News Without Borders: Domain Adaptation of Multilingual Sentence Embeddings for Cross-lingual News Recommendation

Andreea Iana[1], **Fabian David Schmidt**[2], **Goran Glavaš**[2], **Heiko Paulheim**[1]

[1]Data and Web Science Group, University of Mannheim, Germany

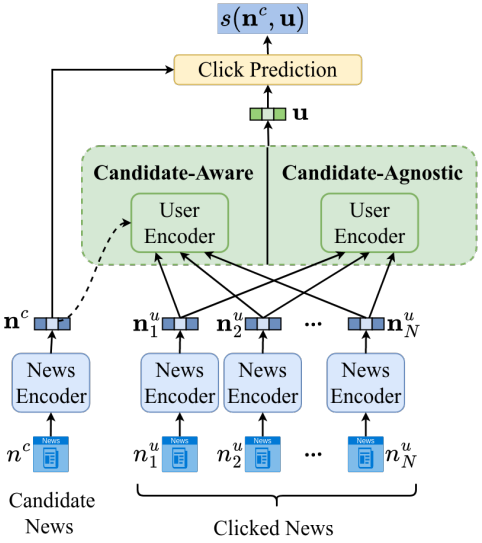
[2]Center for Artificial Intelligence and Data Science, University of Würzburg, Germany

UNIVERSITÄT MANNHEIM

Julius-Maximilians-
UNIVERSITÄT WÜRZBURG

News Recommendation

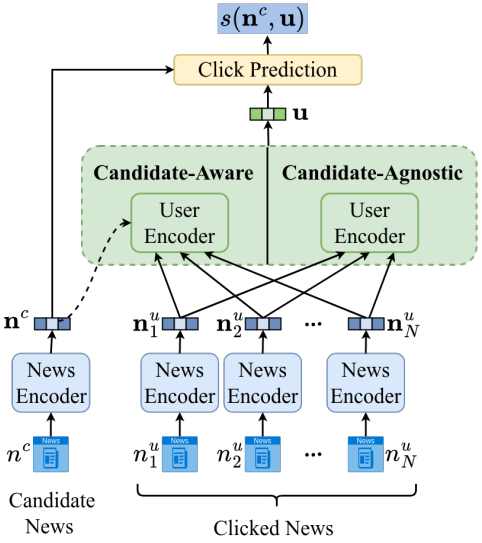
Personalized (neural) news recommendation:
articles tailored to users' preferences



Increasingly language-diverse & polyglot online user community

News Recommendation

Personalized (neural) news recommendation:
articles tailored to users' preferences



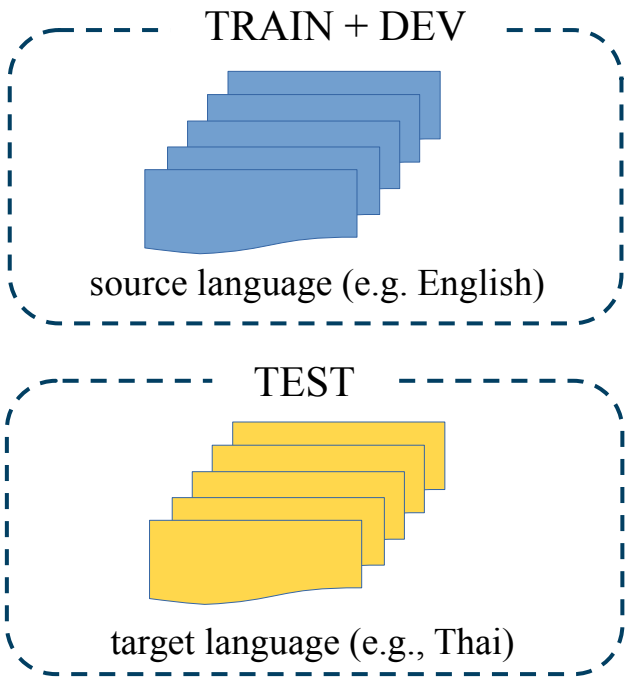
Increasingly language-diverse & polyglot online user community

