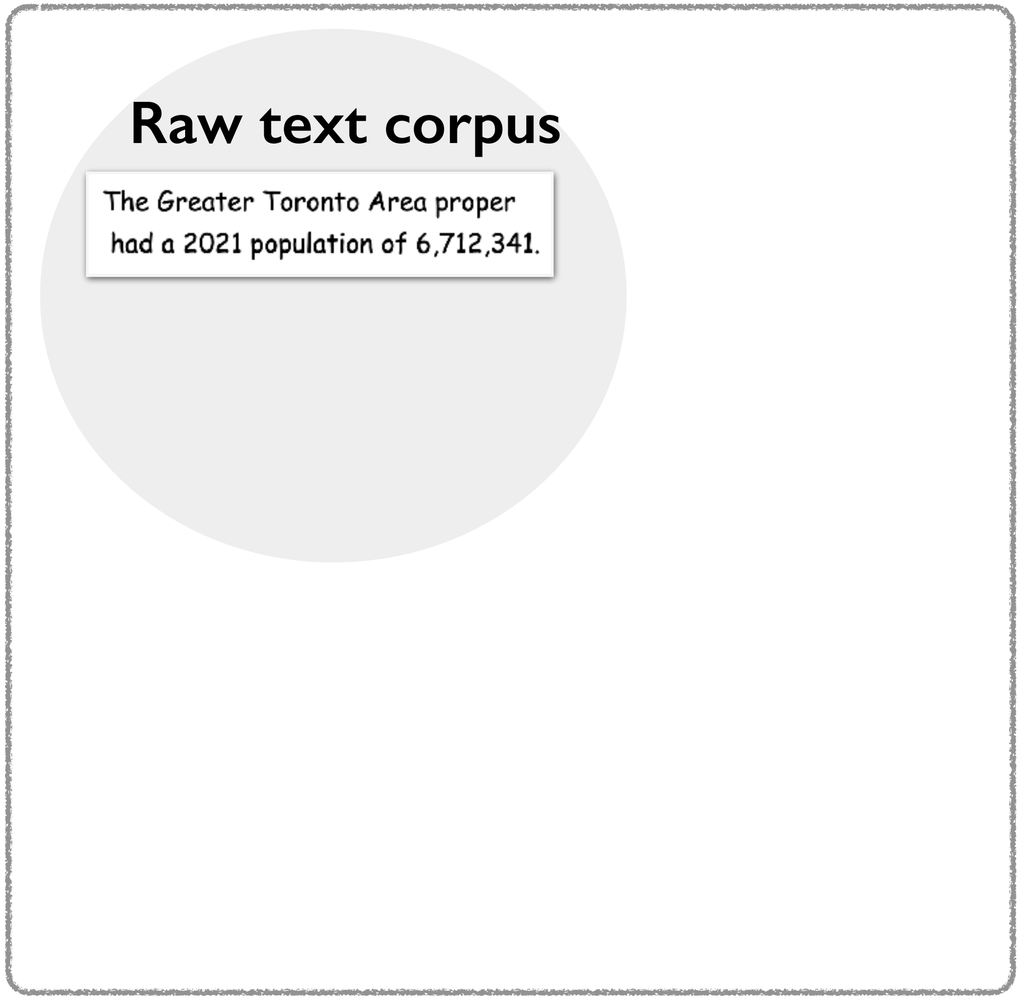
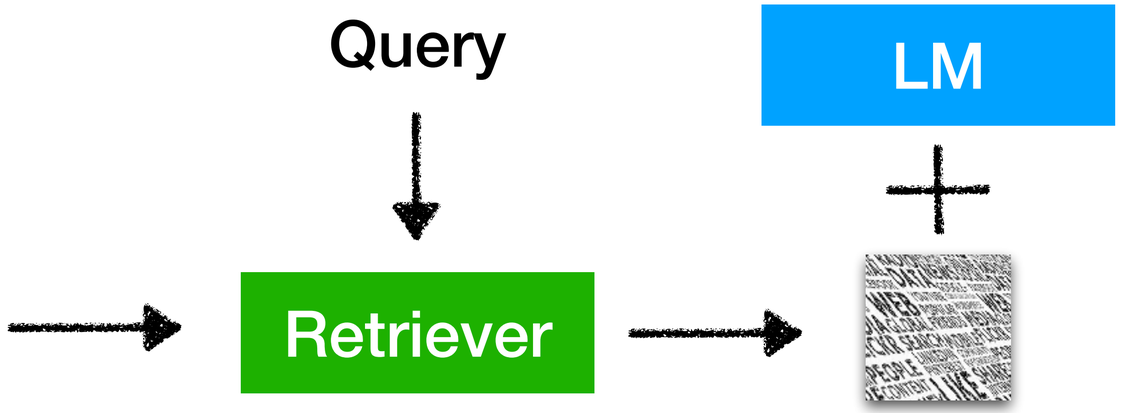
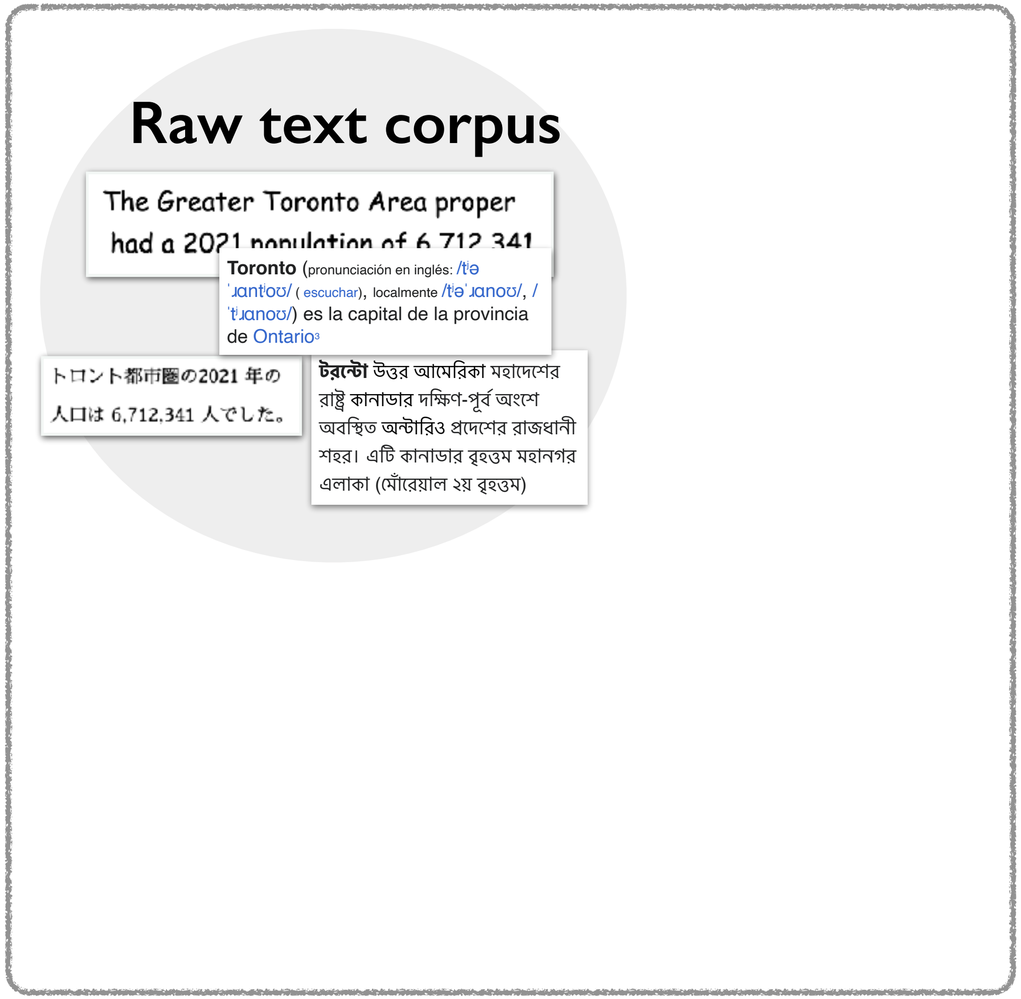
Section 6:
Multilingual & Multimodal

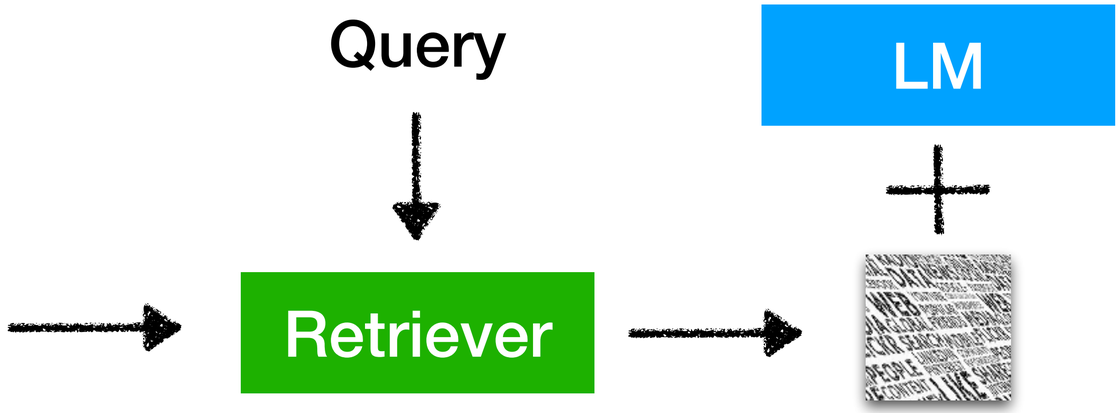
Retrieval-based LM for diverse knowledge sources



