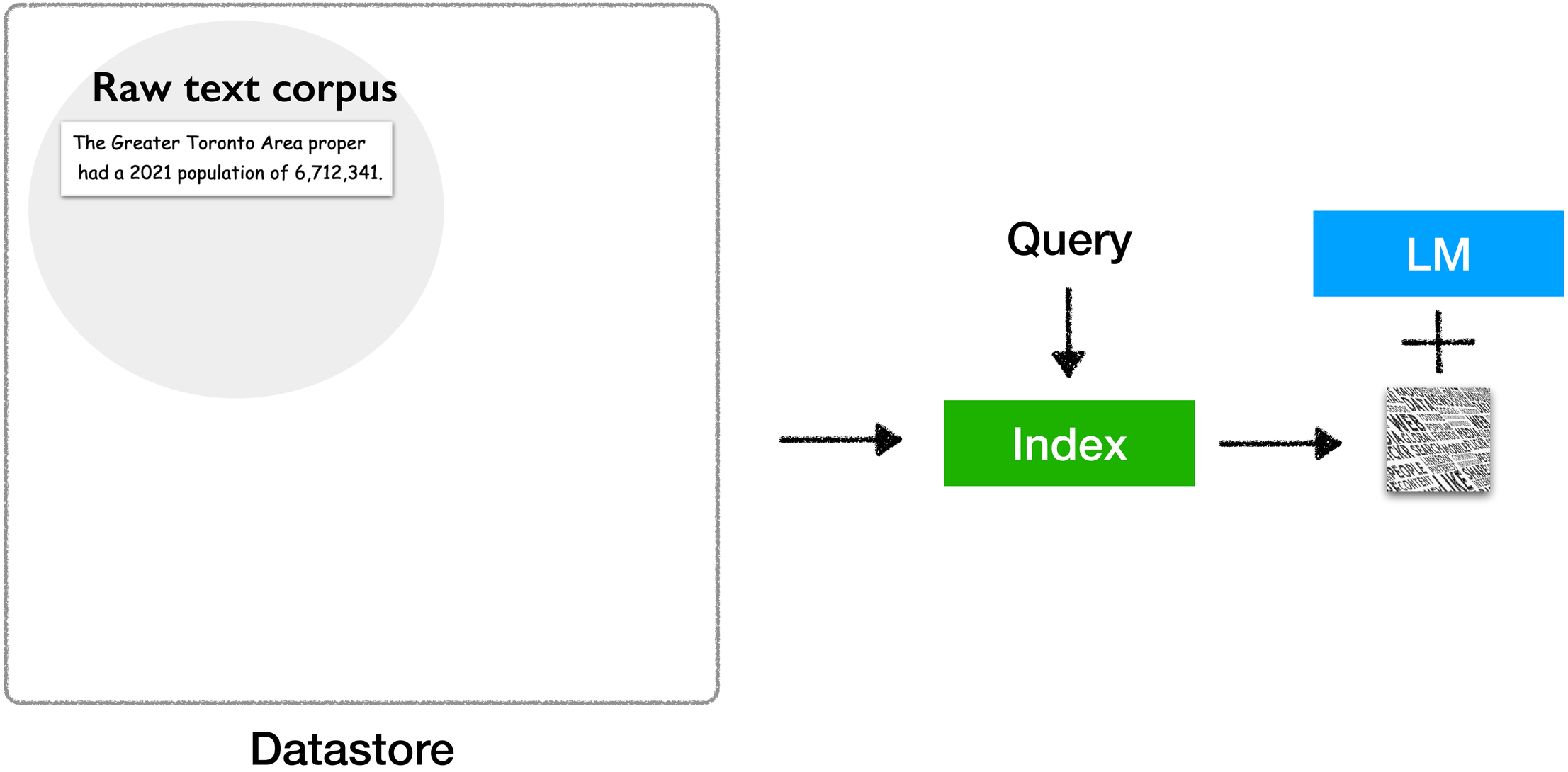
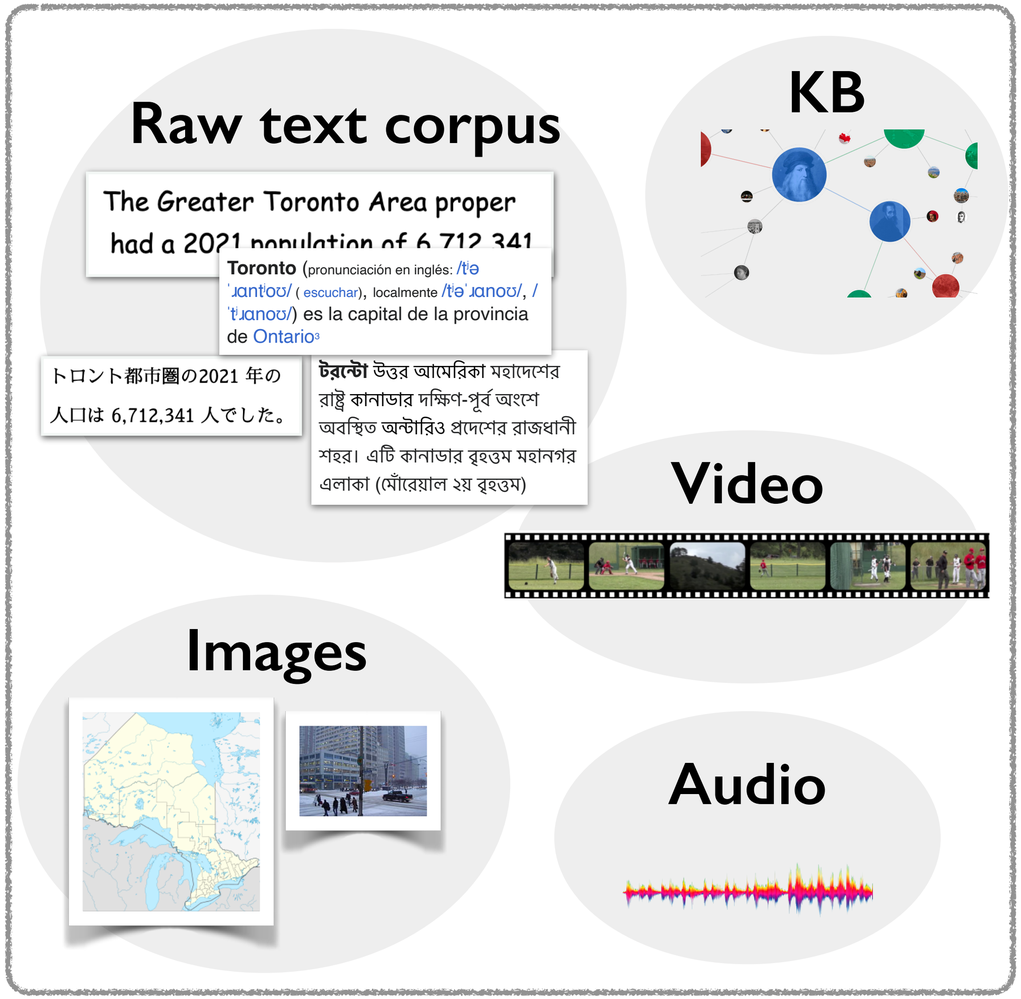
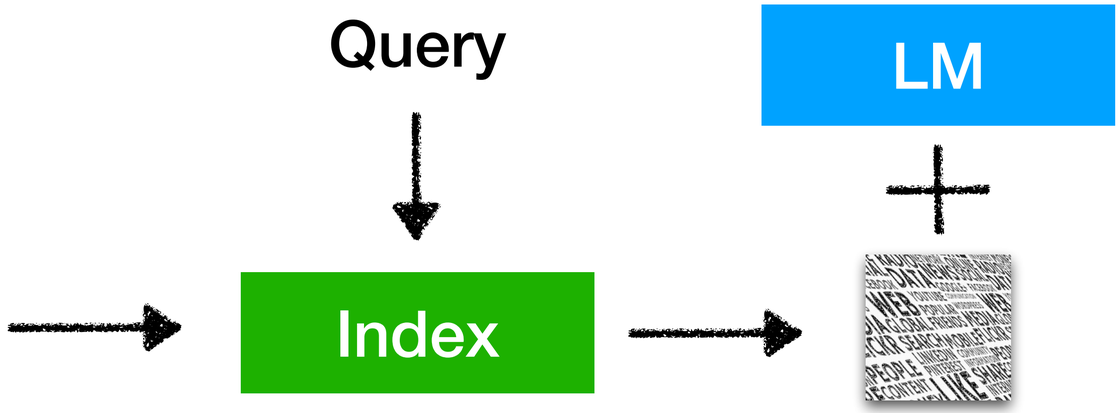
Section 7:
Multilingual & Multimodal

