Incorporating Lexical and Syntactic Knowledge for Unsupervised Cross-Lingual Transfer

Jianyu Zheng, Fengfei Fan and Jianquan Li

zheng_jianyu@126.com {fanfengfei, lijianquan2}@ultrapower.com.cn

Presenter: Jianyu Zheng

Introduction

- Unsupervised cross-lingual transfer refers to the process of leveraging knowledge from one language, and applying it to another language without explicit supervision (Conneau et al., 2019)

- Many studies try to improve performance by focusing on cross-lingual knowledge, particularly lexical and syntactic knowledge

- Current approaches are limited as they only incorporate syntactic or lexical information

Introduction

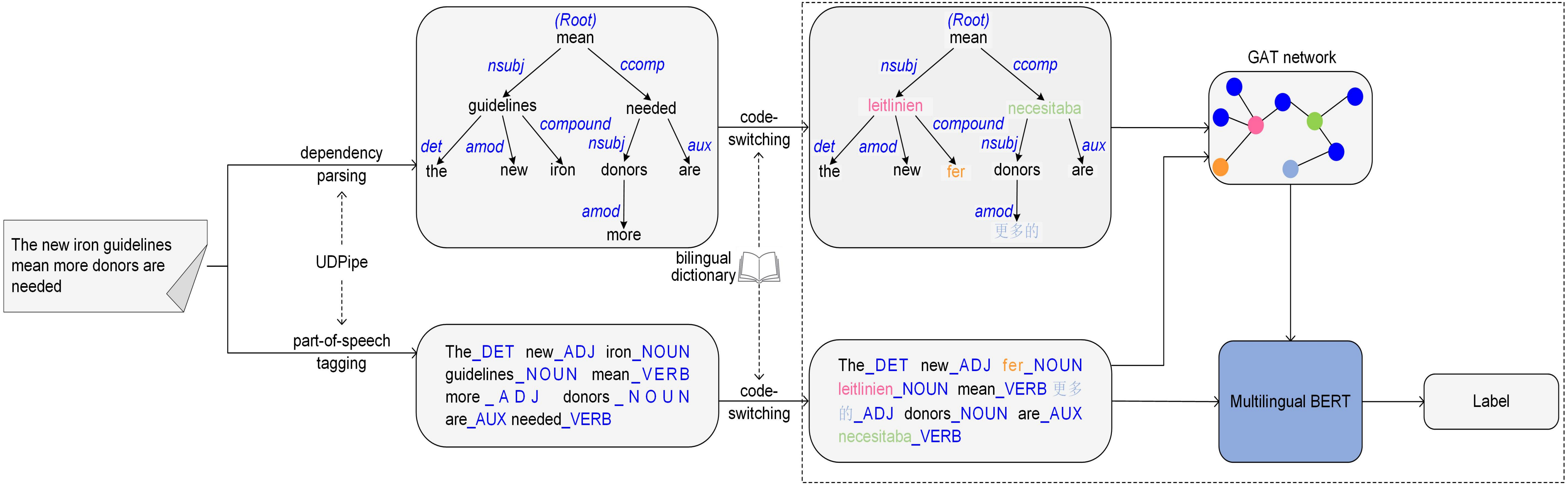
- Each type of information offers unique advantages and no previous attempts have combined both.

- We aim to enhance unsupervised cross-lingual transfer by integrating knowledge from different linguistic levels.

- To achieve this, a framework called ”Lexicon-Syntax Enhanced Multilingual BERT” (”LS-mBERT”) is proposed, based on a pre-trained multilingual BERT model.