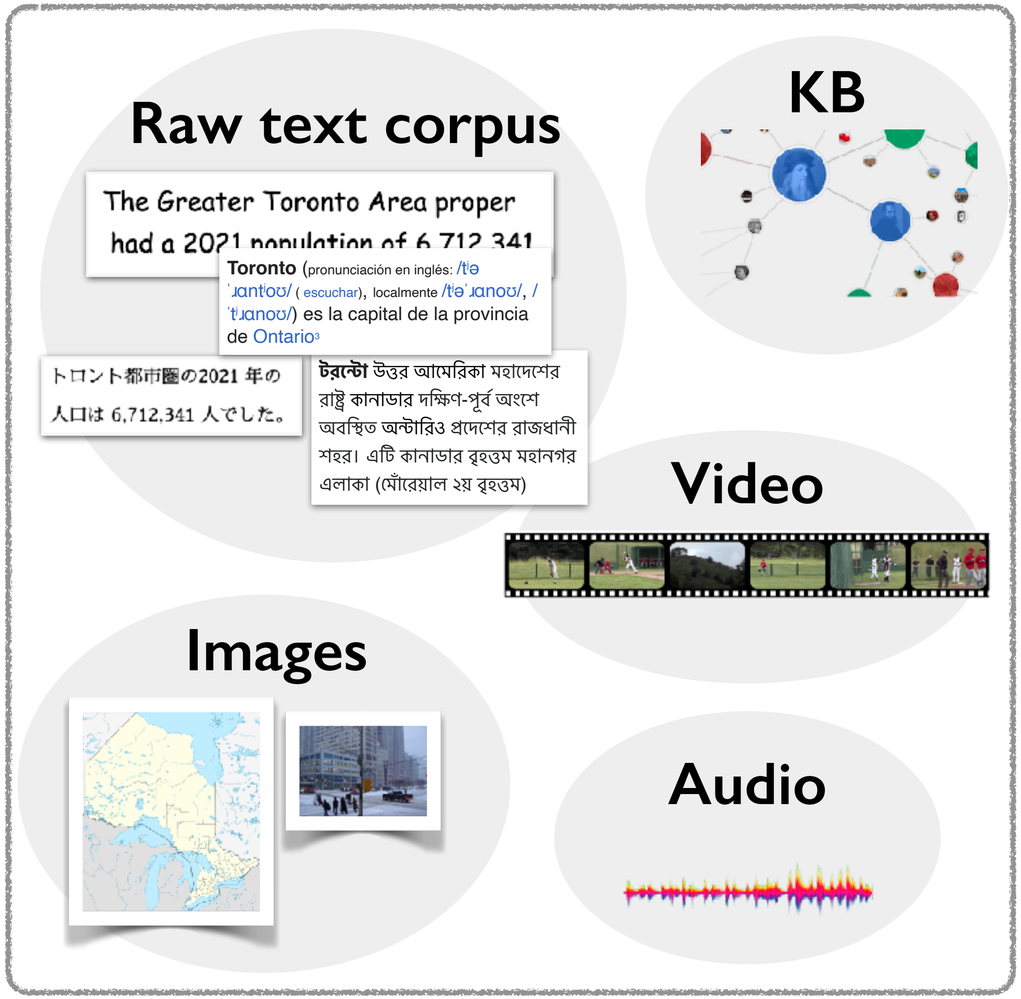


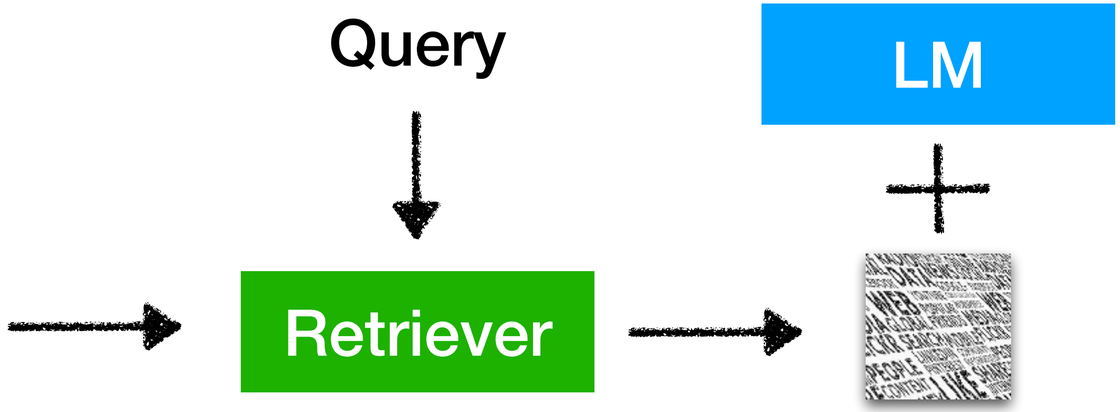
Retrieval-based LM for diverse knowledge sources



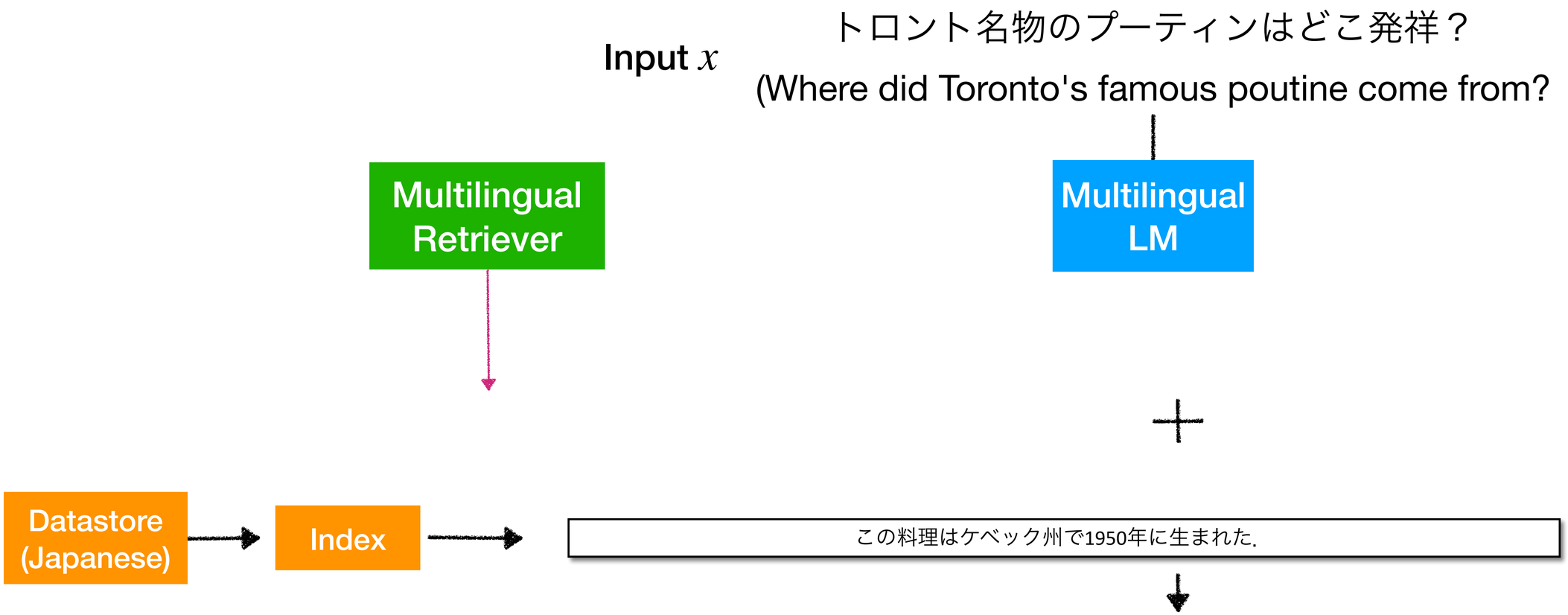


Retrieval-based LM for diverse knowledge sources





Multilingual Retrieval-based LM



(Limited amount of data is available in Japanese datastore…)

