

# Improving Word Translation via Two-Stage Contrastive Learning

Language Technology Lab, University of Cambridge, UK

Yaoyiran Li, Fangyu Liu, Nigel Collier, Anna Korhonen, Ivan Vulić

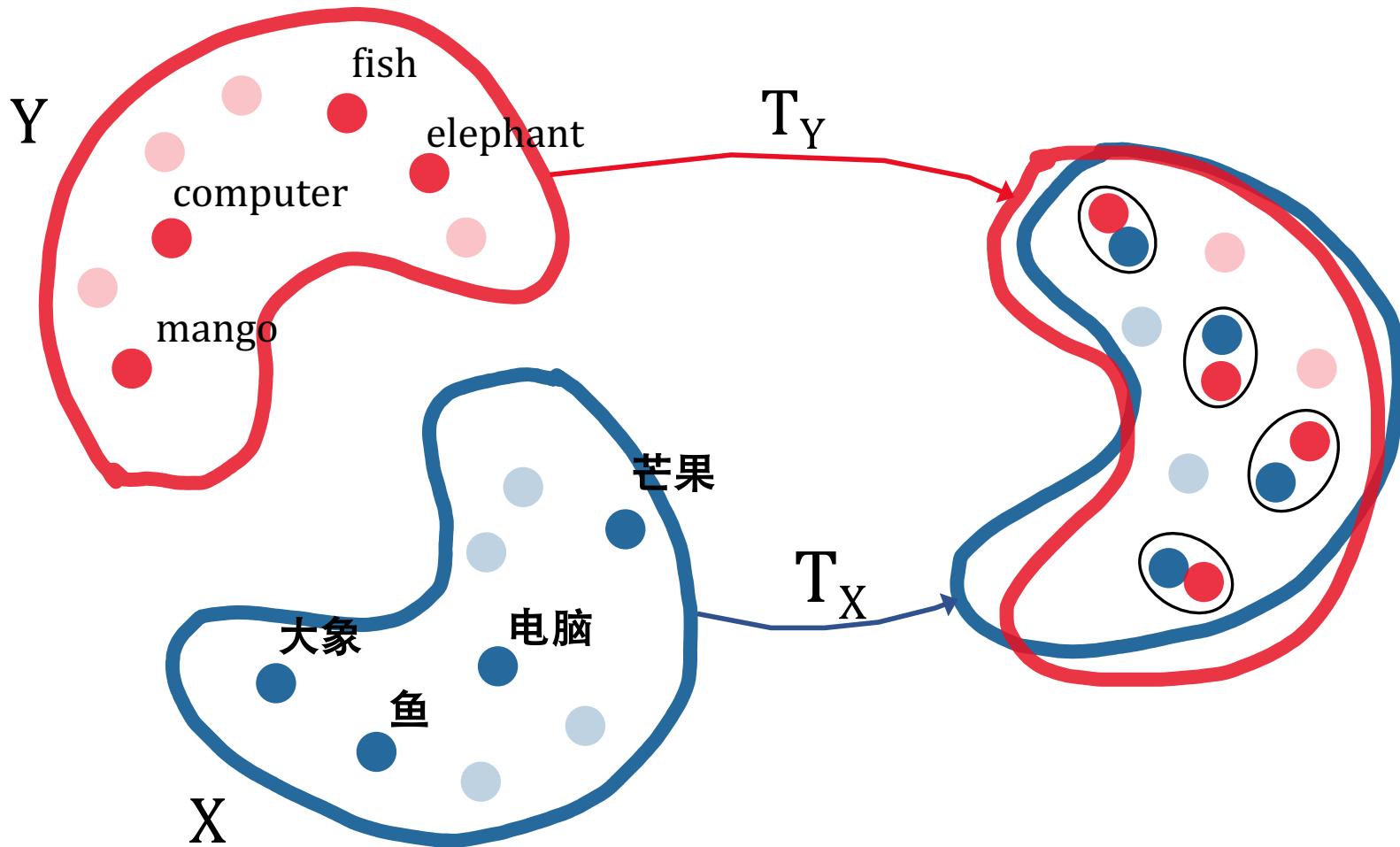


UNIVERSITY OF  
CAMBRIDGE

λ ä Language  
宇 汉 Technology  
w й Lab

ACL 2022  
22ND – 27TH MAY | 60TH MEETING | DUBLIN

# The Task: Bilingual Lexicon Induction



# The Task: Bilingual Lexicon Induction

- An important and long-standing task
  - Applications of BLI
    - Language learning & language acquisition
    - Machine Translation
    - Cross-lingual Transfer Learning
    - Low-resource NLP
- .....

# The Task: Bilingual Lexicon Induction

- Supervised
  - 5k pairs
- Semi-supervised
  - 1k pairs
- Unsupervised

# The Task: Bilingual Lexicon Induction

- Supervised
  - 5k pairs
- Semi-supervised
  - 1k pairs
- Unsupervised

# Previous Approaches

- Mapping-based
  - Orthogonal (Xing et al., 2015)
  - Linear (Joulin et al., 2018)
  - Non-linear (Mohiuddin et al., 2020)

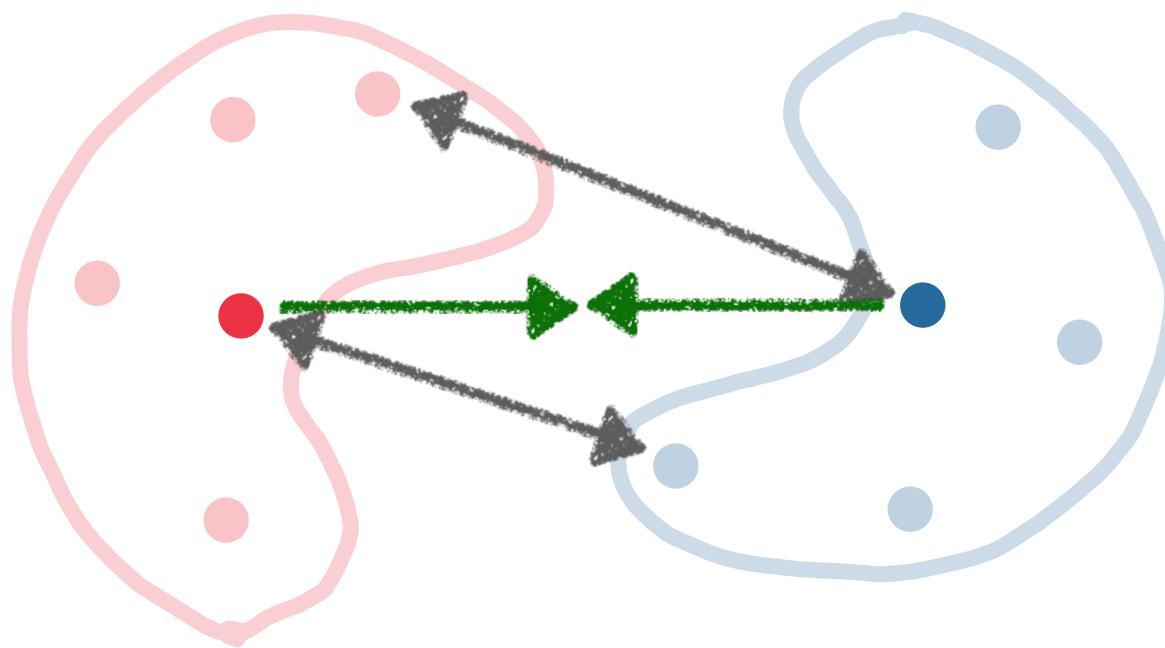
# Previous Approaches

- Mapping-based
  - Orthogonal (Xing et al., 2015)
  - Linear (Joulin et al., 2018)
  - Non-linear (Mohiuddin et al., 2020)
- Static WEs > pretrained LMs
  - fastText > monolingual BERT (Vulić et al., 2020)
  - fastText > multilingual LMs (e.g. mBERT) (Gonen et al., 2020)
  - Why?

# Motivation

- Contrastive Learning (CL)
  - Success in sentence encoders (Gao et al., 2021; Liu et al., 2021)
  - BLI ?

