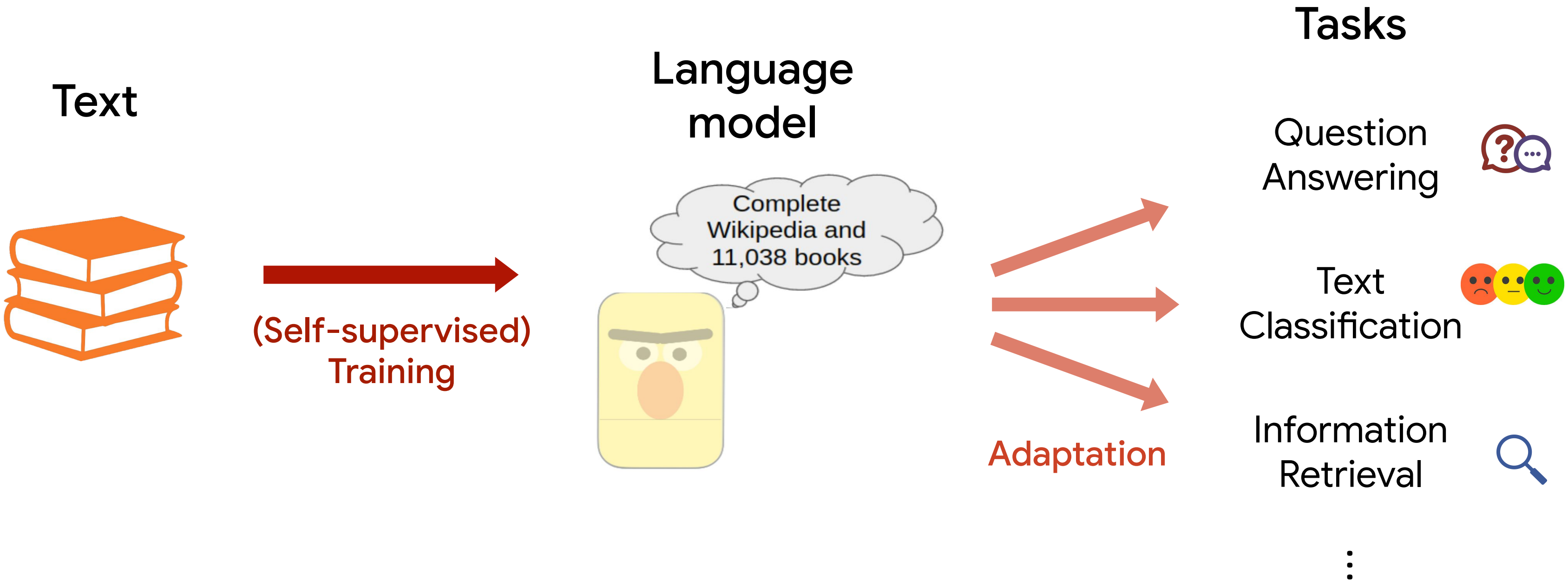
DRAGON:
Deep Bidirectional Language-Knowledge Pretraining

NeurIPS 2022

Michihiro Yasunaga, Antoine Bosselut, Hongyu Ren, Xikun Zhang, Chris Manning, Percy Liang*, Jure Leskovec*
Stanford University

Foundation Model Pretraining



Text & KG offer complementary information

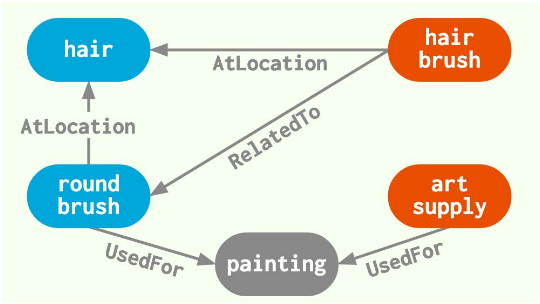
Text & Pretrained Language Model (LM)

- Broad coverage (e.g. Gao+2020)
- Captures rich context

WIKIPEDIA

Knowledge Graph (KG)

- Latent, structured relations
- Multihop reasoning (e.g. Yasunaga+2021)