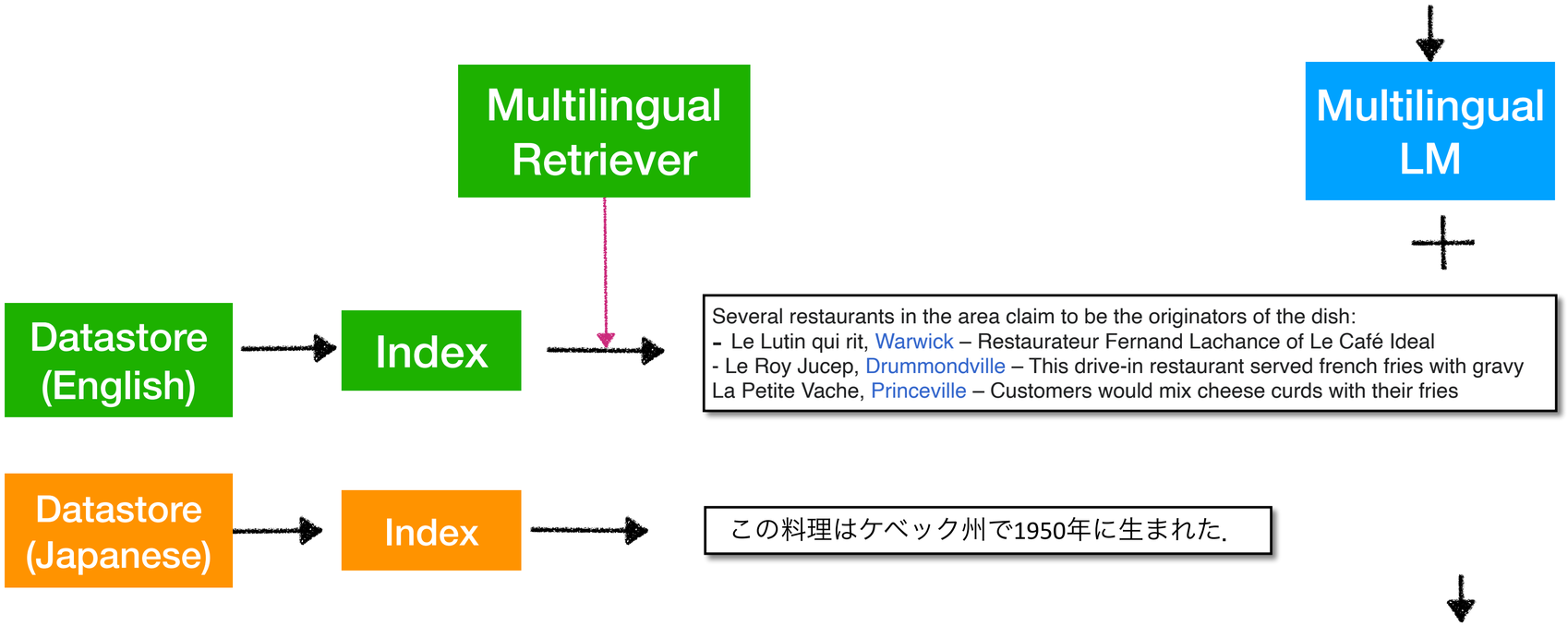
Output y

プーティンはケベック州で1950年代に生まれた.

Multilingual Retrieval-based LM

Input x トロント名物のプーティンはどこ発祥？

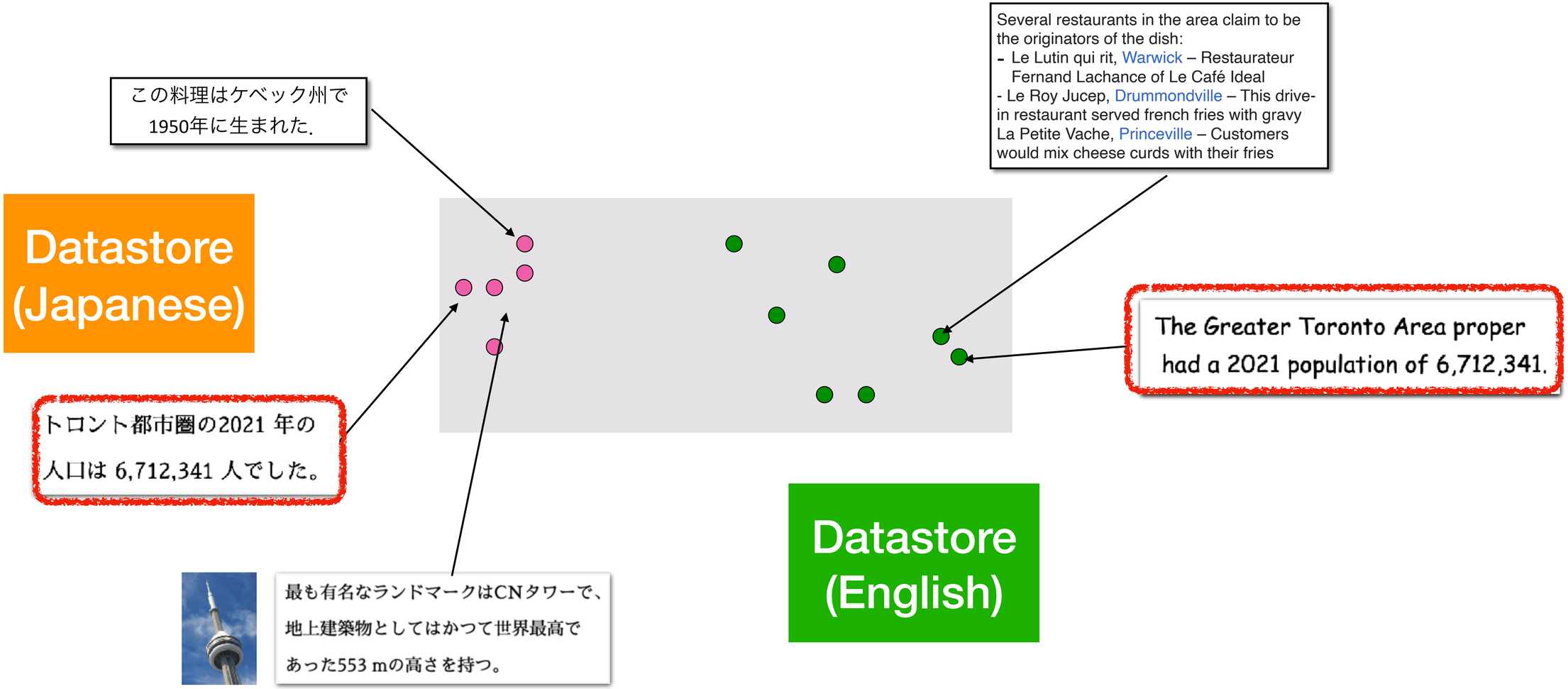
(Where did Toronto's famous poutine come from?



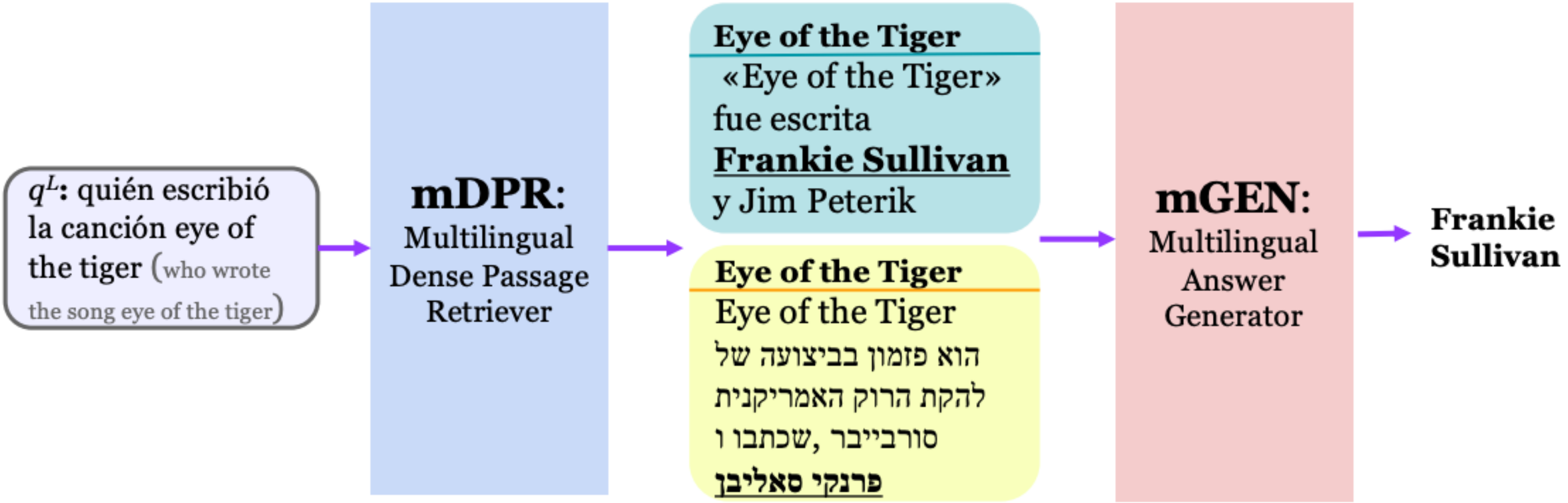
プーティンはケベック州で1950年代に生まれた．発祥については諸説あるが,

Output y Le Lutin qui rit, Le Roy Jucep, La Petite Vacheの三つのレストランが一番最初にプーティンを提供したレストランとして有力である.

