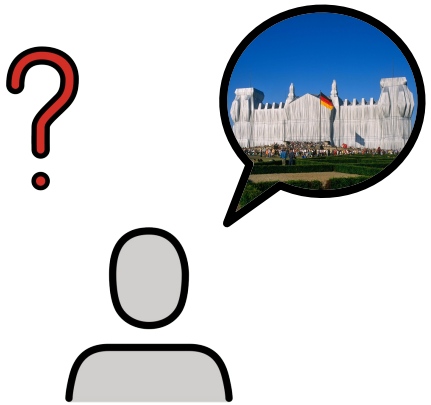
Resource-Lean Transfer Methods for Cross-Lingual Information Retrieval

Disputation, 11.07.24, Robert Litschko

Photo Credit: Picturebank / Alamy Stock Photo


Photo Credit: Picturebank / Alamy Stock Photo

Information Need

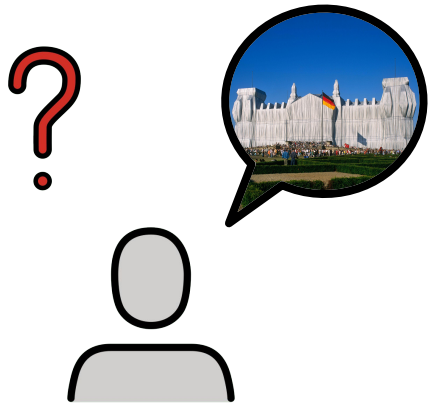


What?

When?

Why?

Information Need



What?

When?

Why?

Christo and Jeanne-Claude

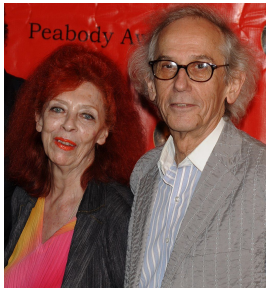
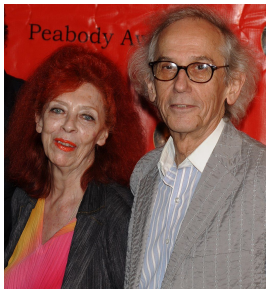


Image Source: https://en.wikipedia.org/wiki/Christo_and_Jeanne-Claude, CC-BY-2.0

Query

Christo and
Jeanne-Claude

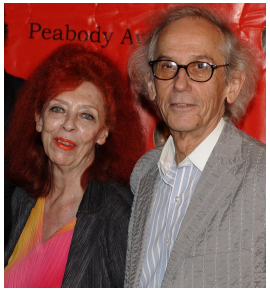


Example taken from the CLEF 2003 dataset (Braschler, 2004).

English Document

Christo wraps German Reichstag.

Christo and Jeanne-Claude



Glasgow Herald (20.06.1995)

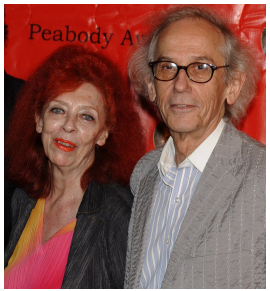
WORKERS lower a **giant panel of cloth** over the entrance to the Reichstag in Berlin, helping Hungarian artist Christo to fulful a dream of 24 years. Christo and his wife Jeanne Claude are using a **#4.6m loan secured on their private art collection** to fund the work of covering the former German Parliament in silver fabric. […]

Example taken from the CLEF 2003 dataset (Braschler, 2004).

German Document

Christo wraps German Reichstag.

Christo and Jeanne-Claude

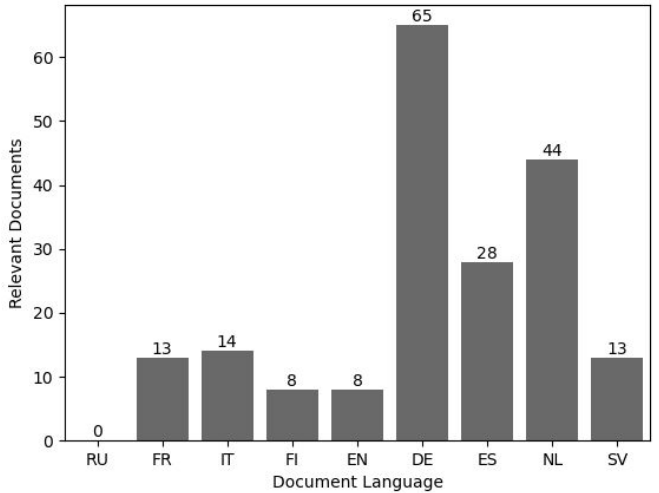


Frankfurter Rundschau (23.01.1994)

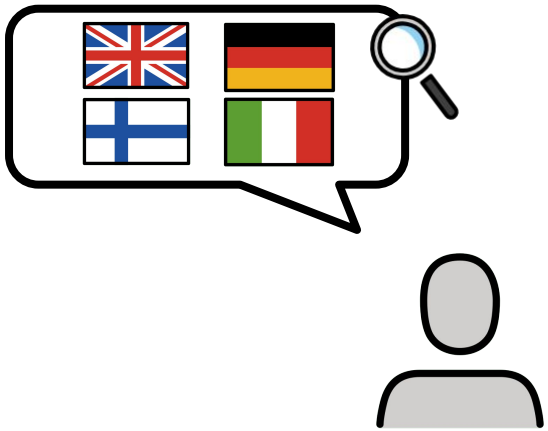
Des Künstlers Plan, den Berliner Reichstag zu verpacken, **ist bei Umweltschützern auf Kritik gestoßen**. Unter Umweltgesichtspunkten, so urteilt Michael Braungart, Vorsitzender des Hamburger Umweltinstituts, ist das Projekt eine Schweinerei. [...]

Example taken from the CLEF 2003 dataset (Braschler, 2004).

Information Asymmetry



Linguistic Diversity



EU Charter of Fundamental Right

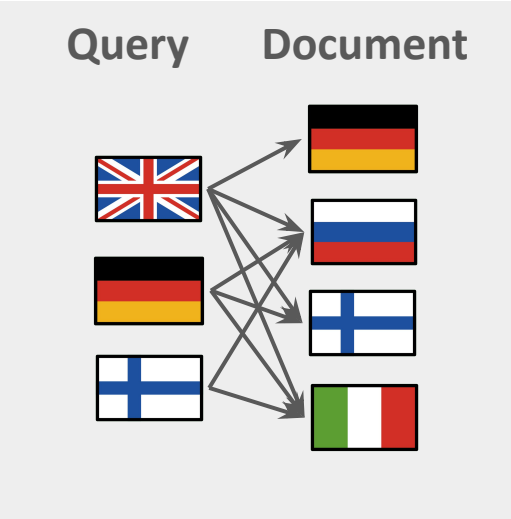
"The Union shall respect cultural, religious and linguistic diversity"

Universal Declaration of Human Rights

"everyone has the right to [...] seek, receive and impart information and ideas through any media and regardless of frontiers."

Cross-Lingual Information Retrieval (CLIR)

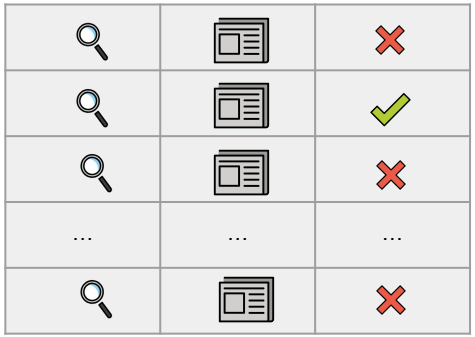
-  Language ≠  Language
- Bridge Information Asymmetry
- Promote Linguistic Diversity 



Why Resource-Lean Transfer?

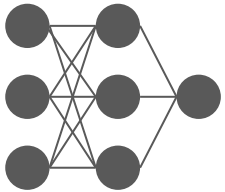
CLIR Training data

Direct Supervision



Supervised CLIR*

Resource-hungry

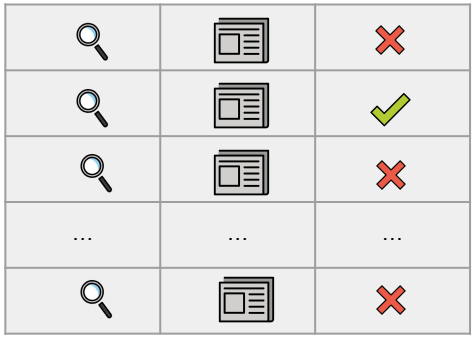


*and Neural Machine Translation

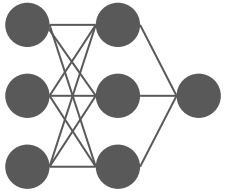
Why Resource-Lean Transfer?

CLIR Training data

Direct Supervision

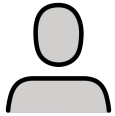


>7k languages



Supervised CLIR*

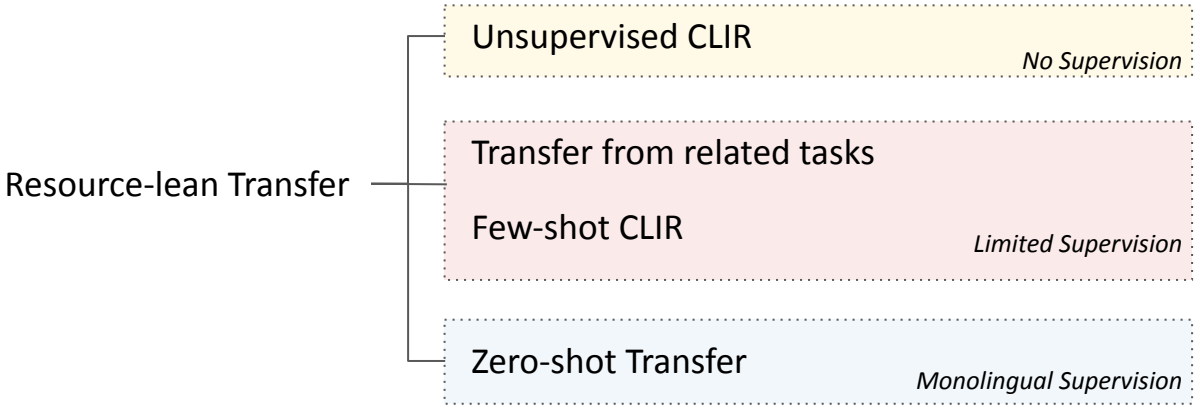
Resource-hungry



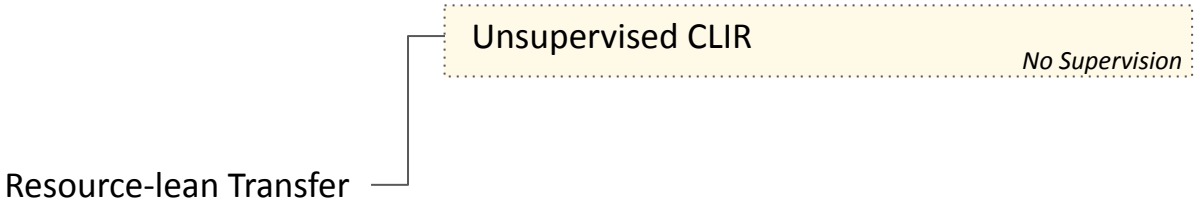
Human relevance annotations are too expensive.

*and Neural Machine Translation

Contribution: Large-Scale Empirical Evaluation



Contribution: Large-Scale Empirical Evaluation



Contribution: Large-Scale Empirical Evaluation

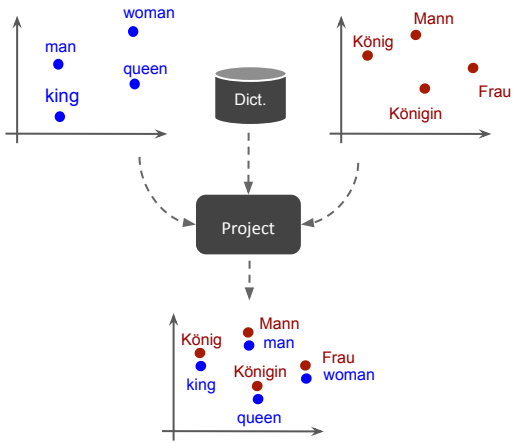
Resource-lean Transfer

Unsupervised CLIR

- Cross-lingual Word Embeddings (CLWE)

No Supervision

- Litschko, R., Glavaš, G., Ponzetto, S. P., & Vulić, I. *Unsupervised cross-lingual information retrieval using monolingual data only.* In Proceedings of **SIGIR'18**.
- Litschko, R., Glavaš, G., Vulic, I., & Dietz, L.. *Evaluating resource-lean cross-lingual embedding models in unsupervised retrieval.* In Proceedings of **SIGIR'19**.



Contribution: Large-Scale Empirical Evaluation

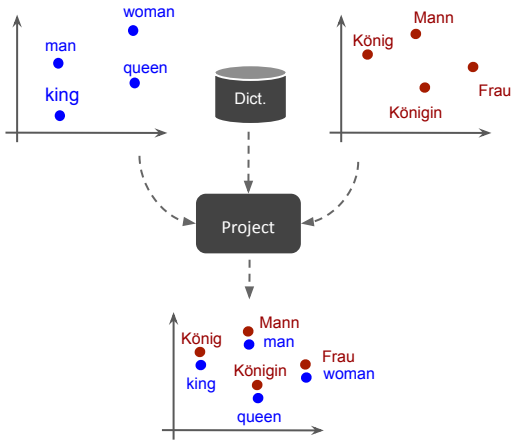
Resource-lean Transfer

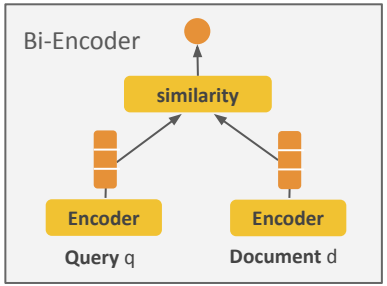
Unsupervised CLIR

- Cross-lingual Word Embeddings (CLWE)

No Supervision

- Litschko, R., Glavaš, G., Ponzetto, S. P., & Vulić, I. *Unsupervised cross-lingual information retrieval using monolingual data only.* In Proceedings of **SIGIR'18**.
- Litschko, R., Glavaš, G., Vulic, I., & Dietz, L.. *Evaluating resource-lean cross-lingual embedding models in unsupervised retrieval.* In Proceedings of **SIGIR'19**.