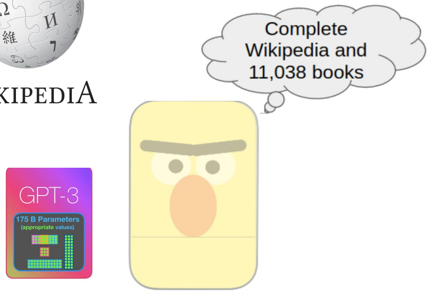


Goal: Combine text & KG for pretraining

Text

- Broad coverage (e.g. [Gao+2020])
- Captures rich context

WIKIPEDIA

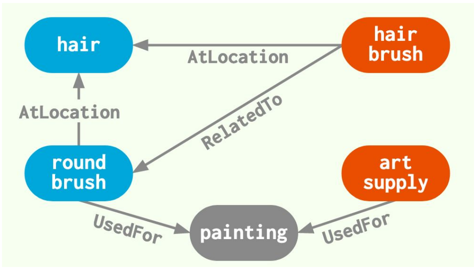


Joint Pretraining

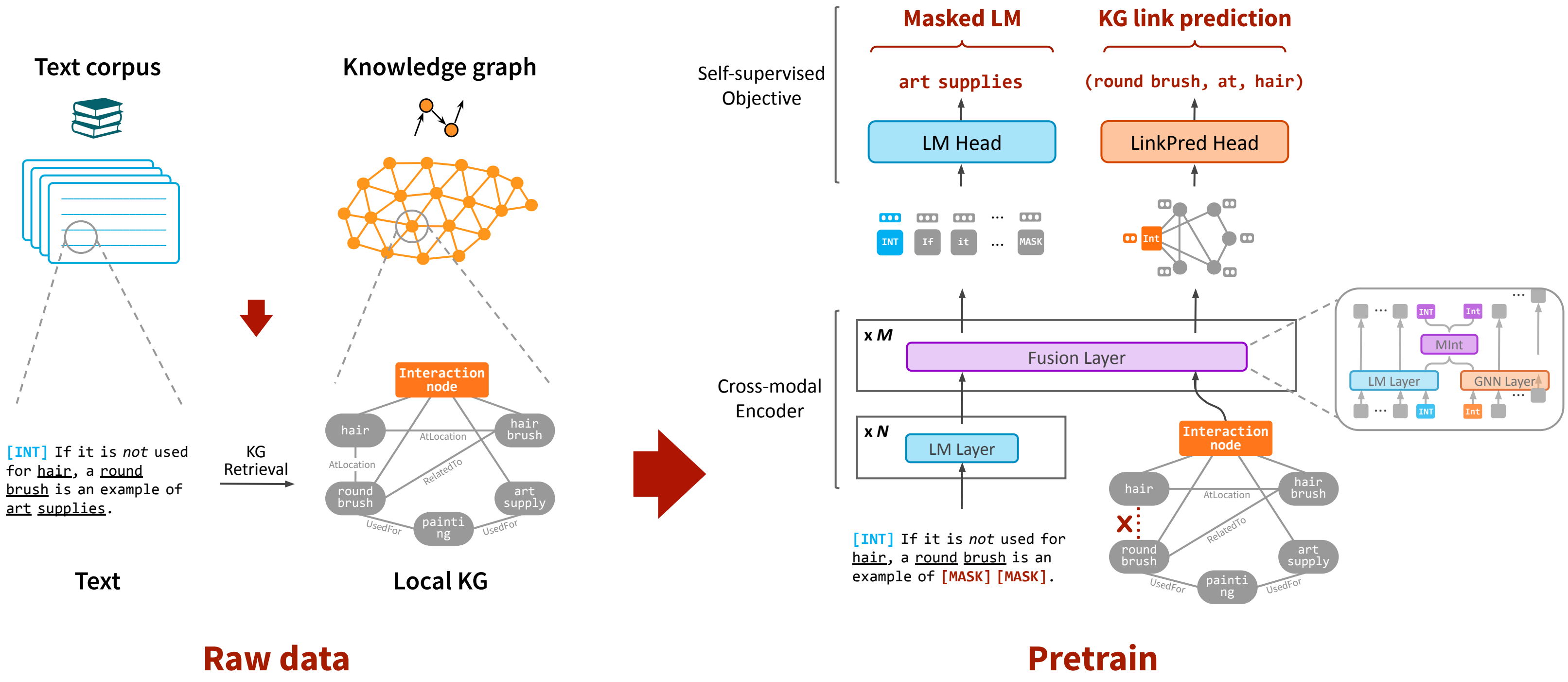
Language-Knowledge Foundation Model

Knowledge Graph (KG)

