# Efficient Unseen Language Adaptation for Multilingual Pre-Trained Language Models

Po-Heng Chen and Yun-Nung (Vivian) Chen

National Taiwan University 國立臺灣大學

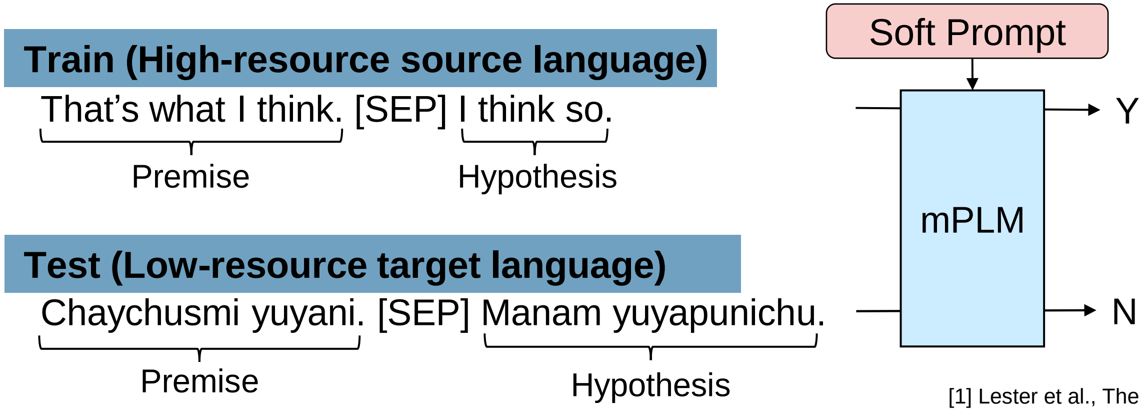
MIULAB

# Introduction

- Motivation:
  - Apply machine learning application to **low-resource** language
- Issues
  - The lack of labeled data in low-resource language
    - Zero-shot cross-lingual transfer is needed
  - Low-resource languages are **unseen** by most mPLMs
    - Language adaptation is needed
  - The corpora of low-resource languages are small and mPLMs are large
    - A data efficient and parameter efficient method is needed

# Introduction

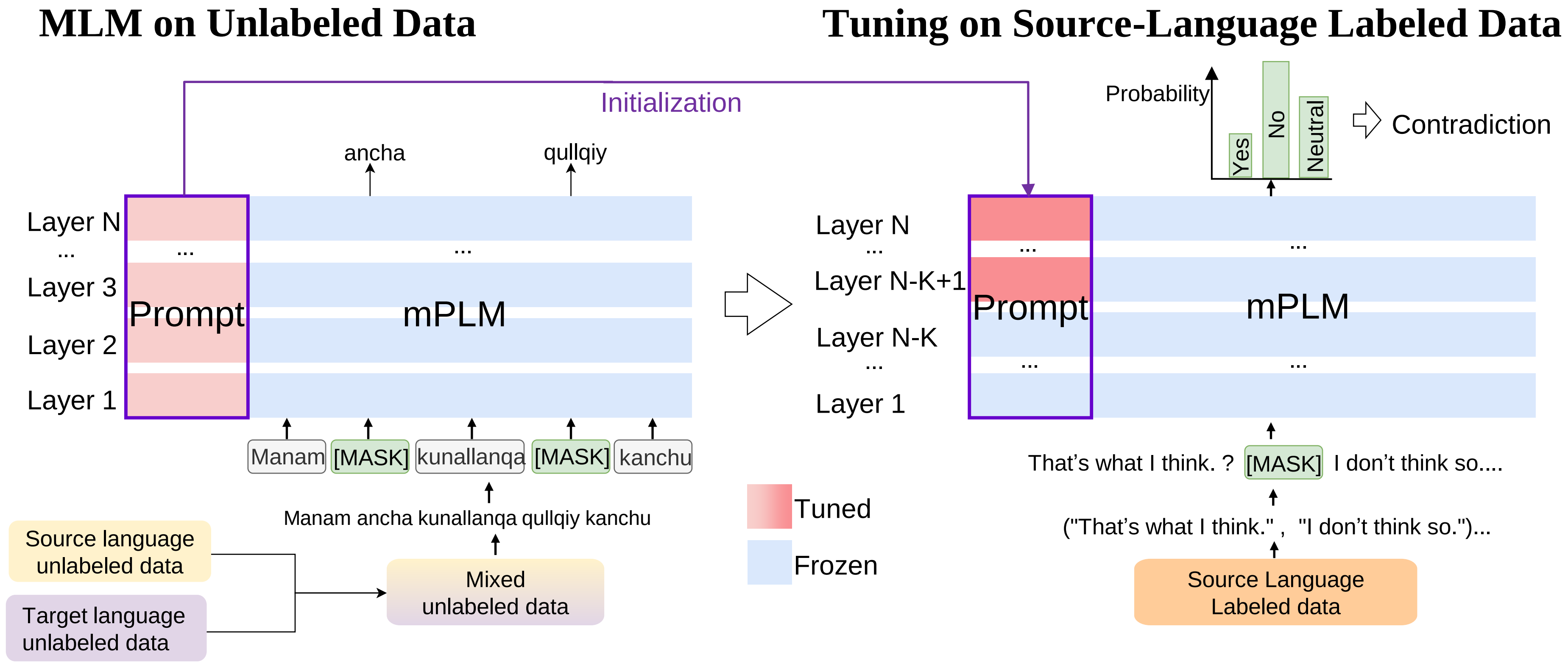
- Idea:
  - Design **parameter-efficient** and **data-efficient** framework to **adapt** mPLMs to **unseen** languages and achieve strong cross-lingual transfer performance
  - In this work, we employ **prompt-tuning**[1] [2] to achieve our goal