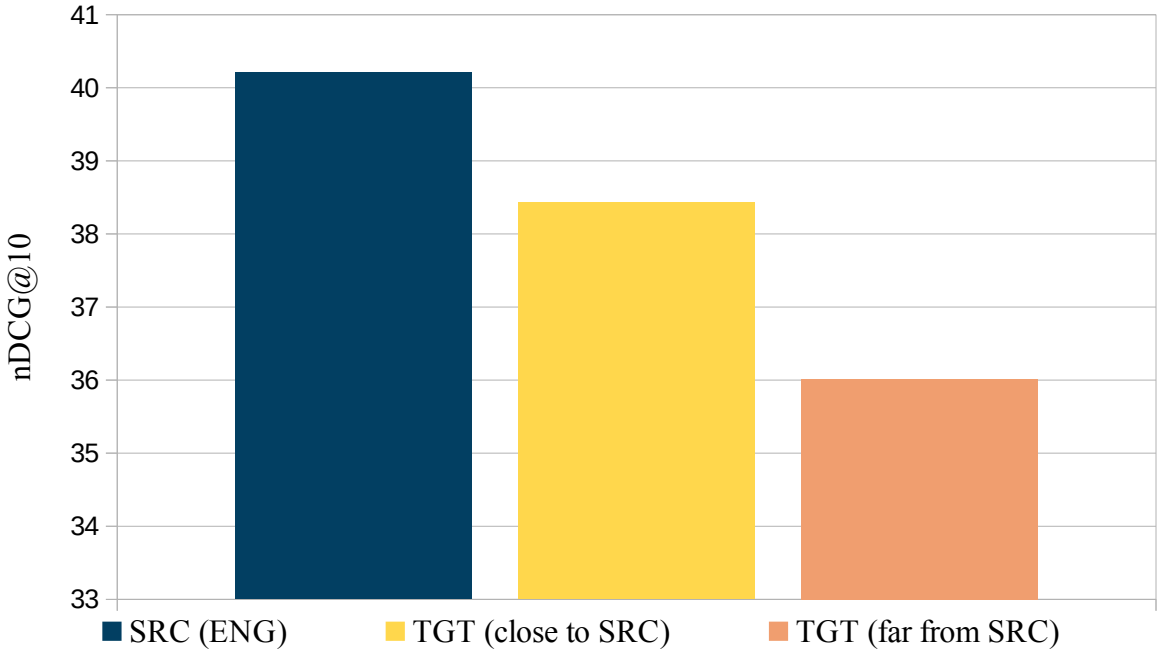
Neural news recommenders (NNRs) need to:

1. Generate **suitable, balanced, diverse recommendations** for **users irrespective of language**
2. Perform **accurately** in **cold-start scenarios** (e.g., no news data, no user click logs)

Multilinguality in News Recommendation

Zero-shot Cross-lingual Transfer (ZS-XLT)





Considerable performance loss in ZS-XLT recommendation

Multilinguality in News Recommendation

NNRs typically fine-tune the backbone language model (LM) on task-specific data

Resource-intensive task (e.g., fine-tuning for too many languages)

Cold-start scenario: little / no news-click data available about new users

Guaraní

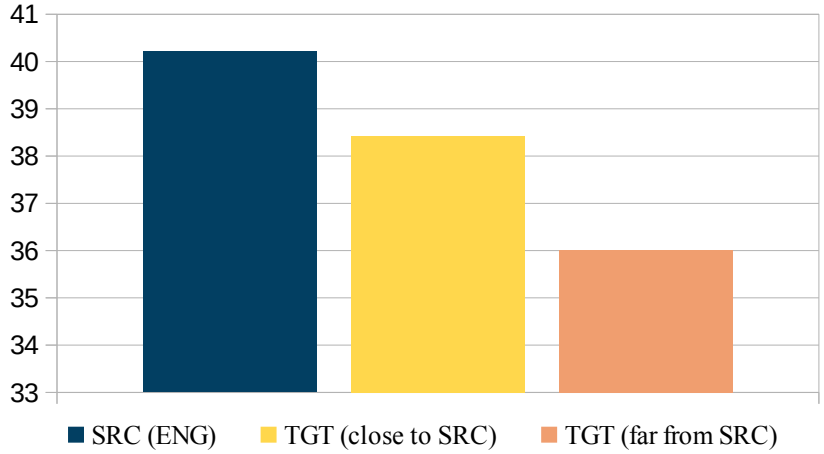
In practice, little data is available for some target languages (e.g., low-resource ones)

Fine-tuning the backbone LM is often infeasible in practice

Multilinguality in News Recommendation

News Recommendation Needs Specialized LMs!

