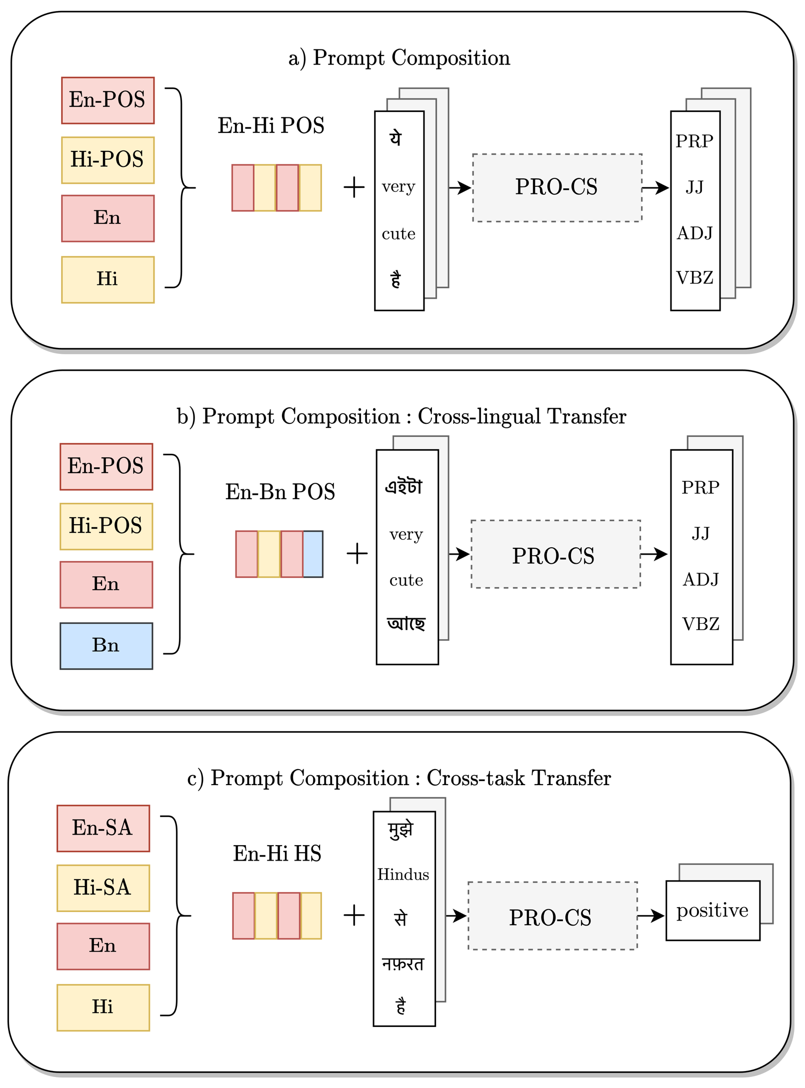
PRO-CS : An Instance-Based Prompt Composition Technique for Code-Switched Tasks

Srijan Bansal* Suraj Tripathi* Sumit Agarwal* Teruko Mitamura Eric Nyberg
Language Technologies Institute, Carnegie Mellon University

Introduction

Prompt Composition for Code Switching

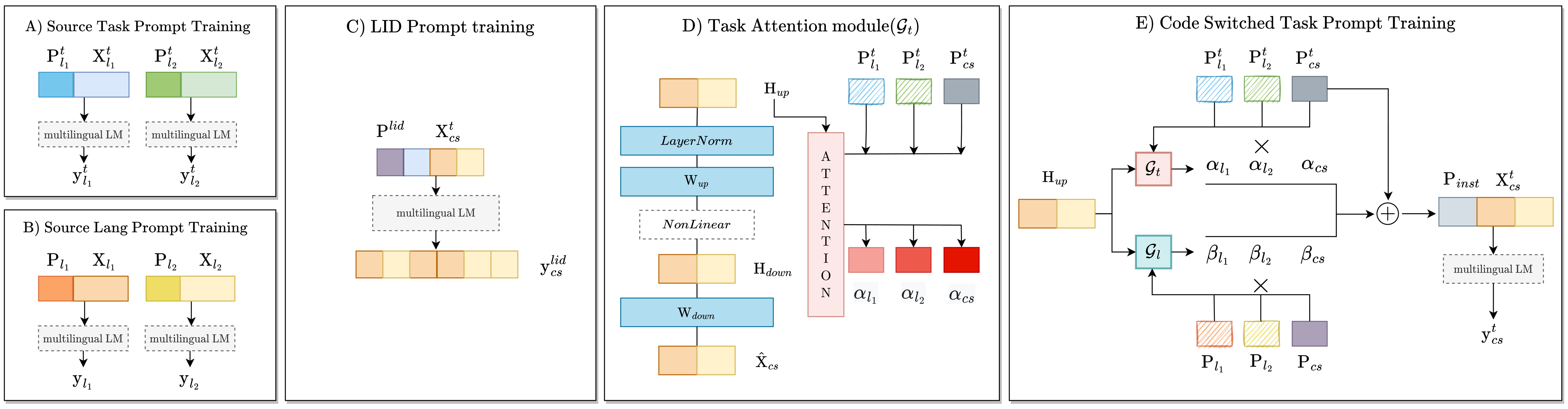
- Parameter-efficient training is essential for low-resource code-switching domain
- Prompt-composition of **task** and **language**-specific prompts to compose instance-level code-switched task prompts.
- Leverage monolingual task and language corpora to transfer knowledge to downstream code-switched task
- Evaluated on 10 classification and sequence tagging datasets from **GLUECoS** and **LinCE** benchmarks across four language pairs (En-Hi, En-Es, En-Bn, En-Ta).