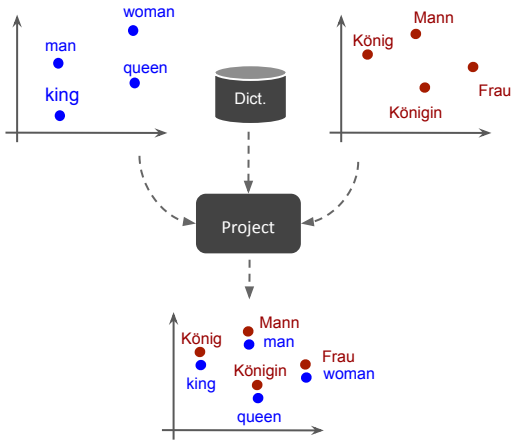
Contribution: Large-Scale Empirical Evaluation

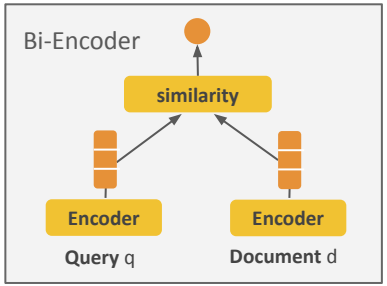
Resource-lean Transfer

Unsupervised CLIR
- Cross-lingual Word Embeddings (CLWE)

No Supervision

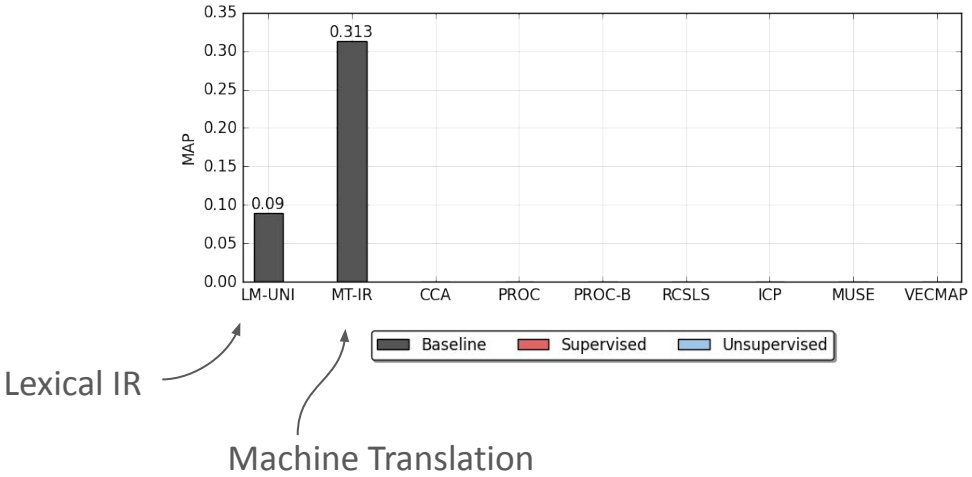
- Litschko, R., Glavaš, G., Ponzetto, S. P., & Vulić, I. *Unsupervised cross-lingual information retrieval using monolingual data only.* In Proceedings of **SIGIR'18**.
- Litschko, R., Glavaš, G., Vulic, I., & Dietz, L.. *Evaluating resource-lean cross-lingual embedding models in unsupervised retrieval.* In Proceedings of **SIGIR'19**.



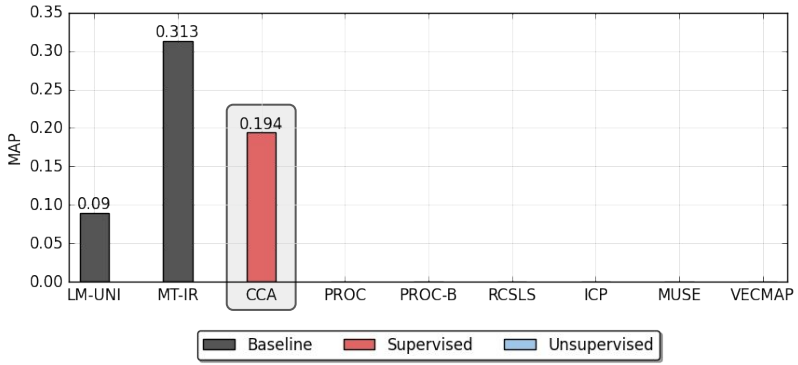


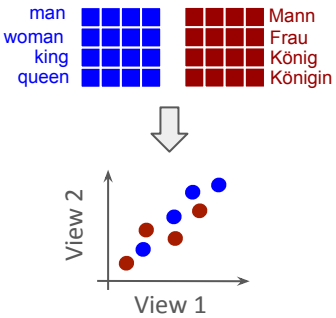
RQ-1: How well do CLWEs work?

Unsupervised CLIR with CLWEs



Unsupervised CLIR with CLWEs

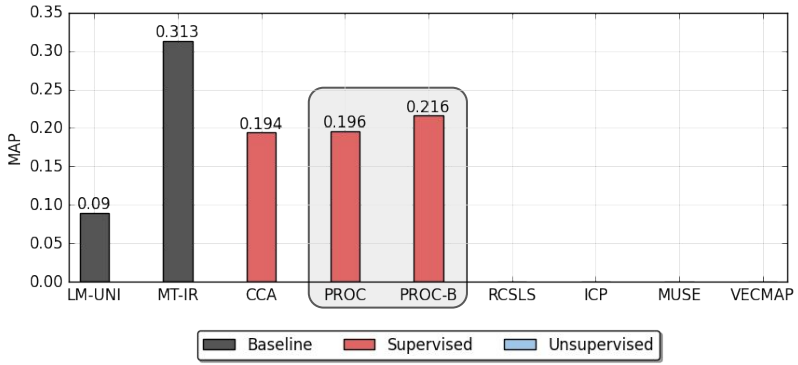


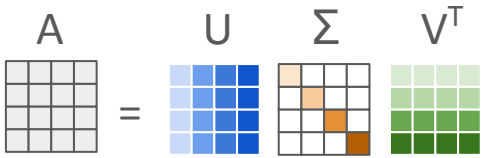


CCA: Maximize Correlation between X_S and X_T.

Model proposed by Fauqi and Dyer (2014). Illustration adopted from (Bansal and Gimpel, 2015).

Unsupervised CLIR with CLWEs

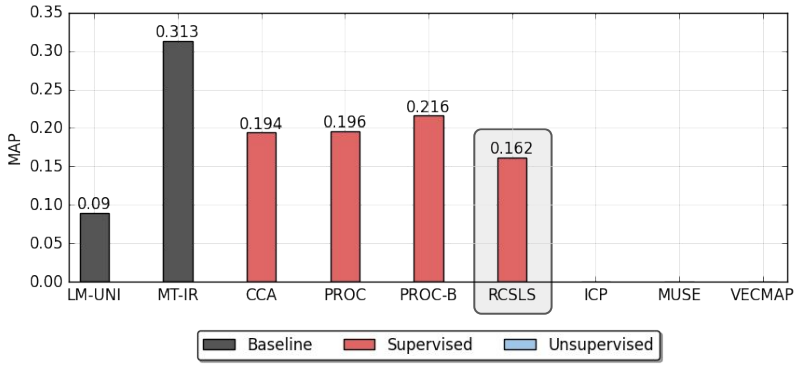


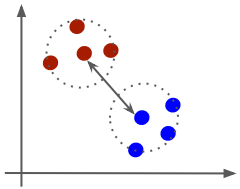


Proc(-B): Decompose Similarity Matrix.

Model proposed by Smith et al. (2017)

Unsupervised CLIR with CLWEs

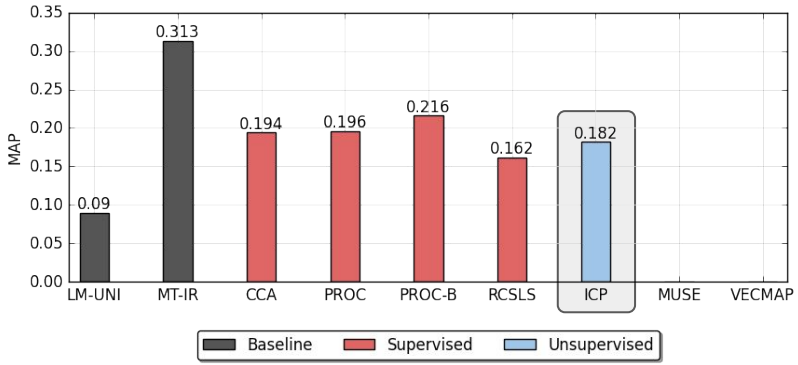


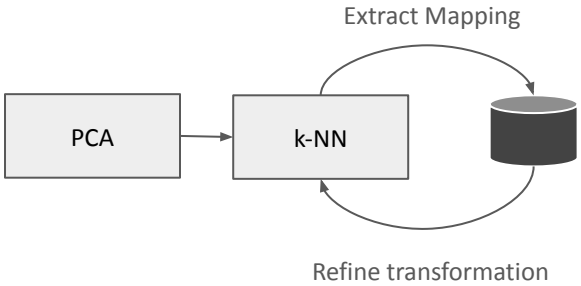


RCSLS: Minimize cosine distance + adjust for hubness.

Model proposed by Hoshen et al. (2018)

Unsupervised CLIR with CLWEs

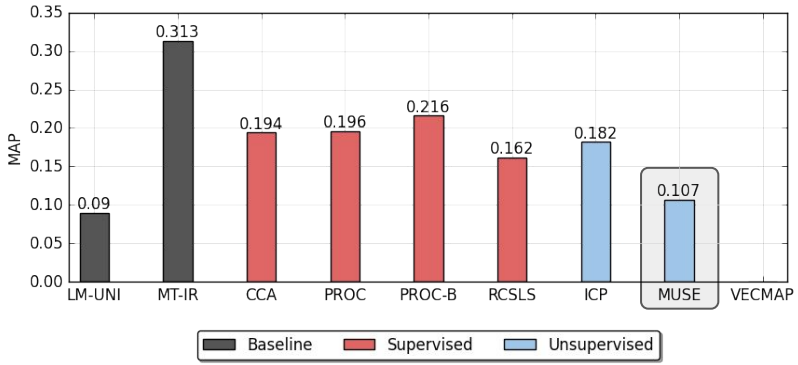


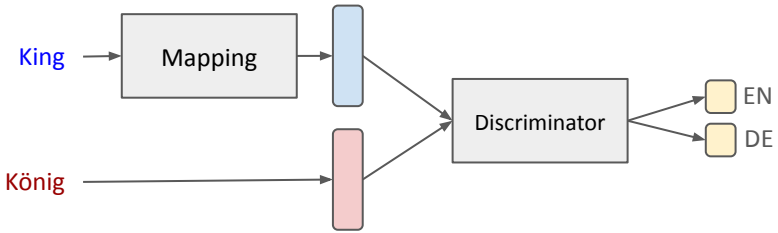


ICP: Cross-lingual k-NN, refine mapping, repeat.

Model proposed by Hoshen et al. (2018)

Unsupervised CLIR with CLWEs

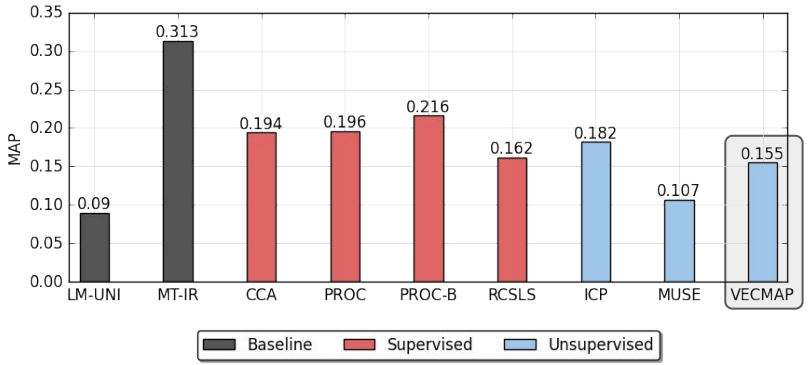


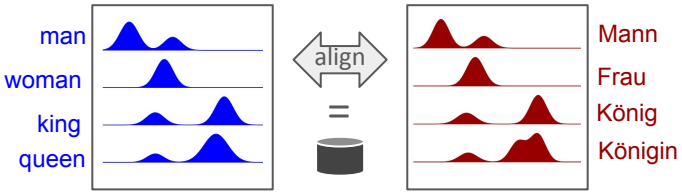


MUSE: Adversarial Learning

Model proposed by Conneau et al. (2018)

Unsupervised CLIR with CLWEs

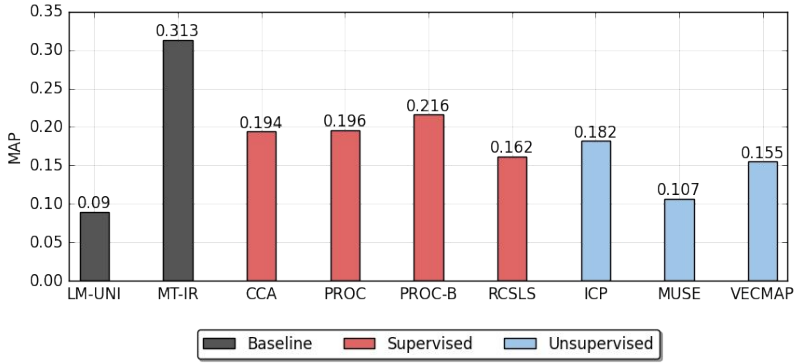




VecMap: Align monolingual similarity distributions.

Model proposed by Artetxe et al. (2018)

Results



- CLWEs outperform lexical baseline and fall behind Machine Translation.
- Supervised CLWE outperform unsupervised CLWE methods.

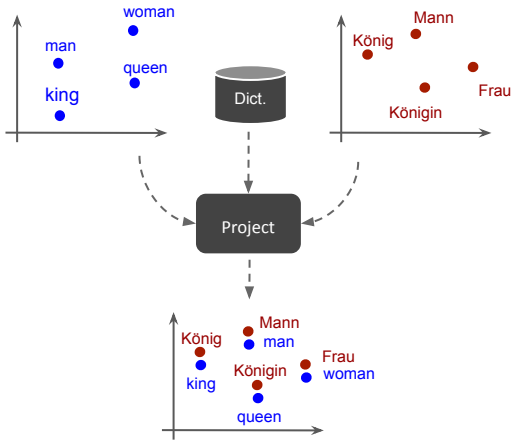
Contribution: Large-Scale Empirical Evaluation

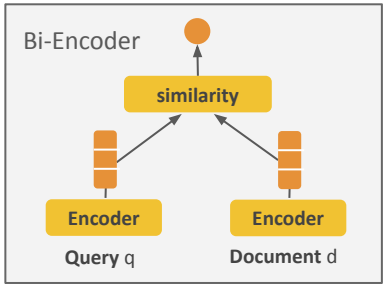
Resource-lean Transfer

Unsupervised CLIR

- Cross-lingual Word Embeddings (CLWE)

No Supervision





Contribution: Large-Scale Empirical Evaluation

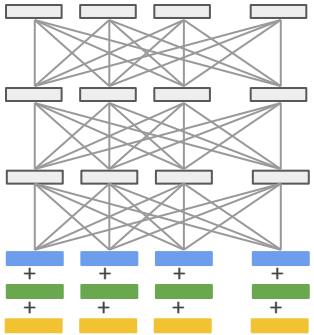
Resource-lean Transfer

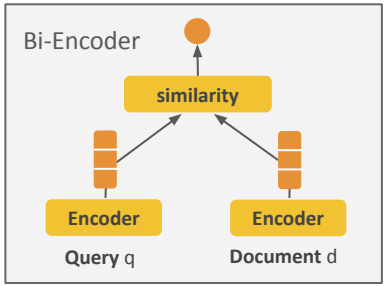
Unsupervised CLIR
- Cross-lingual Word Embeddings (CLWE)
- multilingual Pre-trained Language Models (mPLM)

No Supervision

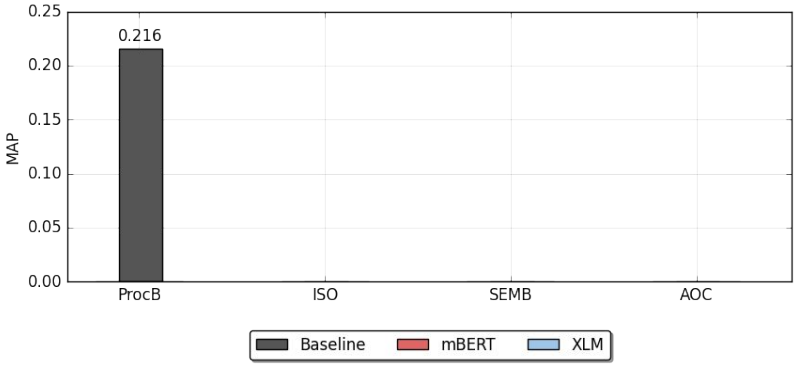
- Litschko, R., Vulić, I., Ponzetto, S. P., & Glavaš, G. (2021). *Evaluating multilingual text encoders for unsupervised cross-lingual retrieval.* In Proceedings of **ECIR'21**.
- Litschko, R., Vulić, I., Ponzetto, S. P., & Glavaš, G. (2022). *On cross-lingual retrieval with multilingual text encoders.* **Information Retrieval Journal**, 25.2 (2022).