Considerable performance loss in cross-lingual transfer



Fine-tuning the backbone LM is often infeasible in practice

Resource-intensive task

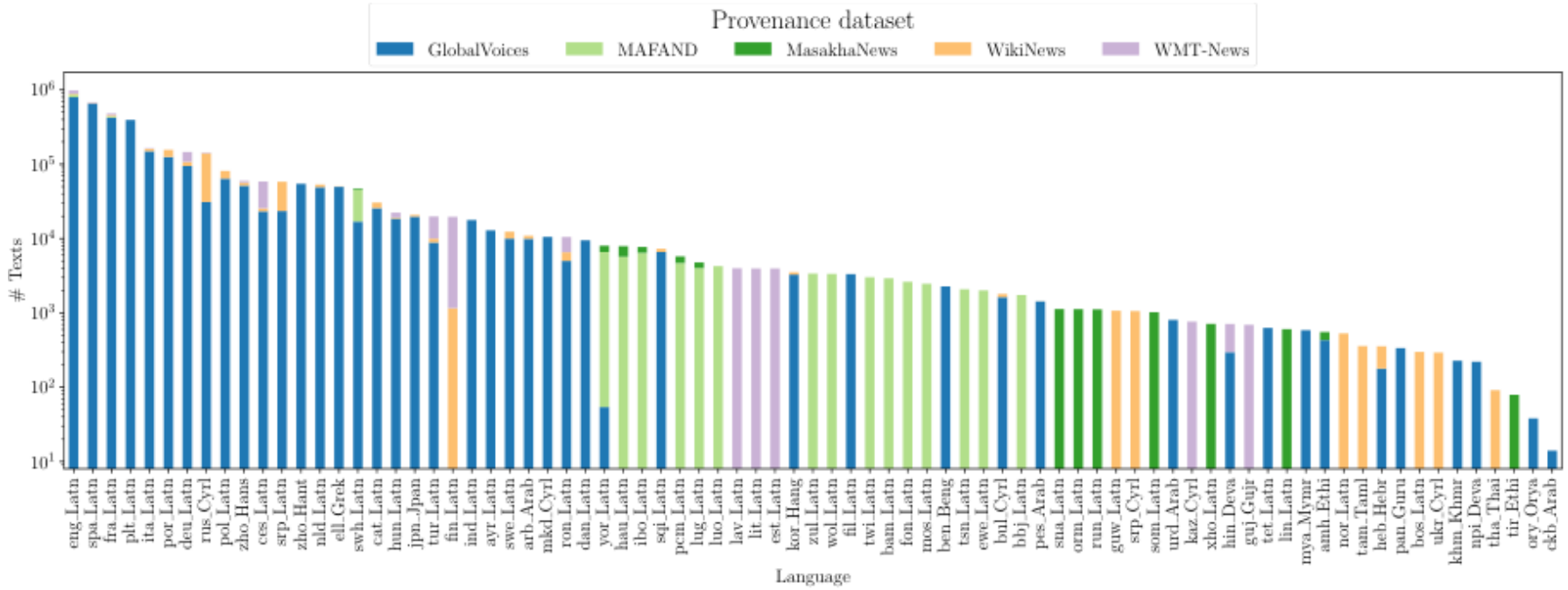
Requires task-specific data

NaSE: massively multilingual sentence encoder (LaBSE), adapted to the news domain with auto-encoding & machine translation objectives

Multilingual Corpora for Domain Adaptation

Polynews

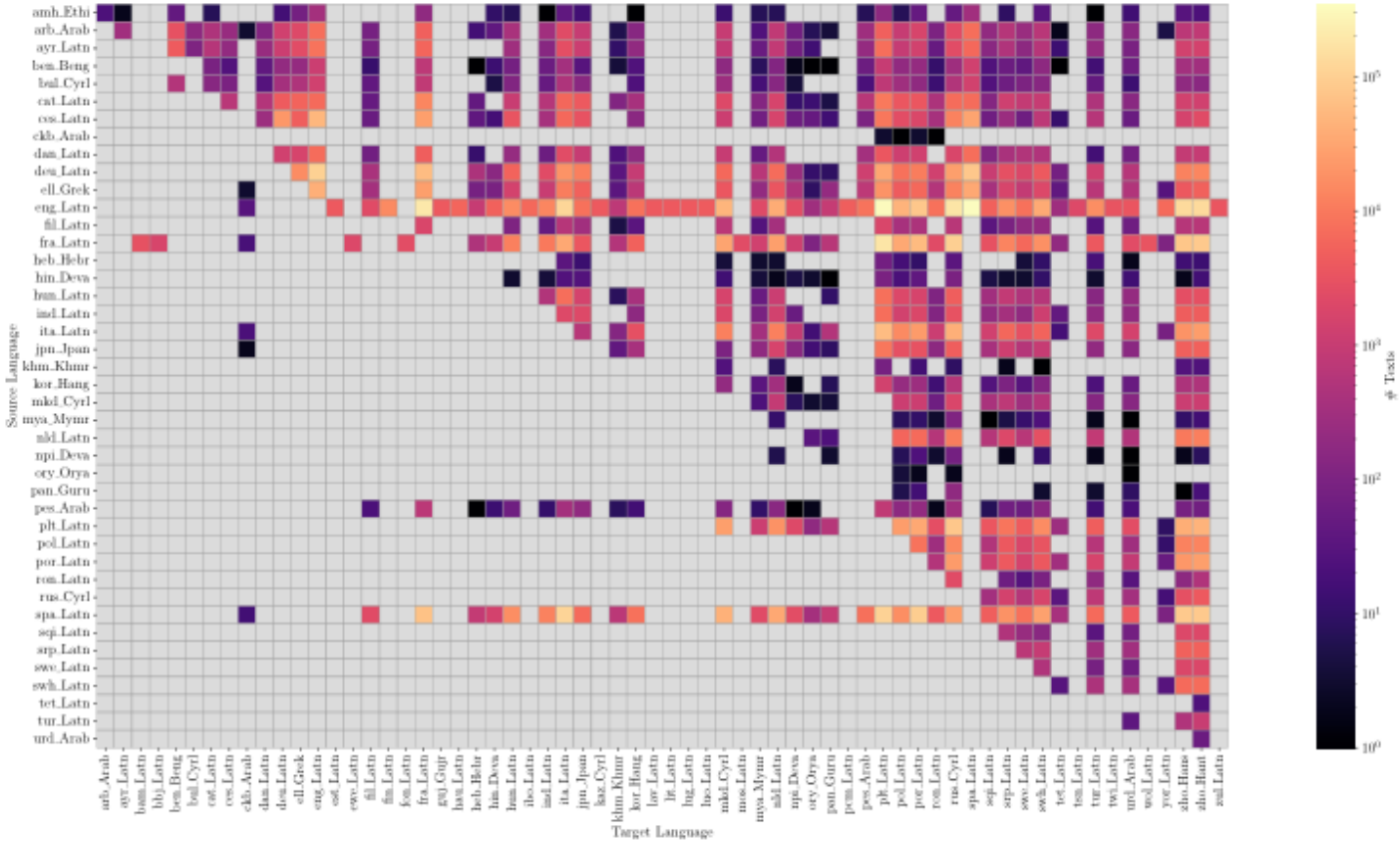
- Compiled from 5 sources
- Data cleaning: duplicate removal, language detection, short text removal, MinHash near de-duplication
- Size: 3.9 million news
- 77 languages, 19 scripts