Retrieval-based LM for diverse knowledge sources

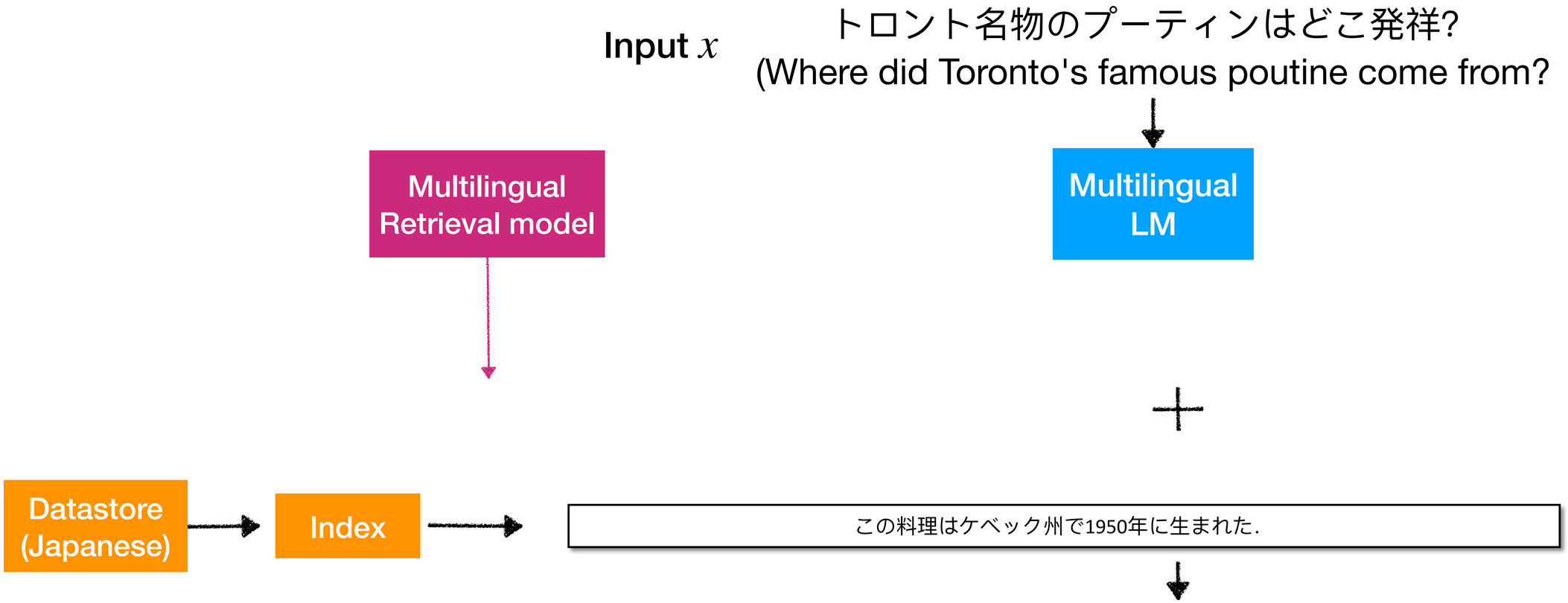


Retrieval-based LM for diverse knowledge sources





Multilingual Retrieval-based LM

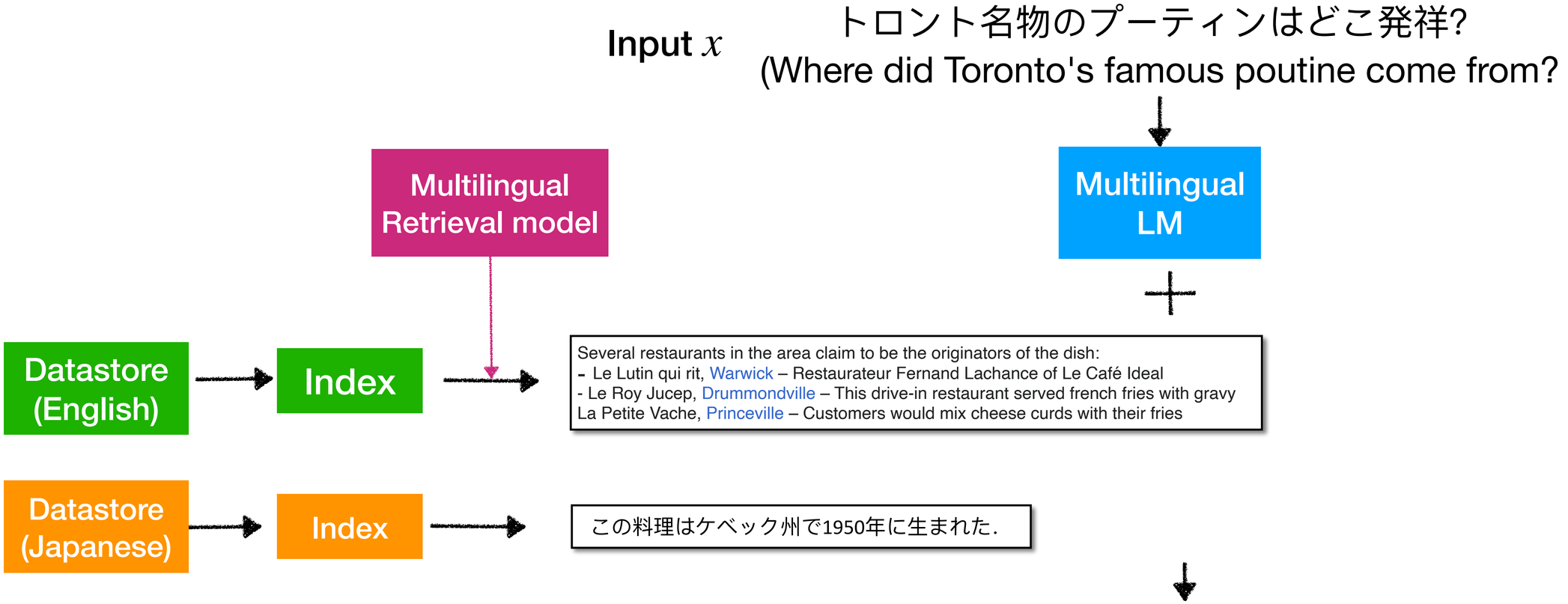


(Limited amount of data is available in Japanese datastore…)

