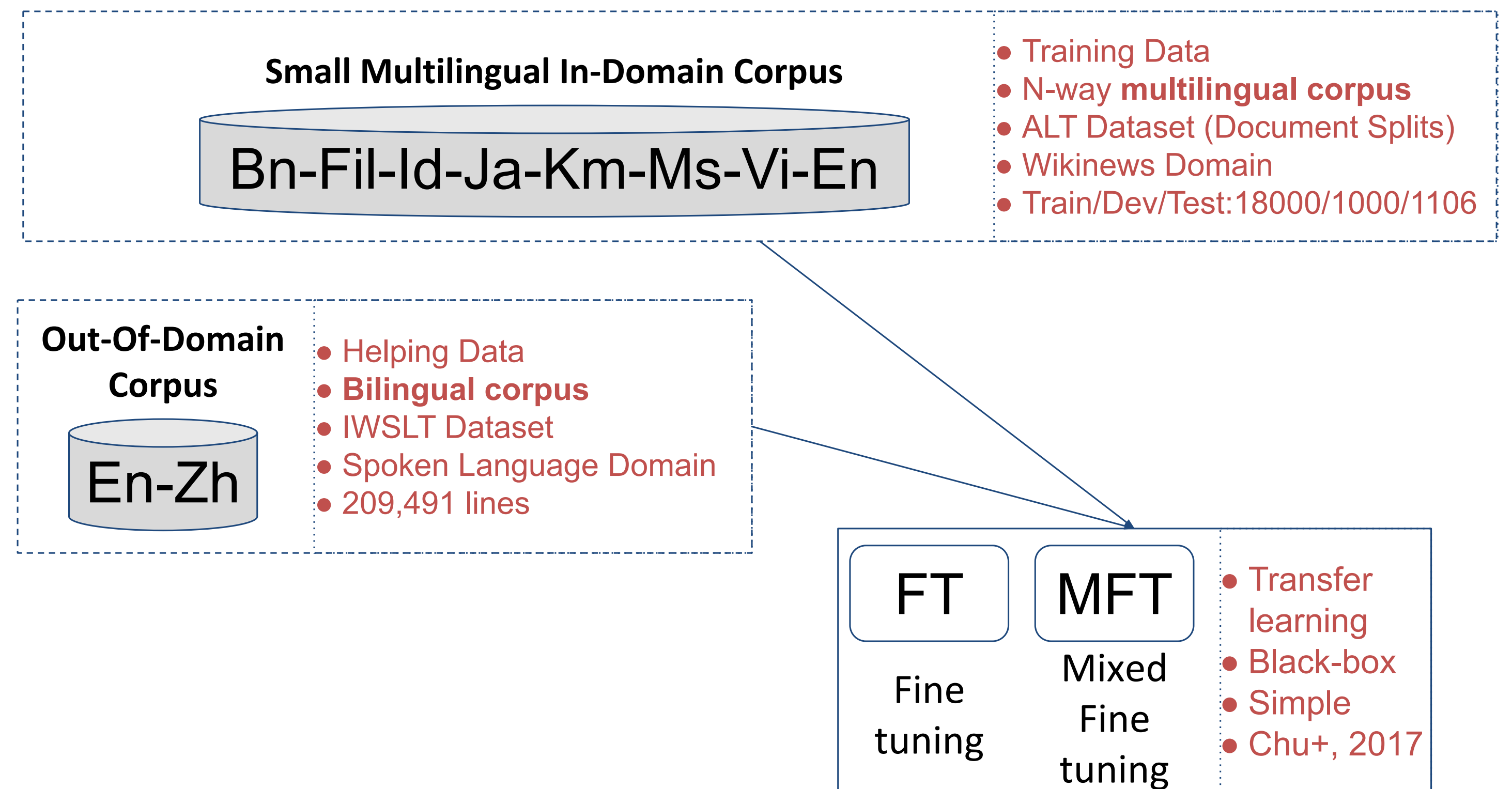
### Pre-processing

- ALT data was used in pre-tokenized form
- Chinese and Japanese segmented with KyotoMorph and JUMAN
- Source sentences pre-pended with “XX” representing target language
- Oversampled smaller corpus to match larger corpus size
- Used internal sub-word segmentation of tensor2tensor