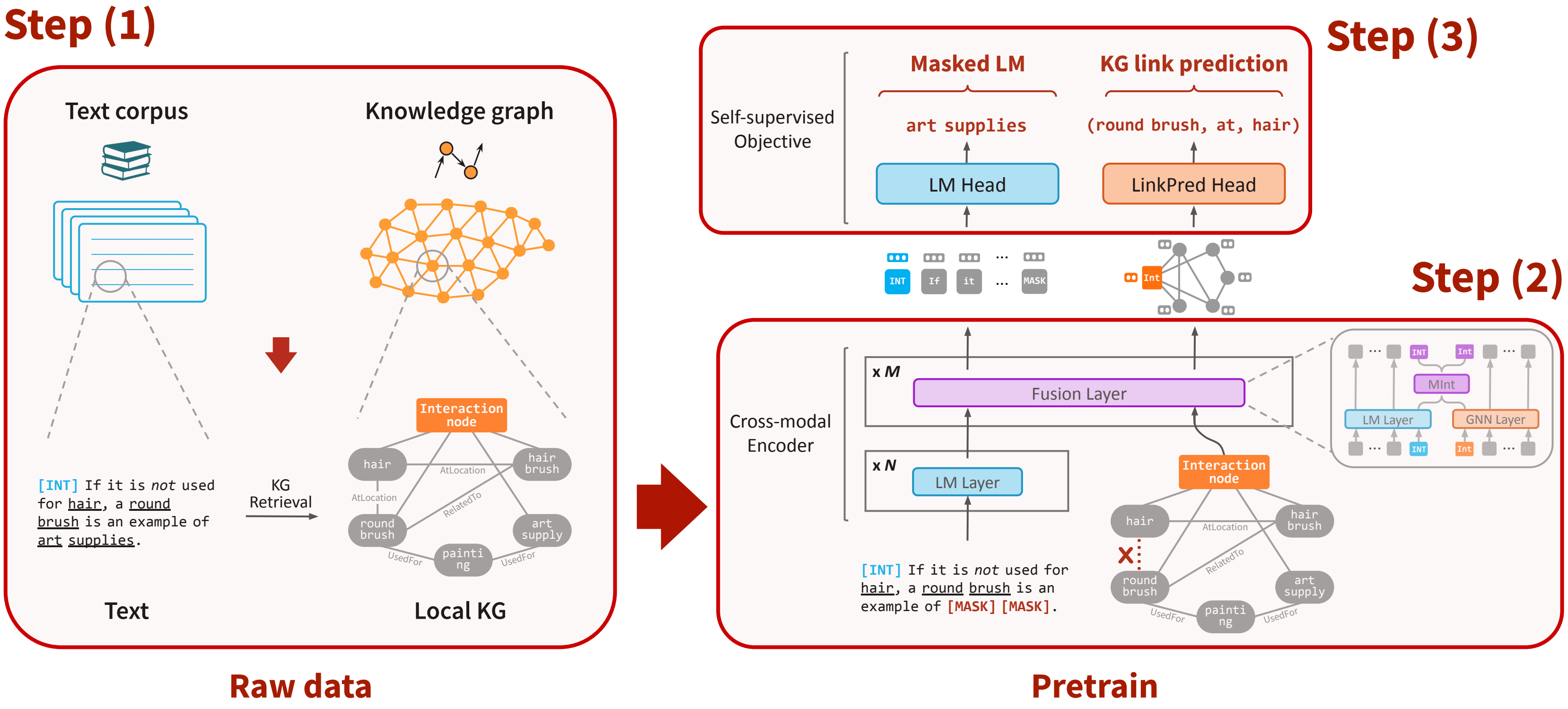
- Latent, structured relations
- Multihop reasoning (e.g. [Yasunaga+2021])



Proposed Method: DRAGON



Proposed Method: DRAGON



(1) Text-KG Input

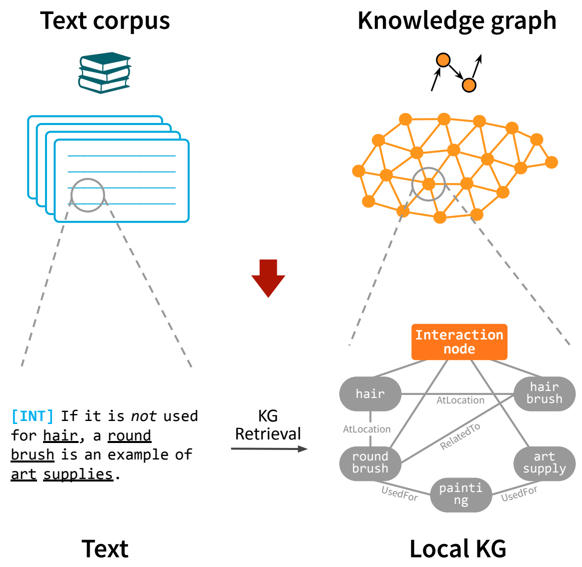
Motivation

- Informative pair of (text, local KG):
  - Text can contextualize the KG
  - KG can ground the text

Idea

- Given text corpus and KG, sample a text segment and retrieve a relevant knowledge subgraph by entity linking