RQ-2: Impact of contextualization?

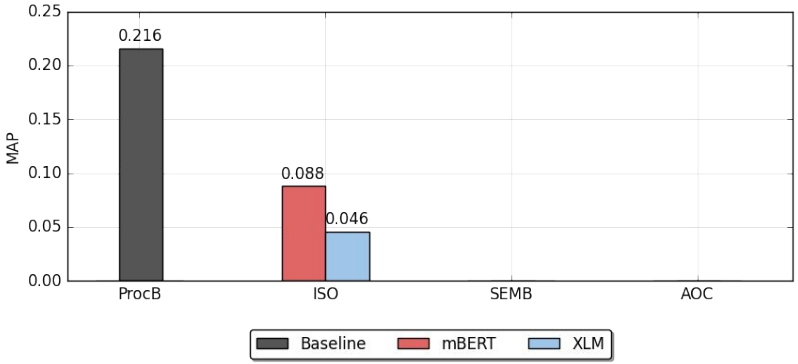


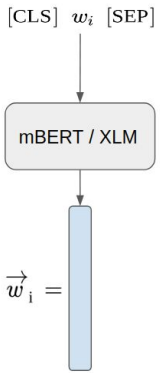


Unsupervised CLIR with mPLM



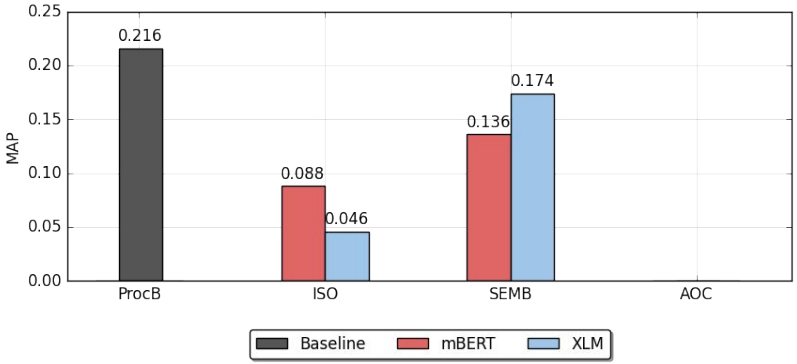
Unsupervised CLIR with mPLM

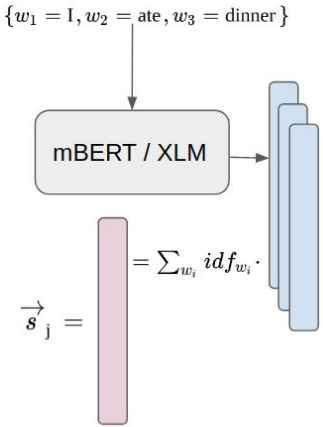




CLWEs from encoding words in **ISO**lation.

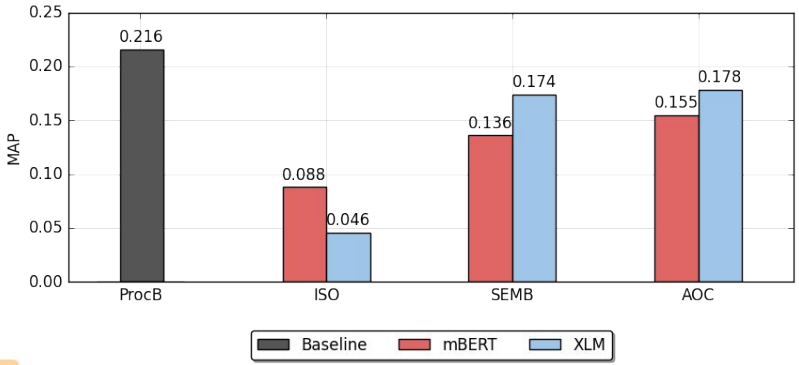
Unsupervised CLIR with mPLM

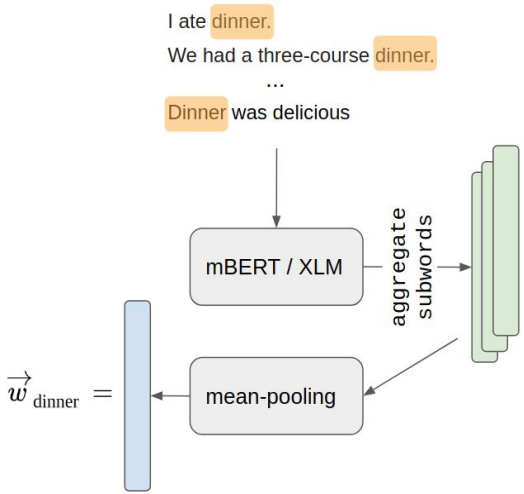




Encode queries and documents similar to **S**entence **EMB**eddings.

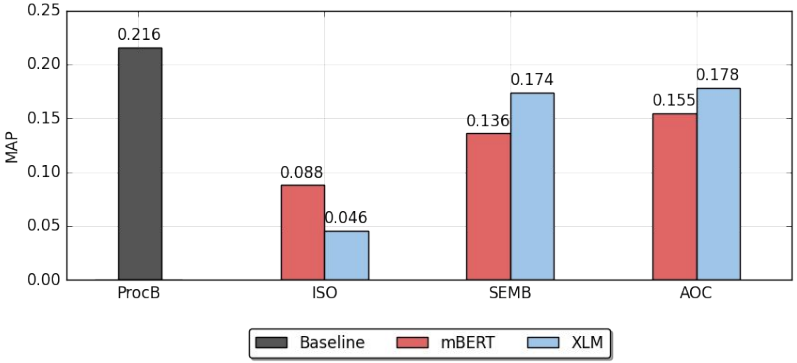
Unsupervised CLIR with mPLM



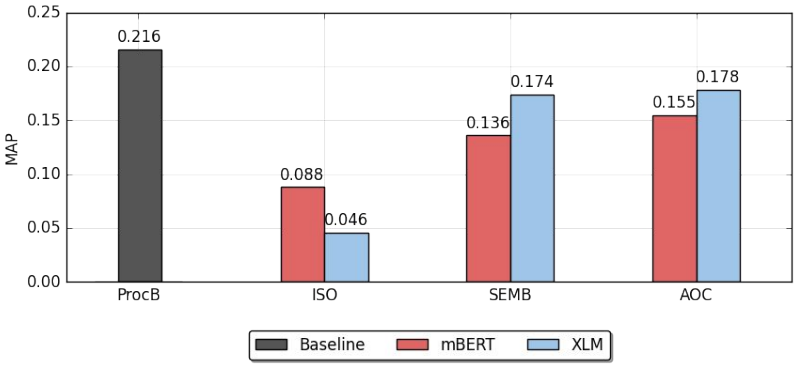


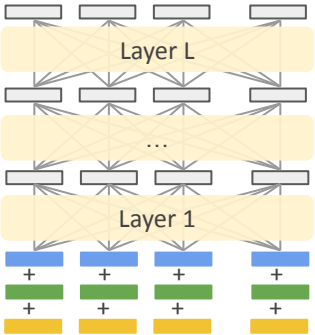
CLWEs from **A**verage **O**ver **C**ontexts Embeddings.

Unsupervised CLIR with mPLM



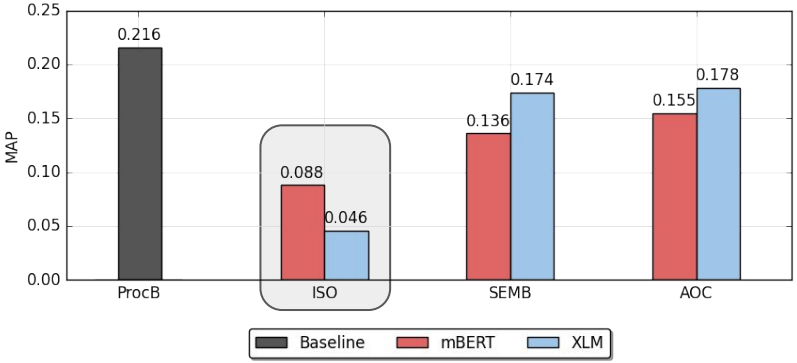
Degree of Contextualization

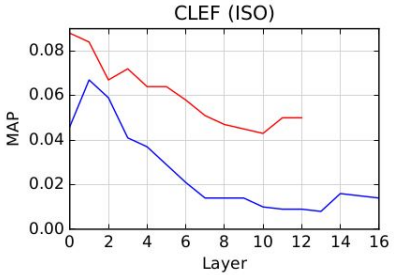




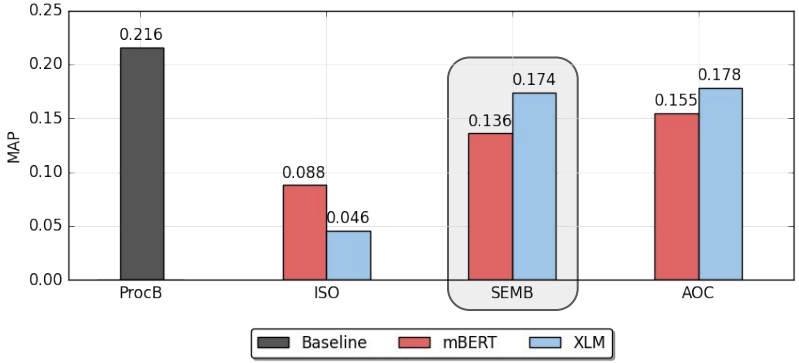
How much **contextualization** do we need?

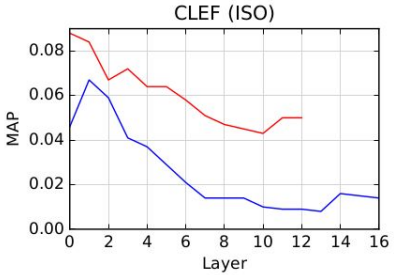
Degree of Contextualization

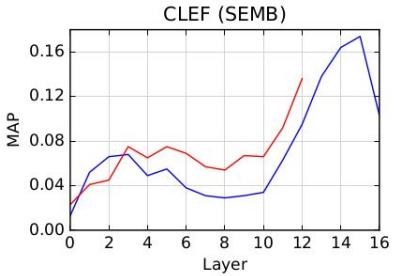




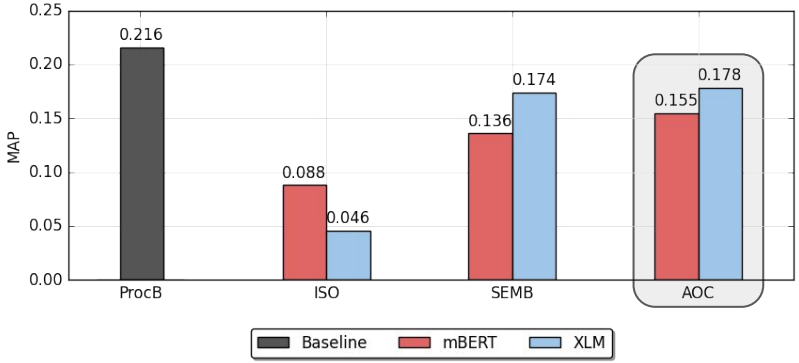
Degree of Contextualization

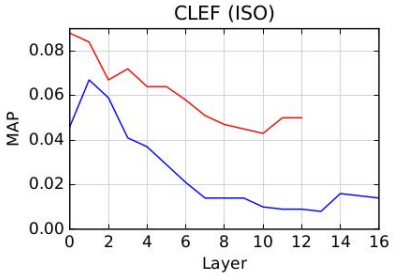


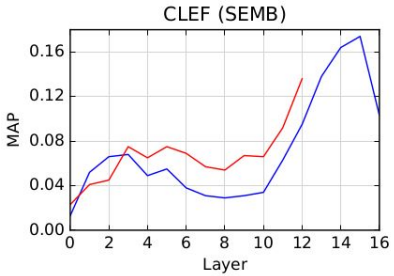


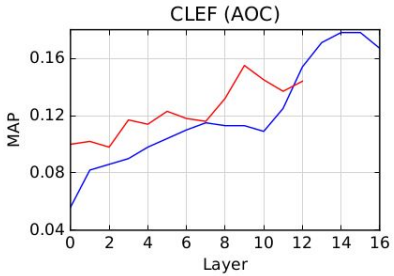


Degree of Contextualization

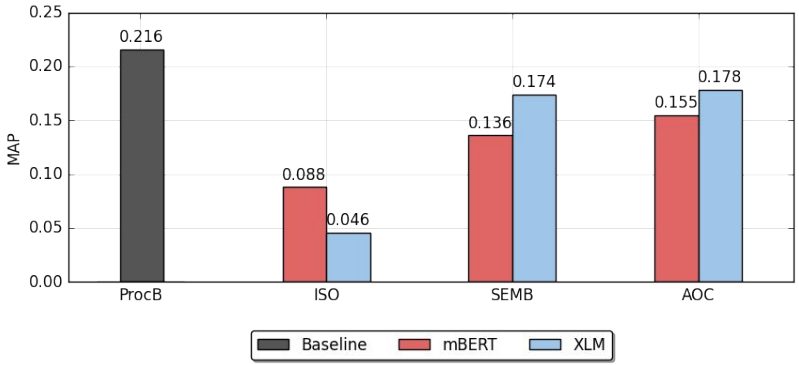






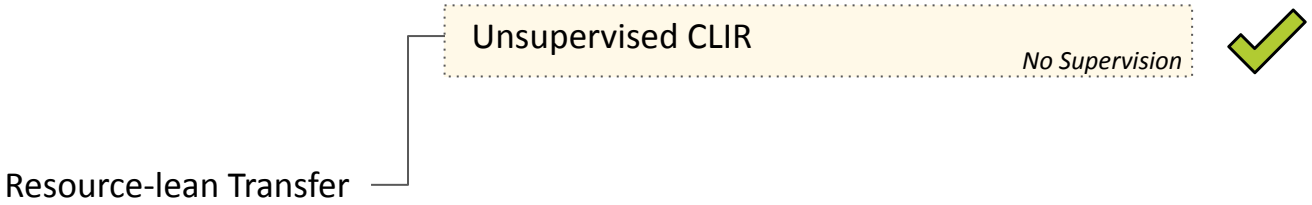


Results

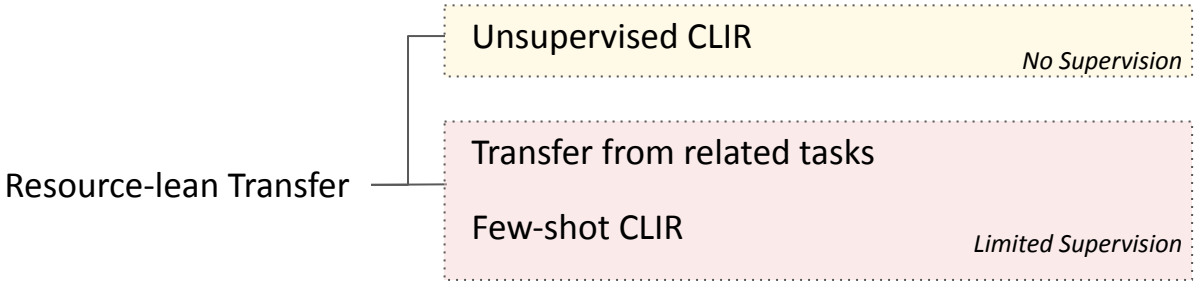


- **RQ-2:** **Weaky aligned** contextual representations do not outperform CLWEs.
  - Lower (upper) layers work better for ISO (AOC, SEMB).