[1] Lester et al., The Power of Scale for Parameter-Efficient Prompt Tuning, EMNLP 2021
[2] Li et al., Prefix-Tuning: Optimizing Continuous Prompts for Generation, ACL 2021

4

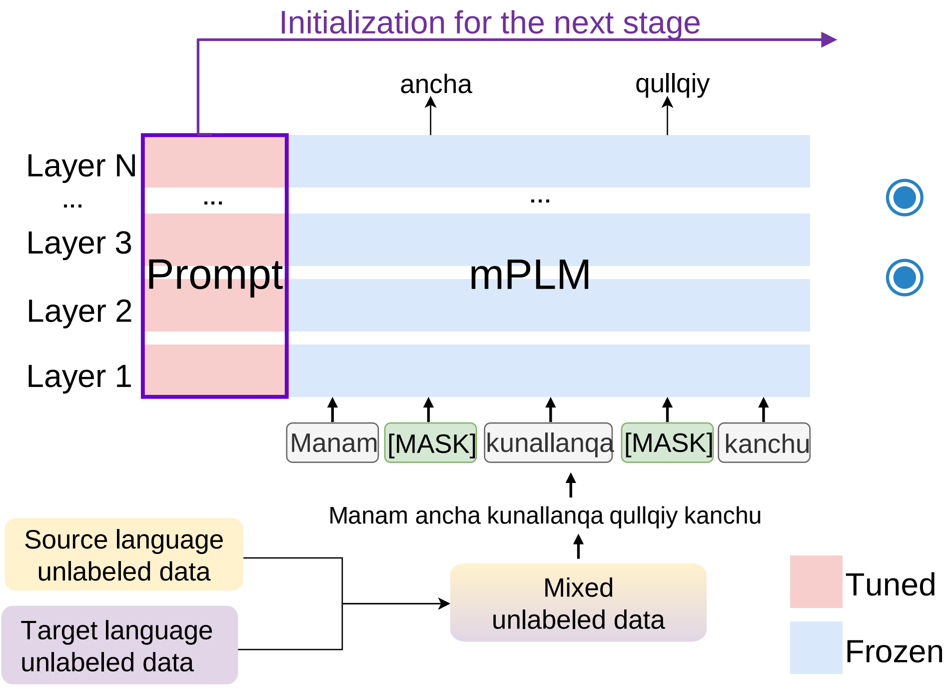
# Framework: Soft-Prompt Language Adaptation