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

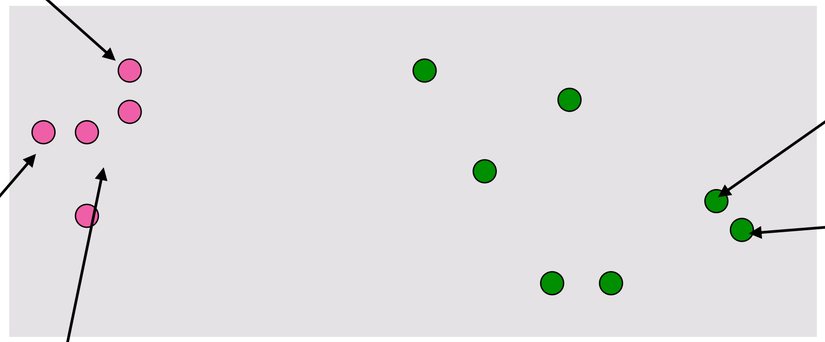


Language biases in representation spaces

この料理はケベック州で
1950年に生まれた.

Several restaurants in the area claim to be the originators of the dish:
- Le Lutin qui rit, Warwick – Restaurateur Fernand Lachance of Le Café Ideal
- Le Roy Jucep, Drummondville – This drive-in restaurant served french fries with gravy
La Petite Vache, Princeville – Customers would mix cheese curds with their fries

Datastore (Japanese)



The Greater Toronto Area proper had a 2021 population of 6,712,341.

トロント都市圏の2021 年の人口は 6,712,341 人でした。

Datastore (English)

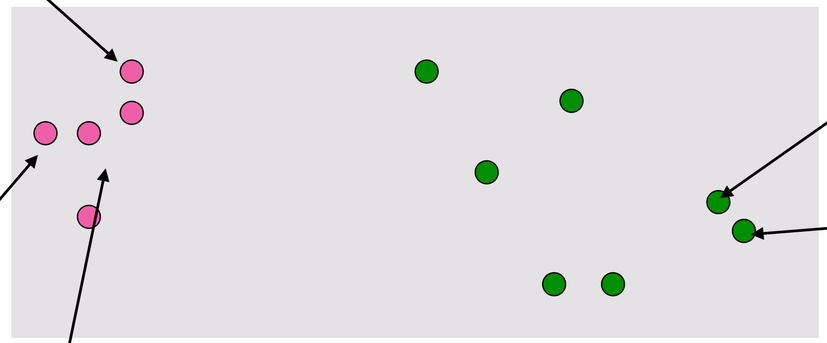
最も有名なランドマークはCNタワーで、地上建築物としてはかつて世界最高であった553 mの高さを持つ。

Language biases in representation spaces

この料理はケベック州で
1950年に生まれた.

Several restaurants in the area claim to be the originators of the dish:
- Le Lutin qui rit, Warwick – Restaurateur Fernand Lachance of Le Café Ideal
- Le Roy Jucep, Drummondville – This drive-in restaurant served french fries with gravy
La Petite Vache, Princeville – Customers would mix cheese curds with their fries

Datastore (Japanese)



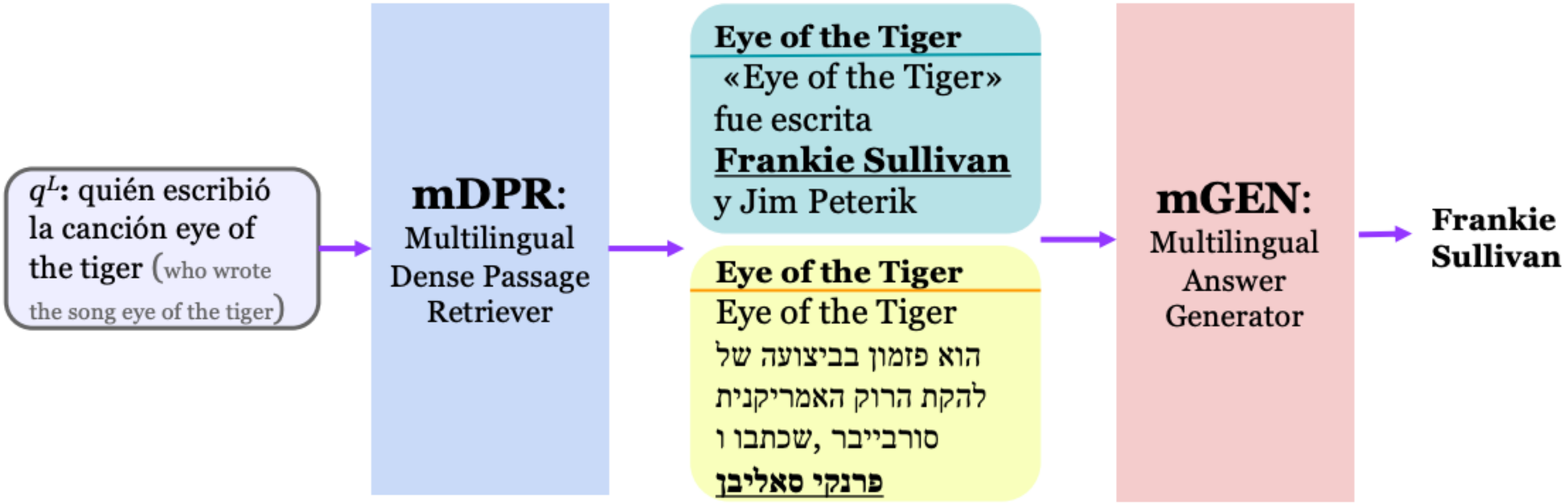
The Greater Toronto Area proper had a 2021 population of 6,712,341.