Contribution: Large-Scale Empirical Evaluation



Contribution: Large-Scale Empirical Evaluation

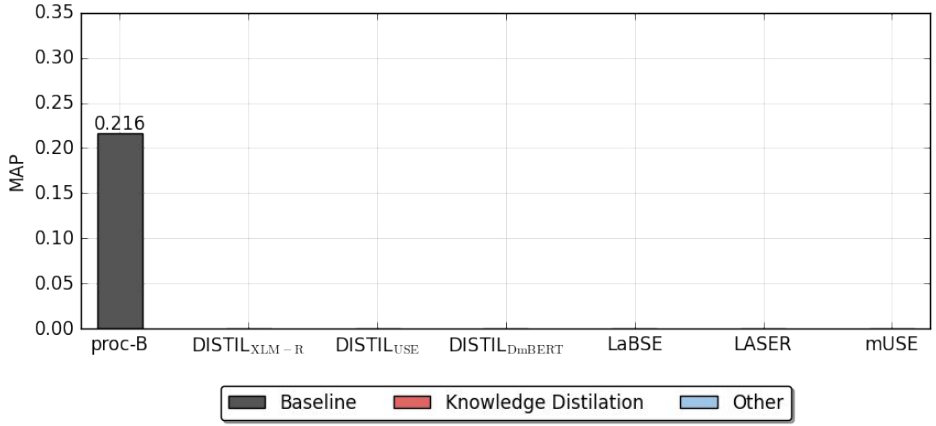


- Litschko, R., Vulić, I., Ponzetto, S. P., & Glavaš, G. (2021). *Evaluating multilingual text encoders for unsupervised cross-lingual retrieval.* In Proceedings of **ECIR'21**.
- Litschko, R., Vulić, I., Ponzetto, S. P., & Glavaš, G. (2022). *On cross-lingual retrieval with multilingual text encoders.* **Information Retrieval Journal**, 25.2 (2022).

RQ-3: How well do multilingual Sentence Encoders perform?

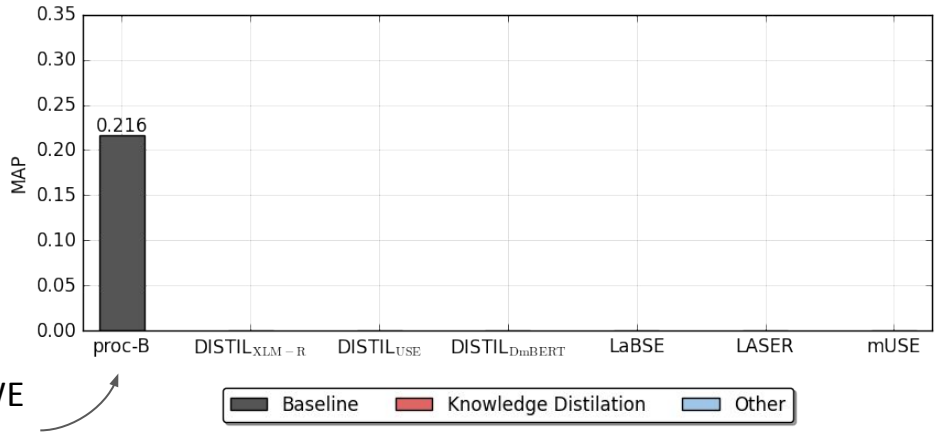
RQ-4: Can we improve their performance with few in-domain data?

Transfer from Related Tasks



- Knowledge Distilation (DISTIL)
- Contrastive Loss (LaBSE)
- Machine Translation (LASER)
- Multi-task Learning (mUSE)

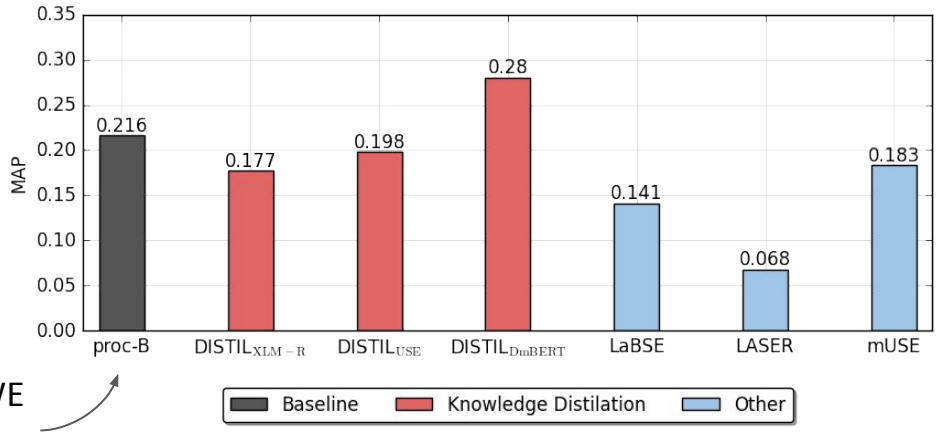
Transfer from Related Tasks



best-performing CLWE method (Proc-B)

- Knowledge Distilation (DISTIL)
- Contrastive Loss (LaBSE)
- Machine Translation (LASER)
- Multi-task Learning (mUSE)

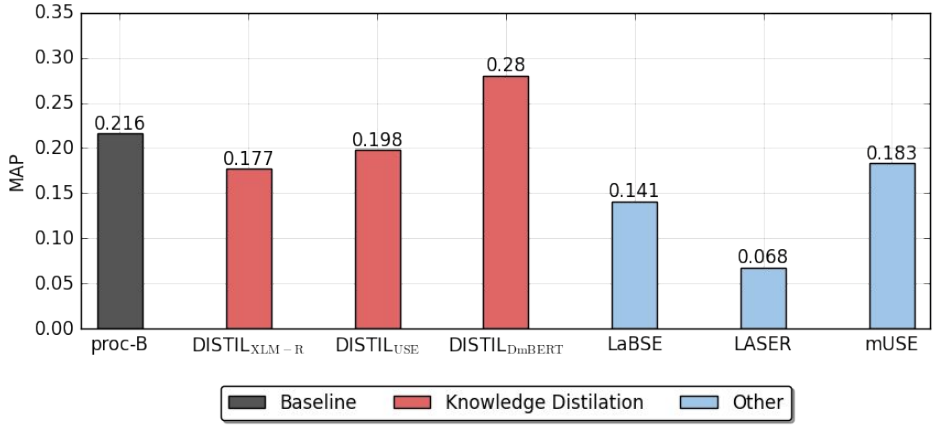
Transfer from Related Tasks



best-performing CLWE method (Proc-B)

- **RQ-4:** On average, multilingual sentence encoders outperform CLWEs.
  - Mixed results w.r.t. best-performing CLWE-based approach.

Context Outside Maximum Sequence Length?

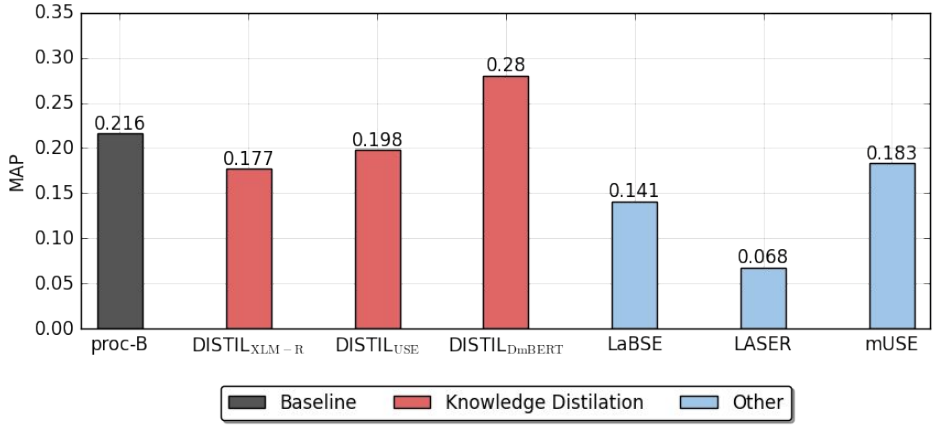


Proc-B

Sentence Encoder

Unfair comparison?

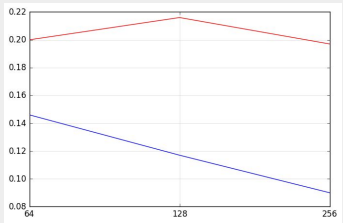
Increase Maximum Sequence Length?



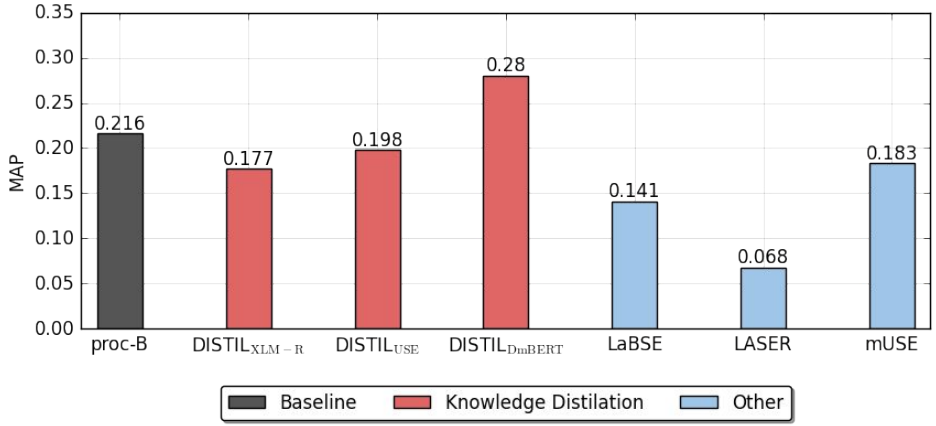
Proc-B

Sentence Encoder

Increase seq. length



Score Top-k Sentences

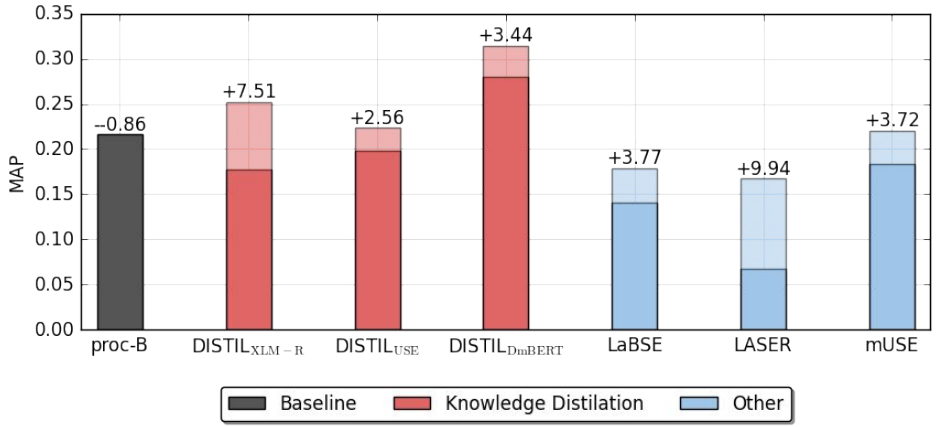


Proc-B

Sentence Encoder

Split Sent.

Score Top-k Sentences

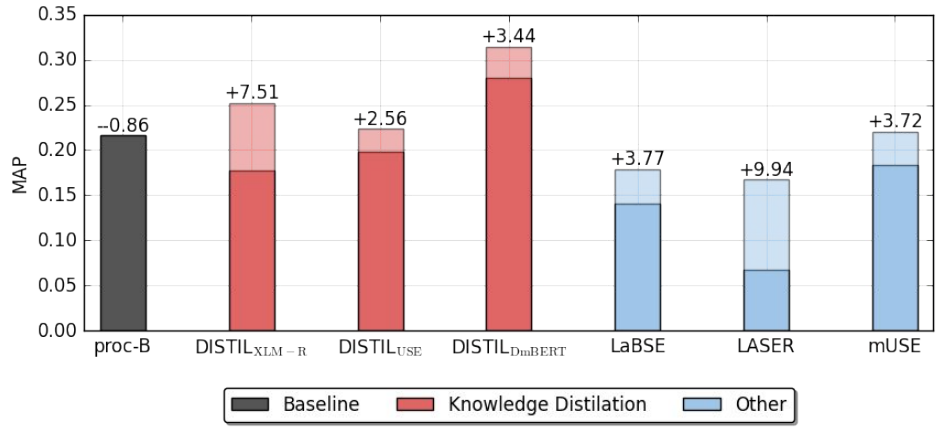


Proc-B

Sentence Encoder

Split Sent.

Score Top-k Sentences



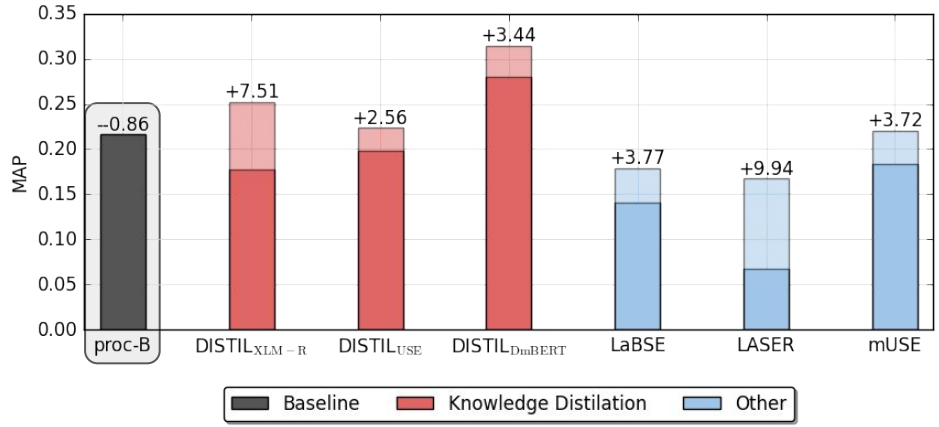
Proc-B

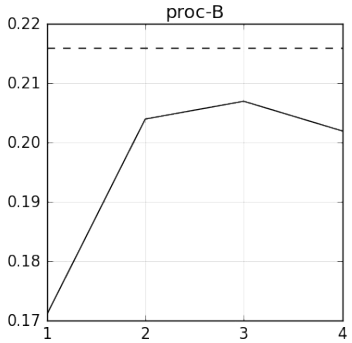
Sentence Encoder

Split Sent.