Language biases in representation spaces



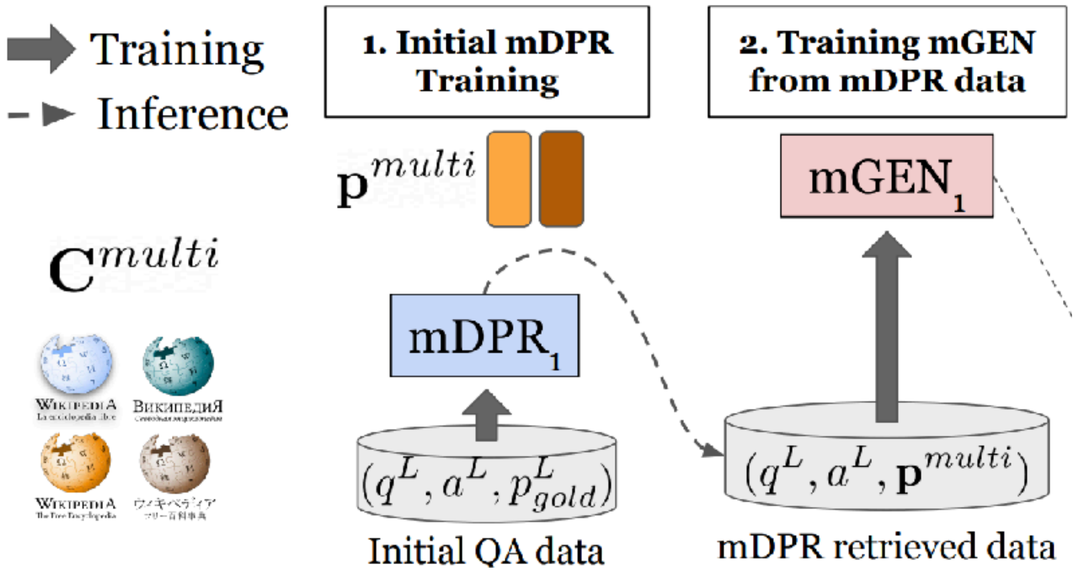
CORA (Asai et al., 2021)



Asai et al. 2021. “One question answering model for many languages with cross-lingual dense passage retrieval”

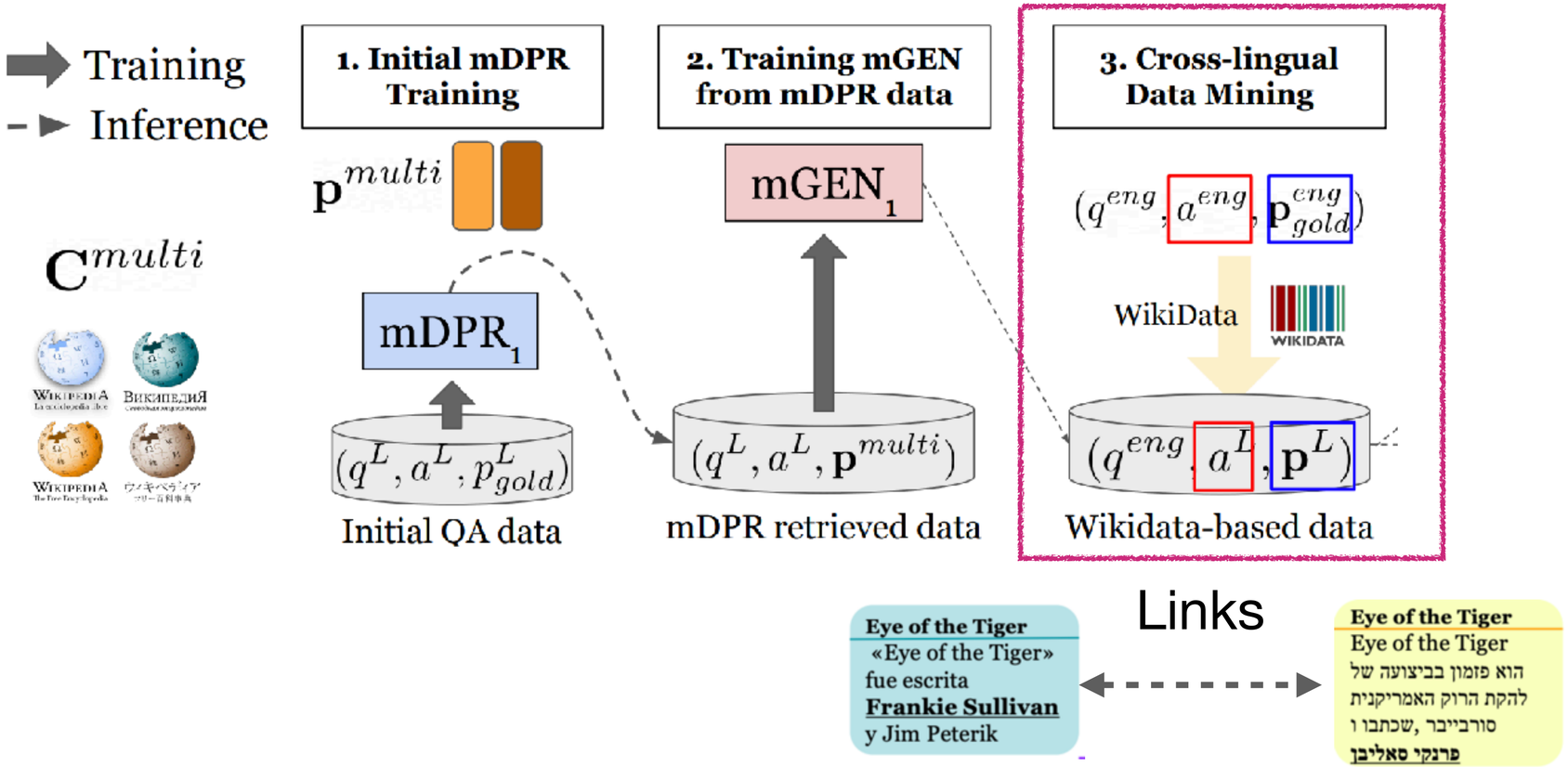
CORA: Iteratively training multilingual LM & retriever

Initial fine-tuning of retriever and LM using task data



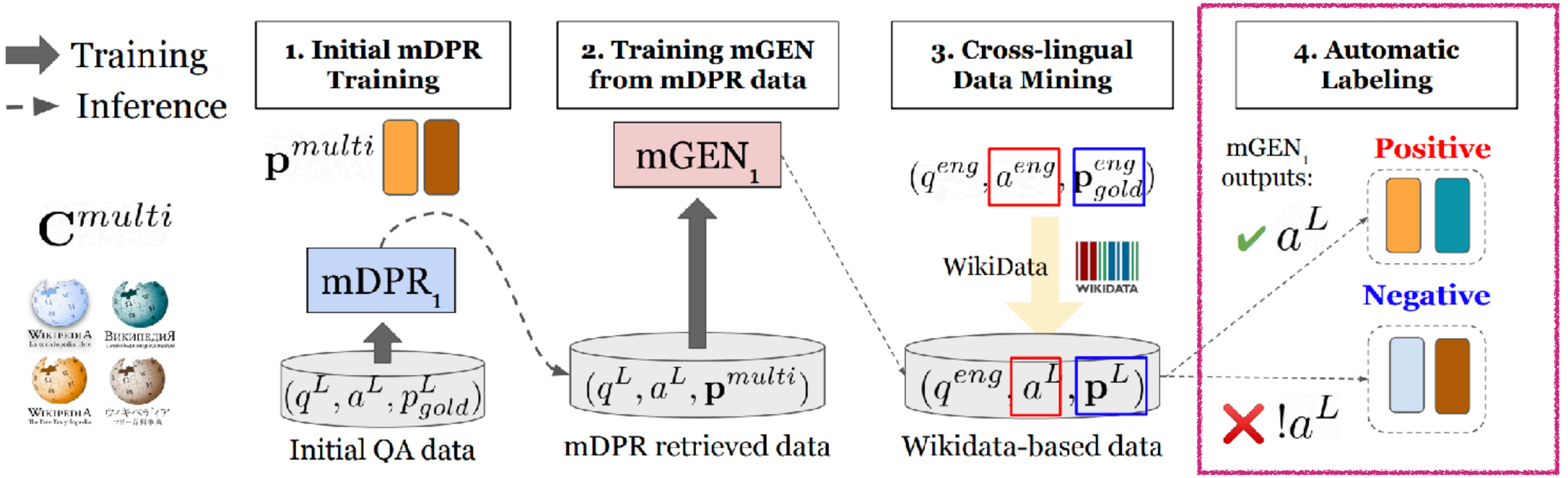
CORA: Iteratively training multilingual LM & retriever

Retrieve positive paragraphs in other languages using language links



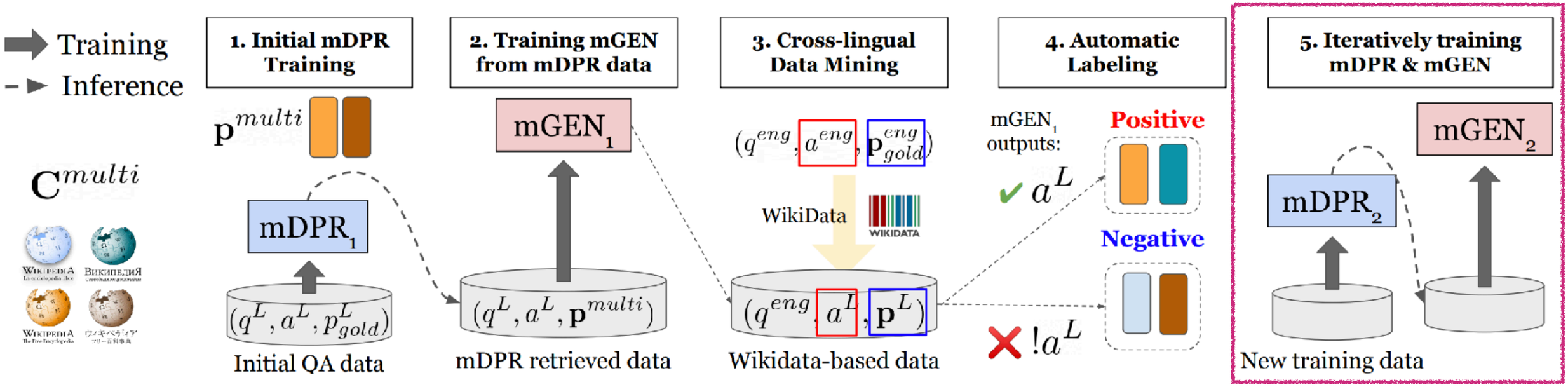
CORA: Iteratively training multilingual LM & retriever