トロント都市圏の2021 年の人口は 6,712,341 人でした。

Datastore (English)

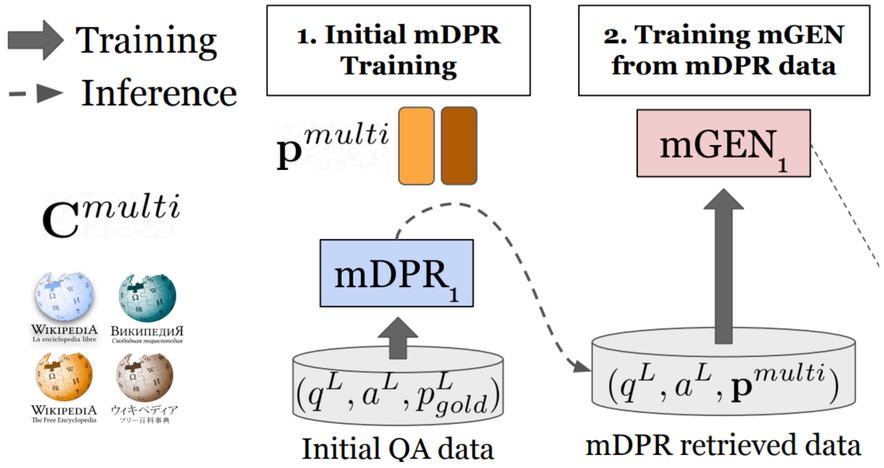
最も有名なランドマークはCNタワーで、地上建築物としてはかつて世界最高であった553 mの高さを持つ。

CORA (Asai et al., 2021)



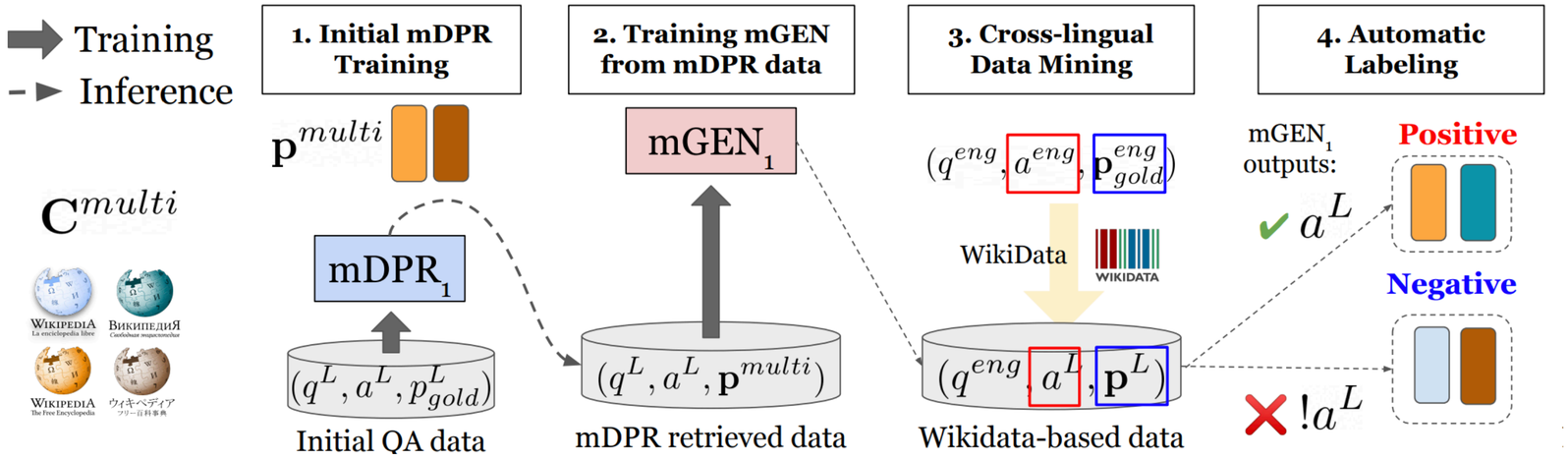
Asai et al. 2021. “One question answering model for many languages with cross-lingual dense passage retrieval”

CORA: iteratively training multilingual LM & retriever



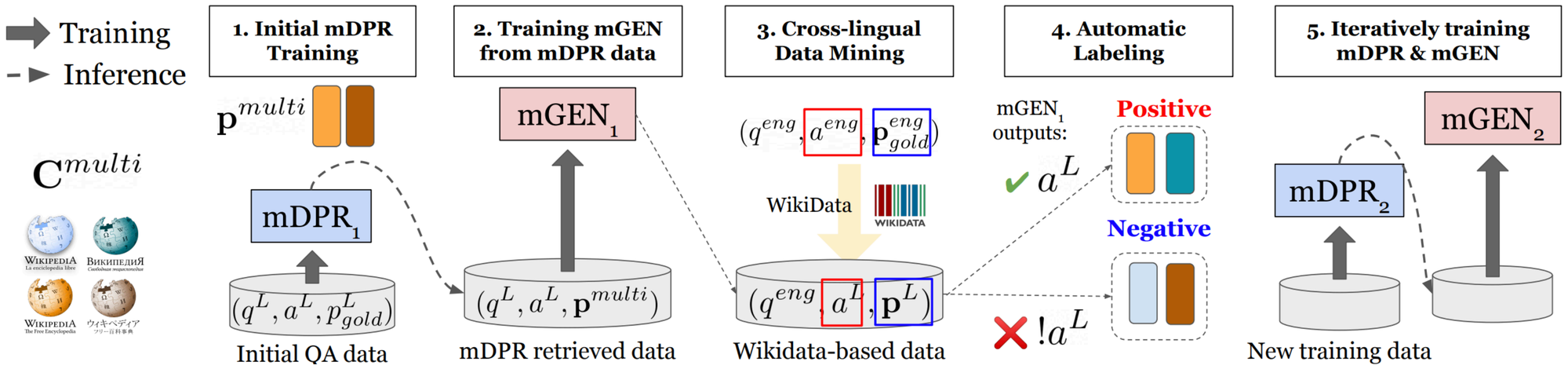
Initial training of retriever and LM

CORA: iteratively training multilingual LM & retriever



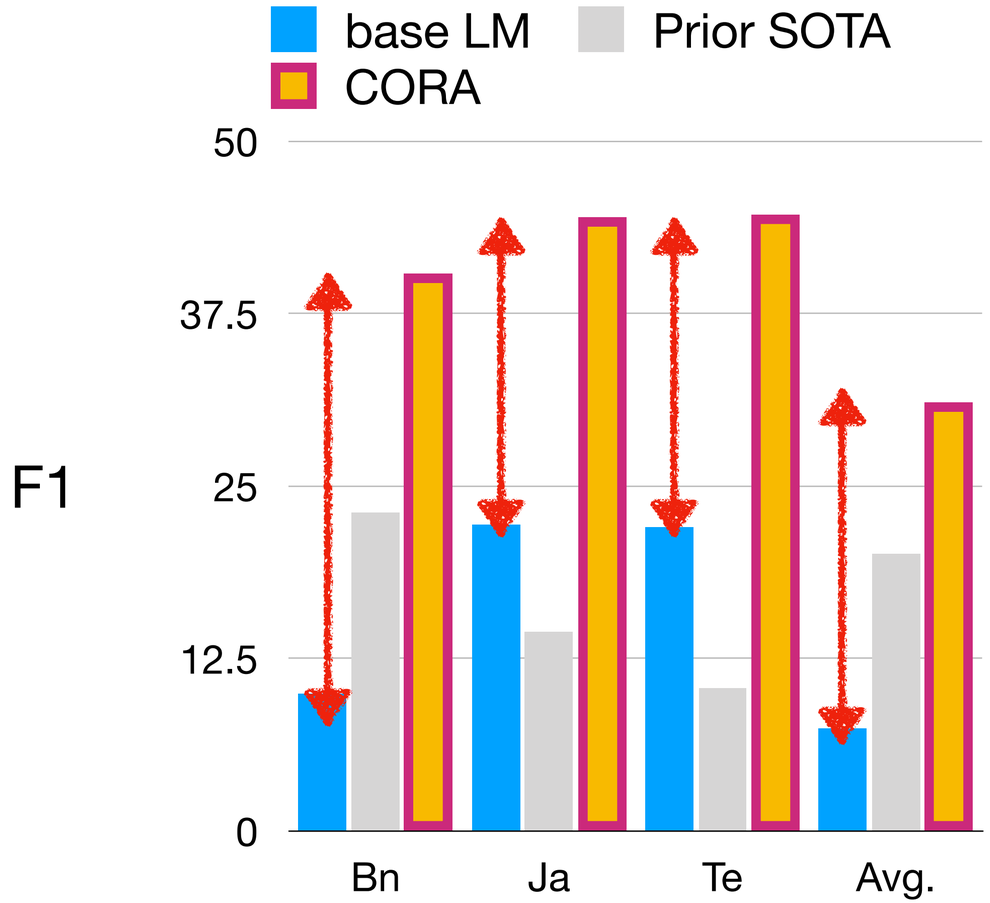
Expand training data using trained models as well as structured cross-lingual data.