### Settings

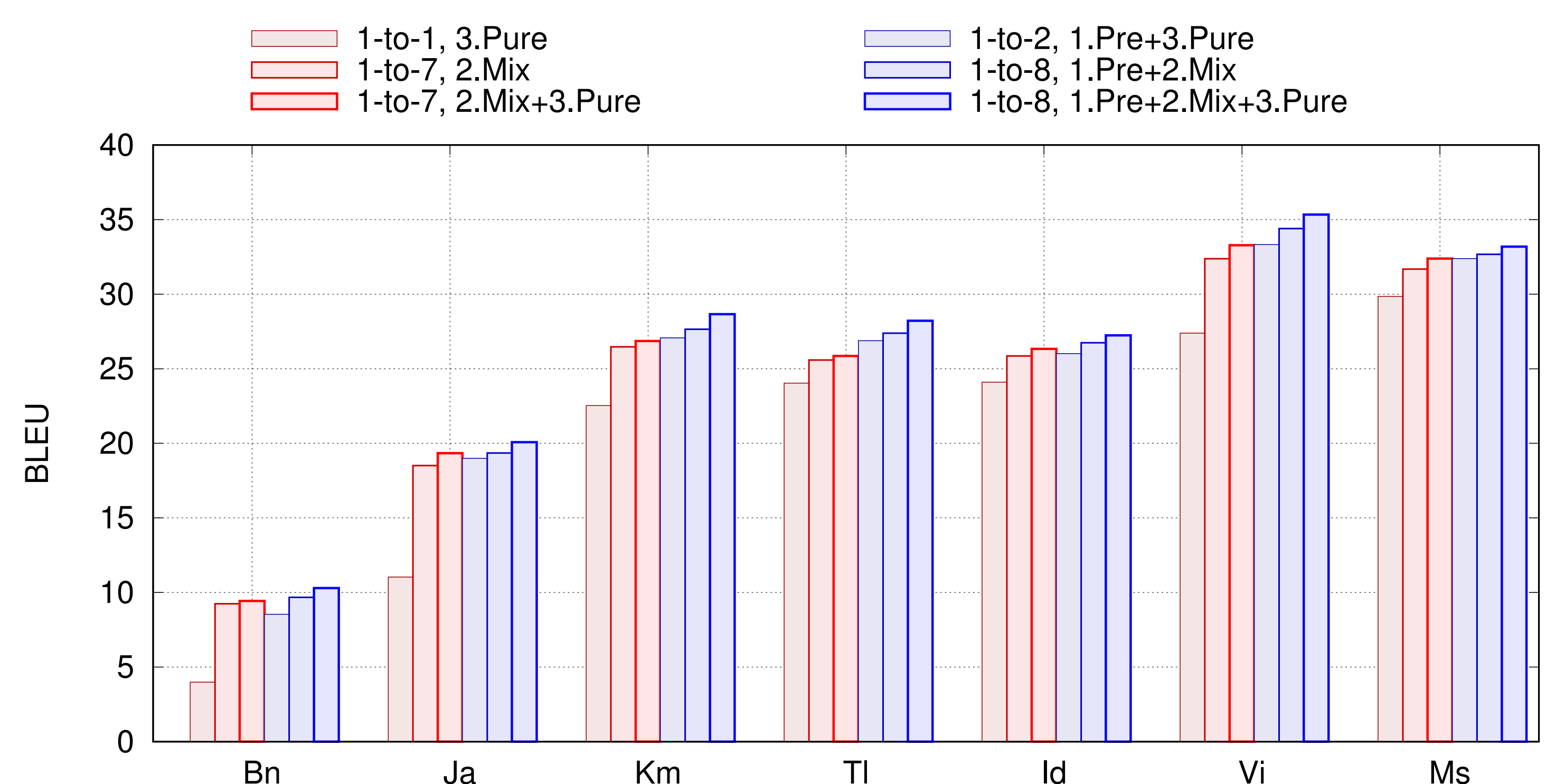
- All implementations done using tensor2tensor v1.6
- Vocabulary sizes chosen according to dataset size (see paper)
- Training and decoding (“transformer\_base”)
  - Trained till convergence on development data
  - Averaged the last 10 checkpoints for decoding
  - Beam size and length penalty of 4 and 0.6, respectively

## 2. KEY ASPECTS OF SETTING



- Target language mismatch between helping and training corpora
- Reliance on off-the-shelf techniques

## 4. EXPERIMENTAL RESULTS



### Observations

- Improvements despite different target languages between datasets
- The size of the helping corpus has a significant impact
- Multilingual models are better for fine-tuning
  - Multilingual MFT gives robust pre-trained models
  - N-way corpora help despite content redundancy

## 6. CONCLUSION AND FUTURE WORK

### Conclusion

- Robust multilingual models are useful for transfer learning
  - Multistage fine-tuning is better than direct fine-tuning
  - Division of labor leading to improved final translation quality
- Strong evidence for the importance of multilingualism
  - N-way training data is content redundant
  - External corpus is useful despite mismatched target language
- Improvements of 3–9 BLEU for all low resource languages

### Future Work

- Experiment with several helping corpora
- Study impact of domain and helping corpora sizes on final translation
- Apply the approach to many-to-many translation