Multilingual Corpora for Domain Adaptation

PolynewsParallel

- Compiled from 3 **parallel** sources (e.g., MAFAND, WMT-News, Global Voices)
- Data cleaning: duplicate removal, language detection, short text removal, MinHash near de-duplication
- Size: 5.3 million news
- 64 languages, 17 scripts



NaSE: News-adapted Sentence Encoder

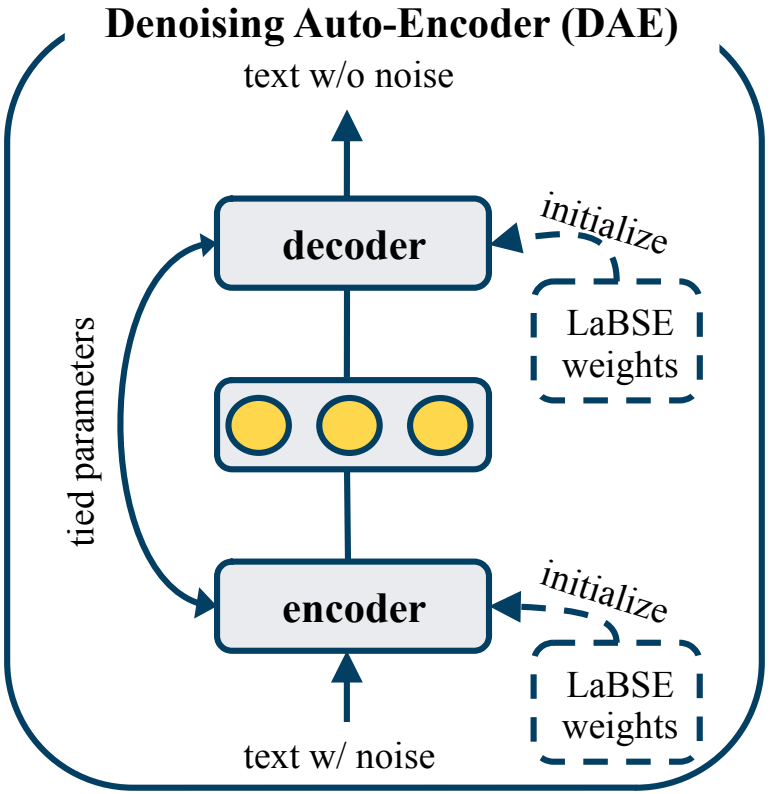
Domain Adaptation

- Sequence-to-sequence training of sentence encoder (initialized with LaBSE weights) on the multilingual corpora

NaSE: News-adapted Sentence Encoder

Domain Adaptation

- Sequence-to-sequence training of sentence encoder (initialized with LaBSE weights) on the multilingual corpora