CORA: iteratively training multilingual LM & retriever



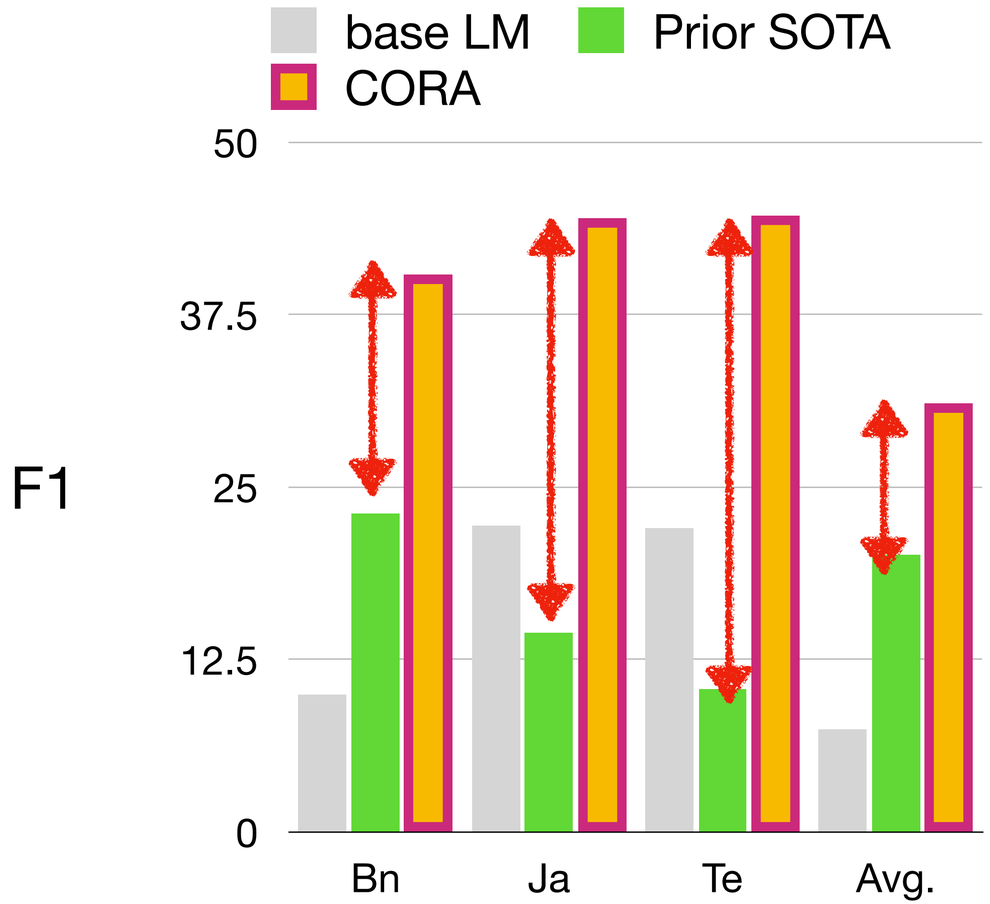
Repeat training

Results



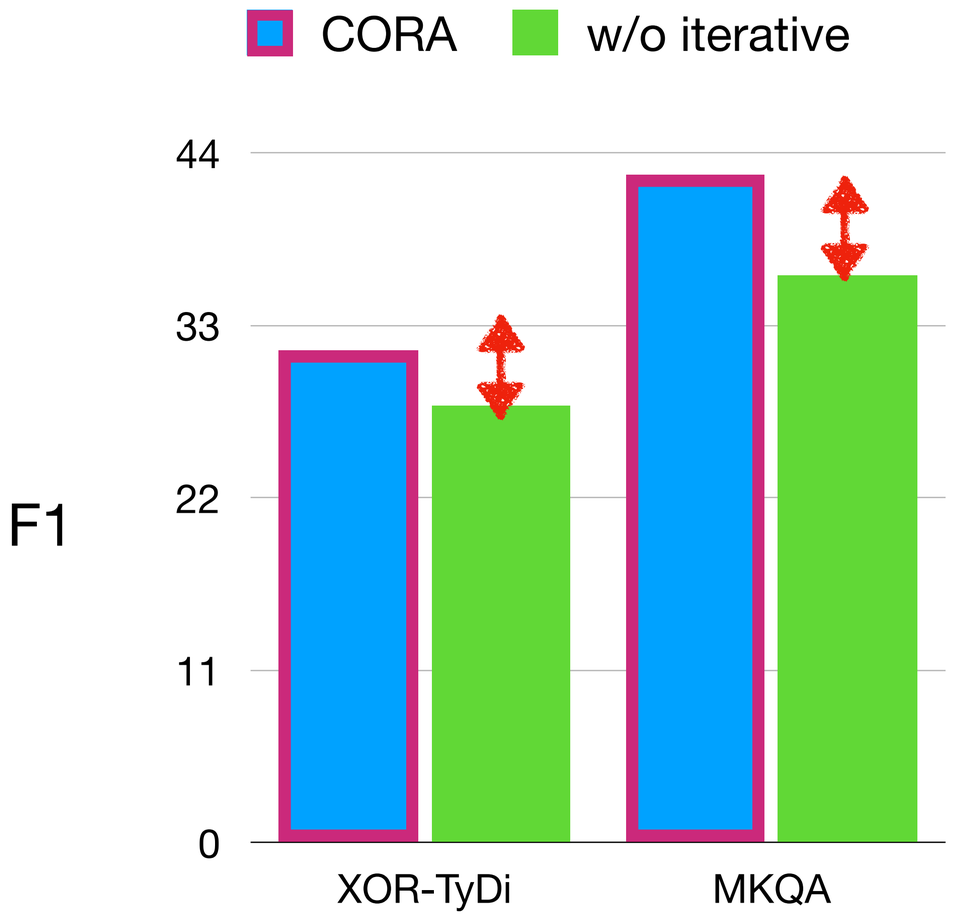
Large gains from fine-tuned LM without retrieval

Results



Significantly Outperforms prior SOTA

Results



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.