Main Contributions

- First work to explore prompt-tuning with multilingual LMs for code-switching.
- Reducing the gap between finetuning and prompt tuning by proposing a prompt composition technique that leverages language and task-specific knowledge captured through language and task-based prompts.
- Outperforms prompt-tuning across all datasets and exceeds or remains at par with fine-tuning by using just **0.18% of total parameters.**
- Data efficient and effective in low-resource **cross-lingual** and **cross-task settings**.

Methodology



Source Task Prompt Training : Monolingual task corpora consists of same tasks as the code-switching target tasks (like POS, SA). Source task prompts are obtained through prompt-tuning on monolingual source task corpora.

Source Language Prompt Training : Source language prompts are trained on wikipedia data for each source language to capture language-specific knowledge. Inspired by the discriminative pre-training of language models, we used a distilled mBERT as a generator to generate tokens from masked input. The frozen mBERT discriminator with soft prompt learns to distinguish between fake and real tokens.

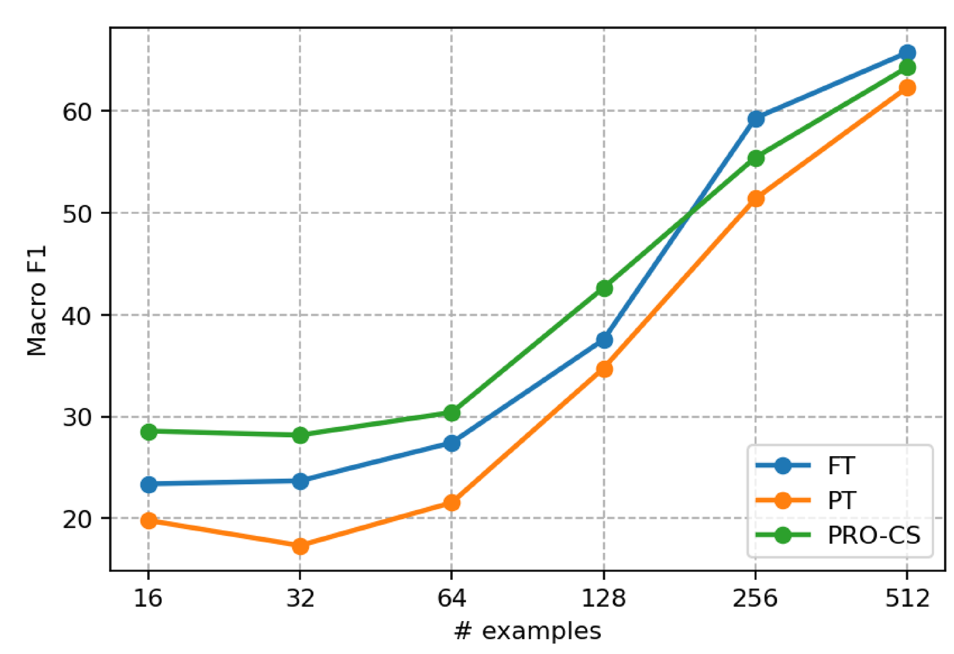
Code-Switched Target Prompt Training : Learn task and language-specific target prompts to capture task and language-specific information of the code-switched task. Separate attention modules are trained for each target prompt to learn the contribution of task and language prompts. Attention modules are then used to compose an instance based prompt using source and target prompts.

Target Prompt Initialization : Target language prompt is initialized using a prompt trained on language identification task for the code-switched language pair. This is done to add inductive bias for the downstream task. Target task prompt is initialized from random words from the vocabulary.

Results

		full-data						*cross-lingual* (Hi → X)		*cross-task* (SA → X)	
		POS		NER		SA		POS	SA	HS	IN
Setting	Params	En-Hi	En-Es	En-Hi	En-Es	En-Hi	En-Es	En-Bn	En-Ta	En-Hi	En-Hi
FT	178M	**60.76**	81.14	**75.09**	**49.93**	66.06	42.15	**65.75**	28.38	**63.12**	43.74
PT	77K	54.43	77.99	70.94	45.98	63.34	43.00	62.33	25.38	61.33	39.95
PRO-CS	314K	60.33	**82.98**	71.98	46.02	**67.88**	**45.52**	64.30	**29.57**	62.82	**45.03**

- PRO-CS achieves an improvement of **3** and **3.5** points in Macro-F1 scores on sequence tagging and classification tasks compared to the prompt-tuning approach in the full data setting.
- For cross-lingual transfer, PRO-CS is significantly more data efficient compared to both fine-tuning and prompt-tuning approaches. Experimented with total training instances ranging from 16 to 512.



Cross-Lingual transfer (En-Hi → En-Bn) POS