Add new positive / negative paragraphs
based on whether LMs can answer correctly

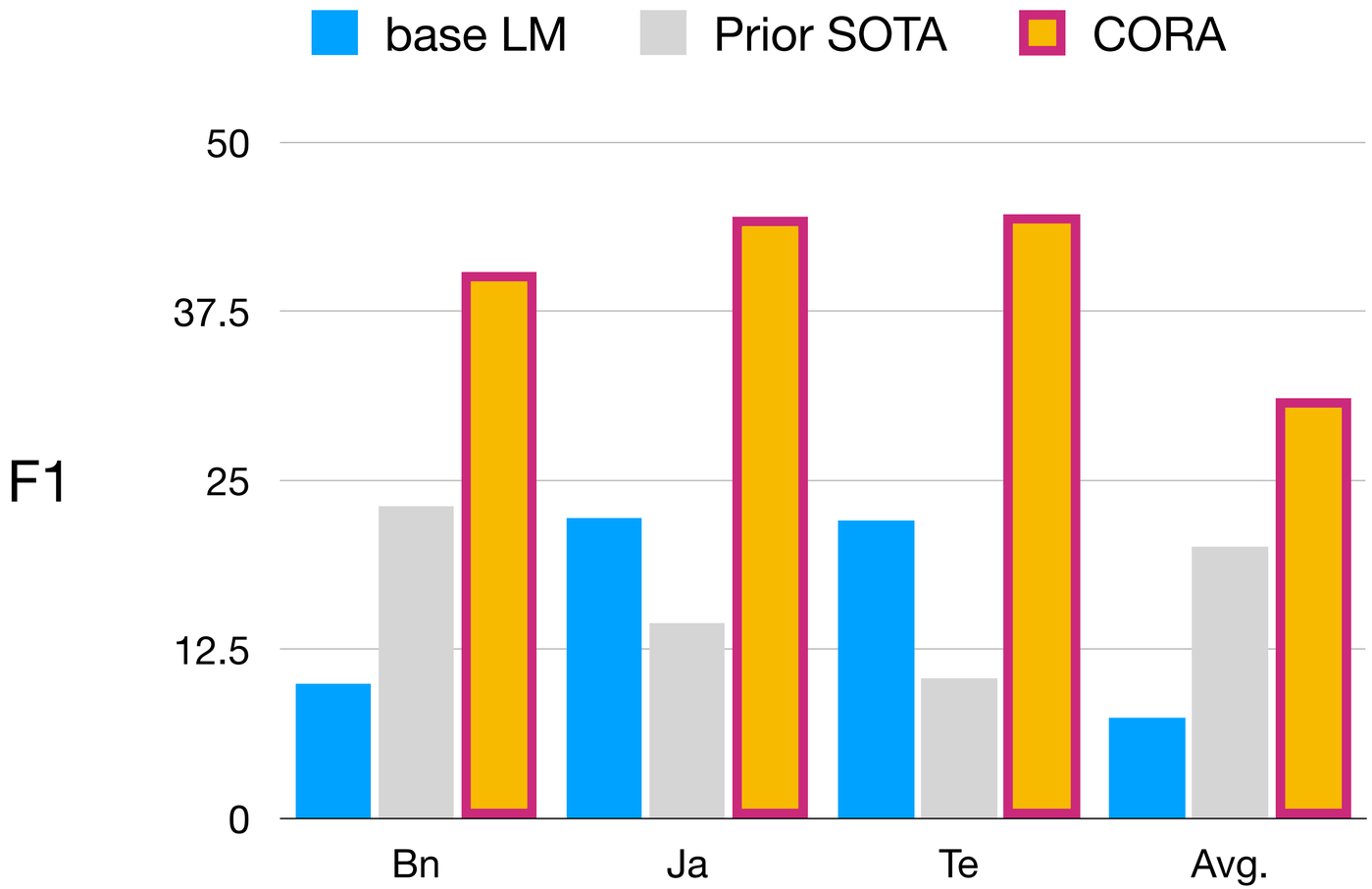


CORA: Iteratively training multilingual LM & retriever

Repeat DA+training

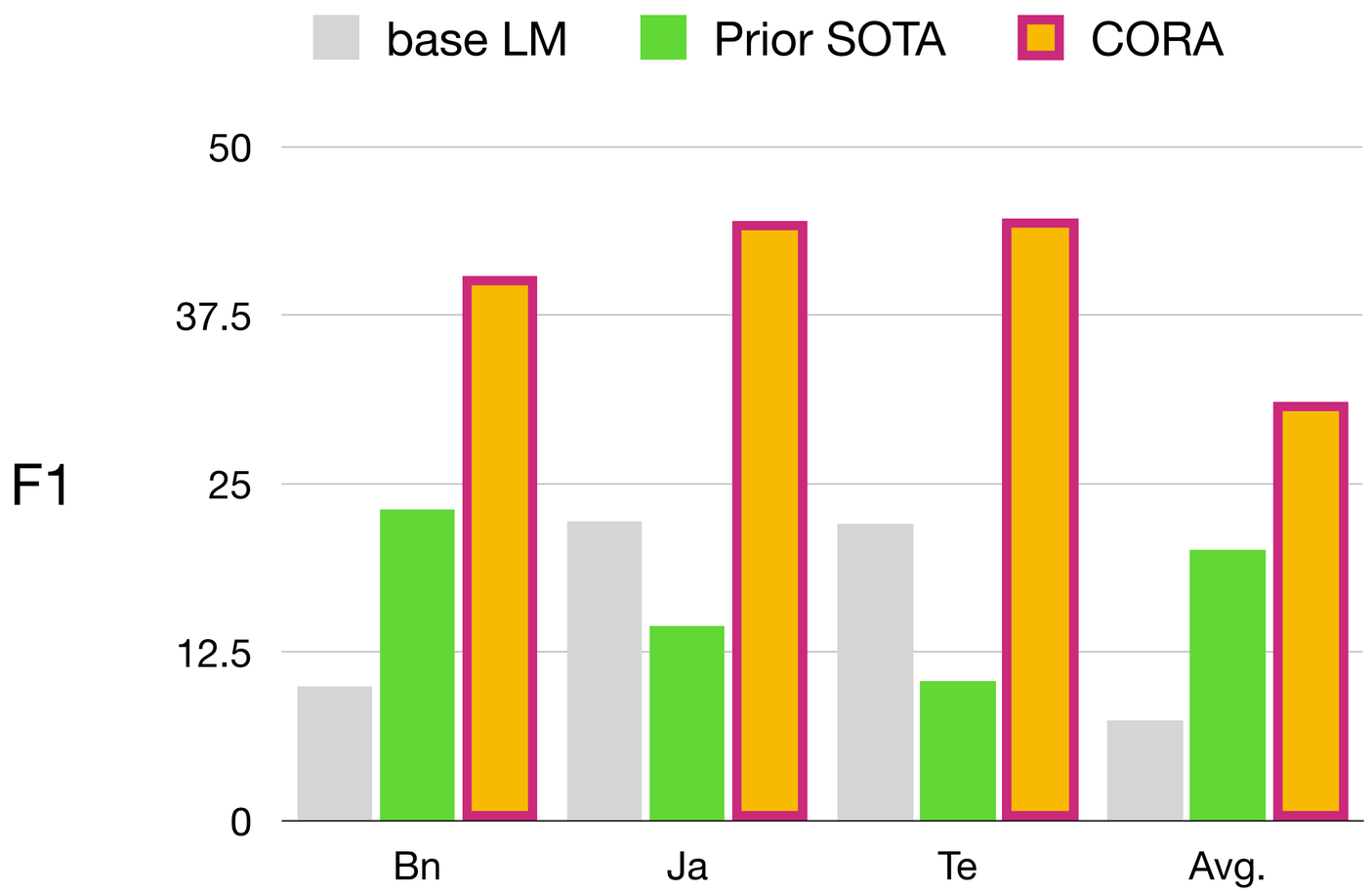


Results



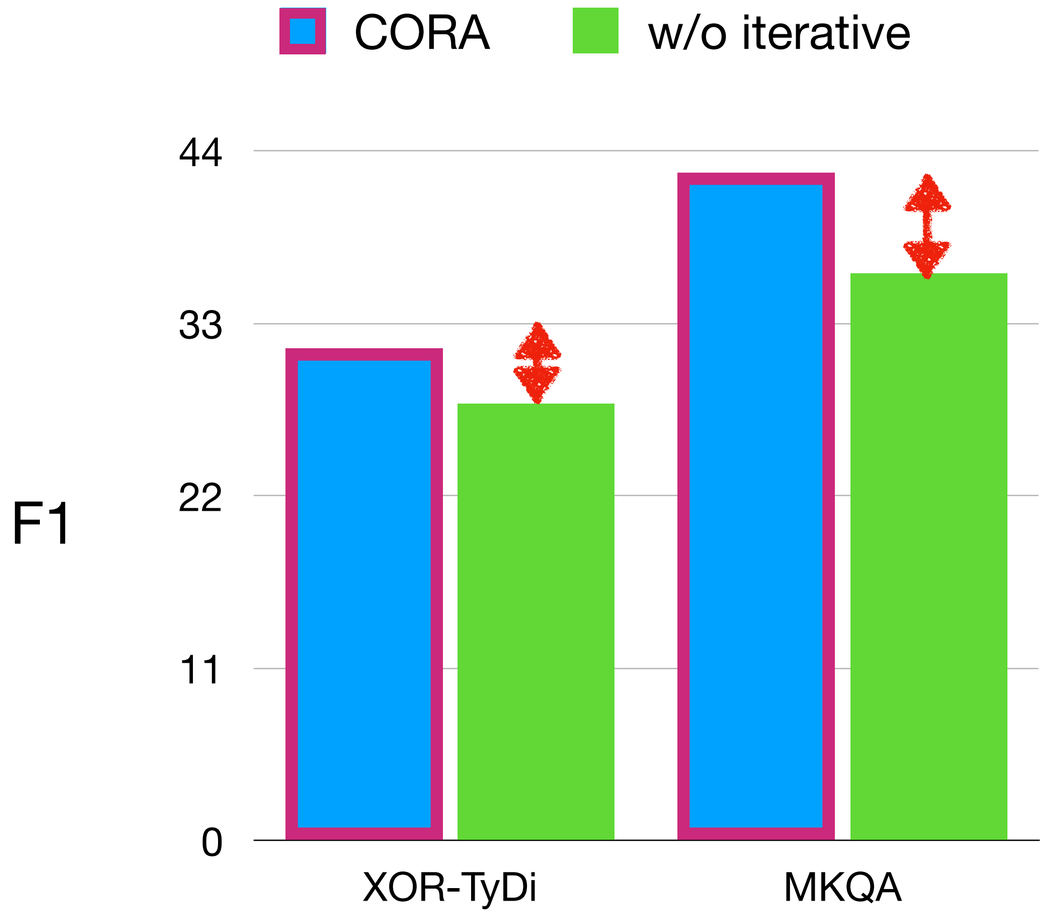
Large gains from fine-tuned LM without retrieval

Results



Significantly outperforms prior SOTA

Ablations: Effects of iterative retrieval



Iterative training of retriever and LM gives large performance improvements

Multilingual retrieval-based LMs for diverse tasks

Question Answering

* CL-ReLKT (Limkonchotiwat et al., 2022): knowledge transfer for better cross-lingual retrieval training
* Gen-TyDi QA (Muller et al., 2023): generate full sentence answers for cross-lingual QA.