→ ← Attract  
← → Repel



# Motivation

- We propose a novel two-stage CL approach.
  - Stage C1 can be evaluated independently
  - Stage C2, further improvement

# Motivation

- We propose a novel two-stage CL approach.
  - Stage C1 can be evaluated independently
  - Stage C2, further improvement
- Stage C1
  - Static WEs + CL = SotA (RQ 1)

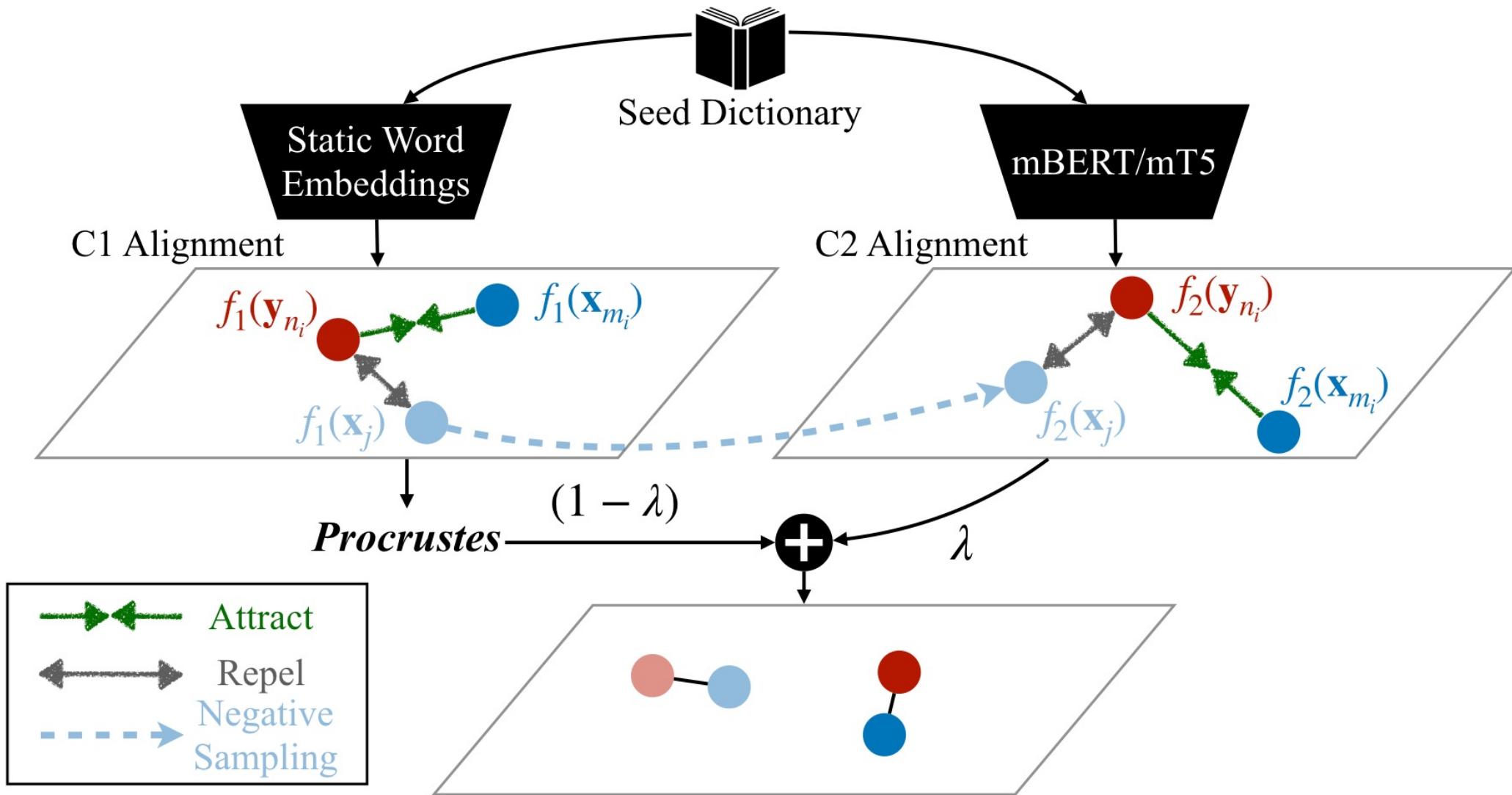


# Motivation

- We propose a novel two-stage CL approach.
  - Stage C1 can be evaluated independently
  - Stage C2, further improvement
- Stage C1
  - Static WEs + CL = SotA (RQ 1)
- Stage C2
  - LMs + CL = Lexical Encoders (RQ 2)
  - CL exposes translation knowledge (RQ 2)
  - Combine LMs with static Wes (RQ 3)



# Method: Two-Stage CL



# Method: CL in Stage C1 (Static WEs)

Linear map on  $L_x$  WE

Linear map on  $L_y$  WE

$$s_{i,j} = \exp(\cos(x_i \mathbf{W}_x, y_i \mathbf{W}_y) / \tau)$$

$$p_i = \frac{s_{m_i, n_i}}{\sum_{w_j^y \in \{w_{n_i}^y\} \cup \bar{w}_{n_i}^y} s_{m_i, j} + \sum_{w_j^x \in \bar{w}_{m_i}^x} s_{j, n_i}}$$

$$\min_{\mathbf{W}_x, \mathbf{W}_y} - \mathbb{E}_{(w_{m_i}^x, w_{n_i}^y) \in \mathcal{D}_{CL}} \log(p_i)$$

A positive pair

Hard negative pairs

## Method: CL in Stage C2 (multilingual LMs)

Encoder  $L_x$  Word with LM

Encode  $L_y$  Word with LM

$$s'_{i,j} = \exp(\cos(f_\theta(w_i^x), f_\theta(w_j^y))/\tau)$$

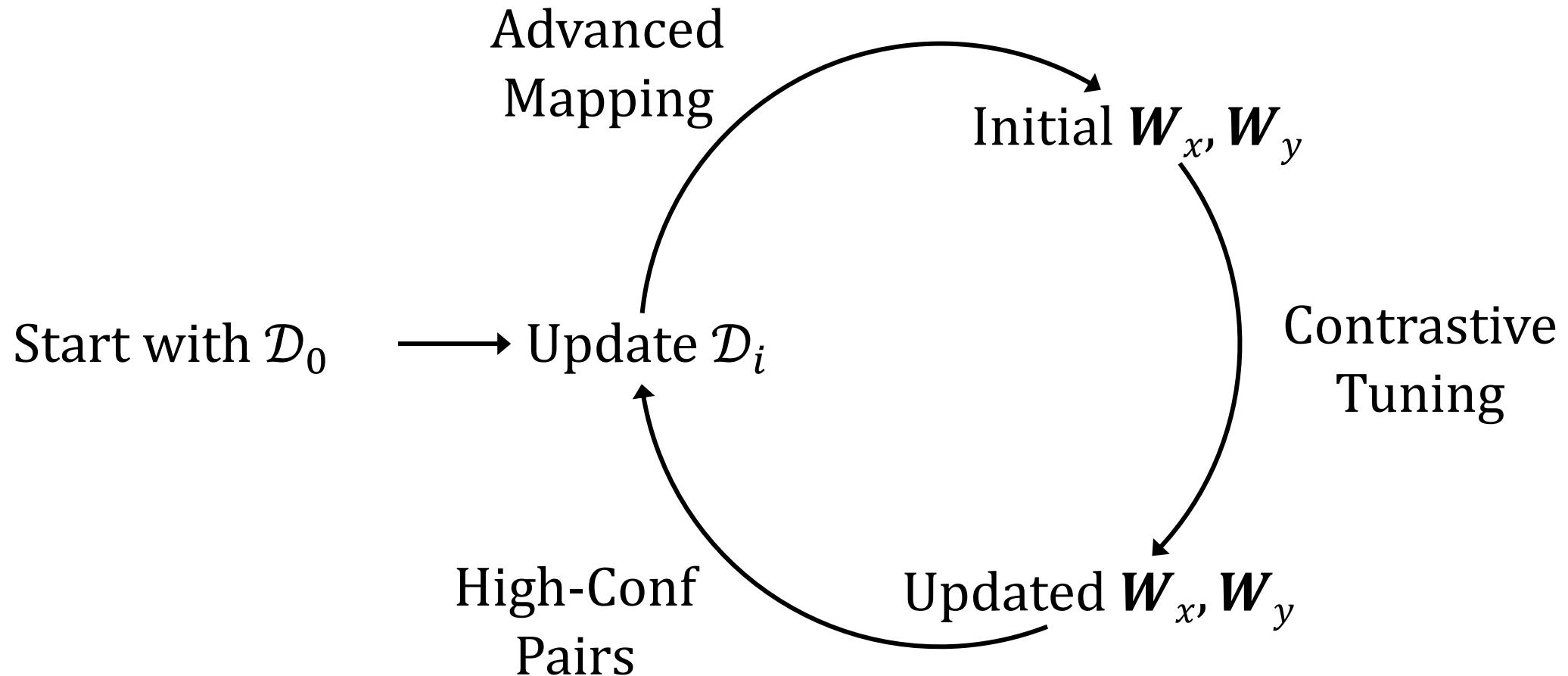
$$p'_i = \frac{s'_{m_i, n_i}}{\sum_{w_j^y \in \{w_{n_i}^y\} \cup \bar{w}_{n_i}^y} s'_{m_i, j} + \sum_{w_j^x \in \bar{w}_{m_i}^x} s'_{j, n_i}}$$

$$\min_{\theta} - \mathbb{E}_{(w_{m_i}^x, w_{n_i}^y) \in \mathcal{D}_{CL}} \log(p'_i)$$