Method: LS-mBERT

- Lexicon-Syntax Enhanced Multilingual BERT (“LS-mBERT”)



- Pre-processing input sequence
- Lexical knowledge: code-switching for text (Qin et al., 2021)
- Syntactic knowledge: graph attention network (Ahmad et al., 2021)

Tasks

- Text classification: XNLI (Conneau et al., 2018), PAWS-X (Yang et al., 2019)
- Named entity recognition: Wikiann (Pan et al., 2017)
- Task-oriented semantic parsing: mTOP (Li et al., 2021)

Task	Dataset	|Train|	|Dev|	|Test|	|Lang|	Metric
Classification	XNLI	392K	2.5K	5K	13	Accuracy
Classification	PAWS-X	49K	2K	2K	7	Accuracy
NER	Wikiann	20K	10K	1-10K	15	F1
Semantic Parsing	mTOP	15.7K	2.2K	2.8-4.4K	5	Exact Match

Baselines

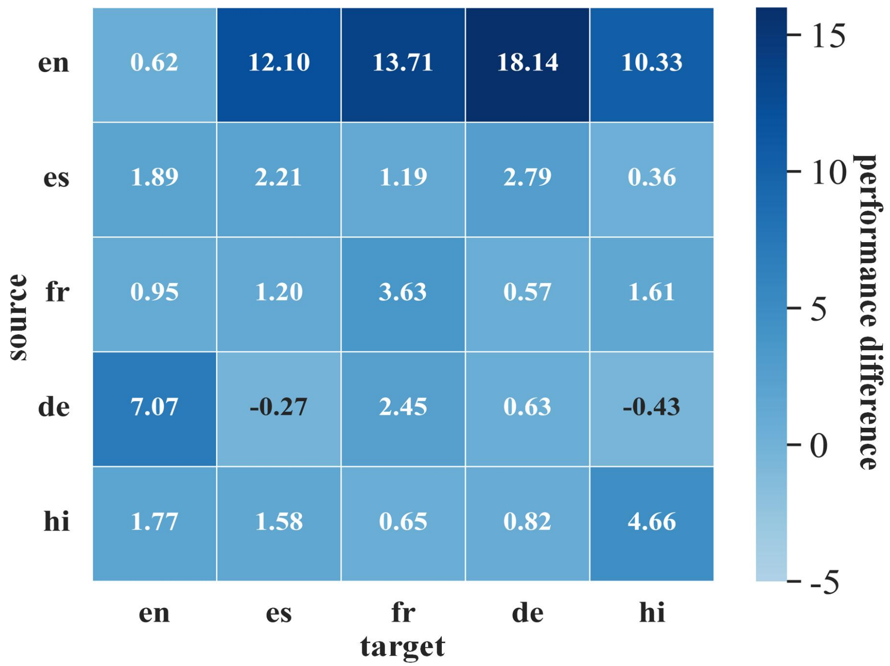
- mBERT: exclusively utilize the multilingual BERT model to perform zero-shot cross-lingual transfer for these tasks
- mBERT + syn: A graph attention network(GAT) is integrated with multilingual BERT
- mBERT + code-switch: The multilingual BERT model is fine-tuned with the code-switched text