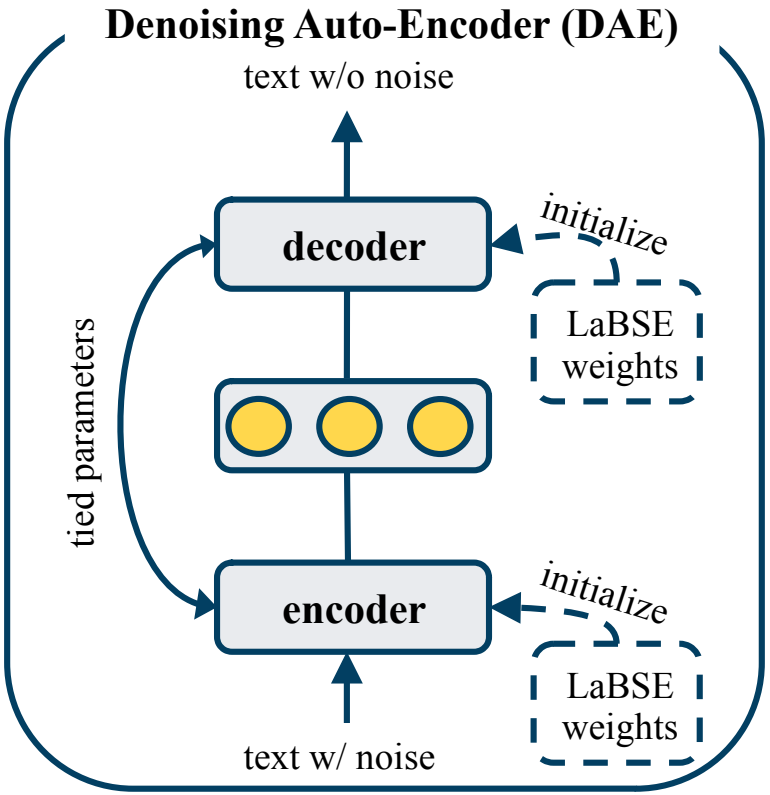


Training Objectives

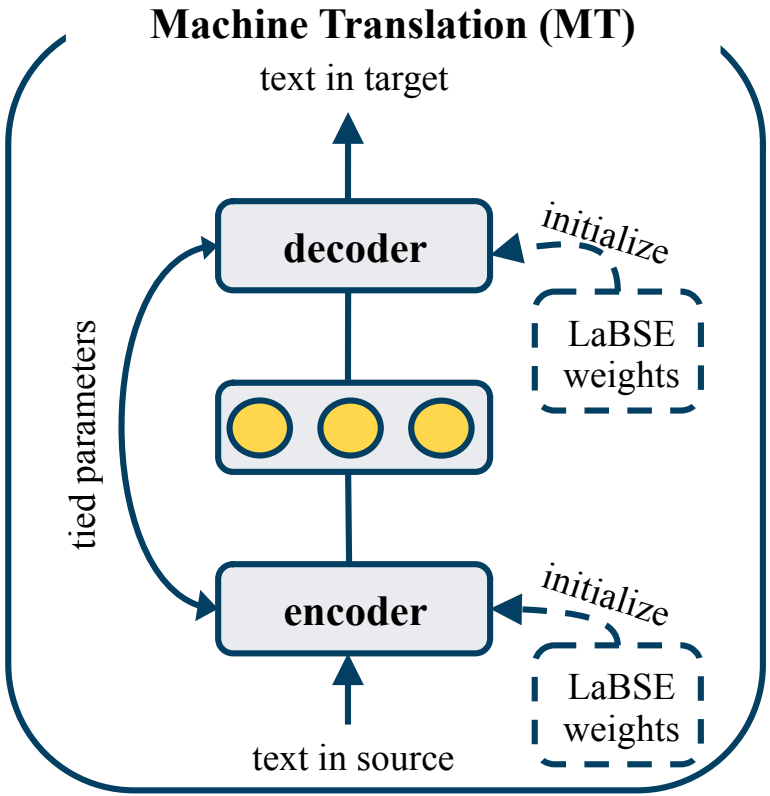
NaSE: News-adapted Sentence Encoder

Domain Adaptation

- Sequence-to-sequence training of sentence encoder (initialized with LaBSE weights) on the multilingual corpora



Training Objectives



NaSE: News-adapted Sentence Encoder

Training Details

- Validation on cross-lingual news recommendation:
  - News encoder: *frozen* NaSE encoder
  - User encoder: late fusion (mean-pooling of dot-product scores between candidate and clicked news embeddings)

- Validation data: small MIND (English) and multilingual xMIND (14 languages, machine-translated news from MIND)

NaSE: News-adapted Sentence Encoder

Training Details

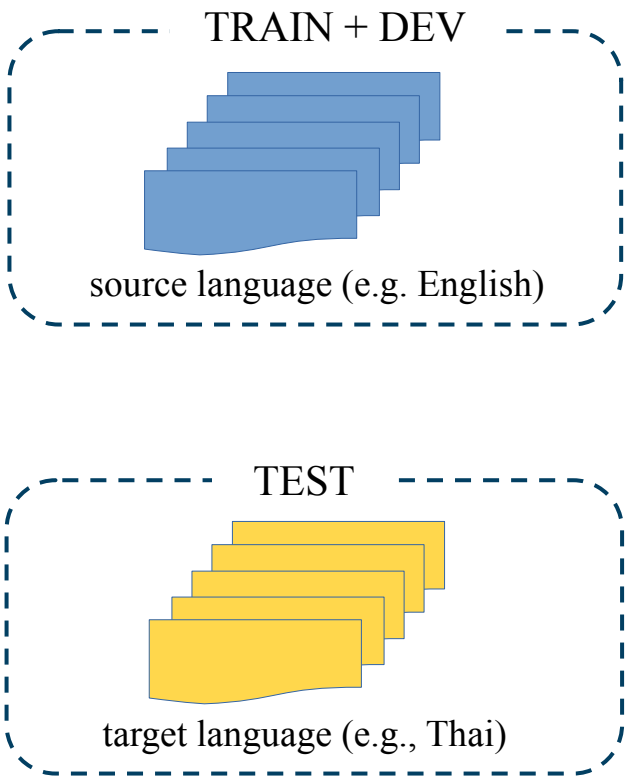
- Validation on cross-lingual news recommendation:
  - News encoder: *frozen* NaSE encoder
  - User encoder: late fusion (mean-pooling of dot-product scores between candidate and clicked news embeddings)