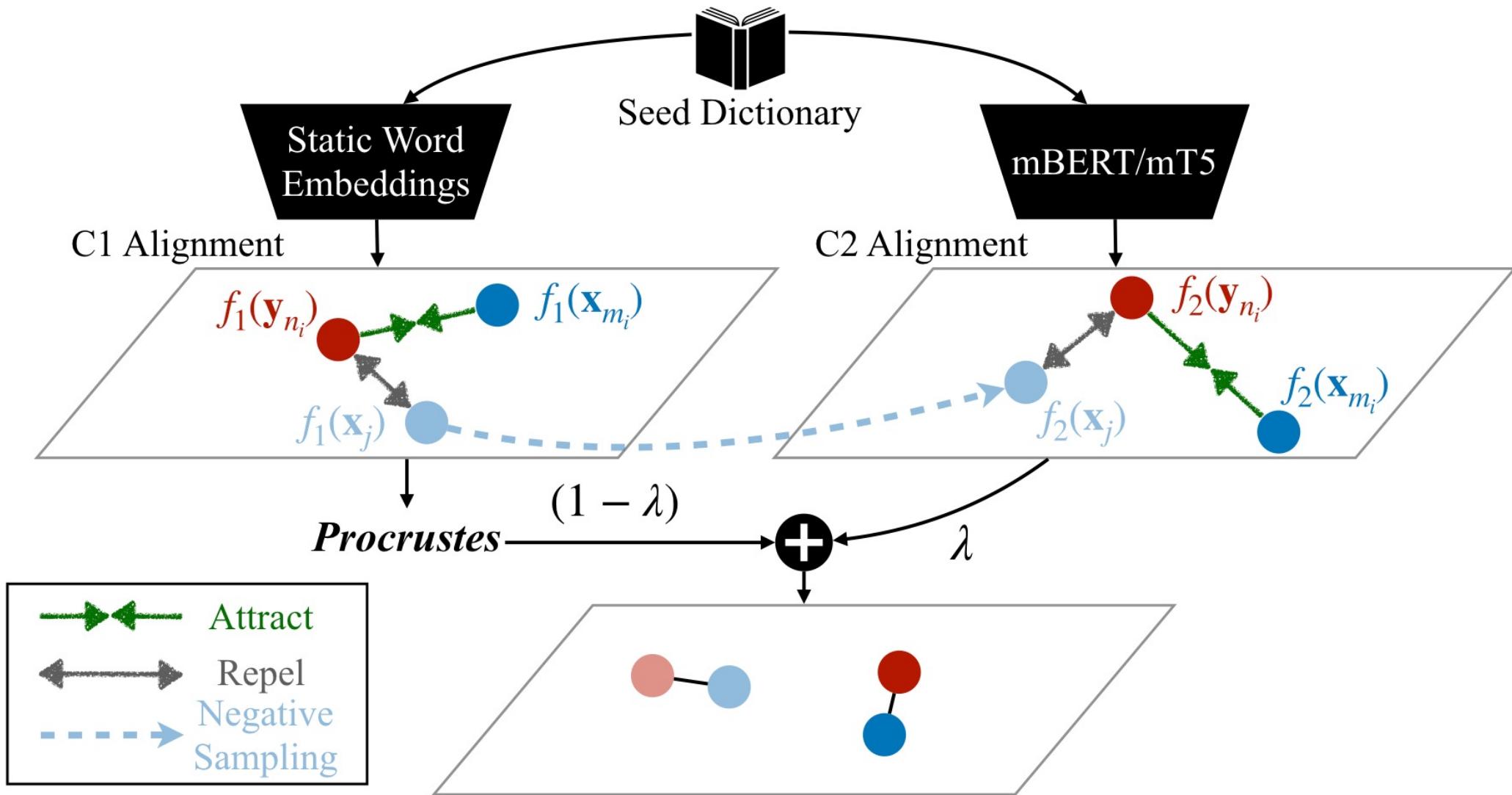
A positive pair

Hard negative pairs

## Method: Stage C1



# Method: Stage C2



# Evaluation

- Datasets
  - XLING (112 setups) (Glavaš et al., 2019)
  - PanLex-BLI (lower-resource languages) (Vulić et al., 2019)
  - Vocabulary size: 200K

# Evaluation

- Datasets
  - XLING (112 setups) (Glavaš et al., 2019)
  - PanLex-BLI (lower-resource languages) (Vulić et al., 2019)
  - Vocabulary size: 200K
- Baselines
  - RCSLS (w/o SL) (Joulin et al., 2018)
  - VecMap (w/ SL) (Artetxe et al., 2018)
  - LNMap (w/ SL) (Mohiuddin et al., 2020)
  - FIPP (w/ SL) (Sachidananda et al., 2021)

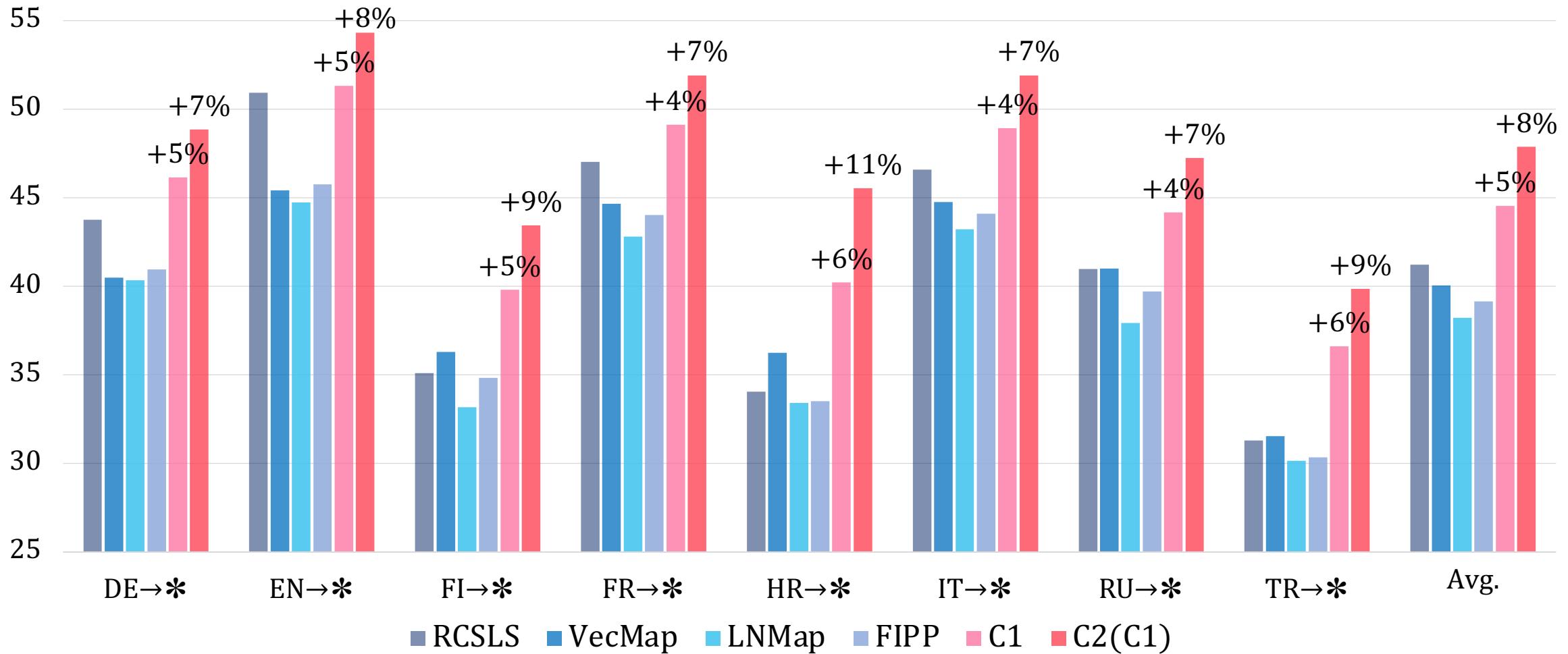
# Evaluation

- Statics WEs & Multilingual LMs
  - fastText
  - mBERT (main); XLM, mT5 (comparison)