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

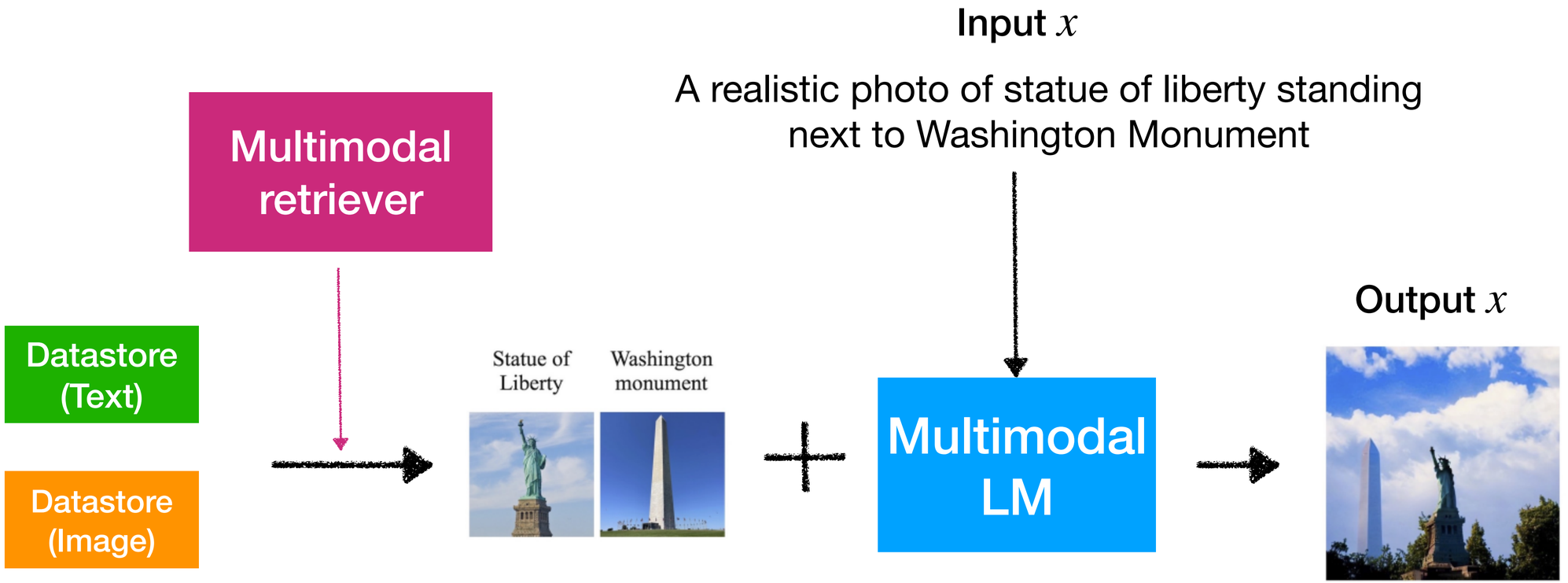
Event Extraction

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

Key-phrase generations

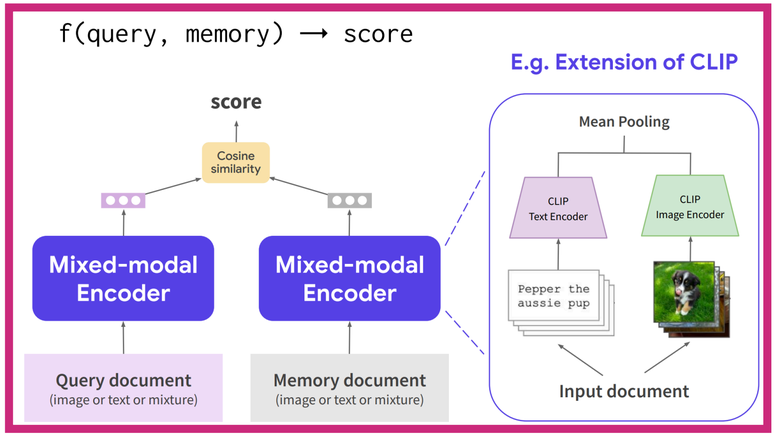
* Retrieval-augmented Multilingual Key phrase Generation (Gao et al 2022): Using iterative training to improve retrieval & LM for key phrase generations

Multi-modal retrieval-augmented generations



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

RA-CM3 (Yasunaga et al., 2023)



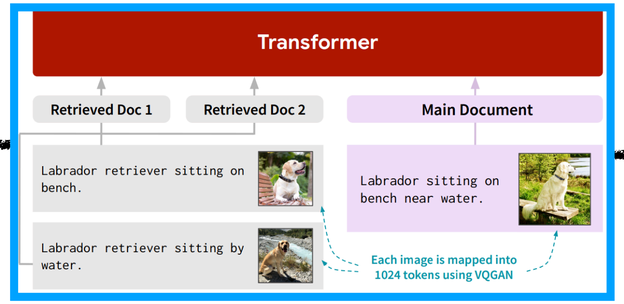
Input x

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

Output x

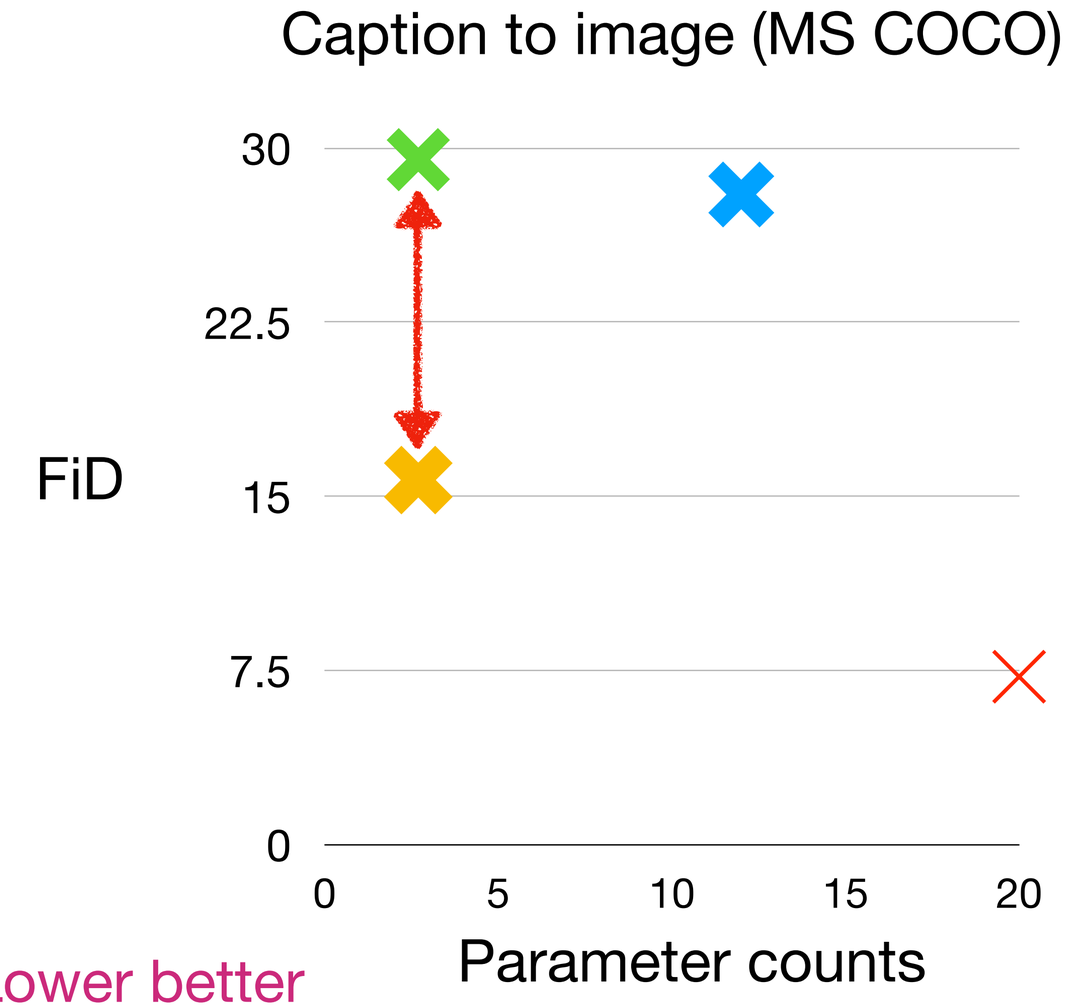
Datastore (Text)