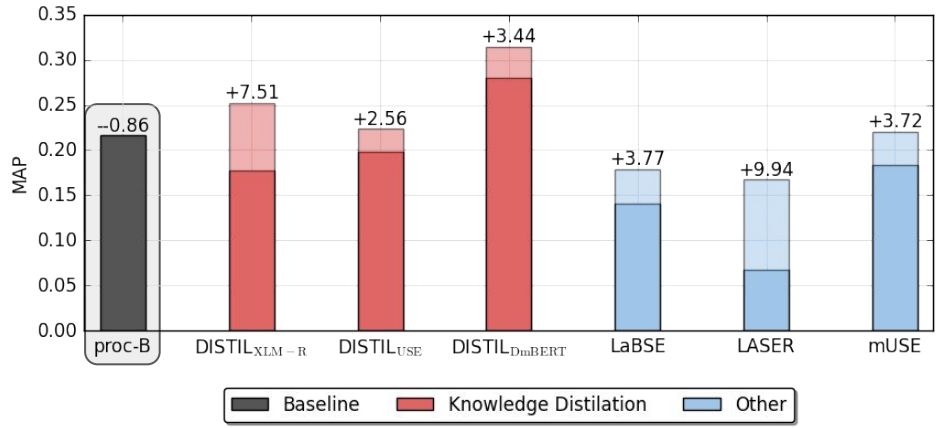
How many?

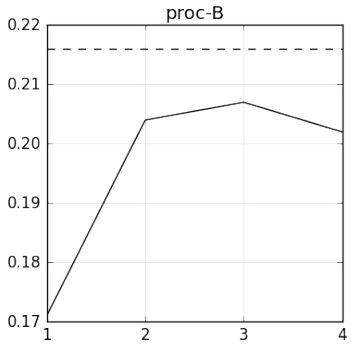
Score Top-k Sentences





Score Top-k Sentences

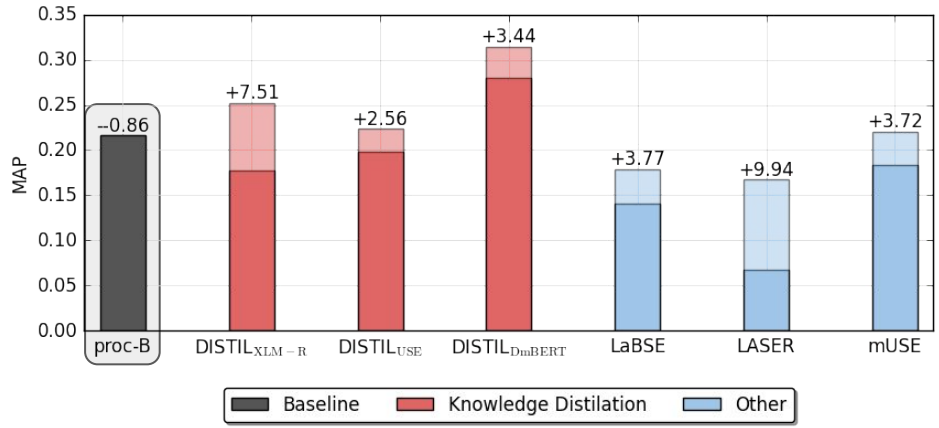


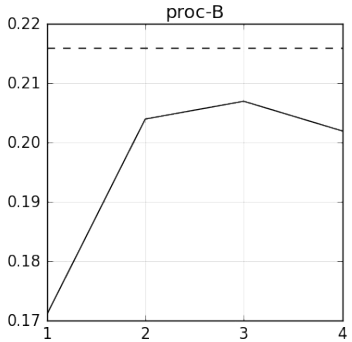


← Score entire document.

← Different number of top-k sentences.

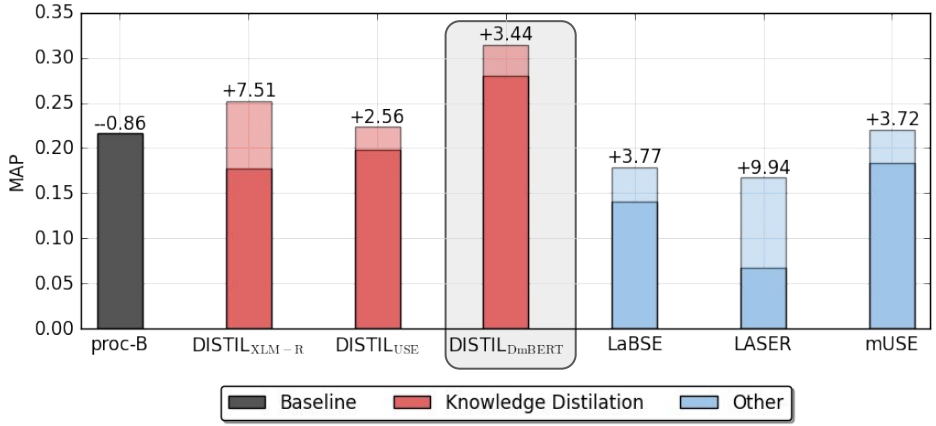
Score Top-k Sentences

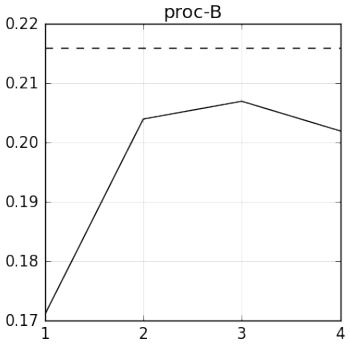


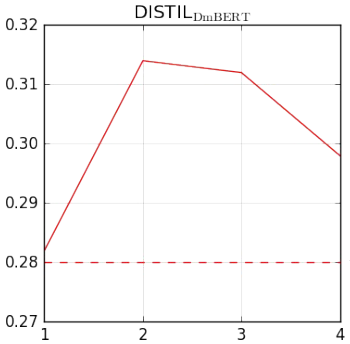


Proc-B does <u>not benefit</u> from sentence splitting.

Score Top-k Sentences

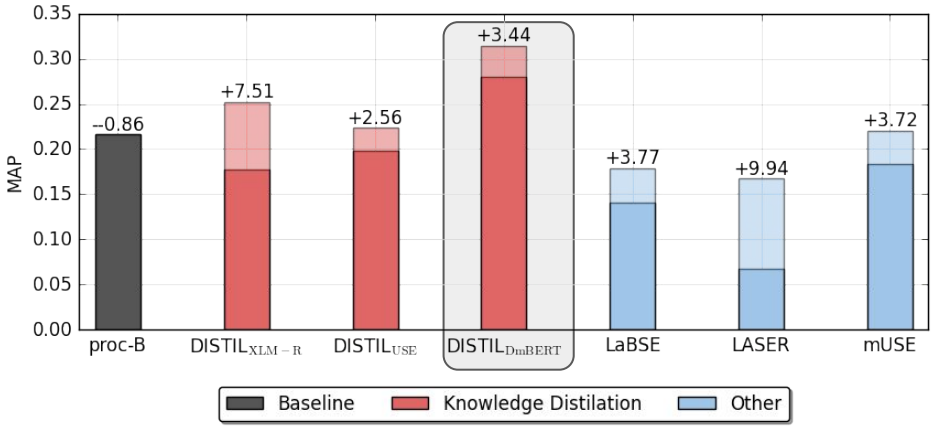


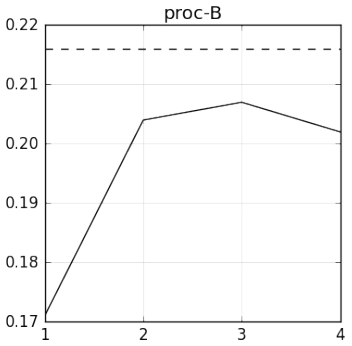


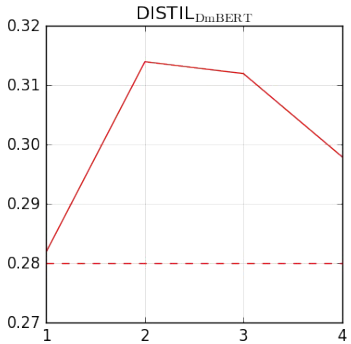


Score first 128 subwords.

Score Top-k Sentences

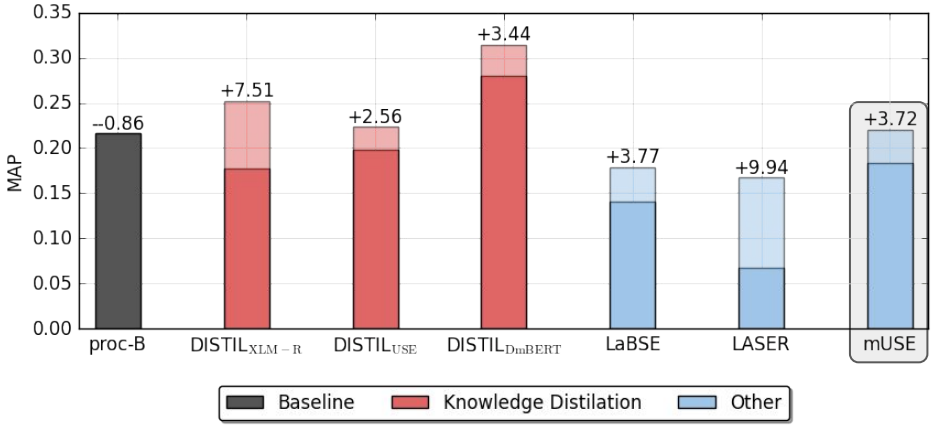


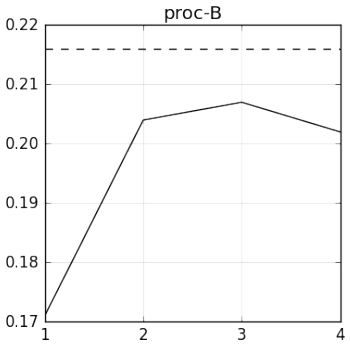


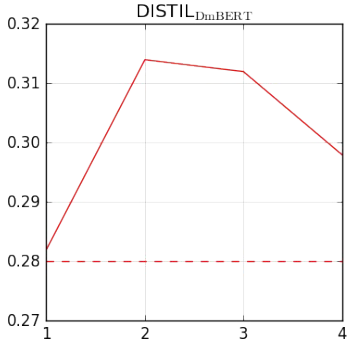


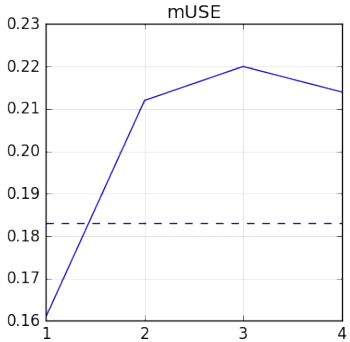
Text outside context window improves CLIR.

Score Top-k Sentences

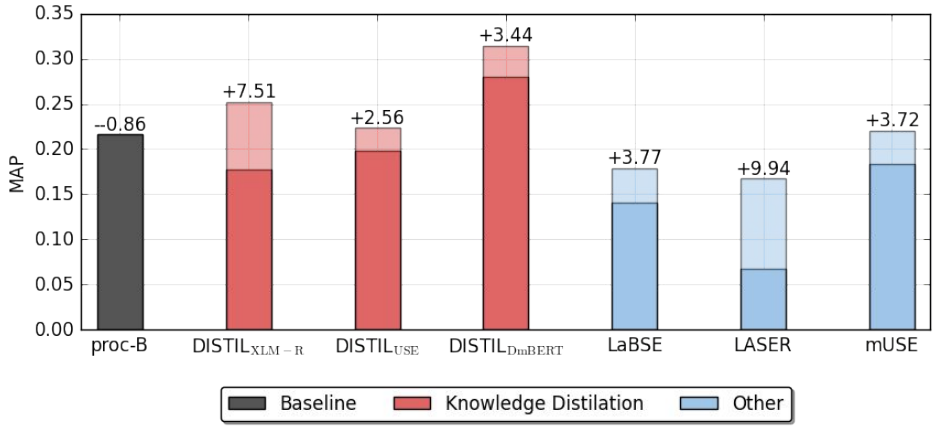






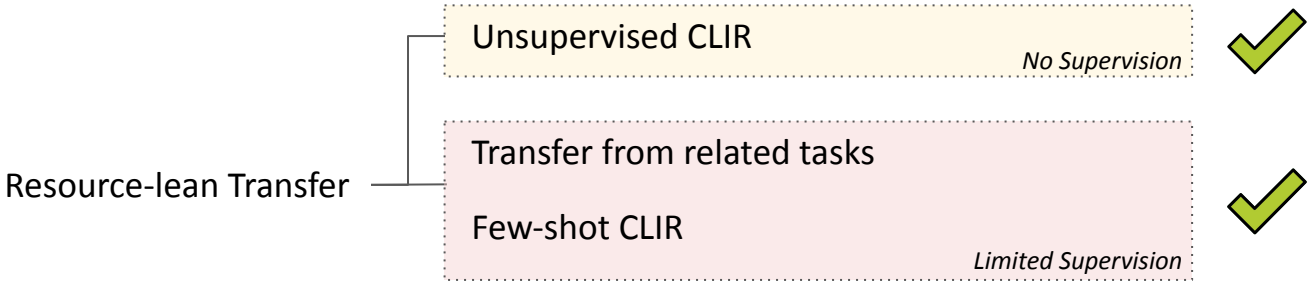


Results

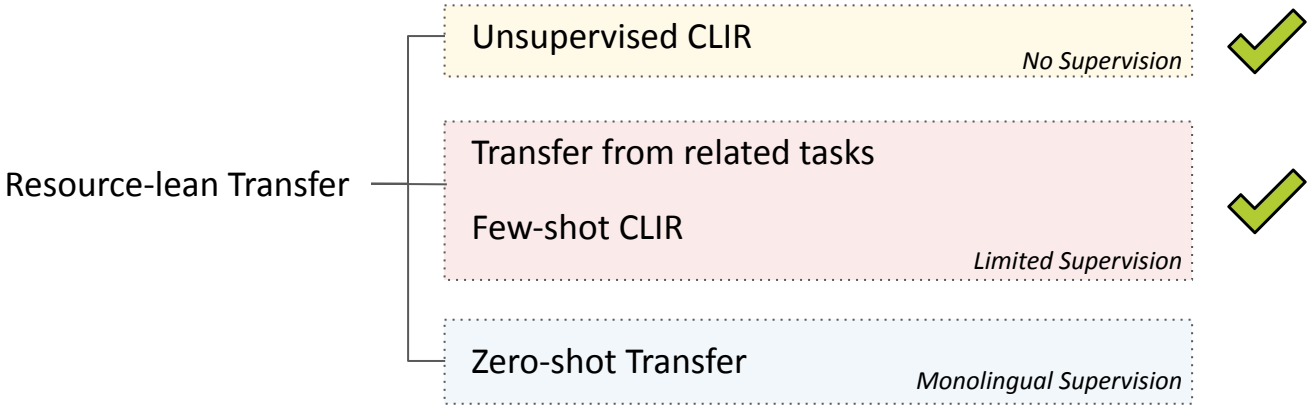


- **RQ-3:** Similarity-specialized sentence encoders outperform CLWEs.
  - Excessive / insufficient context degrade their effectiveness.
- **RQ-4:** Few-shot CLIR improves DISTIL$_{DmBERT}$ by +2.5MAP (see §5.5.5 ).