5

# Framework: Soft-Prompt Language Adaptation

## MLM on Unlabeled Data

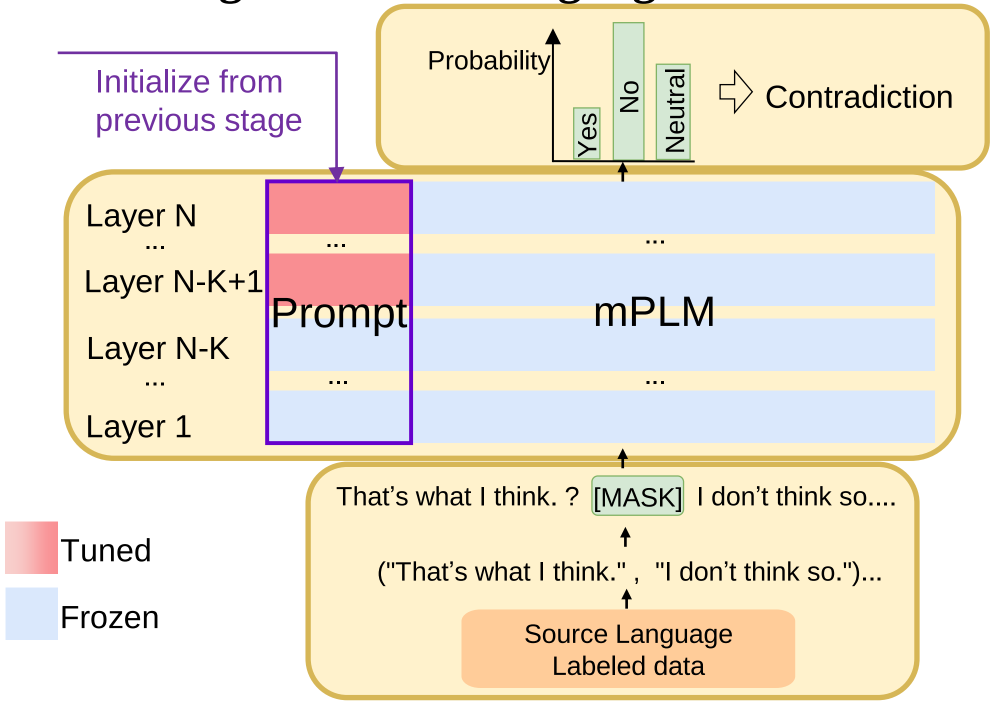



- Mix same amount of source and target language text data
  - Avoid being biased to one of them
- Use MLM as objective
- The target-language adapted soft prompt is then used for downstream tasks

# Framework: Soft-Prompt Language Adaptation

- Init soft-prompt from previous stage
- Only tune the upper K layers of soft-prompt
  - The upper layers are more task-focused and language-independent
- Apply template and verbalizer
  - Transform input into MLM question

**Tuning on Source-Language Labeled Data**



# Datasets

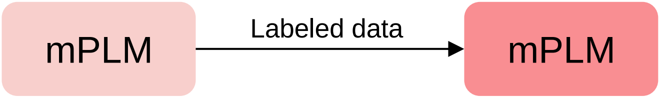
- MasakhaNEWS *(Adelani et al., 2023)* *[3]*
  - Contain 8 African languages unseen by the mPLM (XLM-R)
  - Focus on news topic classification

- AmericansNLI *(Ebrahimi et al., 2022)* *[4]*
  - Contain 10 Indigenous languages unseen by the mPLM (XLM-R)
  - Extended from XNLI

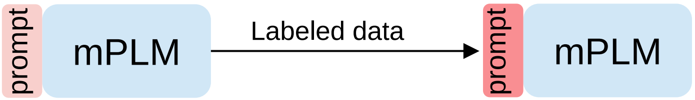
[3] Adelani et al., MasakhaNEWS: News Topic Classification for African languages, IJCNLP-AACL 2023
[4] Ebrahimi et alAmericasNLI: Evaluating Zero-shot Natural Language Understanding of Pretrained Multilingual Models in Truly Low-resource Languages, ACL 2022

# Baselines

- Without language adaptation:
  - Fine-tuning without adaptation

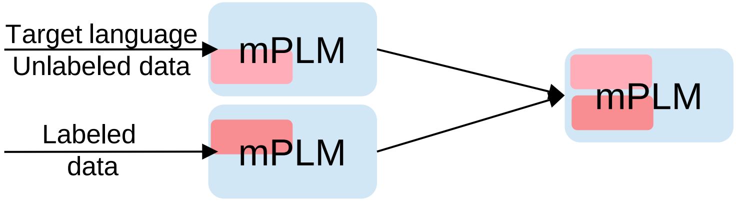



  - Soft-prompt tuning without adaptation




- With language adaptation:
  - Fine-tuning with adaptation


  - MAD-X[5]




[5] Pfeiffer et al., MAD-X: An Adapter-Based Framework for Multi-Task Cross-Lingual Transfer, EMNLP 2020

# Parameter and Storage Efficiency

| Methods | Trainable Parameter | Checkpoint size |
|---|---|---|
| Fine-tuning | 816M | 2.24GB |
| MAD-X | 27M | 103MB |
| Ours | **1.57M** | **6.2MB** |

5.88%

0.28%

# Results - MasakhaNEWS

***Without** target language adaptation*

| Model | ibo | lin | lug | pcm | run | sna | tir | yor | Avg. |
|---|---|---|---|---|---|---|---|---|---|
| Fine-tuning | 67.95 | 74.86 | 60.54 | 93.11 | 69.25 | 58.27 | 66.54 | 67.64 | 69.77 |
| Prompt-tuning | 64.36 | 66.86 | 43.50 | 91.15 | 63.35 | 49.32 | 54.41 | 67.40 | 62.54 |