Datastore (Image)



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

Results



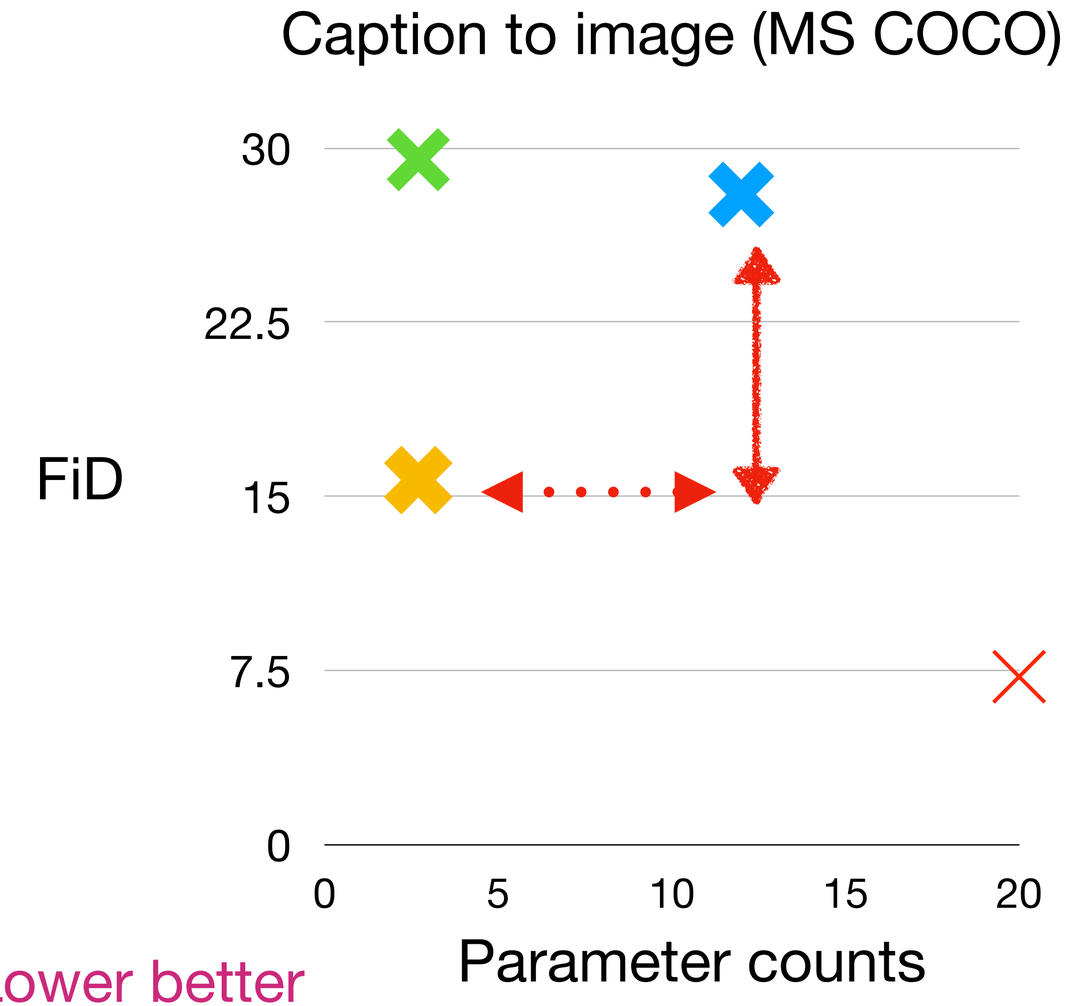
Lower better

Significantly outperforms base LM (CM3)

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

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

Results



Lower better

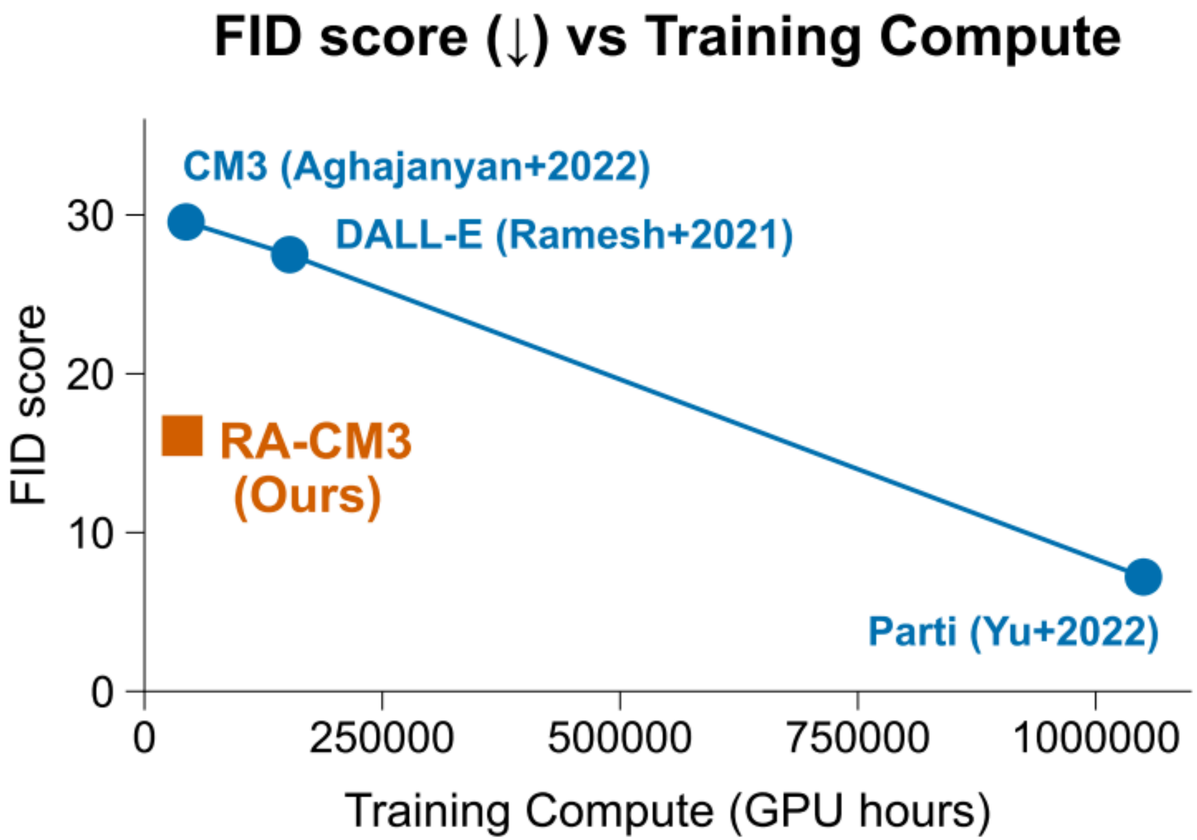
Outperforms DALE (12B) using much less params (2.7 B)