# Evaluation

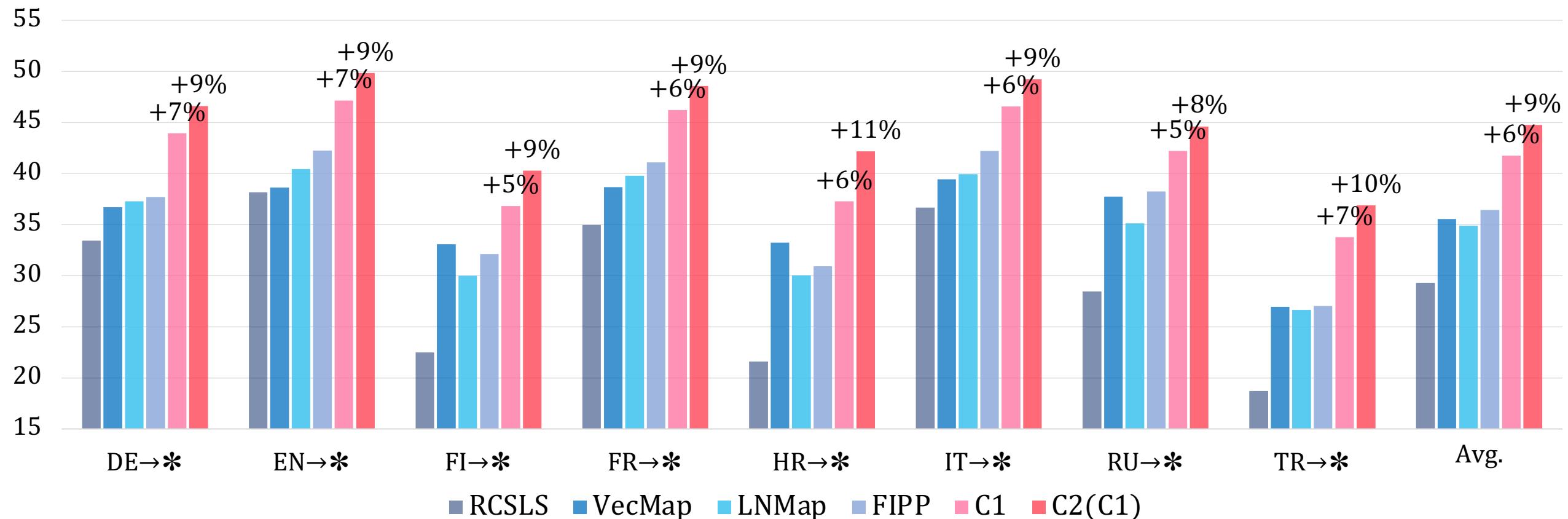
- Statics WEs & Multilingual LMs
  - fastText
  - mBERT (main); XLM, mT5 (comparison)
- Hyperparameters
  - Tuned on EN→TR
  - $\lambda=0.2$  (fixed)

# Evaluation: XLING 5K



- C1:  $\approx 5\%$  higher than SotA (Avg. of 4 Baselines)
- C2:  $\approx 8\%$  higher than SotA (Avg. of 4 Baselines)

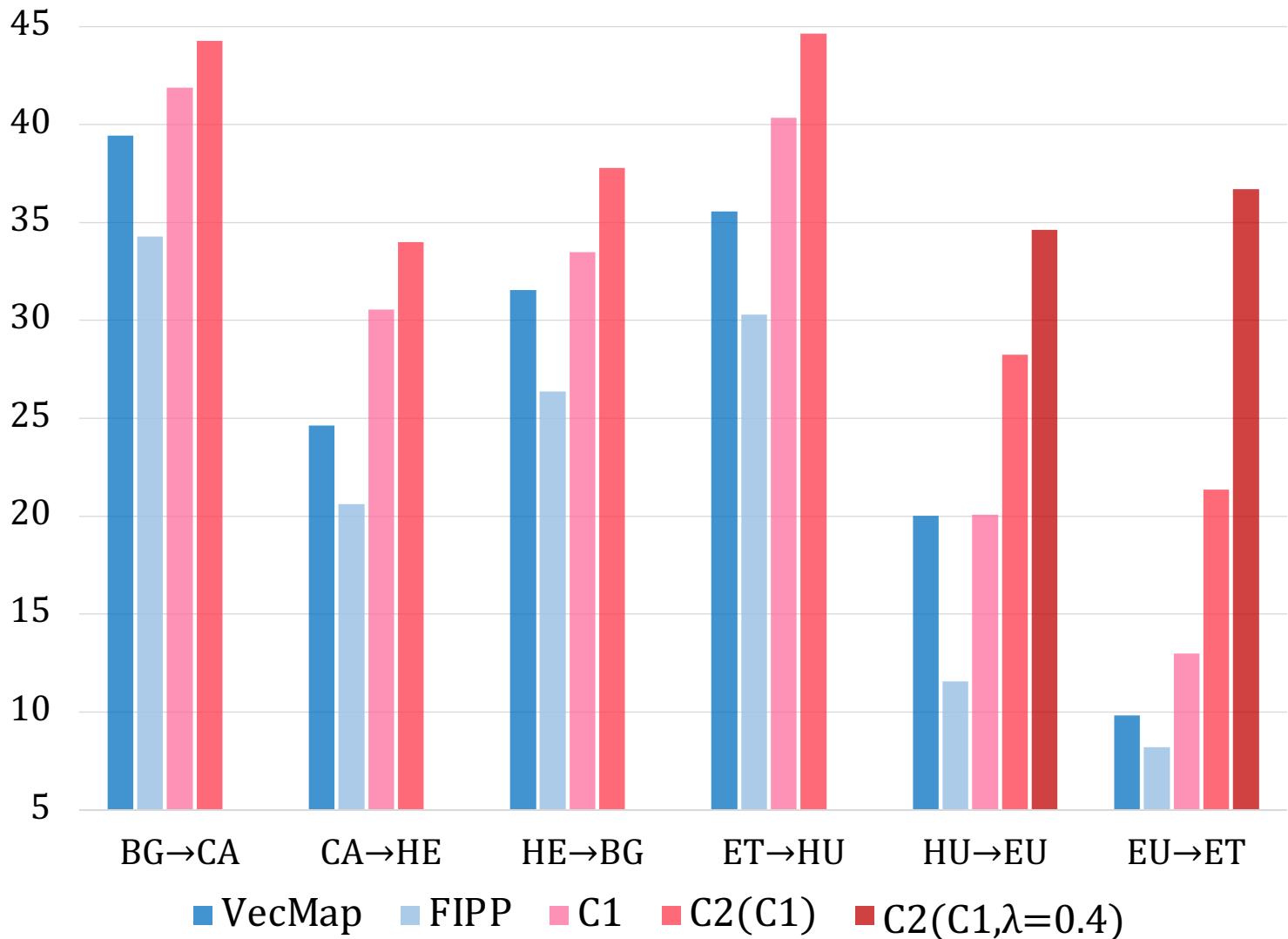
# Evaluation: XLING 1K



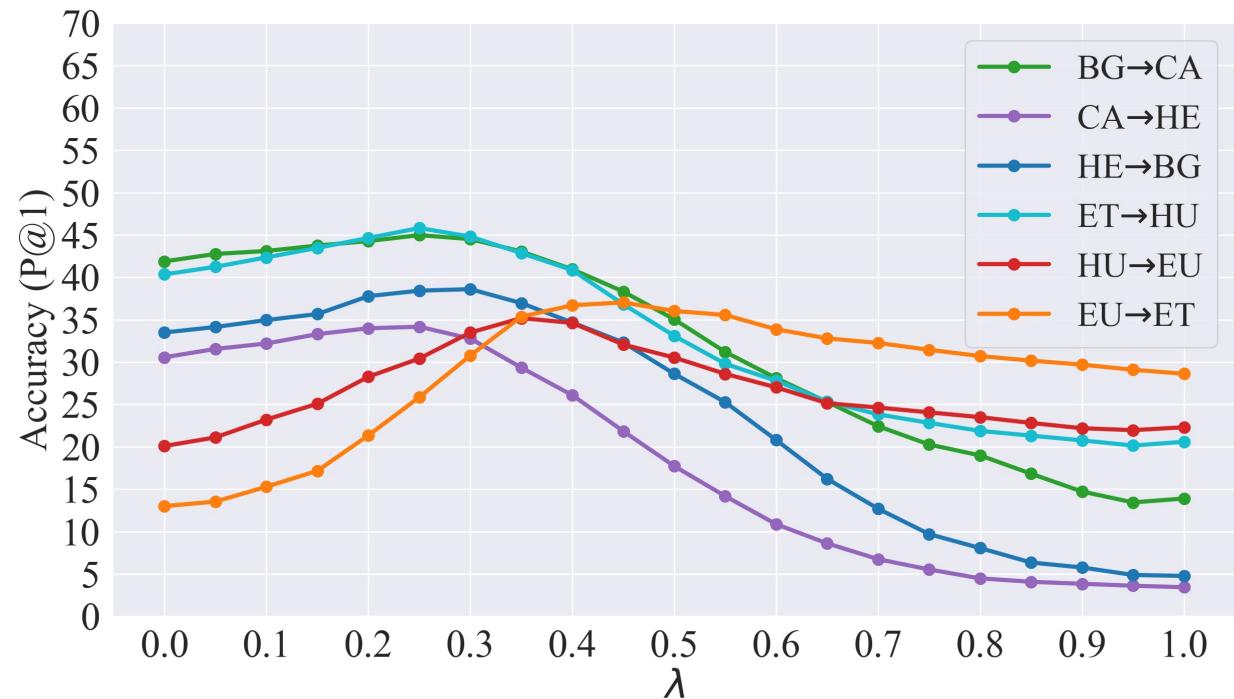
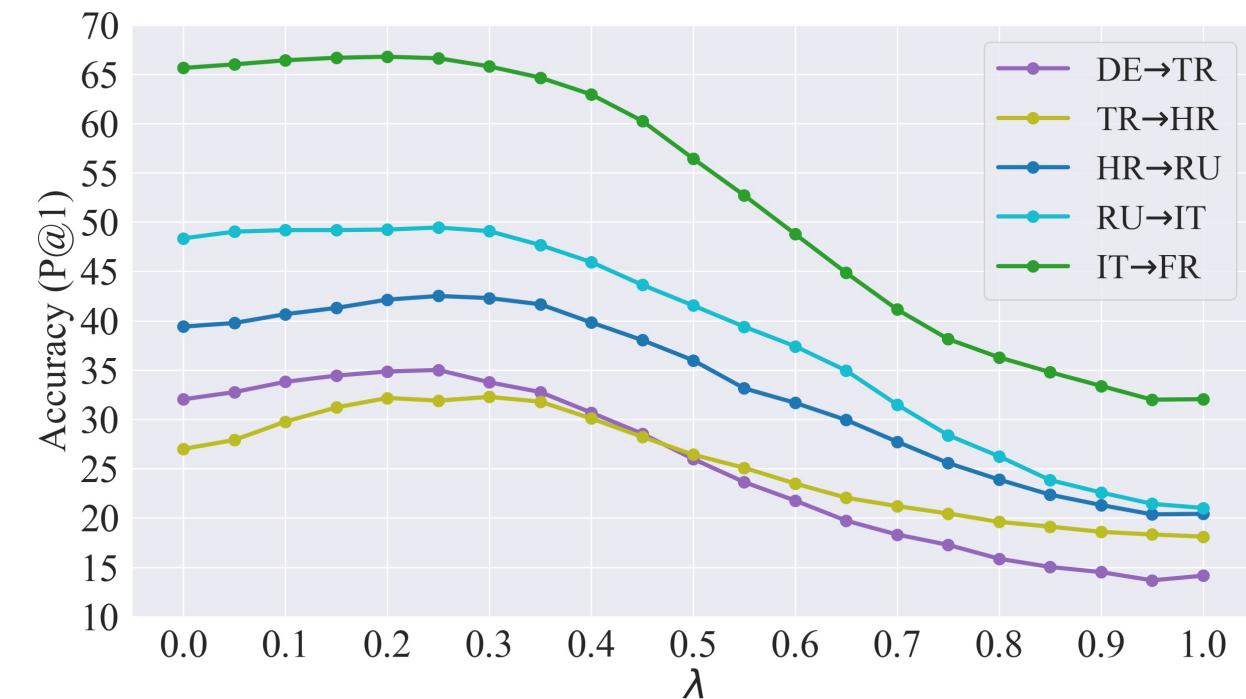
- C1:  $\approx 6\%$  higher than SotA (Avg. of 3 stronger baselines)
- C2:  $\approx 9\%$  higher than SotA (Avg. of 3 stronger baselines)