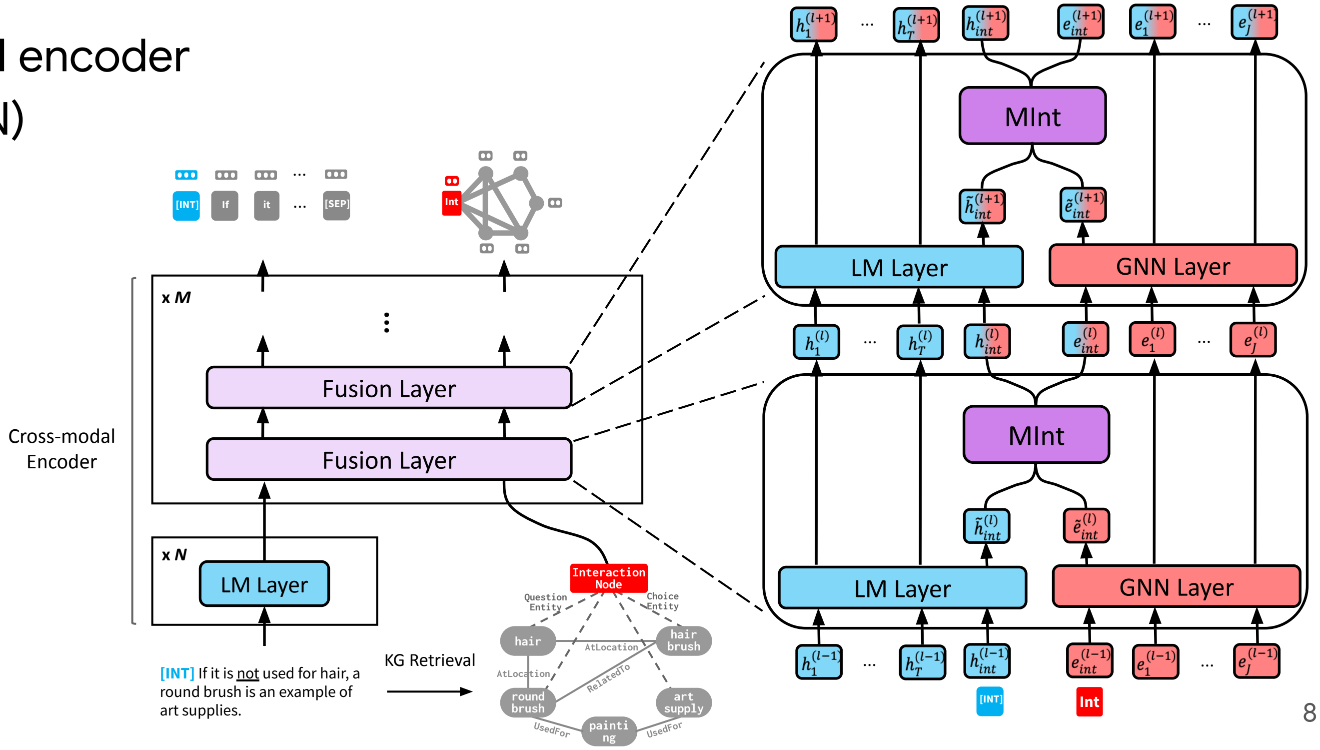
⇒ **Aligned pairs of (text, local KG)**



(2) Deep Bidirectional Cross-Modal Model

Idea

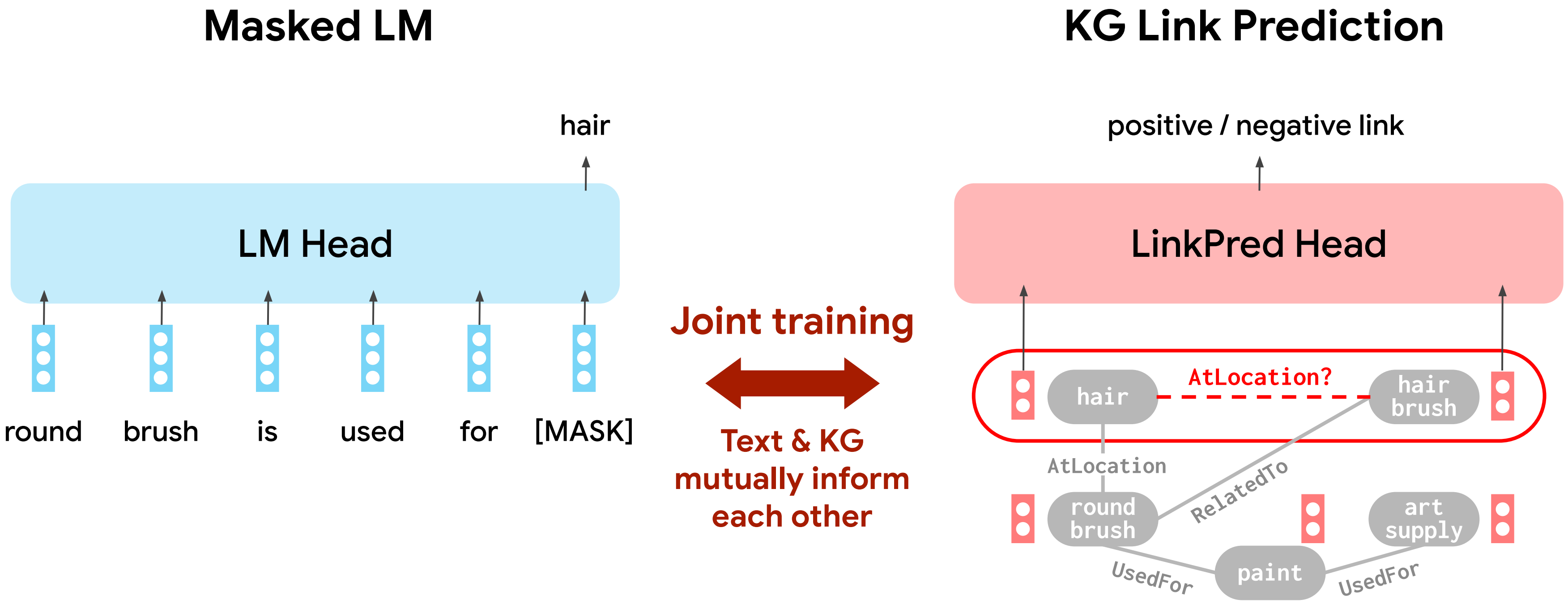
- Use the **GreaseLM** encoder (Transformer+GNN)
- Fuse text tokens & KG nodes bidirectionally for multiple layers