Contribution: Large-Scale Empirical Evaluation



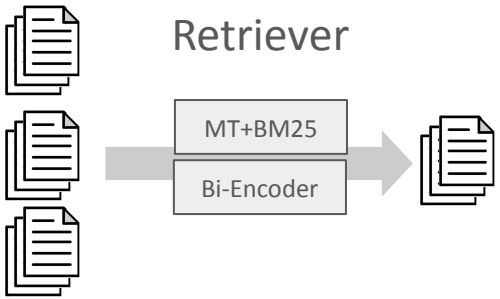
Contribution: Large-Scale Empirical Evaluation



Contribution: Large-Scale Empirical Evaluation

Resource-lean Transfer

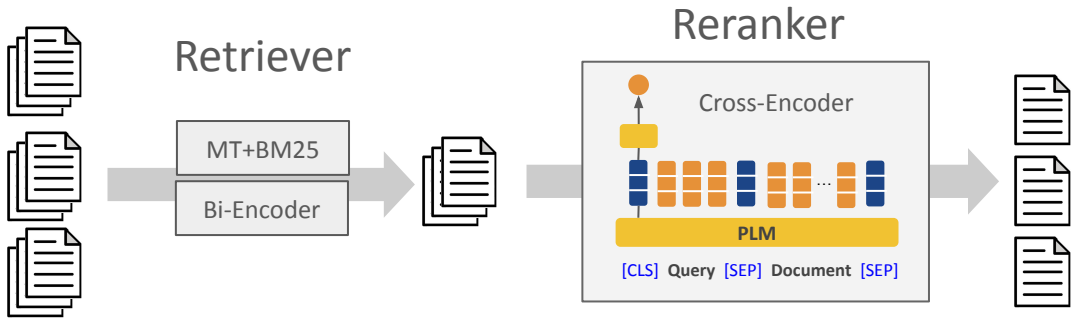
- Unsupervised CLIR — *No Supervision* ✓
- Transfer from related tasks
  Few-shot CLIR — *Limited Supervision* ✓
- Zero-shot Transfer — *Monolingual Supervision*



Contribution: Large-Scale Empirical Evaluation

Resource-lean Transfer

- Unsupervised CLIR — *No Supervision* ✓
- Transfer from related tasks
  Few-shot CLIR — *Limited Supervision* ✓
- Zero-shot Transfer — *Monolingual Supervision*

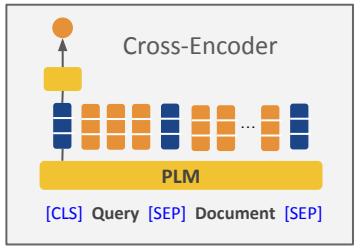


Contribution: Large-Scale Empirical Evaluation

Resource-lean Transfer

- Unsupervised CLIR — *No Supervision* ✓
- Transfer from related tasks
  Few-shot CLIR — *Limited Supervision* ✓
- Zero-shot Transfer — *Monolingual Supervision*

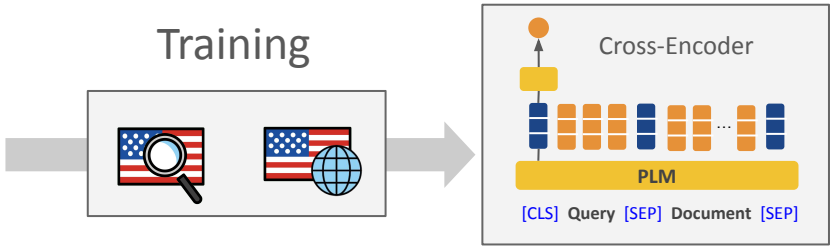
CLIR Training data
Direct Supervision



Contribution: Large-Scale Empirical Evaluation

Resource-lean Transfer

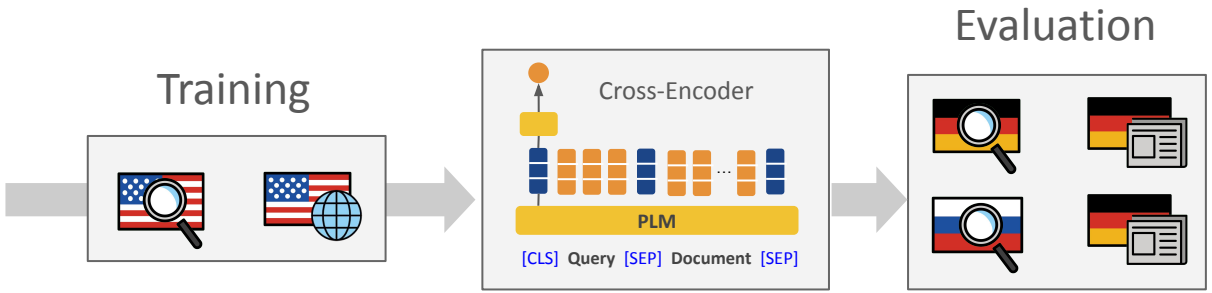
- Unsupervised CLIR — *No Supervision* ✓
- Transfer from related tasks
  Few-shot CLIR — *Limited Supervision* ✓
- Zero-shot Transfer — *Monolingual Supervision*



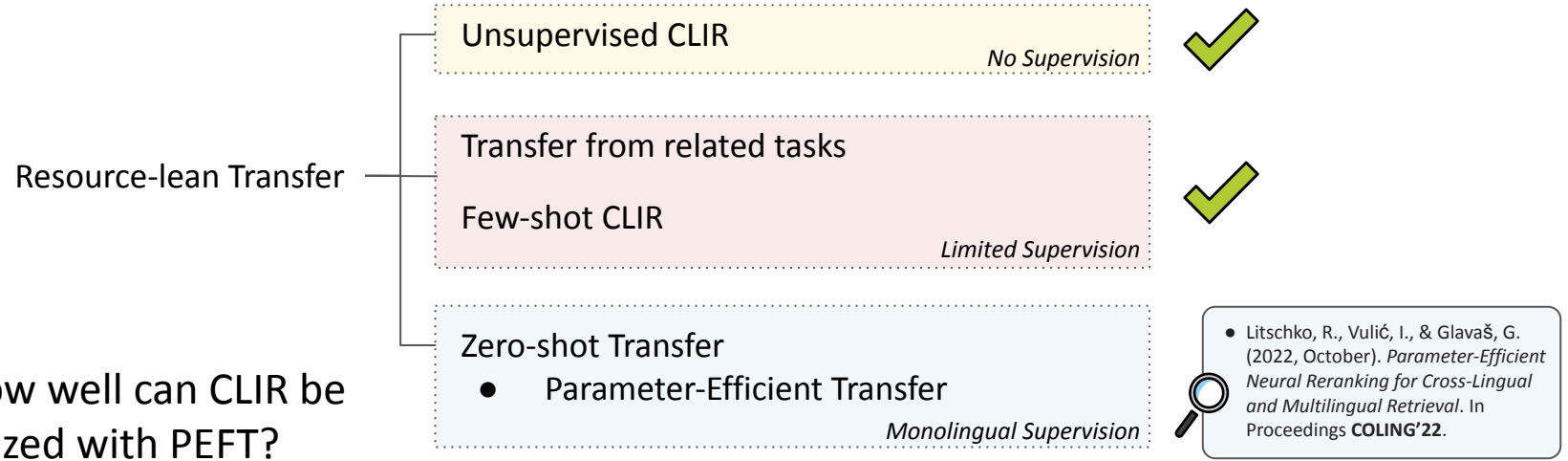
Contribution: Large-Scale Empirical Evaluation

Resource-lean Transfer

- Unsupervised CLIR — *No Supervision* ✓
- Transfer from related tasks
  Few-shot CLIR — *Limited Supervision* ✓
- Zero-shot Transfer — *Monolingual Supervision*



Contribution: Large-Scale Empirical Evaluation

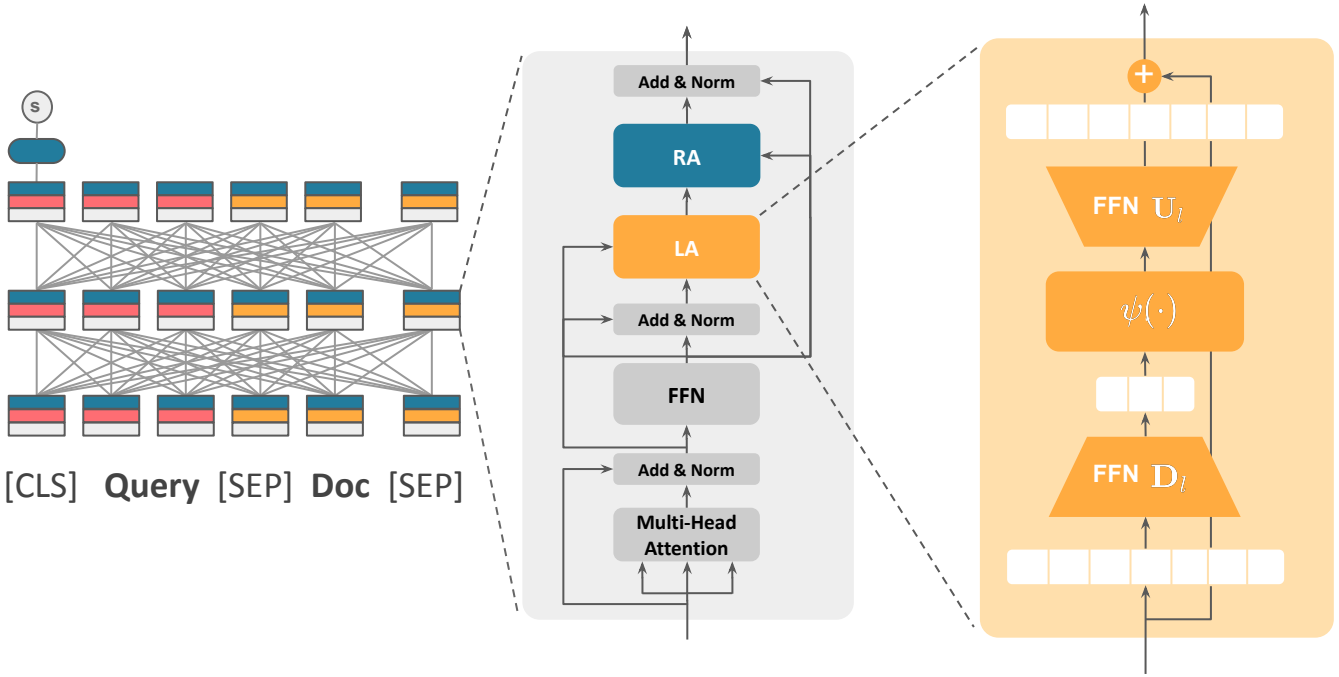


RQ-5: How well can CLIR be modularized with PEFT?

MoIR data (cheap) — self-supervised (cheap)

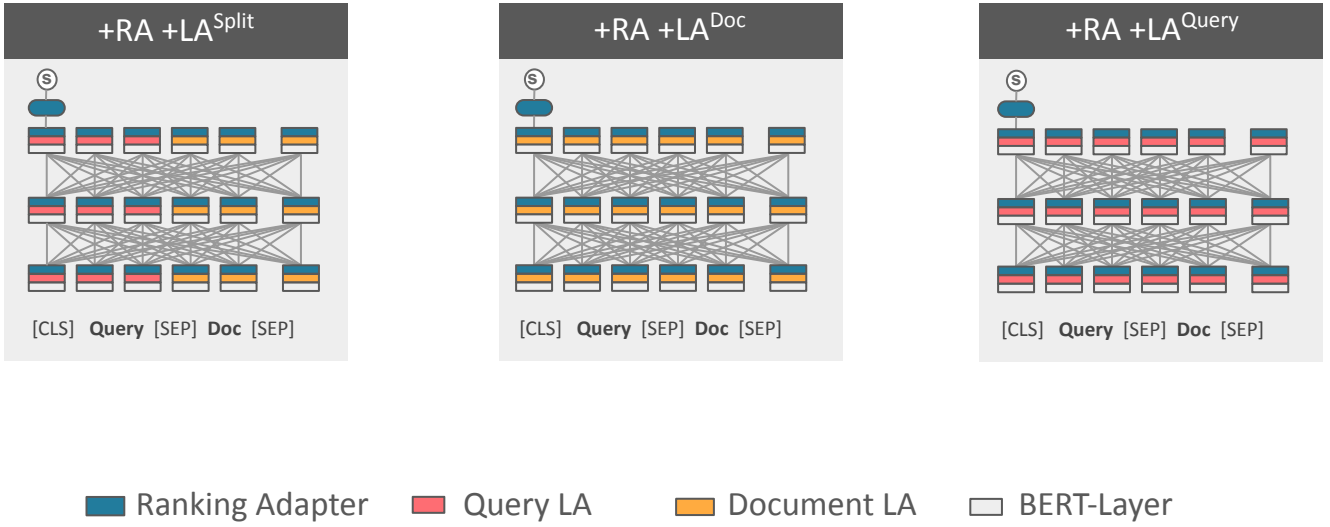
CLIR = L2R + MLM

Contribution: Adapters for Cross-Lingual Reranking



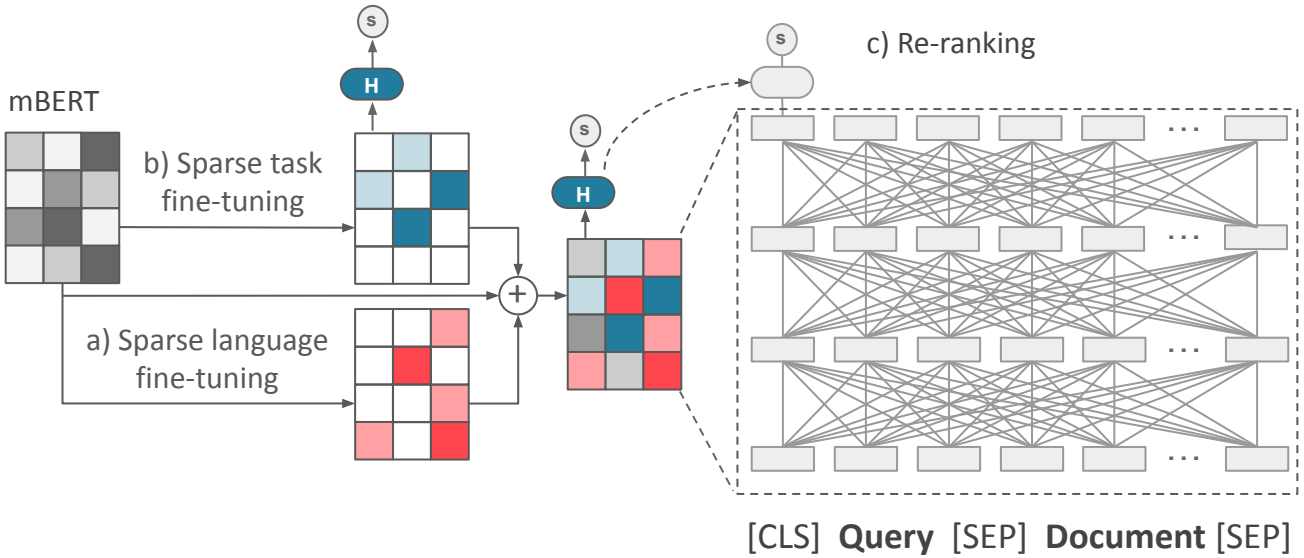
Following Pfeiffer et. al. (2020).