# Evaluation: PanLex-BLI 1K

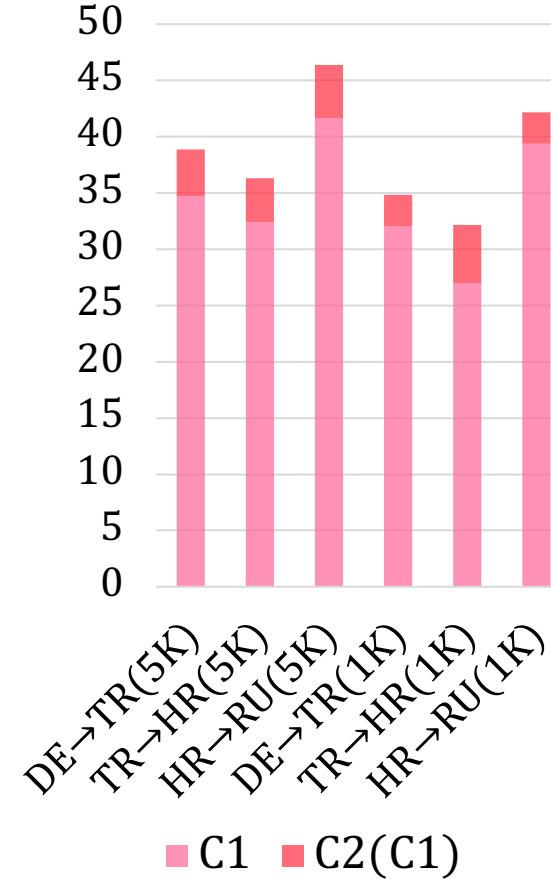
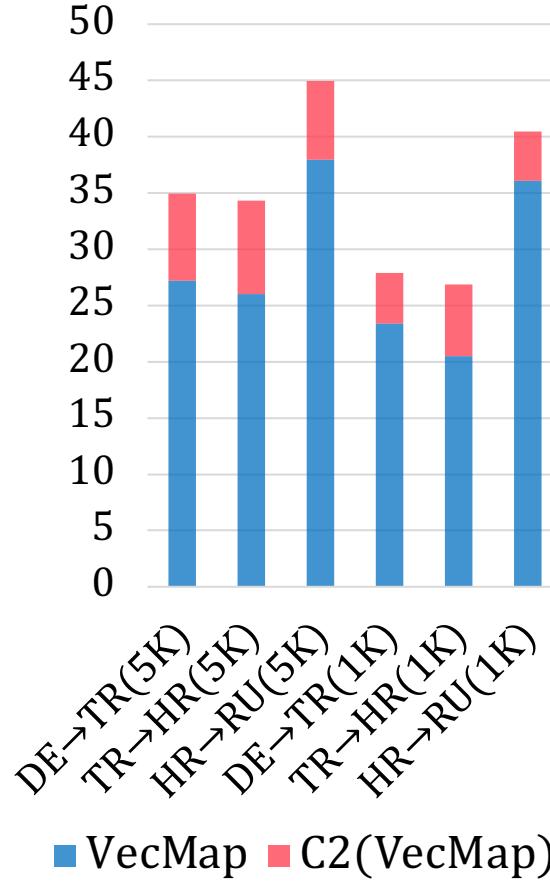
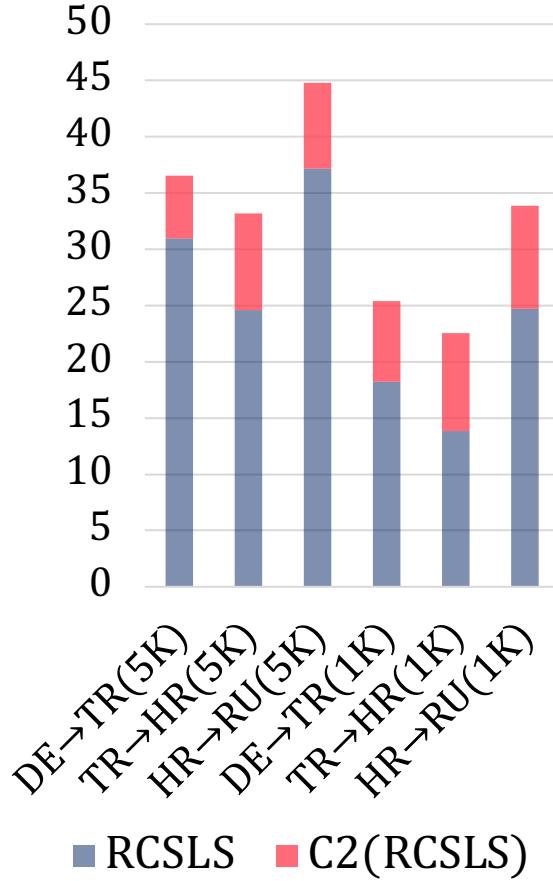
- Same trends
- PLMs more useful?
  - Lower-resource