- Validation data: small MIND (English) and multilingual xMIND (14 languages, machine-translated news from MIND)

Variants	Training objectives	PolyNews	PolyNewsParallel
NaSE_{DAE}	DAE w/ corrupted input	✓	✗
NaSE_{MT}	MT (source → target)	✗	✓
$\text{NaSE}_{\text{DAE+MT}}$	DAE or MT (per batch)	✗	✓
$\text{NaSE}_{\text{DAE} \rightarrow \text{MT}}$ (NaSE)	**DAE, then MT (sequentially)**	✓	✓

Zero-Shot Cross-lingual Transfer

Task: ZS-XLT Recommendation



Setup: Backbone LMs

NE Backbone	Type	#Params
XLM-RoBERTa$_{large}$	Language model	559 M
LaBSE	Sentence encoder	471 M
NaSE		471 M

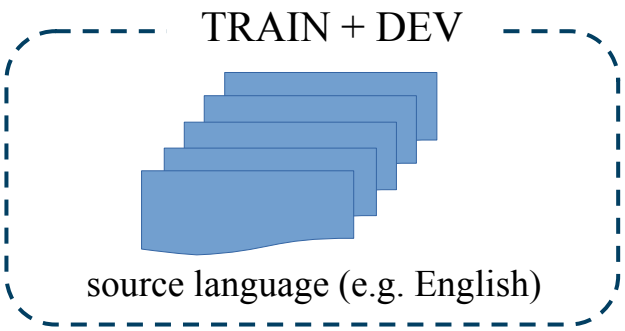
Setup: Data

- MIND (small)
- xMIND (small): statistics *per language* (i.e., 14 languages

	Train	Validation	Test
# News	51,282	51,282	42,416
# Users	45,214	19,703	48,593
# Impressions	124,229	29,498	70,938

Zero-Shot Cross-lingual Transfer

Task: ZS-XLT Recommendation



Setup: Backbone LMs

NE Backbone	Type	#Params
XLM-RoBERTa$_{large}$	Language model	559 M
LaBSE	Sentence encoder	471 M
NaSE		471 M

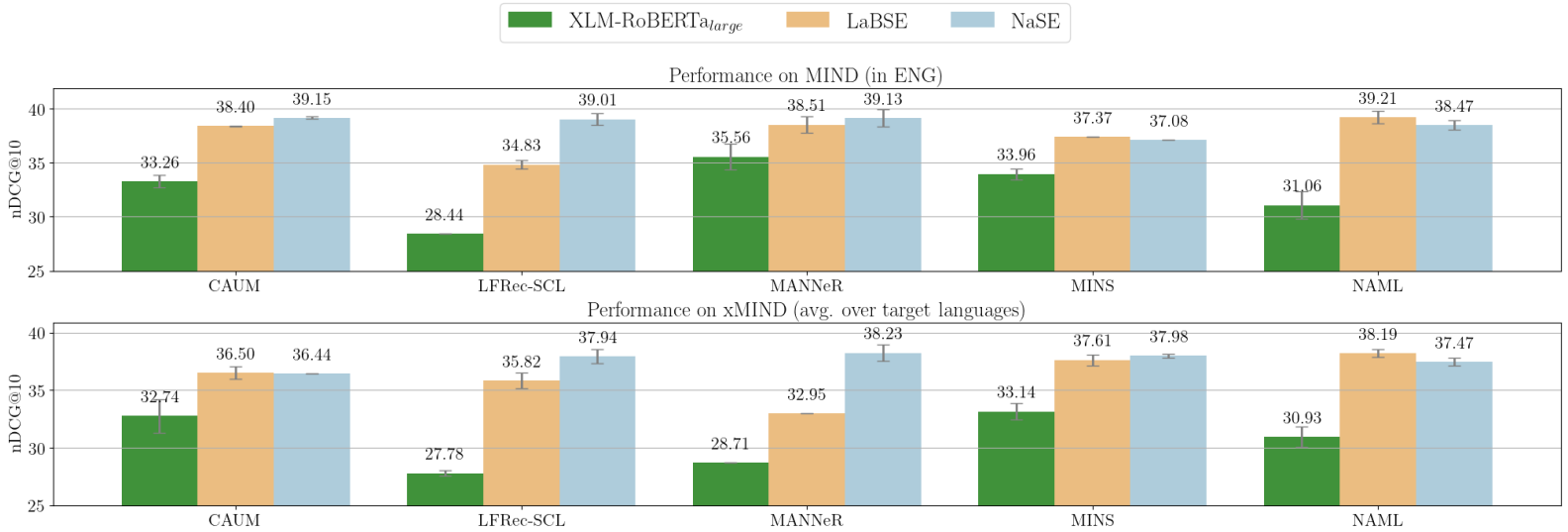
Setup: Models

- Variety of news encoders & user encoders
- LFRec-SCL: **strong baseline** w/ **late fusion (LF)** as user encoder