Contribution: Adapters for Cross-Lingual Reranking



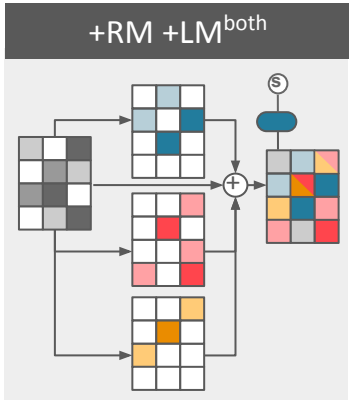
Following Pfeiffer et. al. (2020).

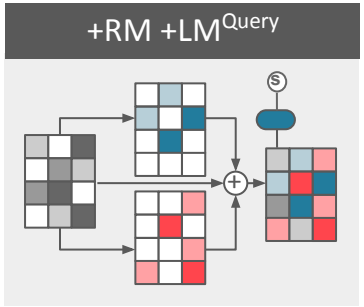
Contribution: Sparse Fine-Tuning Masks for Cross-Lingual Reranking

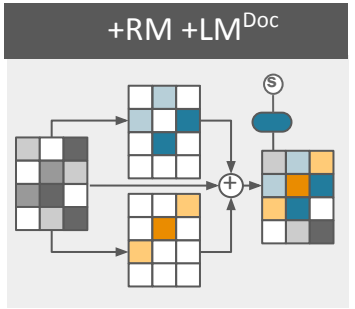


Following Ansell et. al. (2021).

Contribution: Sparse Fine-Tuning Masks for Cross-Lingual Reranking



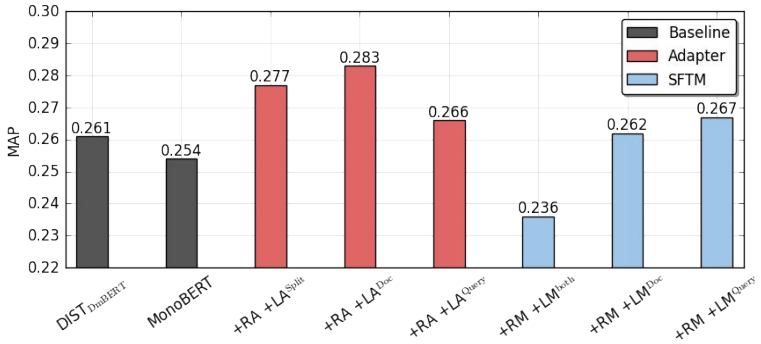




Ranking Mask (**RM**) — Query Language Mask (**LM**) — Document Language Mask (**LM**)

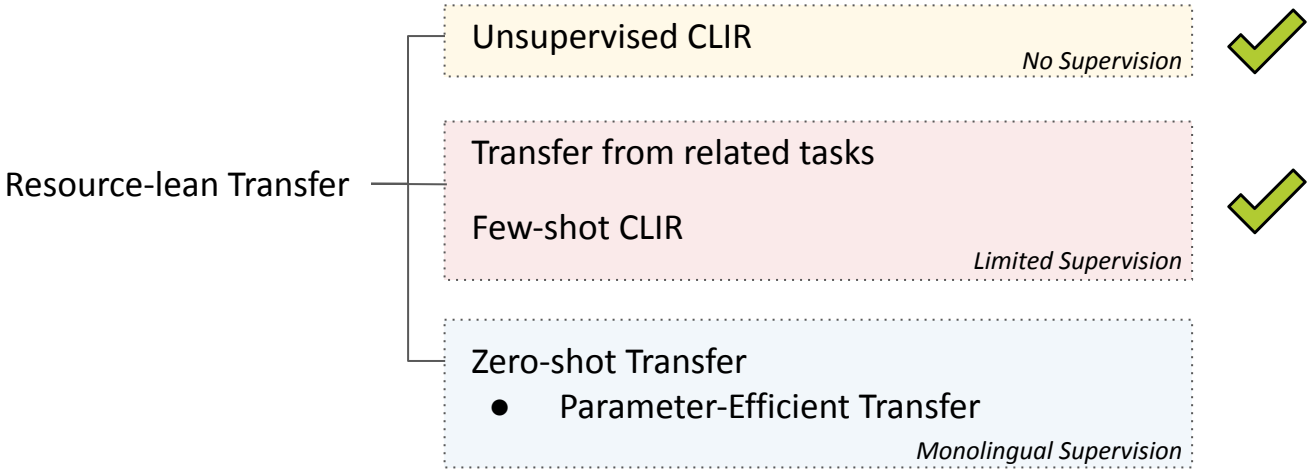
Following Ansell et. al. (2021).

Results: Parameter-Efficient CLIR



- **RQ-5:** Adapters and SFTMs outperform baselines and improve upon MT PR.
  - Performance crucially depends on reduction factor (see §8 📄).

Contribution: Large-Scale Empirical Evaluation



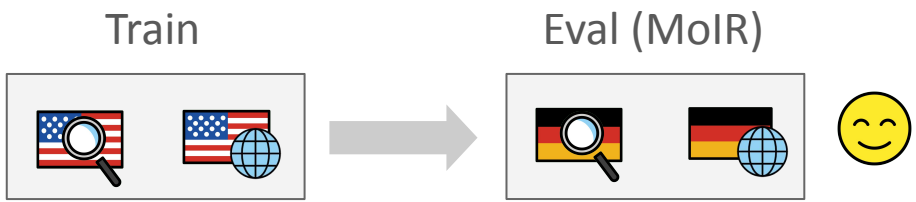
Contribution: Large-Scale Empirical Evaluation

Resource-lean Transfer

- Unsupervised CLIR — *No Supervision* ✓
- Transfer from related tasks
  Few-shot CLIR — *Limited Supervision* ✓
- Zero-shot Transfer
  - Parameter-Efficient Transfer
  - Monolingual Overfitting

  Monolingual Supervision

- Litschko, R., Artemova, E., & Plank, B. *Boosting Zero-shot Cross-lingual Retrieval by Training on Artificially Code-Switched Data*. In **Findings of ACL'23**.

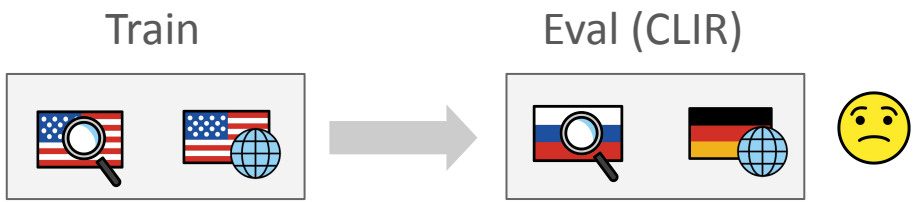


Contribution: Large-Scale Empirical Evaluation

Resource-lean Transfer

- Unsupervised CLIR — *No Supervision*
- Transfer from related tasks
  Few-shot CLIR — *Limited Supervision*
- Zero-shot Transfer
  - Parameter-Efficient Transfer
  - Monolingual Overfitting
  — *Monolingual Supervision*

- Litschko, R., Artemova, E., & Plank, B. *Boosting Zero-shot Cross-lingual Retrieval by Training on Artificially Code-Switched Data*. In **Findings of ACL'23**.



Contribution: Large-Scale Empirical Evaluation

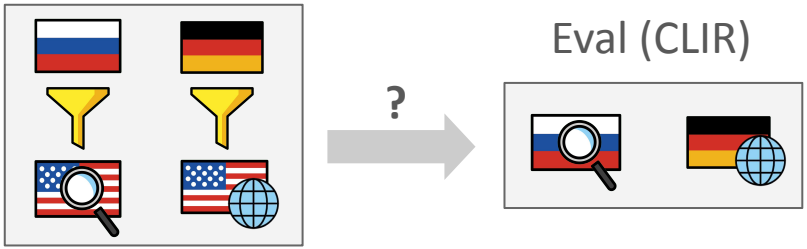
Resource-lean Transfer

- Unsupervised CLIR — *No Supervision*
- Transfer from related tasks
  Few-shot CLIR — *Limited Supervision*
- Zero-shot Transfer
  - Parameter-Efficient Transfer
  - Monolingual Overfitting

  Monolingual Supervision

- Litschko, R., Artemova, E., & Plank, B. *Boosting Zero-shot Cross-lingual Retrieval by Training on Artificially Code-Switched Data*. In **Findings of ACL'23**.

RQ-6: Does Code Switching mitigate Monolingual Overfitting?



Motivation: Monolingual Overfitting

Training

what is a death roll in crocodiles

The death roll performs a number of functions for the Saltwater Crocodile. When it grabs very large prey the crocodile has to drag it into the water and drown it so the crocodile [...] to roll over and over again to drown it's prey.

Keyword Matching

Semantic Matching

We use the mMARCO dataset (Bonifacio et al., 2021).

Motivation: Monolingual Overfitting

Training

what is a death roll in crocodiles

The death roll performs a number of functions for the Saltwater Crocodile. When it grabs very large prey the crocodile has to drag it into the water and drown it so the crocodile [...] to roll over and over again to drown it's prey.

Monolingual IR

Symptome von Fieber

(symptoms of fever)

Die Liste der Anzeichen und Symptome, die in verschiedenen Quellen für Fieber erwähnt werden, umfasst die 8 unten aufgeführten Symptome: Schwitzen. Temperatur. Strenge. Brechreiz. Erbrechen. Durchfall. Lethargie.

We use the mMARCO dataset (Bonifacio et al., 2021).

Motivation: Monolingual Overfitting

Training

what is a death roll in crocodiles

The death roll performs a number of functions for the Saltwater Crocodile. When it grabs very large prey the crocodile has to drag it into the water and drown it so the crocodile [...] to roll over and over again to drown it's prey.

Monolingual IR

Symptome von Fieber

(symptoms of fever)

Die Liste der Anzeichen und Symptome, die in verschiedenen Quellen für Fieber erwähnt werden, umfasst die 8 unten aufgeführten Symptome: Schwitzen. Temperatur. Strenge. Brechreiz. Erbrechen. Durchfall. Lethargie.

Cross-lingual IR

симптомы лихорадки

(symptoms of fever)

Die Liste der Anzeichen und Symptome, die in verschiedenen Quellen für Fieber erwähnt werden, umfasst die 8 unten aufgeführten Symptome: Schwitzen. Temperatur. Strenge. Brechreiz. Erbrechen. Durchfall. Lethargie.

We use the mMARCO dataset (Bonifacio et al., 2021).

Method: Artificial Code-Switching

Zero-Shot Transfer

Query: what is a death roll in crocodiles

Passage: the death roll performs a number of functions for the Saltwater...

Zero-shot rerankers are biased towards lexical matching.

Method: Artificial Code-Switching

Zero-Shot Transfer

Query: what is a death roll in crocodiles

Passage: the death roll performs a number of functions for the Saltwater...

Translate Train (Fine-tuning)

Query: что такое список крокодилов

Passage: Die Todesrolle erfüllt für das Salzwasserkrokodil eine Reihe von Funktionen...

Machine Translation is expensive and prone to error propagation.

Method: Artificial Code-Switching

Zero-Shot Transfer

Query: what is a death roll in crocodiles

Passage: the death roll performs a number of functions for the Saltwater...

Translate Train (Fine-tuning)

Query: что такое список крокодилов

Passage: Die Todesrolle erfüllt für das Salzwasserkrokodil eine Reihe von Funktionen...

Bilingual Code-Switching (CS)*

Cross-lingual Word Embedding Space (Lample et al., 2018)

Query: что is a death roll in крокодилы

Passage: The death roll выполняет a число of функции for в Saltwater...

*Code-Switching similar to Tan and Joty (2021).

Method: Artificial Code-Switching

Zero-Shot Transfer

Query: what is a death roll in crocodiles

Passage: the death roll performs a number of functions for the Saltwater...

Translate Train (Fine-tuning)

Query: что такое список крокодилов

Passage: Die Todesrolle erfüllt für das Salzwasserkrokodil eine Reihe von Funktionen...

Bilingual Code-Switching (CS)*

Cross-lingual Word Embedding Space (Lample et al., 2018)

Query: что is a death roll in крокодилы

Passage: The death roll выполняет a число of функции for в Saltwater...

Multilingual Code-Switching (CS)*

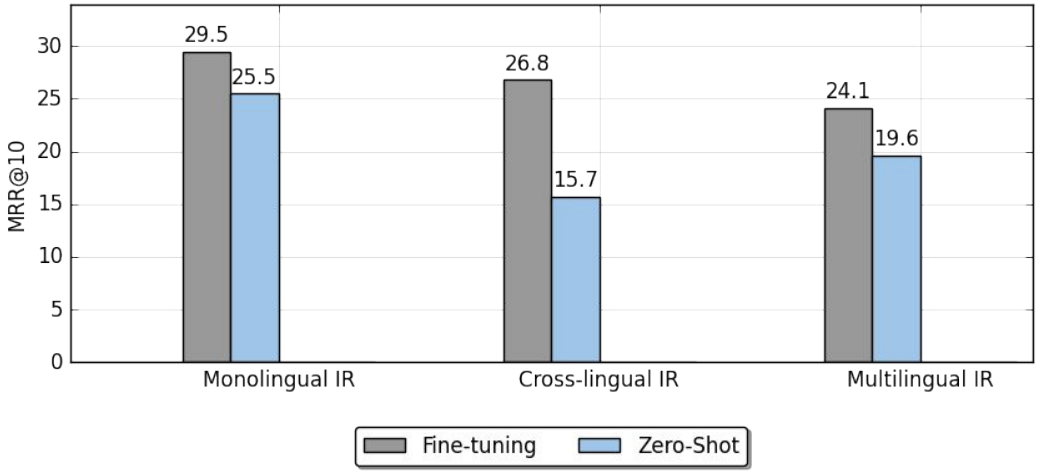
Multilingual Word Embedding Space (Lample et al., 2018)

Query: cosa is a موت rollen in крокодилы

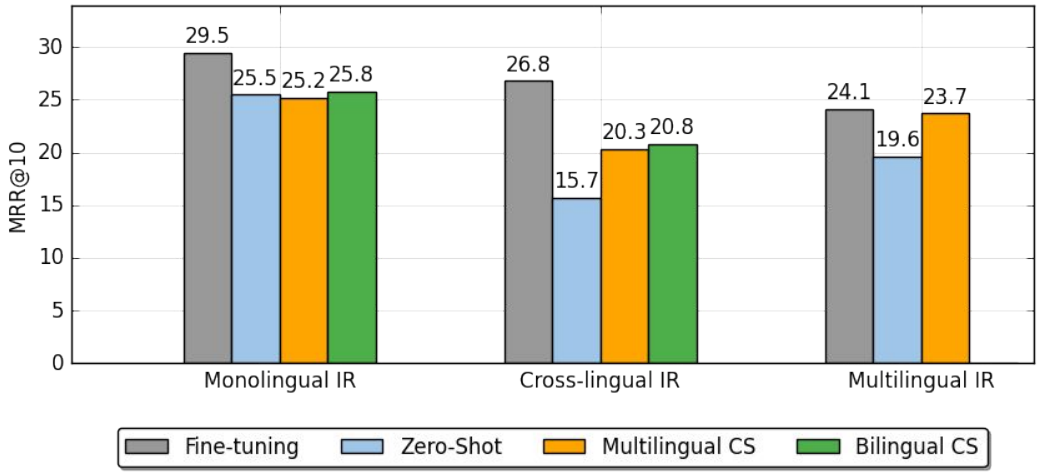
Passage: Der death rotolo performs a число of المهام for в Saltwater...