* AfriQA (Ogundepo et al., 2023):: Cross-lingual Open-Retrieval Question Answering for African Languages: the first open-domain QA datasets for African languages

Fact verification

* CONCRETE (Hung et al., 2022): Improving cross-lingual fact-checking with cross-lingual retrieval

Dialogue

* Cross-lingual Knowledge-grounded Dialogue (Kim et al 2021): a Korean knowledge-grounded dialogue system that learns to generate Korean response given English & Korean knowledge

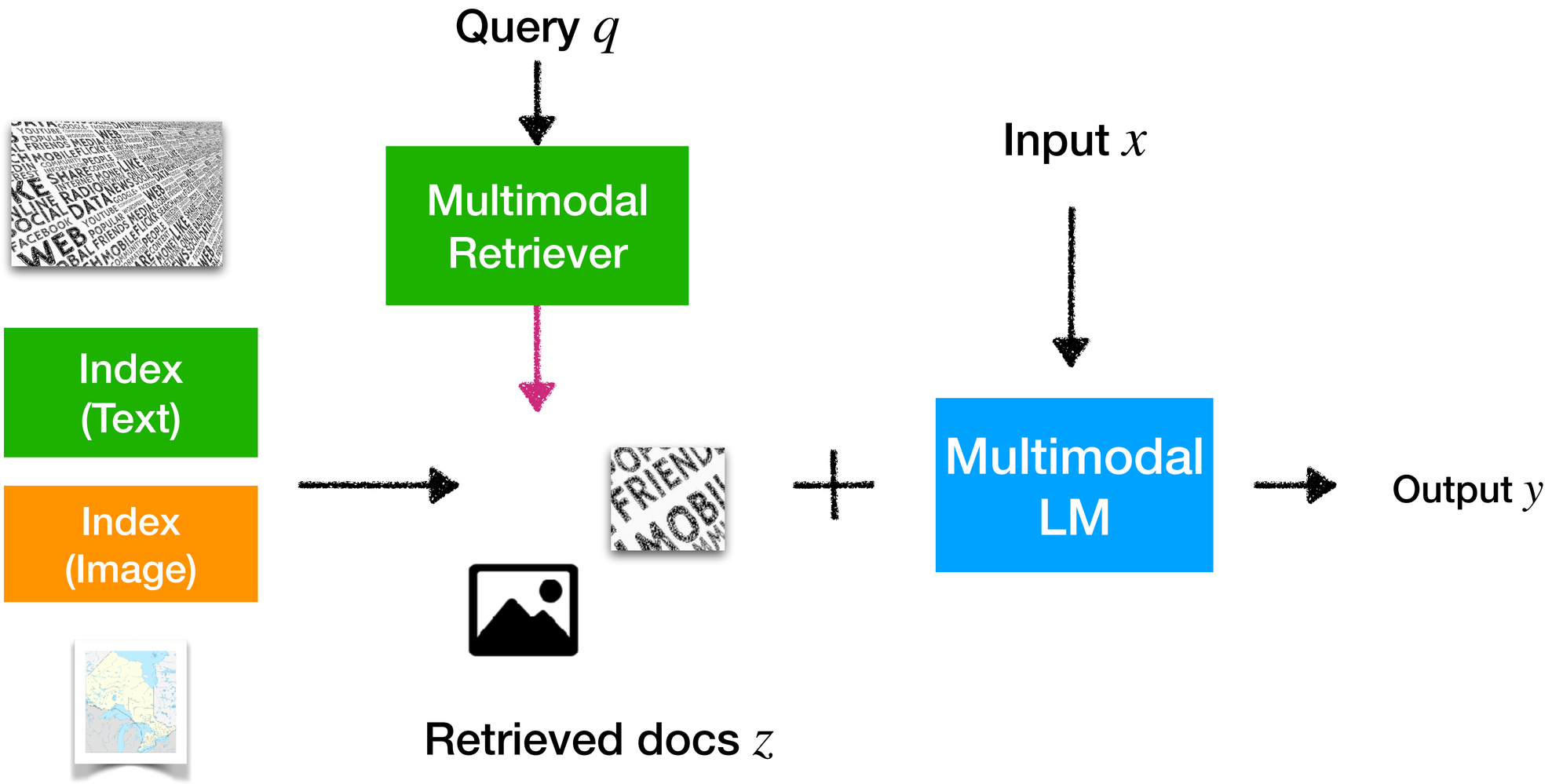
Event extraction

* R-GQA (Du and Ji, 2022): retrieve similar QA pairs for event argument extraction.

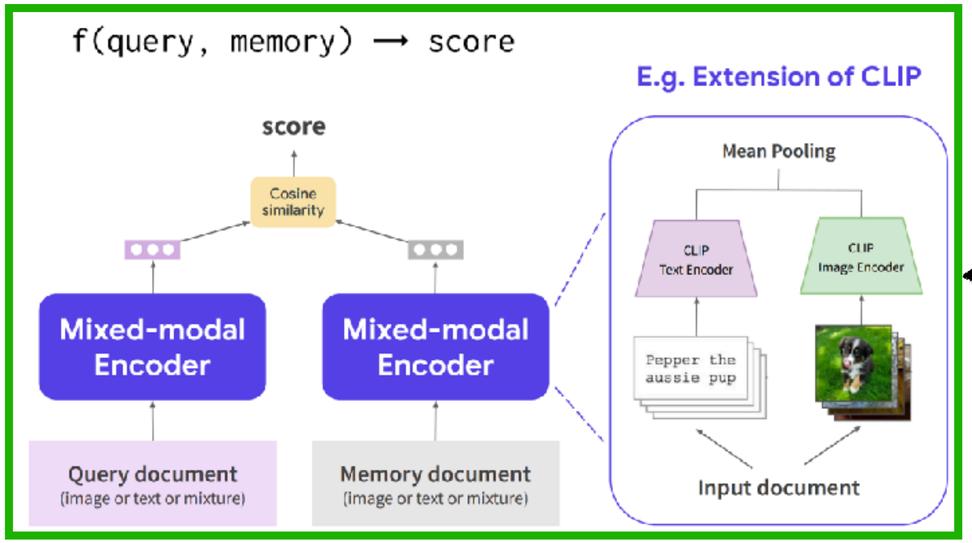
Key-phrase generations

* Retrieval-augmented Multilingual Key phrase Generation (Gao et al 2022): Using iterative training to improve retrieval & LM for key phrase generations