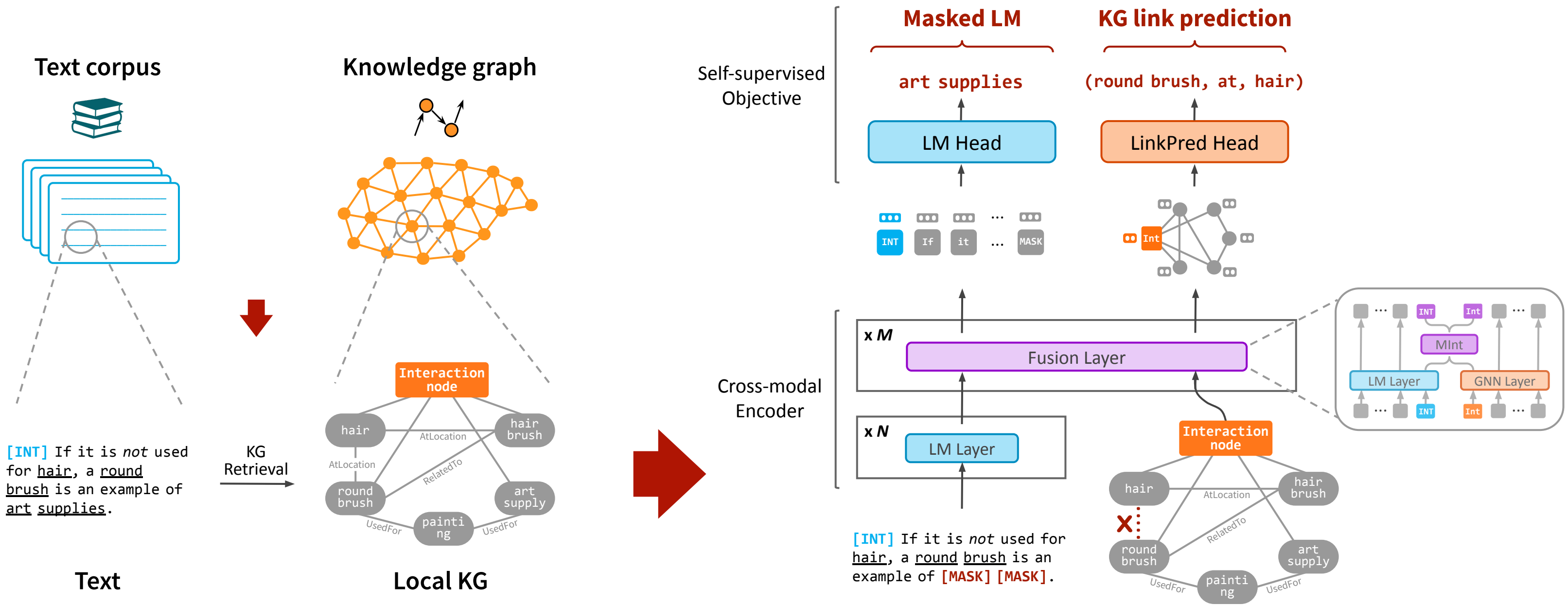
[Zhang et al. 2022](#)

(3) Bidirectional Self-Supervision

Idea: Pretrain with two self-supervised reasoning tasks



Proposed Method: DRAGON



Raw data **Pretrain**

Experiments

	General domain	Biomedical domain
Pretraining data	**Text: BookCorpus** (6GB) **KG: ConceptNet** (800K nodes, 2M edges)	**Text: PubMed** (20GB) **KG: UMLS** (300K nodes, 1M edges)
Downstream tasks	Commonsense reasoning (OBQA, RiddleSense, CommonsenseQA, CosmosQA, HellaSwag, PIQA, SIQA, aNLI, ARC)	Biomedical reasoning (PubMedQA, BioASQ, MedQA-USMLE)
Baseline: LM	RoBERTa (Liu+2019)	BioLinkBERT (Yasunaga+2022)
Baseline: LM finetuned with KG	RoBERTa + GreaseLM	BioLinkBERT + GreaseLM