***With** target language adaptation*

| Model | ibo | lin | lug | pcm | run | sna | tir | yor | Avg. |
|---|---|---|---|---|---|---|---|---|---|
| Fine-tuning | 81.03 | **85.71** | 61.43 | **95.41** | **85.71** | 81.03 | 72.43 | 83.45 | 80.78 |
| MAD-X | 78.97 | 78.86 | 56.05 | 86.23 | 76.71 | 73.98 | **73.16** | 77.86 | 75.23 |
| **Ours** | **81.62** | 82.48 | **71.15** | 91.59 | 85.17 | **86.68** | 72.79 | **85.32** | **82.07** |

# Results - AmericasNLI

## ***Without*** *target language adaptation*

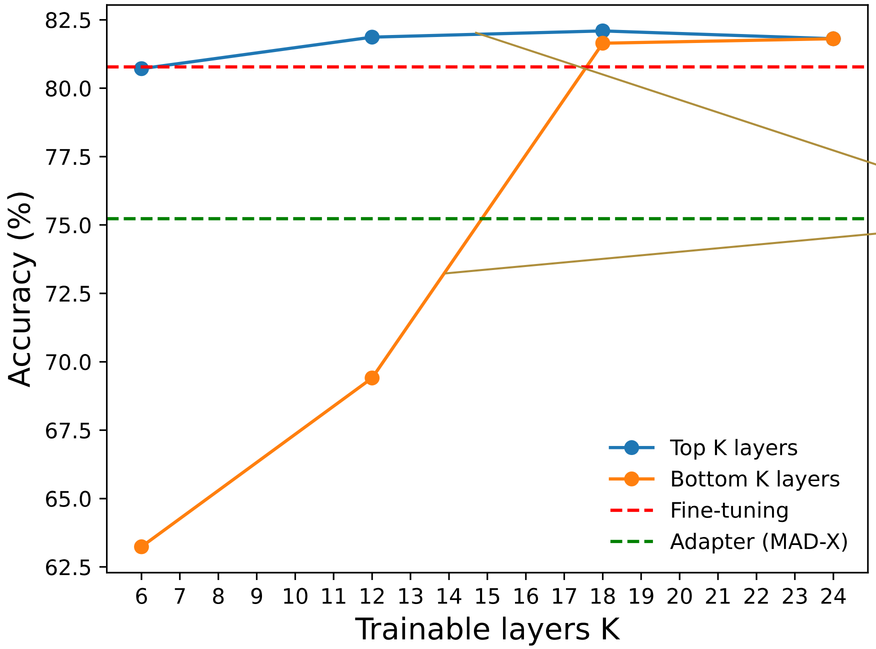
| Model | aym | bzd | cni | gn | hch | nah | oto | quy | shp | tar | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Fine-tuning | 40.67 | 41.33 | 43.07 | 42.93 | 39.20 | 45.39 | 42.25 | 42.13 | 48.27 | 40.53 | 42.58 |
| Prompt-tuning | 42.13 | 41.47 | 44.67 | 44.53 | 39.07 | 45.93 | 43.45 | 44.40 | 48.00 | 39.86 | 43.35 |

## ***With*** *target language adaptation*

| Model | aym | bzd | cni | gn | hch | nah | oto | quy | shp | tar | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Fine-tuning | 48.00 | 44.80 | **44.93** | 56.00 | **42.40** | 47.70 | 42.51 | 49.73 | **46.40** | 42.67 | 46.51 |
| MAD-X | 60.93 | **46.00** | 41.73 | **62.27** | 37.33 | 47.29 | 42.25 | **65.73** | 46.13 | **43.20** | **49.29** |
| **Ours** | 59.51 | 42.84 | 44.04 | 60.31 | 40.71 | **47.97** | **43.09** | 63.60 | 44.67 | 39.15 | 48.59 |

# Trainable Soft-Prompt Layers for Downstream Task

- The average performance(MasakhaNEWS) with varying trainable layers on source-language labeled data




Top K consistently has better results than bottom K

# Conclusion

- First to extend the generalization of mPLMs to unseen languages using only **soft-prompt tuning**
- We demonstrate the efficiency of soft-prompt language adaptation which
  - Outperforms fine-tuning in zero-shot cross-lingual transfer with only **0.28%** of parameter tunable
  - Comparable to MAD-X while utilizing **17 times** fewer parameters

https://github.com/MiuLab/UnseenAdapt