Multi-modal retrieval-based LMs



RA-CM3 (Yasunaga et al., 2023)



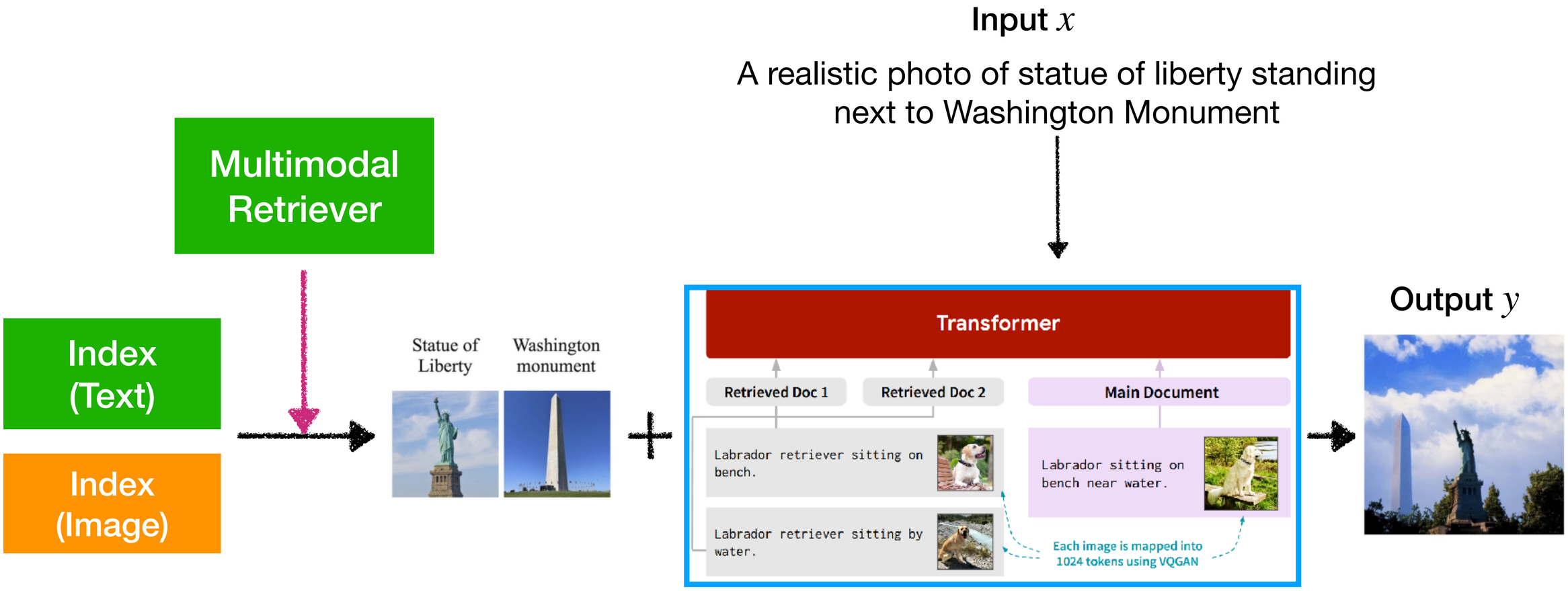
A realistic photo of statue of liberty standing next to Washington Monument

Index (Text)

Index (Image)

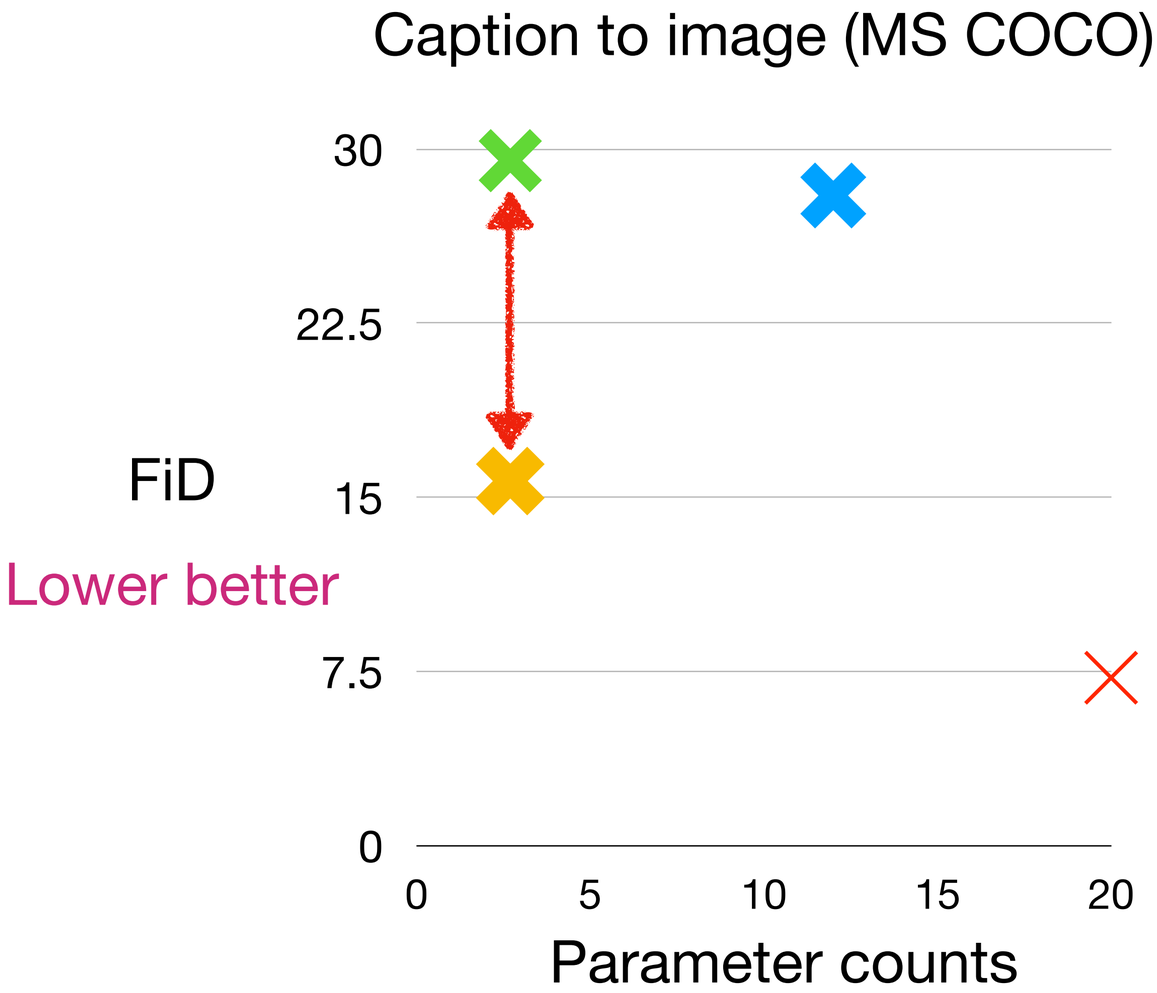
Yasunaga et al. 2023. “Retrieval-Augmented Multimodal Language Modeling”

RA-CM3 (Yasunaga et al., 2023)



Yasunaga et al. 2023. “Retrieval-Augmented Multimodal Language Modeling”