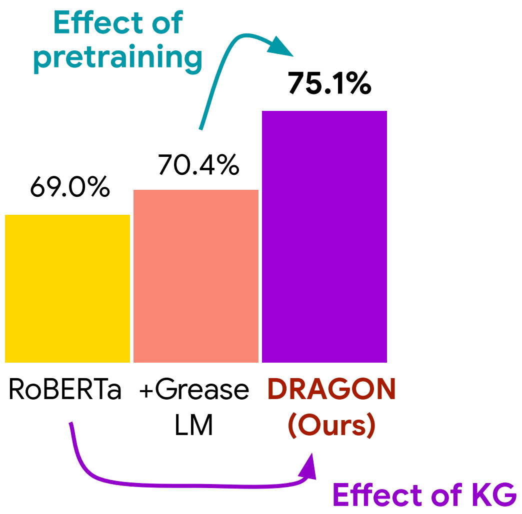
Ours (DRAGON): LM *pretrained* with KG

Performance

DRAGON makes consistent improvement across tasks and domains

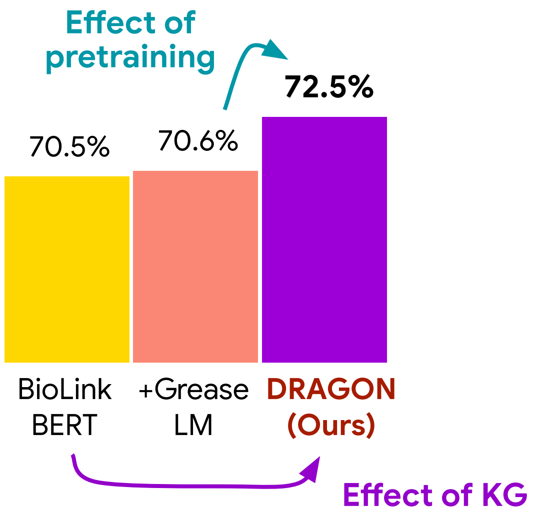
Commonsense reasoning tasks

(e.g. OBQA, RiddleSense)



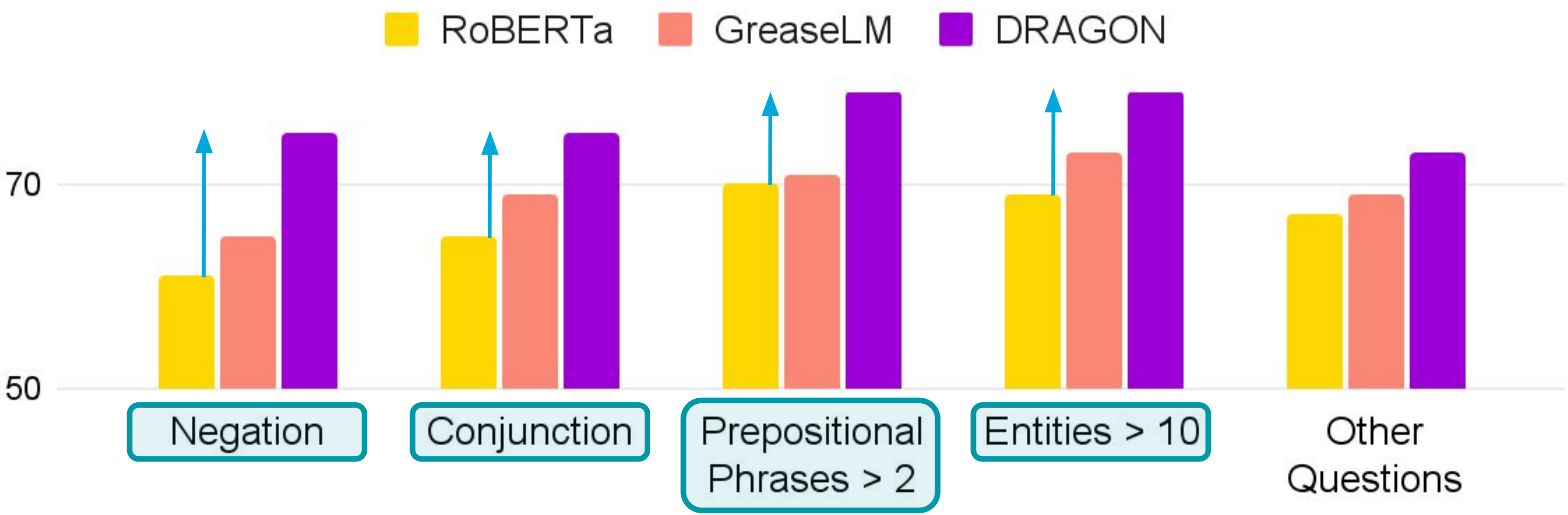
Biomedical reasoning tasks

(e.g. PubMedQA, MedQA)



Benefit 1: Complex Reasoning

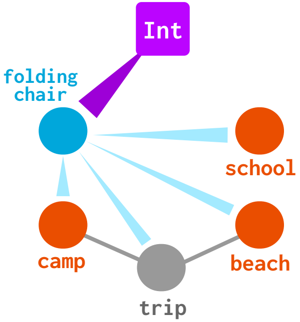
Large gains on QA examples involving complex reasoning



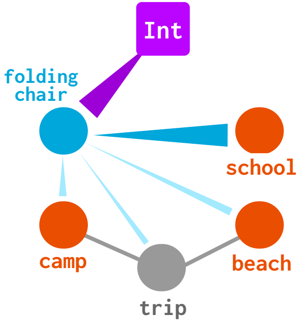
Benefit 1: Complex Reasoning

Conjunction

Where would you use a **folding chair and** store one?
A. camp **B. school** C. beach



DRAGON
GNN **1st** Layer



DRAGON
GNN **Final** Layer

RoBERTa:
A. camp (✗)