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

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

Results

Caption to image (MS COCO)



Retrieval-augmented generation 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): Pretraining 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

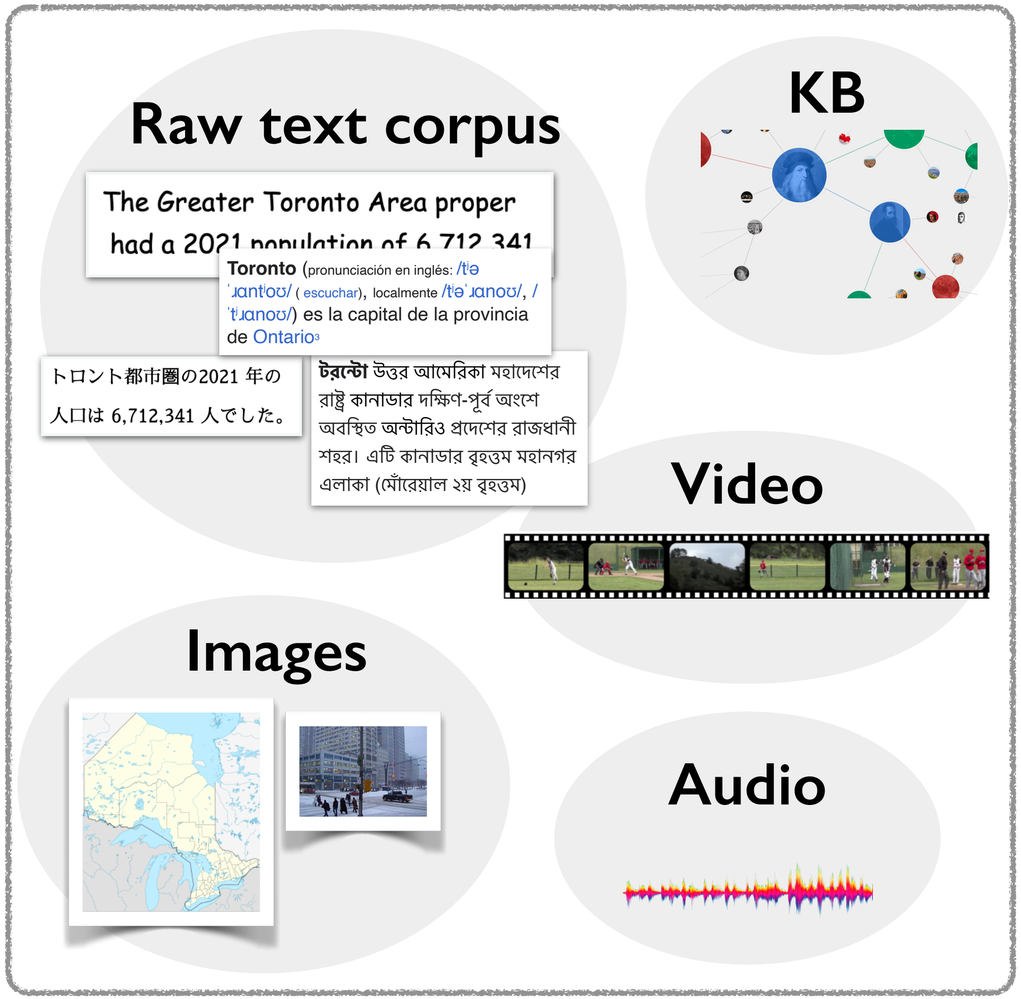
* Retrieval-based Molecule Generation (Wang et al., 2023)

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

References (1)

Akari Asai, Xinyan Yu, Jungo Kasai, Hannaneh Hajishirzi. One Question Answering Model for Many Languages with Cross-lingual Dense Passage Retrieval. *NeurIPS 2021*.

Peerat Limkonchotiwat, Wuttikorn Ponwitayarat, Can Udomcharoenchaikit, Ekapol Chuangsuwanich, Sarana Nutanong. CL-ReLKT: Cross-lingual Language Knowledge Transfer for Multilingual Retrieval Question Answering. *NAACL Findings 2022*.

Benjamin Muller, Luca Soldaini, Rik Koncel-Kedziorski, Eric Lind, Alessandro Moschitti. Cross-Lingual GenQA: A Language-Agnostic Generative Question Answering Approach for Open-Domain Question Answering. *AACL 2022*.

Kung-Hsiang Huang, ChengXiang Zhai, Heng Ji. CONCRETE: Improving Cross-lingual Fact-checking with Cross-lingual Retrieval. COLING *2022*.

Yifan Gao, Qingyu Yin, Zheng Li, Rui Meng, Tong Zhao, Bing Yin, Irwin King, Michael R. Lyu. Retrieval-Augmented Multilingual Keyphrase Generation with Retriever-Generator Iterative Training. NAACL Findings *2022*.

References (2)

Ziniu Hu, Ahmet Iscen, Chen Sun, Zirui Wang, Kai-Wei Chang, Yizhou Sun, Cordelia Schmid, David A. Ross, Alireza Fathi. REVEAL: Retrieval-Augmented Visual-Language Pre-Training with Multi-Source Multimodal Knowledge Memory. *CVPR 2023*.

Wenhu Chen, Hexiang Hu, Xi Chen, Pat Verga, William W. Cohen. MuRAG: Multimodal Retrieval-Augmented Generator for Open Question Answering over Images and Text. *EMNLP 2022*.

Shir Gur, Natalia Neverova, Chris Stauffer, Ser-Nam Lim, Douwe Kiela, Austin Reiter. Cross-Modal Retrieval Augmentation for Multi-Modal Classification. *EMNLP Findings 2021*.

Zichao Wang, Weili Nie, Zhuoran Qiao, Chaowei Xiao, Richard Baraniuk, Anima Anandkumar. Retrieval-based Controllable Molecule Generation. ICLR *2023*.

Mingyuan Zhang, Xinying Guo, Liang Pan, Zhongang Cai, Fangzhou Hong, Huirong Li, Lei Yang, Ziwei Liu. ReMoDiffuse: Retrieval-Augmented Motion Diffusion Model. Arxiv *2023*.