*Code-Switching similar to Tan and Joty (2021).

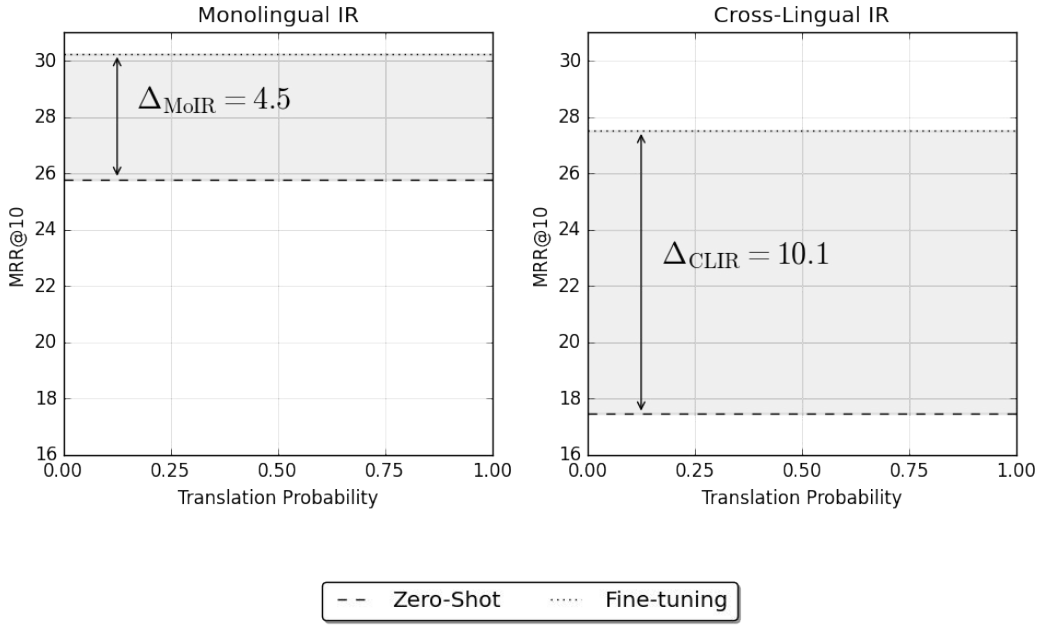
Code-Switching is **Effective**



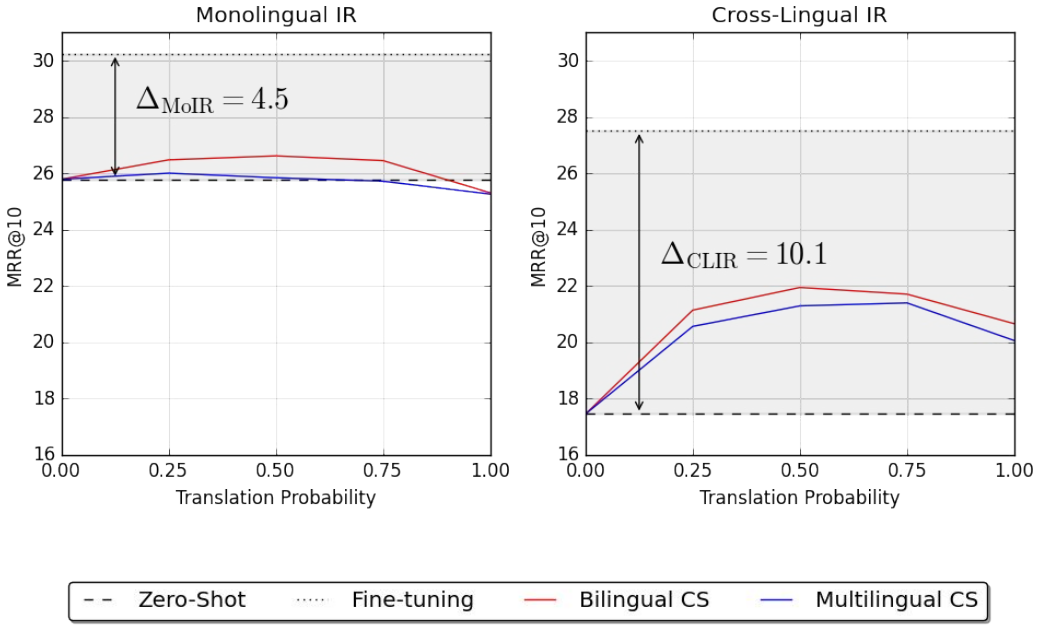
Code-Switching is **Effective**



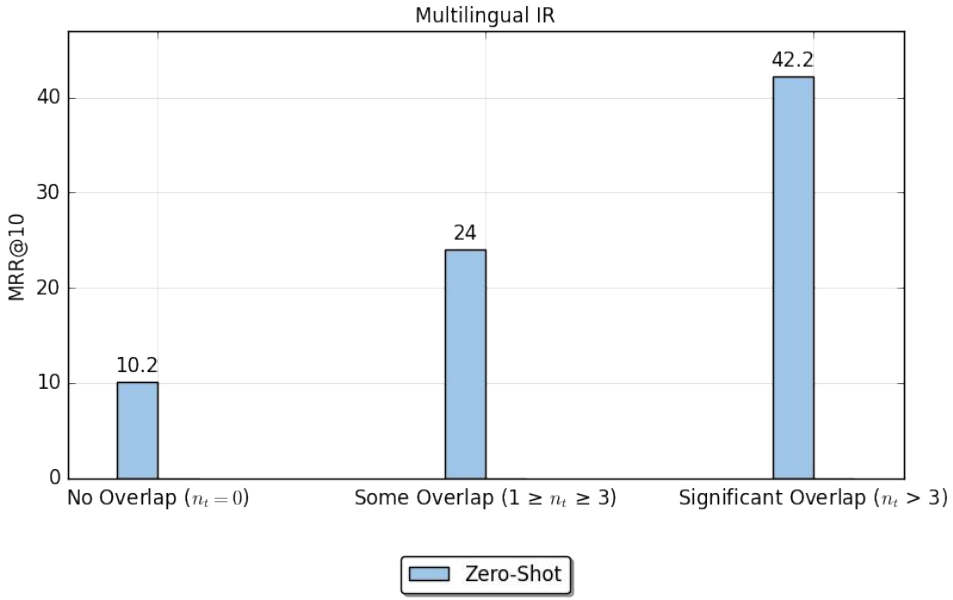
Code-Switching is Robust



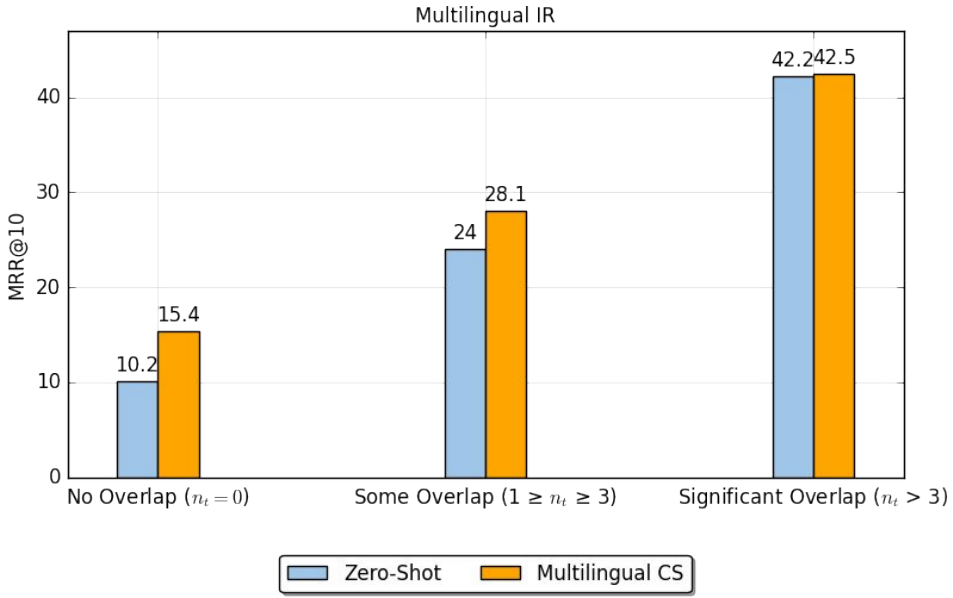
Code-Switching is Robust



…and mitigates Monolingual Overfitting



…and mitigates Monolingual Overfitting



Take-away

Multilingual Code-Switching (CS)*

Multilingual Word Embedding Space (Lample et al., 2018)

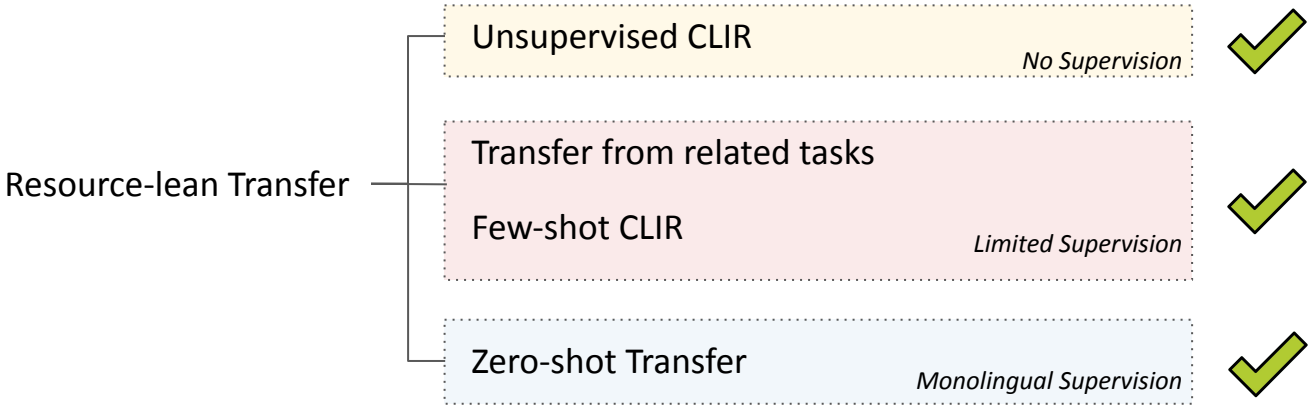
Query: cosa is a موت rollen in крокодилы

Passage: Der death rotolo performs a число of المهام for в Saltwater...

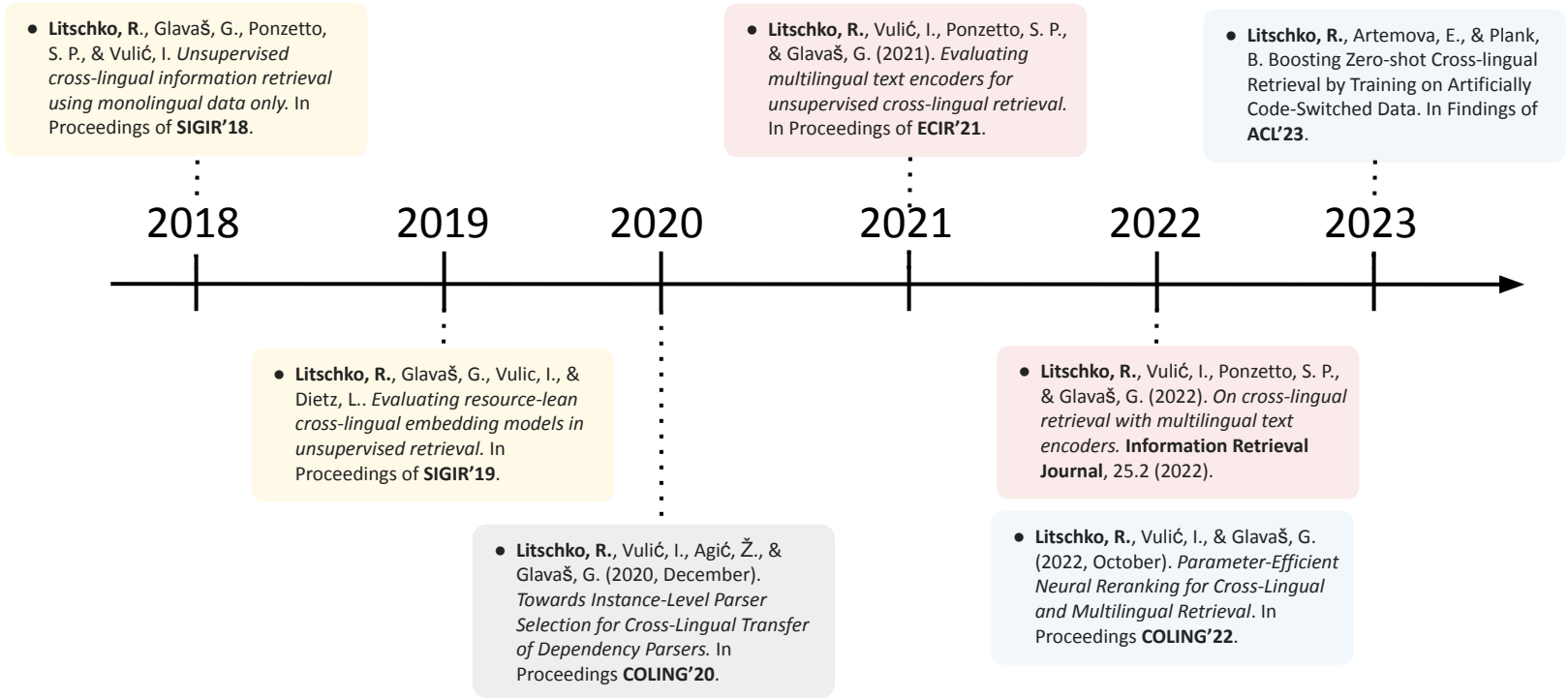
- **RQ-6:** **Monolingual overfitting** negatively impacts zero-shot transfer for CLIR…

  …and can be **regularized by code switching** the training data.

Contribution: Large-Scale Empirical Evaluation



List of Publications



Summary of Findings

Summary of Findings

A. CLWEs are resource-lean and effective.
B. Contextualized representations do not outperform CLWEs.*
C. Too much / too little context harms CLIR performance.

No Supervision & Limited Supervision

**Specializing mPLMs for sentence-similarity helps.*

Summary of Findings

A. CLWEs are resource-lean and effective.
B. Contextualized representations do not outperform CLWEs.*
C. Too much / too little context harms CLIR performance.

No Supervision & Limited Supervision

D. Monolingual overfitting harms zero-shot CLIR and can be regularized with code switching.
E. Decomposing CLIR into L2R and language acquisition is resource-lean and effective.

Monolingual Supervision

**Specializing mPLMs for sentence-similarity helps.*

Summary of Findings

A. CLWEs are resource-lean and effective.
B. Contextualized representations do not outperform CLWEs.*
C. Too much / too little context harms CLIR performance.

No Supervision & Limited Supervision

D. Monolingual overfitting harms zero-shot CLIR and can be regularized with code switching.
E. Decomposing CLIR into L2R and language acquisition is resource-lean and effective.

Monolingual Supervision

F. The effectiveness of CLIR varies with language proximity (see §10 ).

**Specializing mPLMs for sentence-similarity helps.*