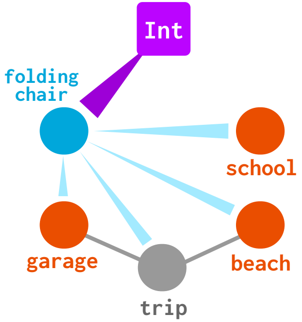
GreaseLM:
C. camp (✗)

DRAGON:
B. school (✓)

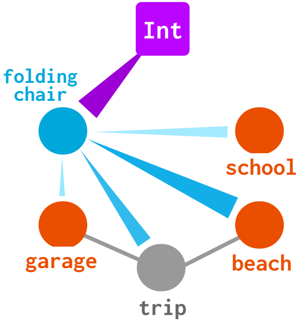
Model
Prediction

Negation + Conjunction

Where would you use a **folding chair but not** store one?
A. garage B. school **C. beach**



DRAGON
GNN **1st** Layer



DRAGON
GNN **Final** Layer

RoBERTa:
B. school (✗)

GreaseLM:
B. school (✗)

DRAGON:
C. beach (✓)

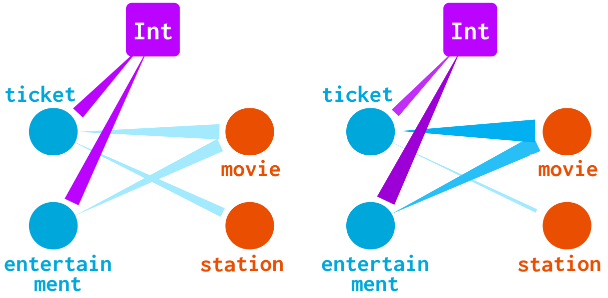
Model
Prediction

In DRAGON, KG serves as **scaffold** for performing **structured reasoning**

Benefit 1: Complex Reasoning

Single context

You will buy a **ticket** for entering what building for **entertainment**?
A. station **B. movie theater**



DRAGON GNN **1st** Layer | DRAGON GNN **Final** Layer

Model Prediction

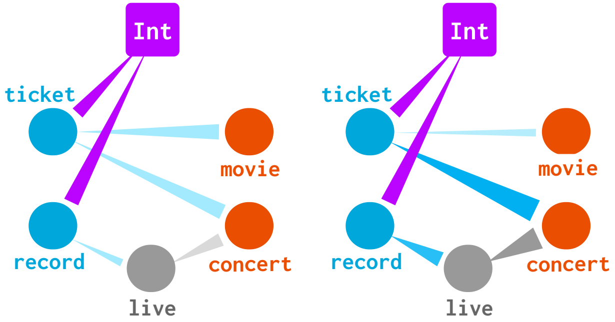
RoBERTa: B. movie theater (✓)

GreaseLM: B. movie theater (✓)

DRAGON: B. movie theater (✓)

Multi context (extra reasoning step)

You don't enjoy watching pre-recorded performance. You will buy a **ticket** for entering what building for **entertainment**?
A. station B. movie theater **C. concert hall**



DRAGON GNN **1st** Layer | DRAGON GNN **Final** Layer

Model Prediction

RoBERTa: B. movie theater (✗)

GreaseLM: B. movie theater (✗)

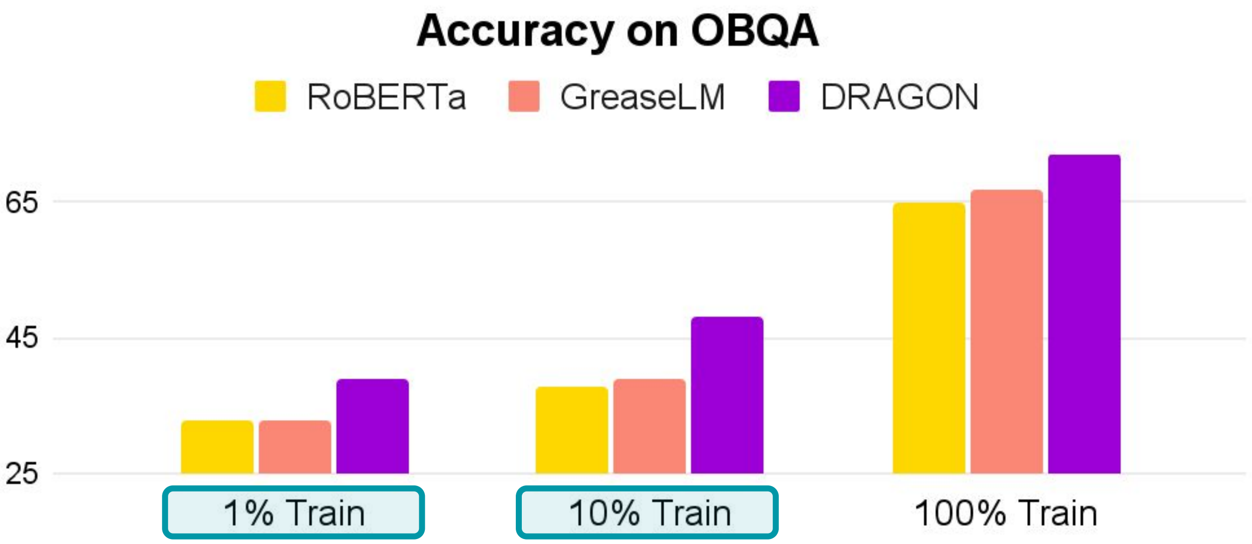
DRAGON: C. concert hall (✓)

Pretraining with KG helps **extrapolate** to harder test examples that need **multi-step reasoning**.