# Analysis: $\lambda$ Values

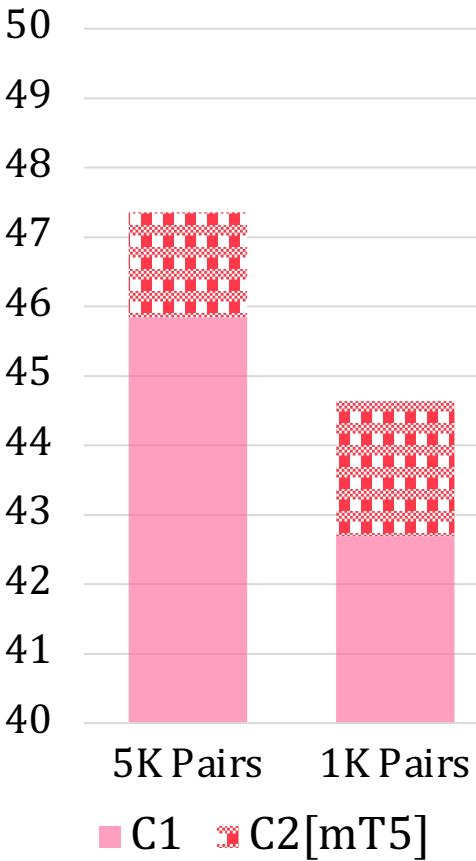
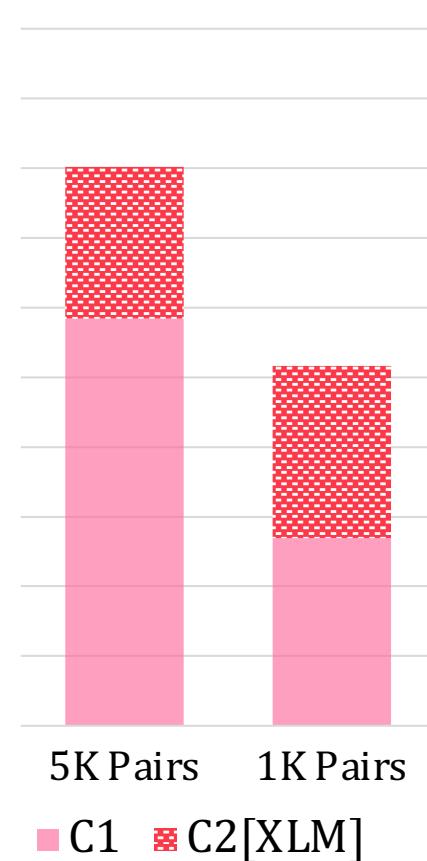
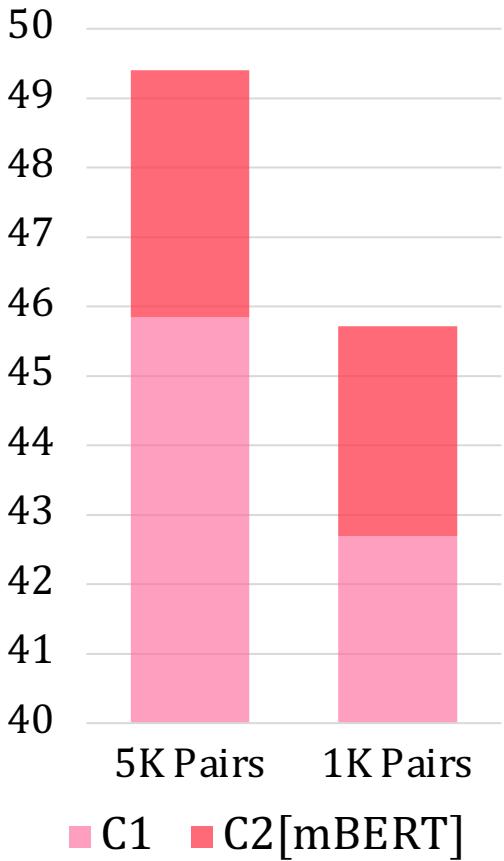