Cross-Lingual Transfer Results

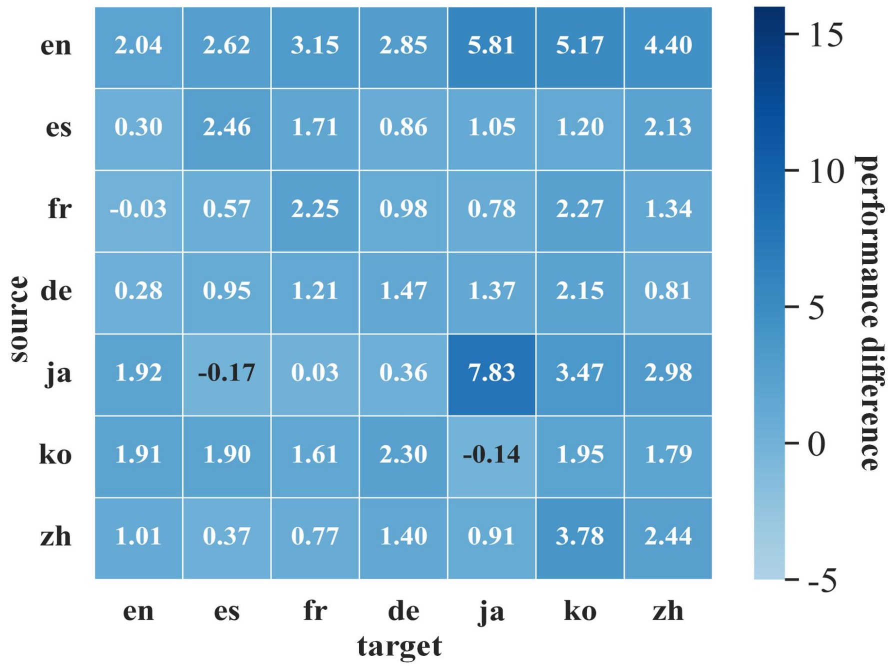
Tasks	Methods	en	ar	bg	de	el	es	fr	hi	ru	tr	ur	vi	zh	ko	nl	pt	ja	AVG / en	AVG
XNLI (Conneau et al., 2018)	mBERT	80.8	64.3	68.0	70.0	65.3	73.5	73.4	58.9	67.8	60.9	57.2	69.3	67.8	-	-	-	-	66.4	67.5
	mBERT+Syn	**81.6**	65.4	69.3	70.7	66.5	74.1	73.2	60.5	68.8	**62.4**	58.7	69.9	69.3	-	-	-	-	67.4	68.5
	mBERT+code-switch	80.9	64.2	70.0	71.5	67.1	73.7	73.2	61.6	68.9	58.6	57.8	69.9	70.0	-	-	-	-	67.2	68.3
	our method	81.3	**65.8**	**71.3**	**71.8**	**68.3**	**75.2**	**74.2**	**62.8**	**70.7**	61.1	**58.8**	**71.8**	**70.8**	-	-	-	-	**68.6**	**69.5**
PAWS-X (Yang et al., 2019)	mBERT	94.0	-	-	85.7	-	87.4	87.0	-	-	-	-	-	77.0	69.6	-	-	73.0	80.2	81.7
	mBERT+Syn	93.7	-	-	86.2	-	89.5	88.7	-	-	-	-	-	78.8	75.5	-	-	75.9	82.7	83.9
	mBERT+code-switch	92.4	-	-	85.9	-	87.9	88.3	-	-	-	-	-	80.2	78.0	-	-	78.0	83.4	84.3
	our method	**93.8**	-	-	**87.2**	-	**89.6**	**89.4**	-	-	-	-	-	**81.8**	**79.0**	-	-	**80.0**	**84.6**	**85.6**
Wikiann(Pan et al., 2017)	mBERT	83.7	36.1	76.0	75.2	68.0	75.8	79.0	65.0	63.9	69.1	38.7	71.0	-	58.9	81.3	79.0	-	66.9	68.1
	mBERT+Syn	84.1	34.6	76.9	75.4	68.2	**76.0**	79.1	64.0	**64.2**	68.7	38.0	**73.1**	-	58.0	81.7	79.5	-	67.0	68.1
	mBERT+code-switch	82.4	39.2	77.1	75.2	68.2	71.0	78.0	66.1	**64.2**	72.4	41.3	69.2		59.9	81.3	78.9	-	67.3	68.3
	our method	**84.5**	**41.4**	**78.9**	**77.3**	**70.2**	75.3	**80.3**	**67.6**	63.9	**73.1**	**46.8**	72.6	-	**62.2**	**81.8**	**80.8**	-	**69.4**	**70.5**
mTOP(Li et al., 2021)	mBERT	81.0	-	-	28.1	-	40.2	38.8	9.8	-	-	-	-	-	-	-	-	-	29.2	39.6
	mBERT+Syn	81.3	-	-	30.0	-	43.0	41.2	11.5	-	-	-	-	-	-	-	-	-	31.4	41.4
	mBERT+code-switch	82.3	-	-	40.3	-	47.5	48.2	16.0	-	-	-	-	-	-	-	-	-	38.0	46.8
	our method	**83.5**	-	-	**44.5**	-	**54.2**	**51.7**	**18.8**	-	-	-	-	-	-	-	-	-	**47.3**	**50.5**