Results

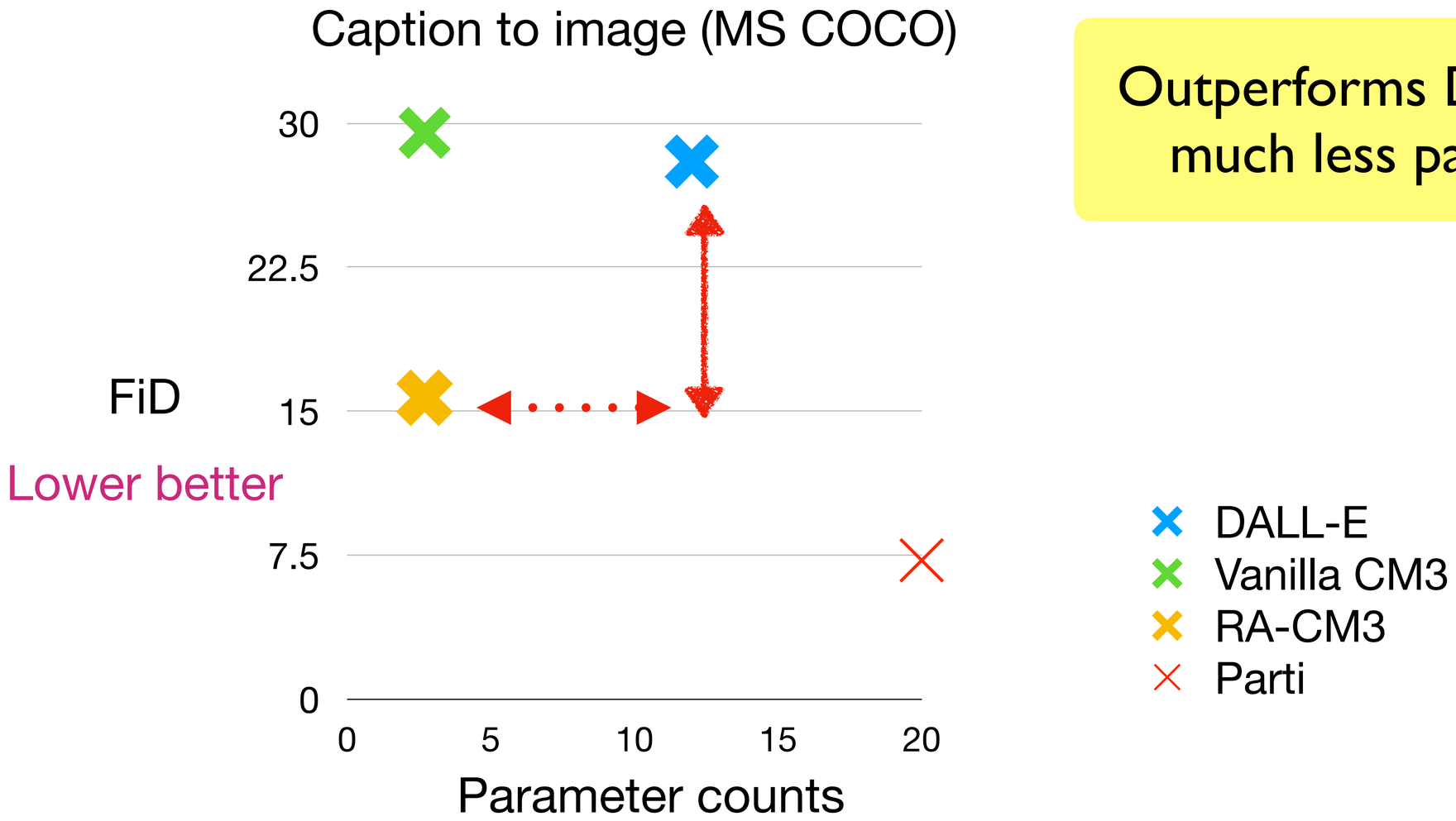


Significantly outperforms base LM (CM3)

- DALL-E
- Vanilla CM3
- RA-CM3
- Parti

Yasunaga et al. 2023. “Retrieval-Augmented Multimodal Language Modeling”

Results



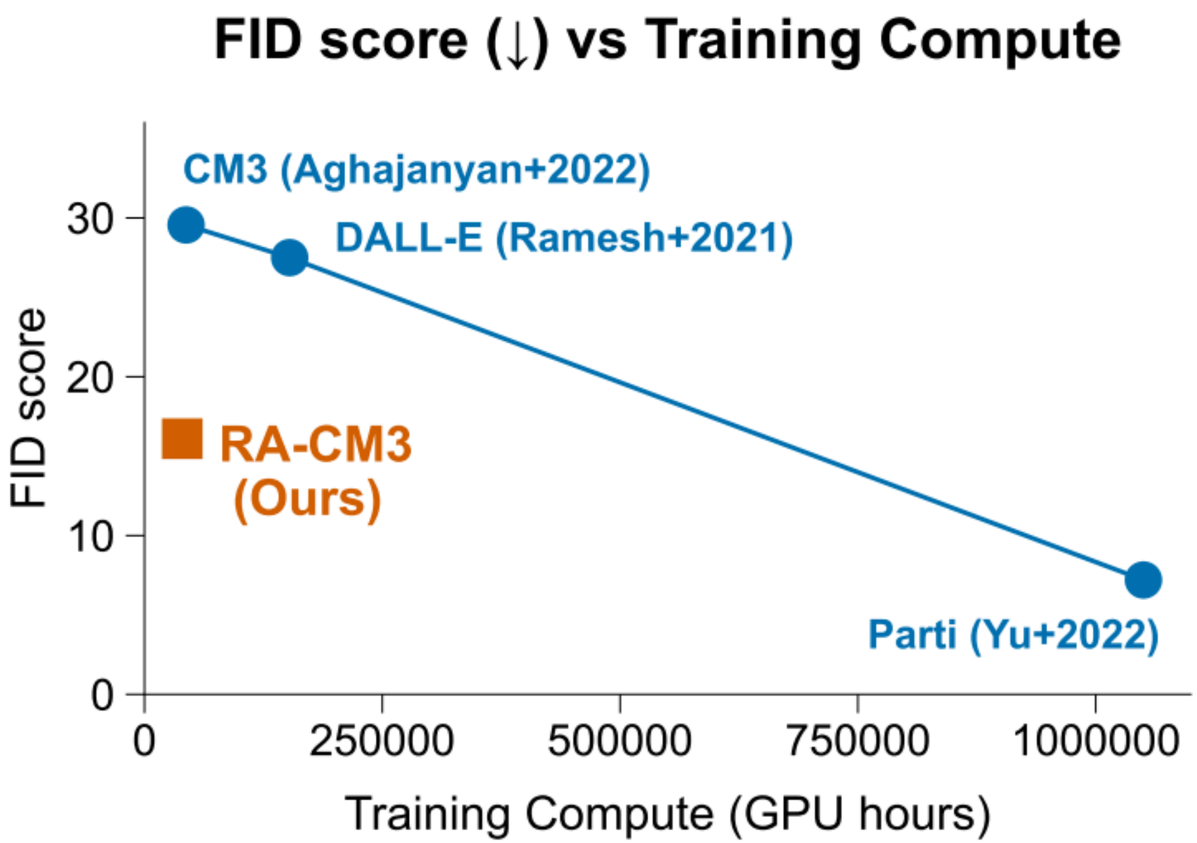
Outperforms DALL-E (12B) using much less parameters (2.7 B)

- DALL-E
- Vanilla CM3
- RA-CM3
- Parti

Yasunaga et al. 2023. “Retrieval-Augmented Multimodal Language Modeling”

Results

Caption to image (MS COCO)



Achieves significantly better training efficiency

Yasunaga et al. 2023. “Retrieval-Augmented Multimodal Language Modeling”

More applications beyond text

Multi-modal Retrieval-augmented Pre-training

* RAVEAL (Hu et al 2023): Pre-training visual-language model using knowledge memory

Multi-modal Question Answering

* MuRAG (Chen et al., 2022)

Multi-modal Classification

* ALIGN (Gur et al., 2021)

Multimodal using image and text have been actively studied

More applications beyond text

Multi-modal Retrieval-augmented Pre-training

* RAVEAL (Hu et al 2023): Pretraining visual-language model using knowledge memory

Multi-modal Question Answering

* MuRAG (Chen et al., 2022)

Multi-modal Classification

* ALIGN (Gur et al., 2021)

Retrieval-augmented training for molecules

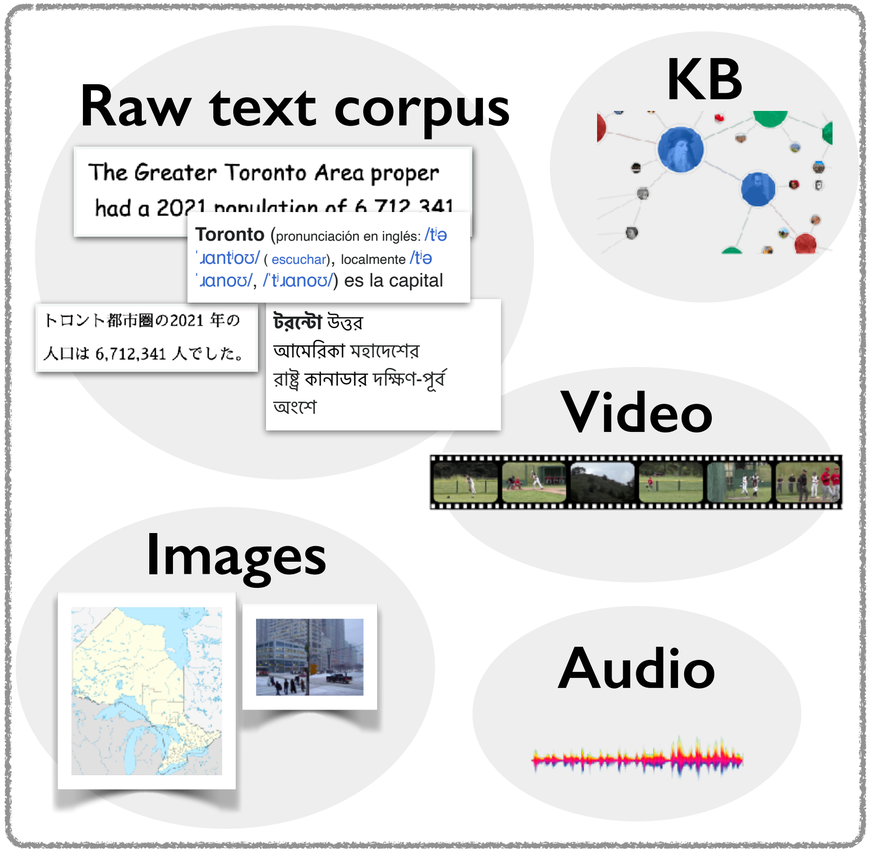
* Retrieval-based Molecule Generation (Wang et al., 2023)

Retrieval-augmented 3D motion generations

* ReMoDiffus (Zhang et al., 2023)

New extensions for new input / output modality!

Wrapping up



Extension to multilingual

Cross-lingual retrieval and generation to overcome **datastore scarcity** in many world languages

Extension to multimodal

Key effectivenesses (Section 5; long-tail, efficiency) apply to diverse modality