Zero-Shot Cross-lingual Transfer

Frozen News Encoder (NE)

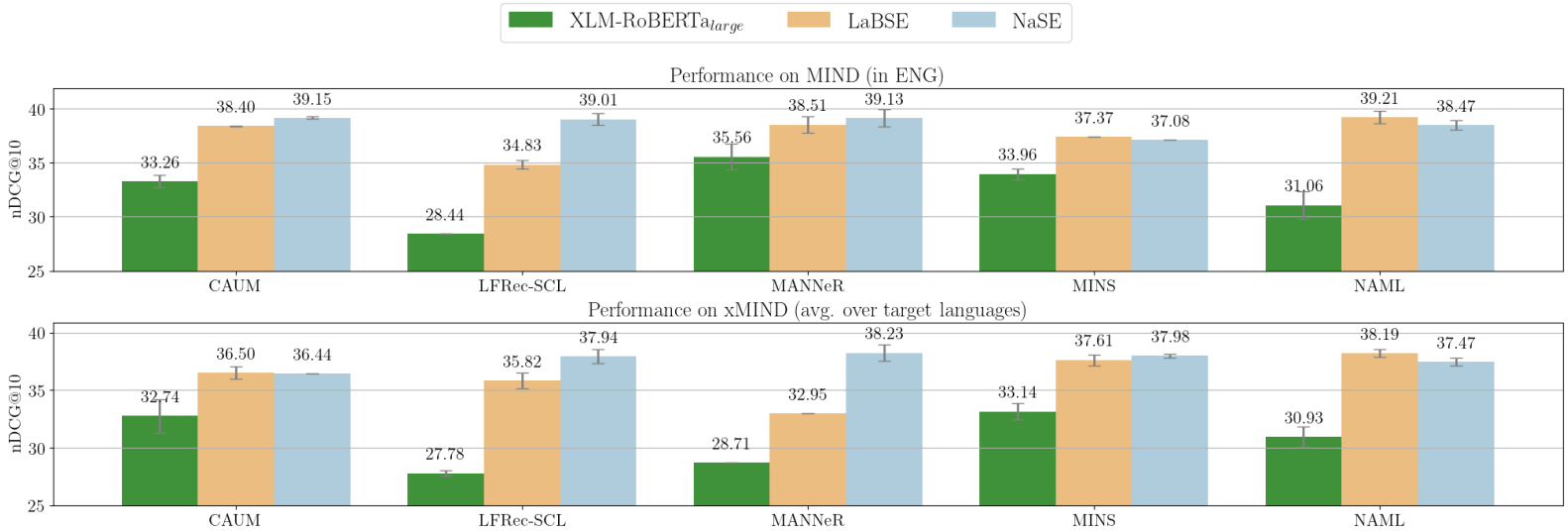
Setup: No updates to backbone LM, only to other (fewer) trainable parameters



Zero-Shot Cross-lingual Transfer

Frozen News Encoder (NE)