- LS-mBERT consistently demonstrates superior performance across all tasks compared to other baselines
- Most of languages can gain improvement by using LS-mBERT

Generalized Cross-Lingual Transfer Results



(a) mTOP

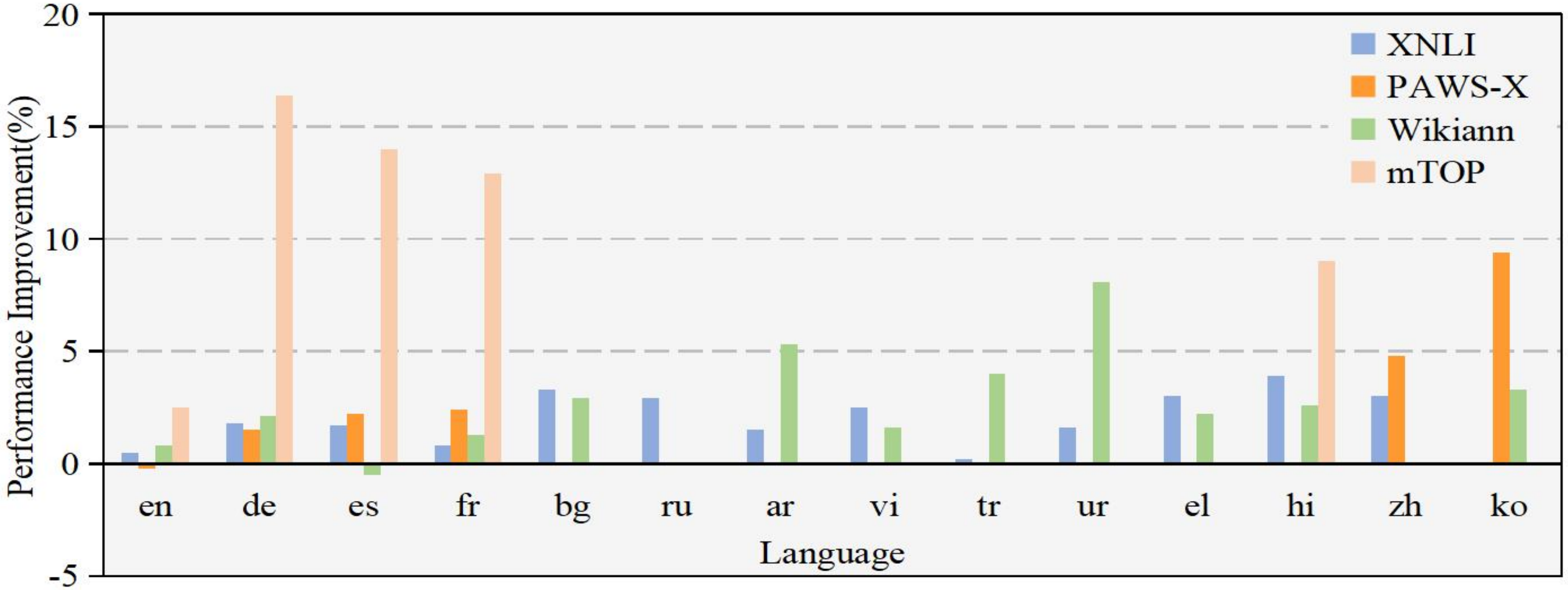


(b) PAWS-X

- Tasks: mTOP and PAWS-X
- Improvements among most language pairs
- A substantial enhancement in performance when English is included in the language pair

Analysis and Discussion:

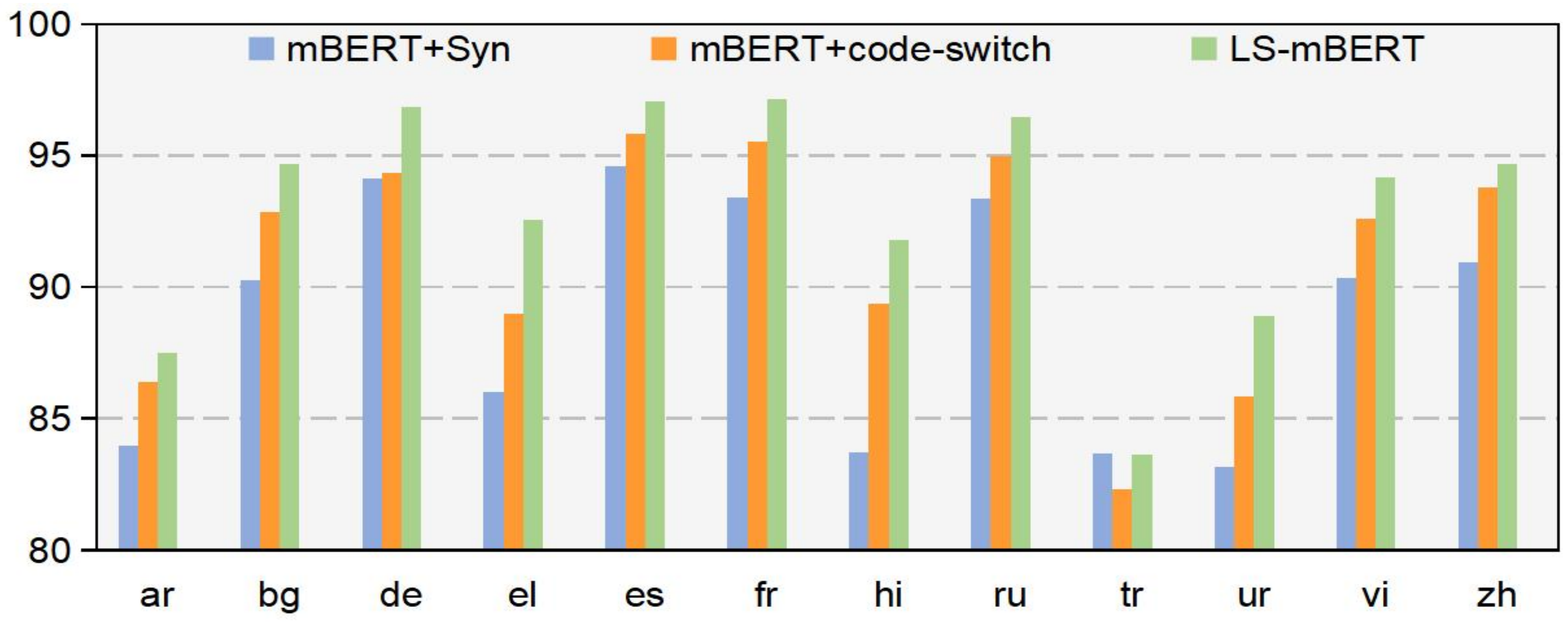
—Impact on Languages



- Almost languages can obtain benefits
- The performance in the mTOP task is improved significantly

Analysis and Discussion:

—Representation Similarities across Languages

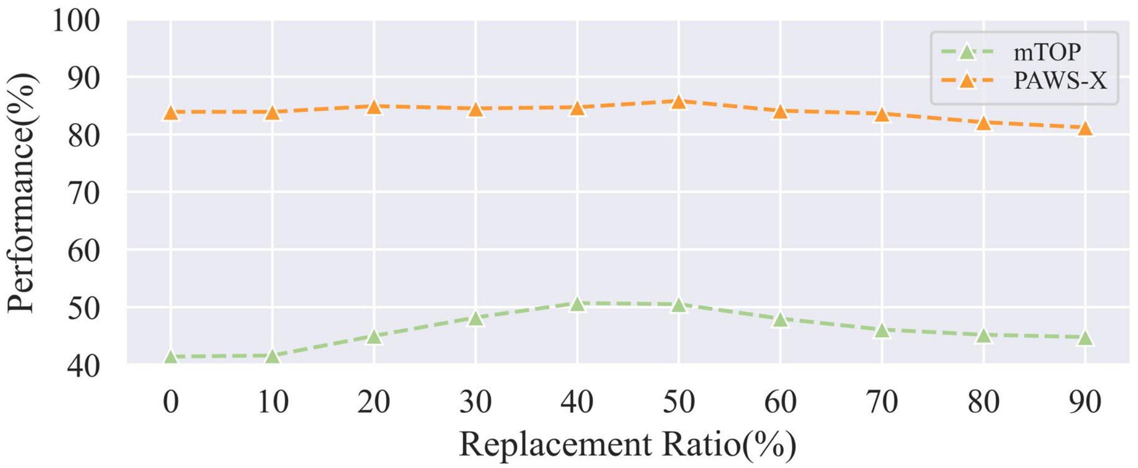


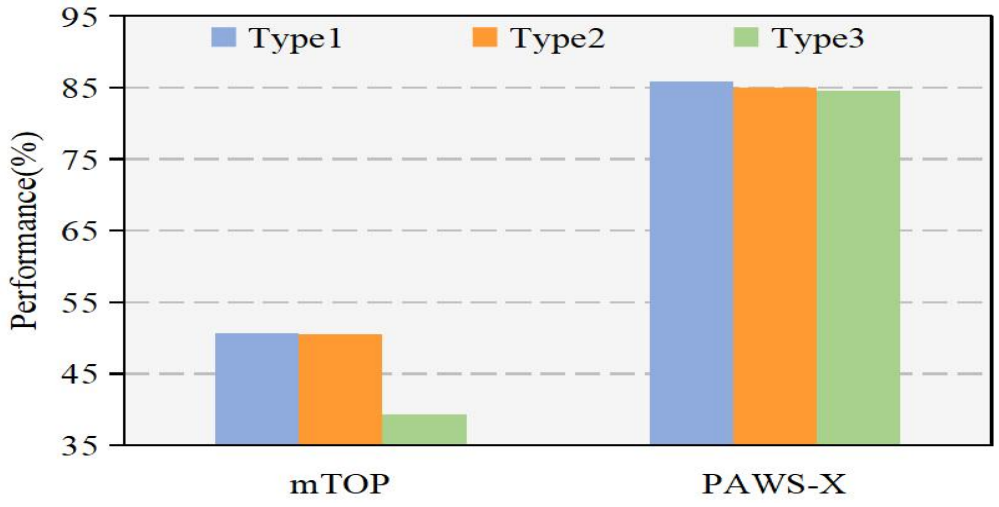
- Utilize the testing set of XNLI
- [CLS] token from the final layer
- Averaging these sentence representations for representing the centroid vector of each language

➡ ”LS-mBERT” outperforms the other two baselines in aligning language representations

Analysis and Discussion:

—Impact of Code-switching





- About half of the words are replaced, the performance reaches their peaks

- Type1: the replacement with the target language
- Type2: the replacement with languages from the same language family as the target language
- Type3: the replacement with randomly selected languages

Analysis and Discussion:

—Performance with XLM-R

Task	Methods	en	ar	bg	de	el	es	fr	hi	ru	tr	ur	vi	ko	nl	pt	AVG
PAWS-X	XLM-R	84.2	48.5	80.5	77.0	77.8	76.1	79.8	67.5	70.4	76.0	54.2	78.5	59.1	83.3	79.3	72.8
	XLM-R+Syn	83.5	46.4	80.1	76.0	78.9	77.6	79.1	72.1	70.6	76.1	55.3	77.6	59.0	83.1	79.2	73.0
	XKLM-R+code-switch	83.4	46.8	81.7	78.2	79.2	71.1	78.6	72.9	70.6	77.2	57.9	76.0	58.2	83.6	80.0	73.0
	our method	83.1	44.9	82.7	76.8	78.4	76.9	79.6	71.1	70.1	76.6	60.4	78.2	58.1	83.5	79.7	73.3

- Our framework outperforms the other three baselines with XLM-R

Thanks for your listening!