Benefit 2: Low-Resource QA

Large gains on few-shot and low-resource QA

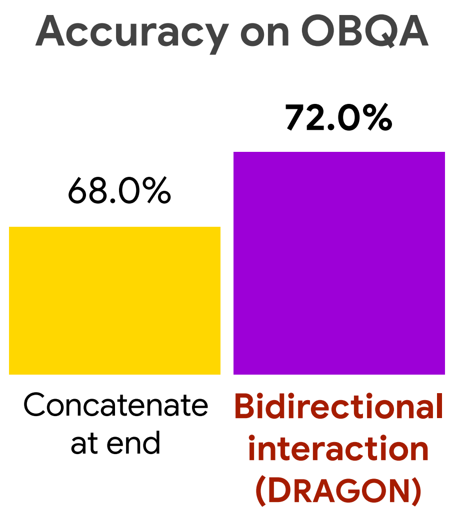
⇒ Intuition: self-supervision helps learn more knowledgec



Key Design Choices: Modeling

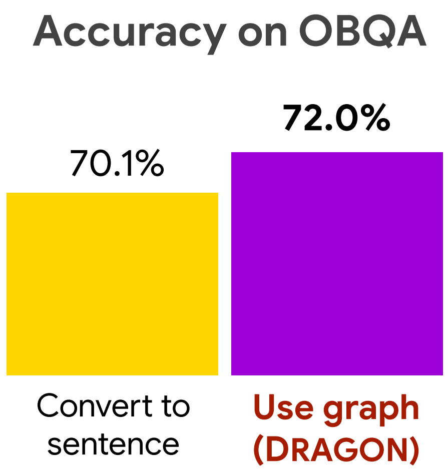
Cross-modal fusion for text+KG

- **Bidirectional interaction (DRAGON)**
- Concatenate representations at end



KG structure

- **Use graph and GNN (DRAGON)**
- Convert to sentence and add to text

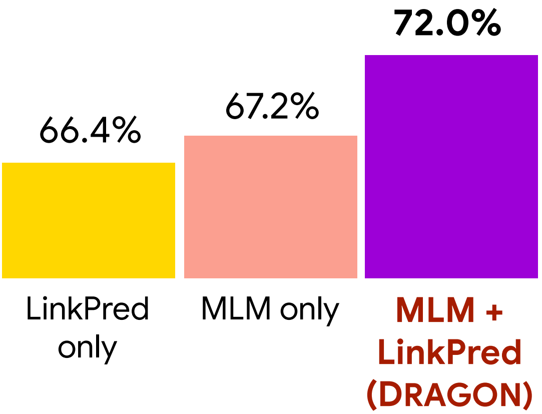


Key Design Choices: Self-Supervision

Pretraining objective

- **Joint MLM + LinkPred (DRAGON)**
- MLM only
- LinkPred only

Accuracy on OBQA

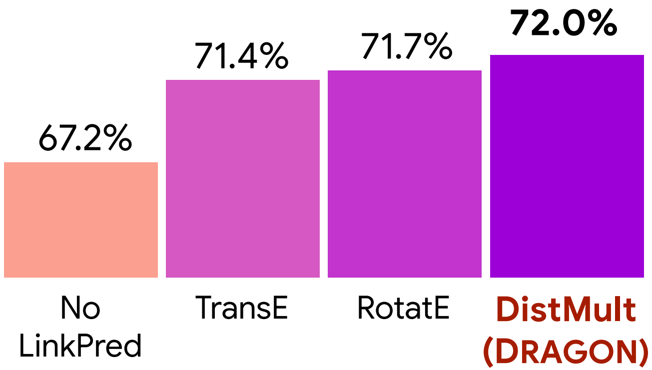


LinkPred head

- DistMult (Final DRAGON)
- TransE
- RotatE

⇒ **All help**

Accuracy on OBQA



Summary

DRAGON: Pretrain a foundation model jointly on text & KG

Approach

- Deeply bidirectional model for the two modalities to interact
- Self-supervised objective to learn joint reasoning over text and KG at scale

Result

- Improved performance on knowledge- and reasoning-intensive applications (e.g. low-resource QA, multi-step reasoning)

Thanks!

Michihiro Yasunaga

Antoine Bosselut

Hongyu Ren

Xikun Zhang

Chris Manning

Percy Liang

Jure Leskovec

We thank the members of the Stanford SNAP / P-Lambda / NLP groups, and the project MOWGLI team, as well as our anonymous reviewers. Funded in part by DARPA MCS.

Paper: Deep Bidirectional Language-Knowledge Graph Pretraining. NeurIPS 2022.

Code: https://github.com/michiyasunaga/DRAGON