# Analysis: Different ‘Support’ Methods



- C2(mod)
  - C2(RCSLS) ✓
  - C2(VecMap) ✓
  - C2(C1) ✓

# Analysis: Different Multilingual LMs

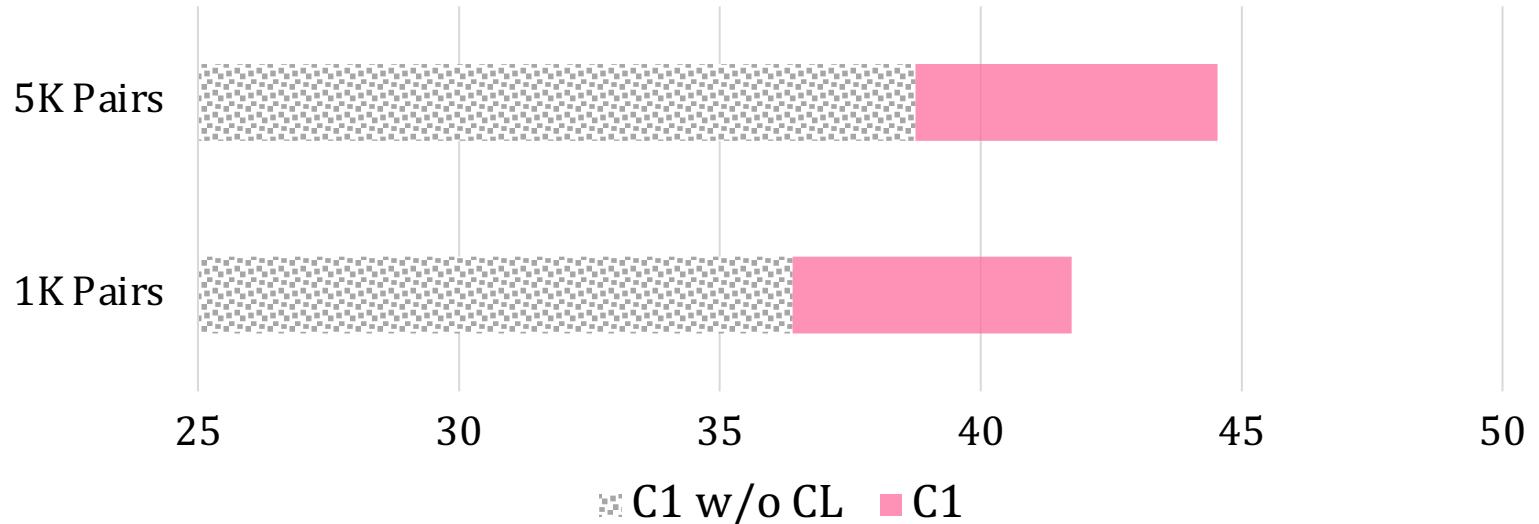


- C2[LM]
  - C2[mBERT] ✓
  - C2[XLM] ✓
  - C2[mT5] ✓

# Analysis: Ablation Study — Stage C1

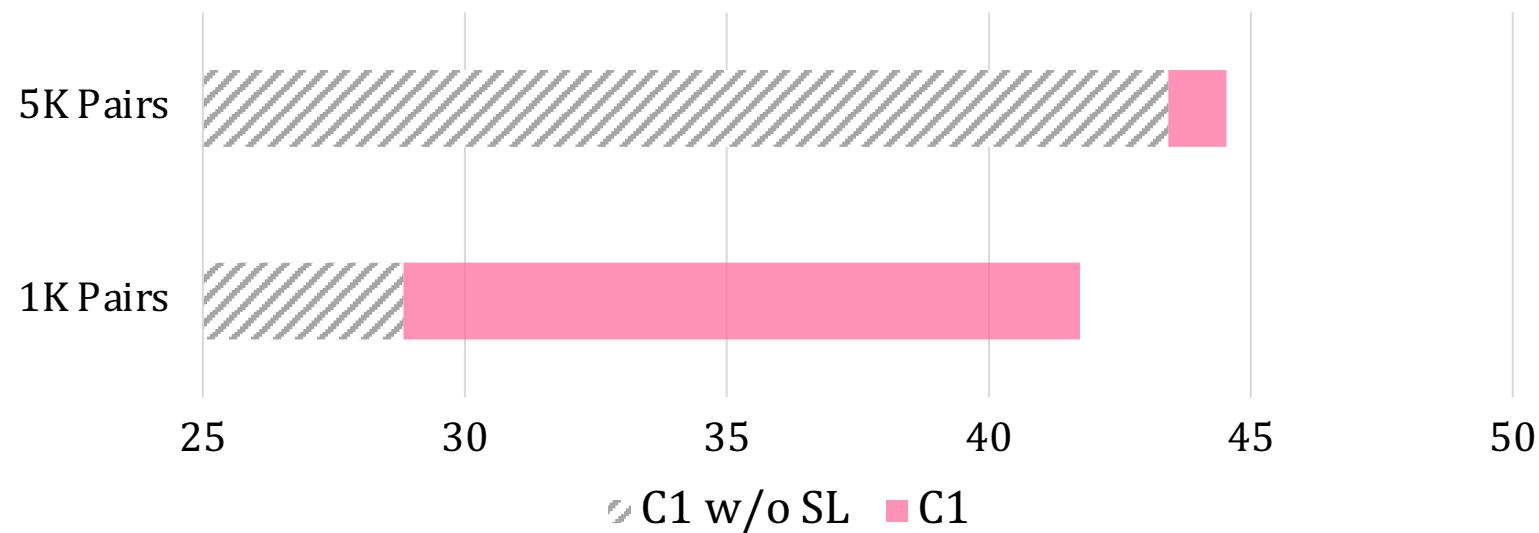
- Contrastive Learning

- 5K ✓
- 1K ✓



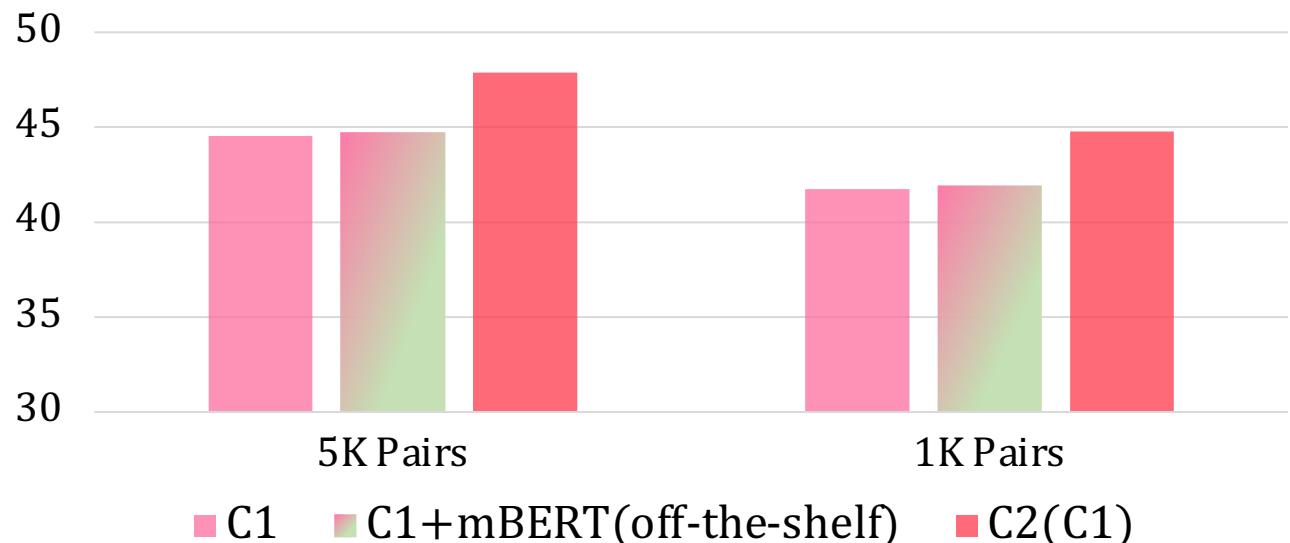
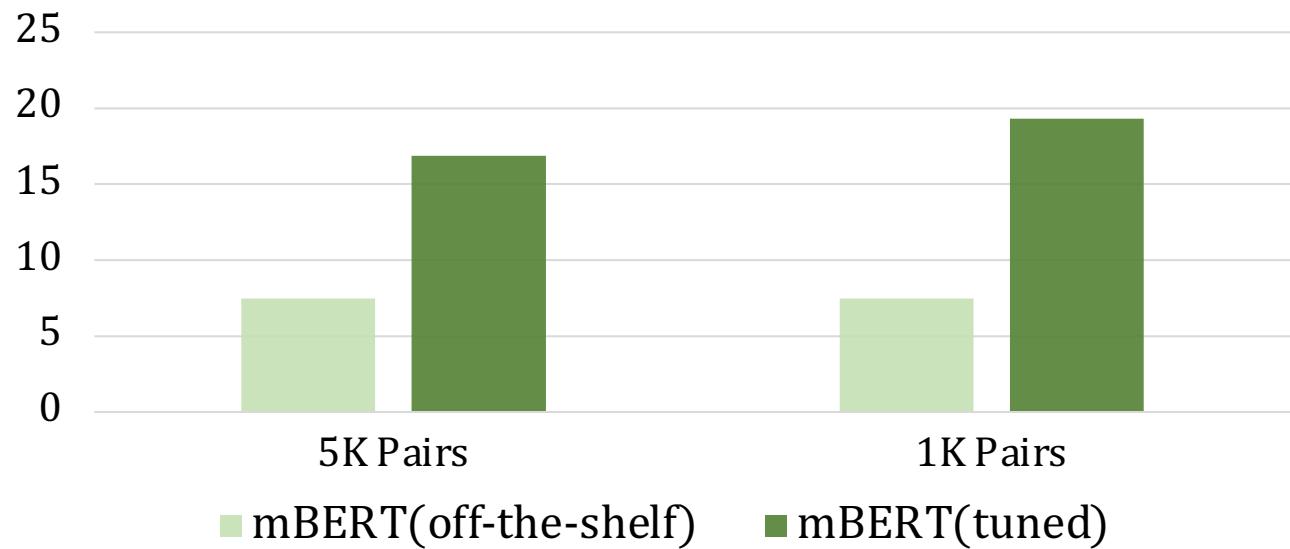
- Self-Learning

- 5K (slight gains)
- 1K ✓

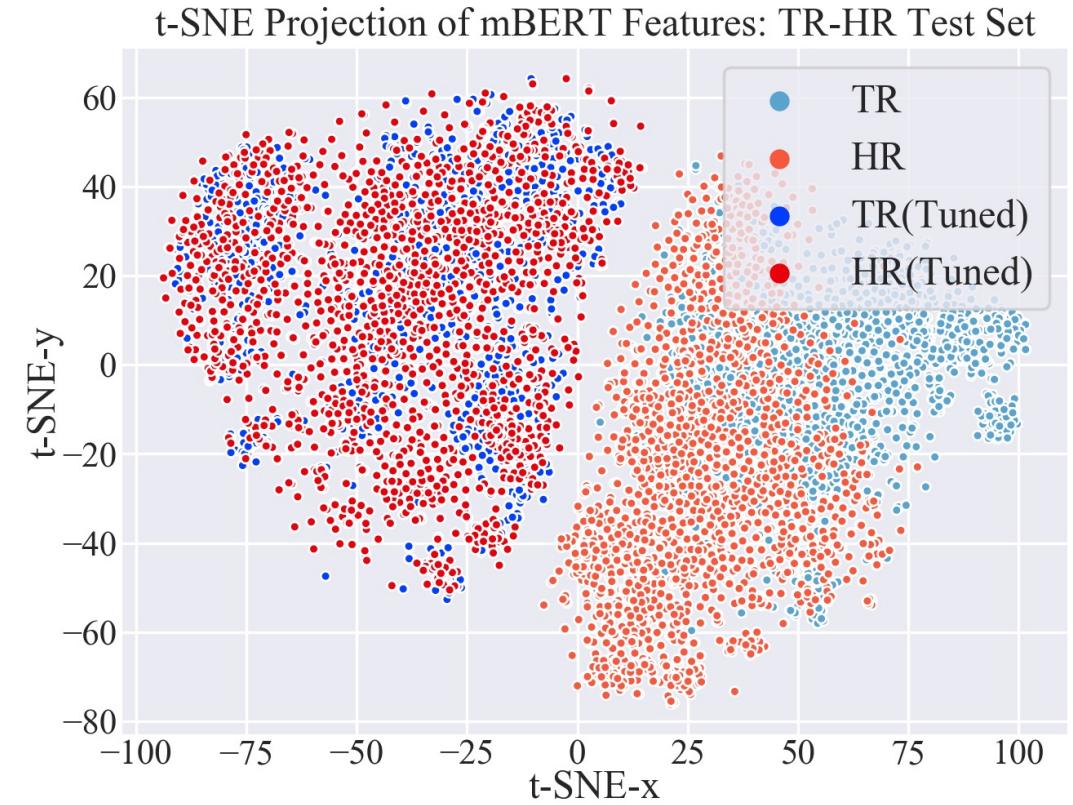
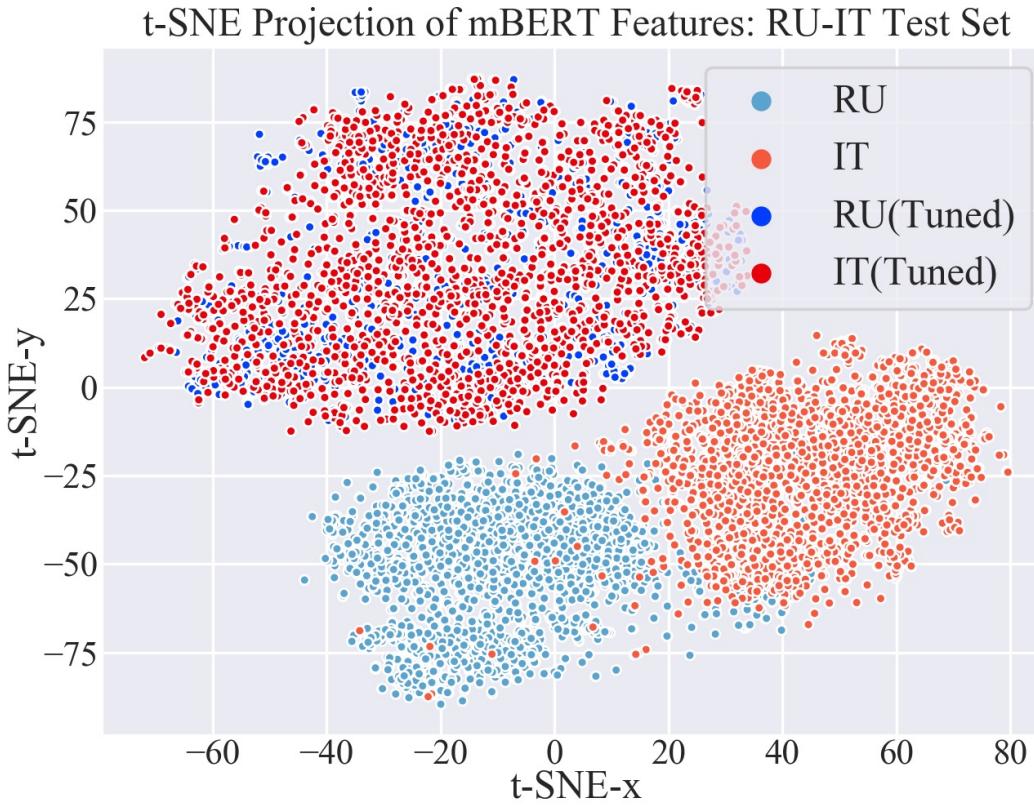