Setup: No updates to backbone LM, only to other (fewer) trainable parameters

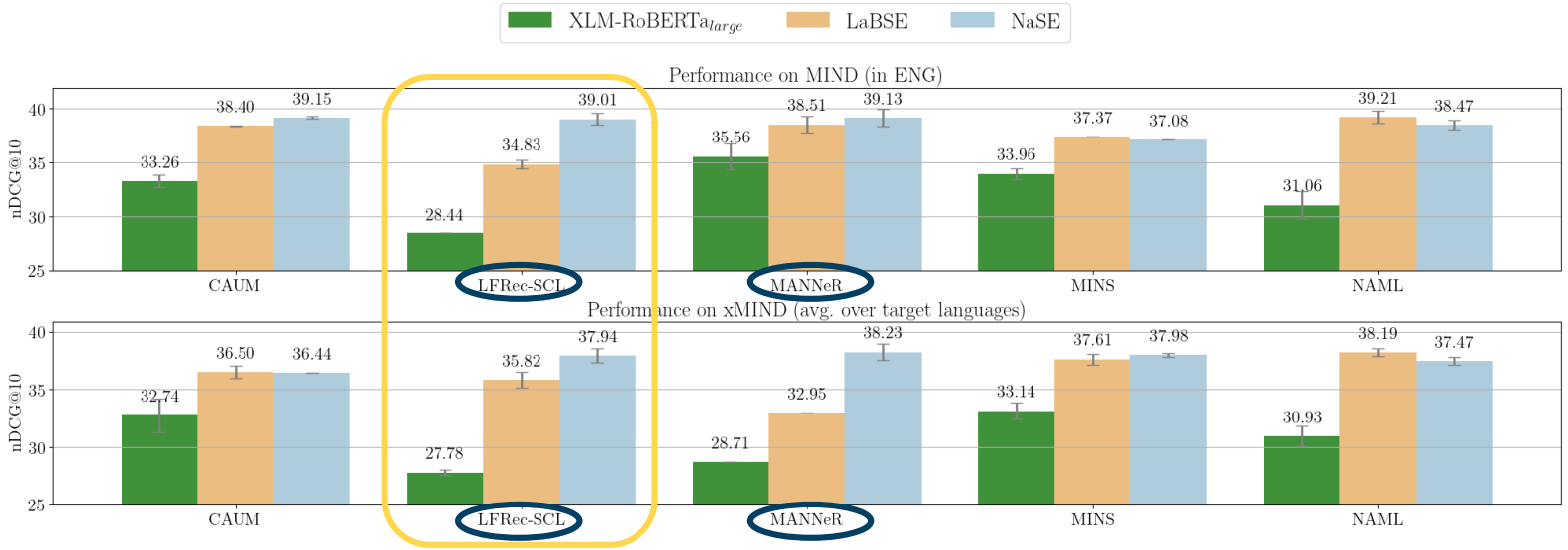


- XLM-RoBERTa$_{large}$-based recommenders yield the weakest performance across all language.
- NaSE vs. LaBSE embeddings: **+ 2.58%** on **English** & **+ 4.17% cross-lingually** (averaged across 14 languages).

Zero-Shot Cross-lingual Transfer

Frozen News Encoder (NE)

Setup: No updates to backbone LM, only to other (fewer) trainable parameters



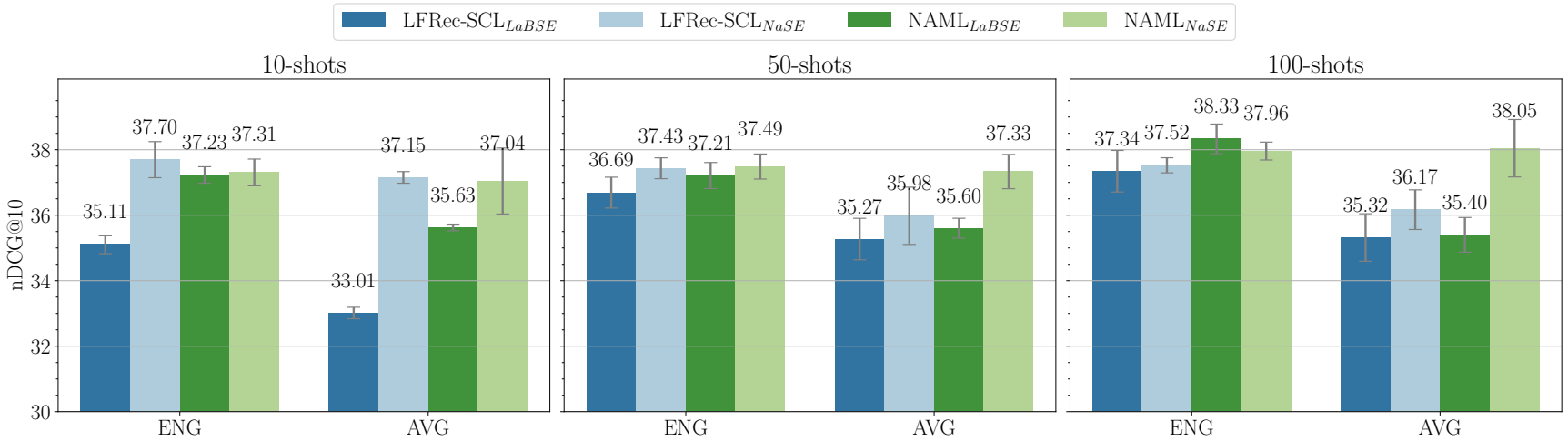
Domain specialization removes the **need for supervised training** of neural news recommenders **on task-specific data.**

Zero-Shot Cross-lingual Transfer

Fine-tuned News Encoder (NE)

Setup: Updates to all trainable parameters on *English task-specific data*

➔ *few* task-specific examples



➔ **NaSE effective** in **ZS-XLT** recommendation in **low-data setups**.

➔ **Fine-tuning** on news recommendation also **leads to domain adaptation**, but **assumes availability** of news & user-click **data**.

Conclusion

1. **Domain-specialization** of a **multilingual sentence encoder** (i.e., **NaSE**) **removes** the **need** for **supervised training** of neural news recommenders.

2. **NaSE** is **highly effective** in **ZS-XLT recommendation** in **cold-start & low-data setups**.

3. **LFRec-SCL: simple & strong baseline** based on **frozen NaSE embeddings & late fusion.**

NaSE **PolyNewsParallel** **PolyNews** **Code** **Contact**