# Analysis: Ablation Study — Stage C2

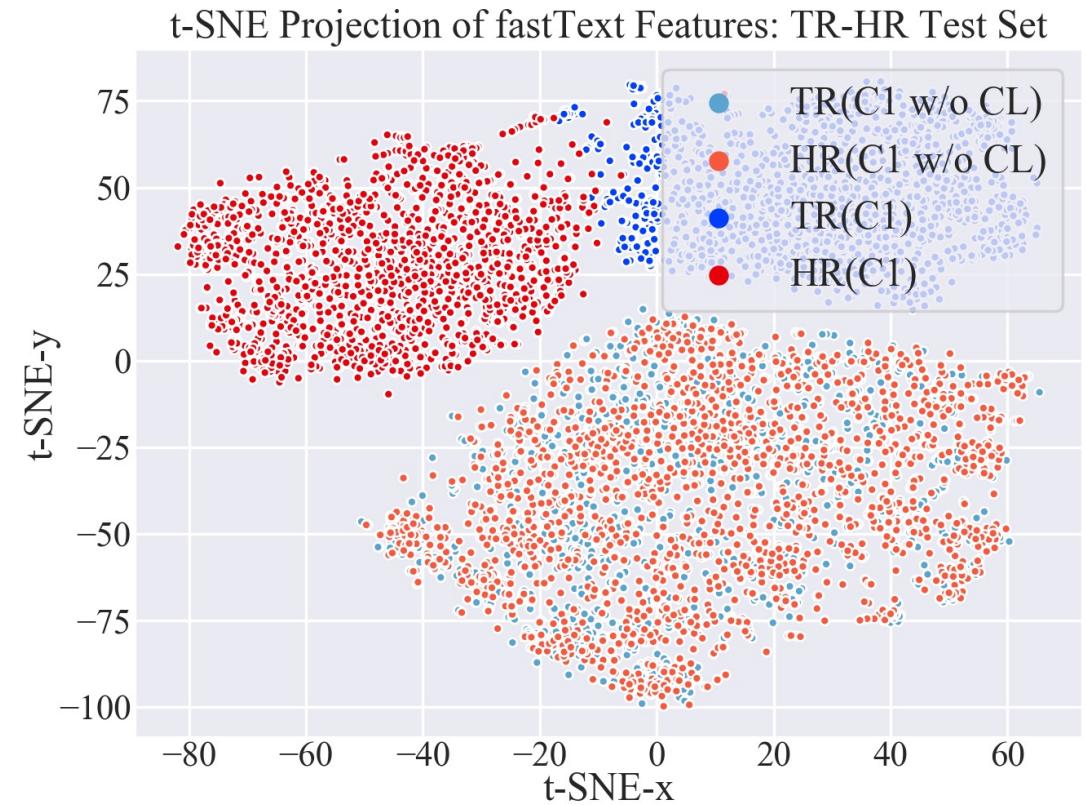
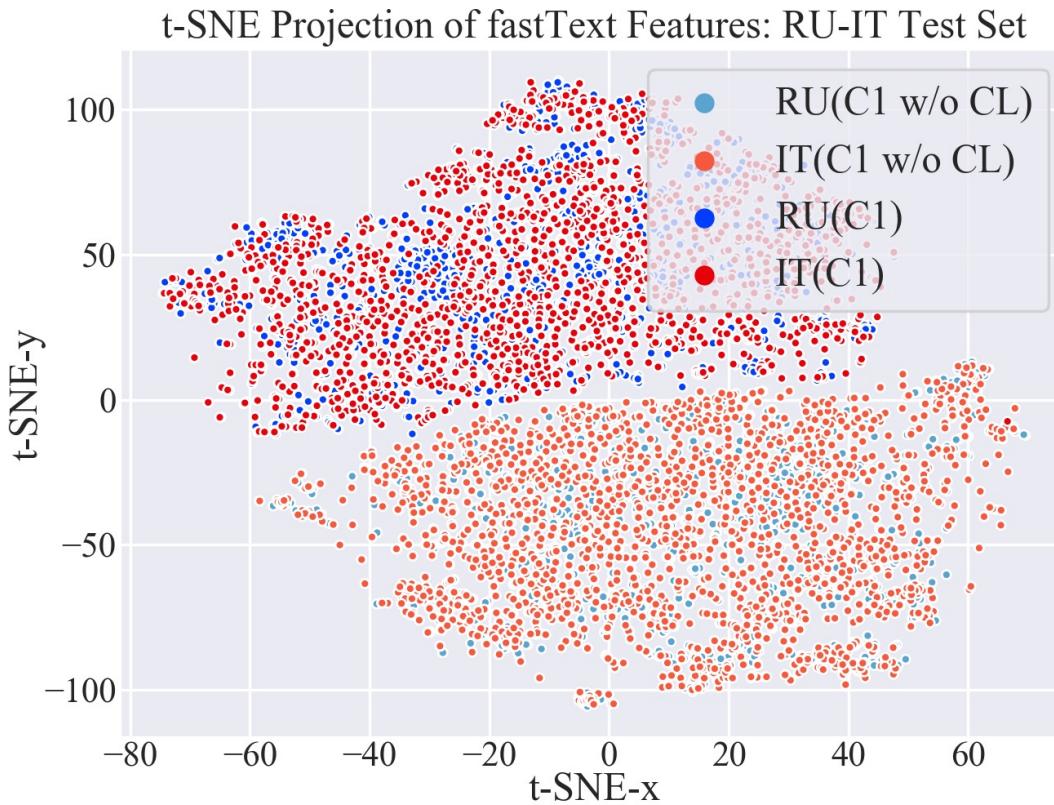
- Contrastive Tuning
  - Tuned > Off-the-shelf
  - fastText (C1) > mBERT
- Combing C1 & mBERT
  - CL exposes word translation knowledge in mBERT



# Analysis: Effectiveness of CL (Stage C2)



# Analysis: Effectiveness of CL (Stage C1)



# Summary

- A two-stage framework for BLI
  - Simple, effective and robust

# Summary

- A two-stage framework for BLI
  - Simple, effective and robust
- Stage C1
  - Static WEs only, e.g., fastText
  - SotA  $\leftarrow$  evaluated independently
- Stage C2
  - CLWEs + Multilingual LMs, e.g., mBERT
  - Extra gains

# Summary

- A two-stage framework for BLI
  - Simple, effective and robust
- Stage C1
  - Static WEs only, e.g., fastText
  - SotA  $\leftarrow$  evaluated independently
- Stage C2
  - CLWEs + Multilingual LMs, e.g., mBERT
  - Extra gains
- A series of analyses & ablation studies

# Summary (Technically)

- Contrastive Learning

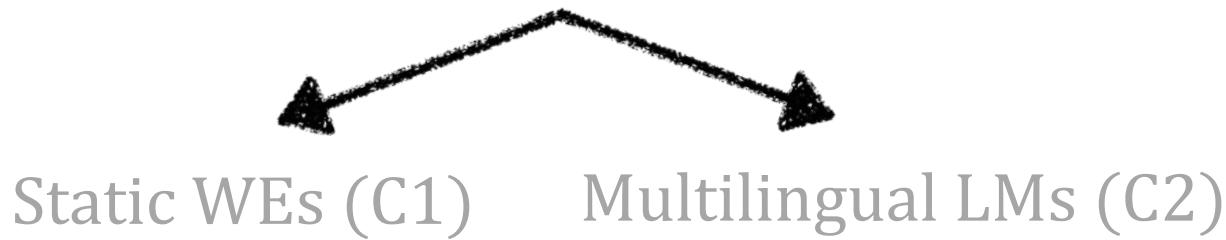


Static WEs (C1)      Multilingual LMs (C2)



# Summary (Technically)

- Contrastive Learning



- Self-Learning
  - Semi-supervised
- Multilingual LMs
  - Word translation knowledge

fastText

word2Vec

Georgian  
Aramaic  
Finnish  
Sindhi  
Koranic  
Chinese  
Simplified Chinese  
Tsonga  
Lithuanian  
Ukrainian  
Serbian  
Portuguese  
Yiddish  
Neapolitan  
Tajik  
Walloon  
Lower Sorbian  
Venda  
Western Sorbian  
Pernyak  
Akan  
Navajo  
Aromanian  
kashmiri  
Southern Sotho  
Maltese  
Cantonese  
Aragonese  
Lezgian  
Croatian  
Komi  
persian  
Urdu  
Hindi  
Bavarian  
Ndonga  
Inuktitut

FastText Vs word2Vec



- Might be useful for future BLI research: CL, SL, and multilingual (also monolingual ) LMs.

# Thanks for Watching!

**Code:** <https://github.com/cambridgetl/ContrastiveBLI>

**Contact:** [yl711@cam.ac.uk](mailto:yl711@cam.ac.uk)