2023년 11월 17일 LAB Meeting

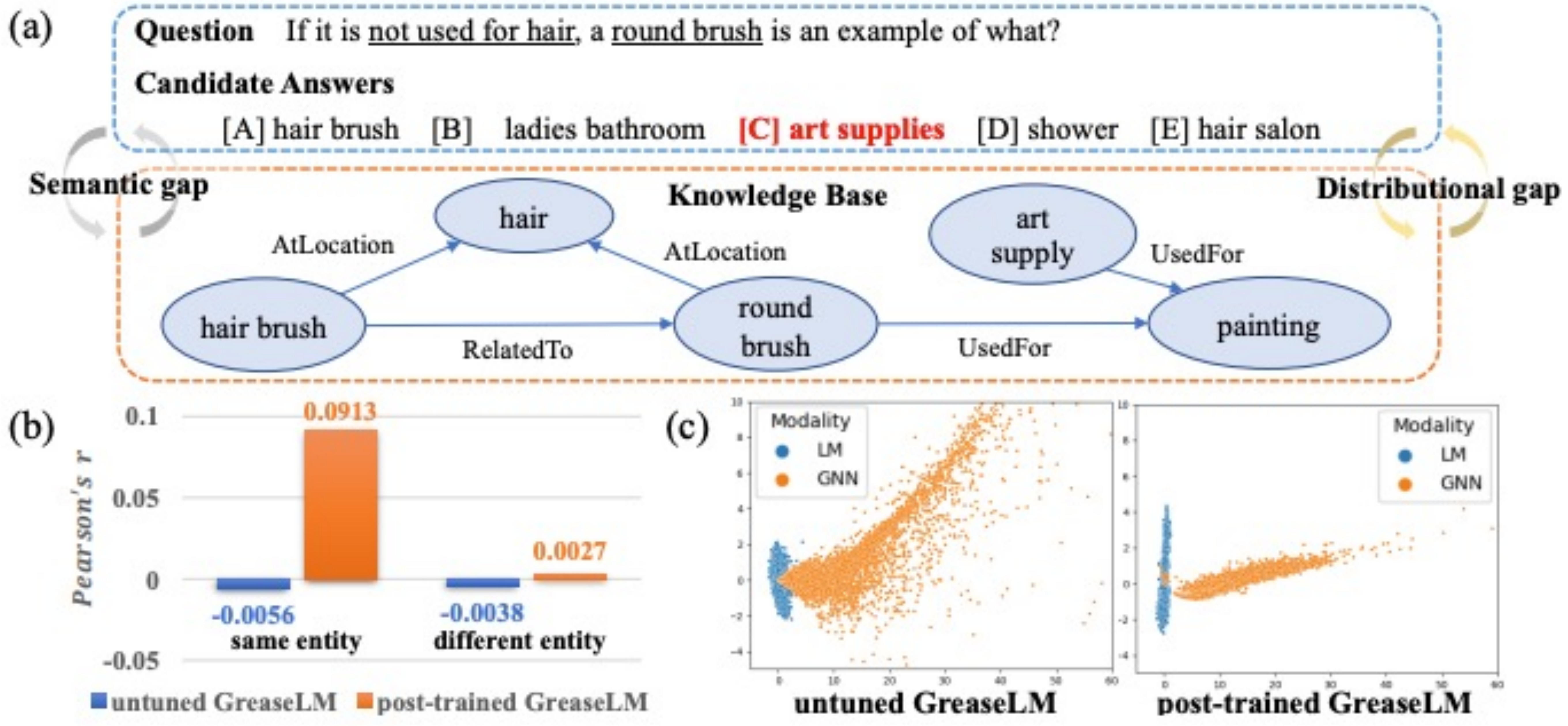
보충 자료

Fits : Fine-grained Two-stage Training for Knowledge-aware Question Answering

정지원
성균관대학교 인공지능학과
석사과정
jwjw9603@g.skku.edu

성균관대학교
SUNG KYUN KWAN UNIVERSITY(SKKU)

Figure 1



Post training 하자

Figure 2

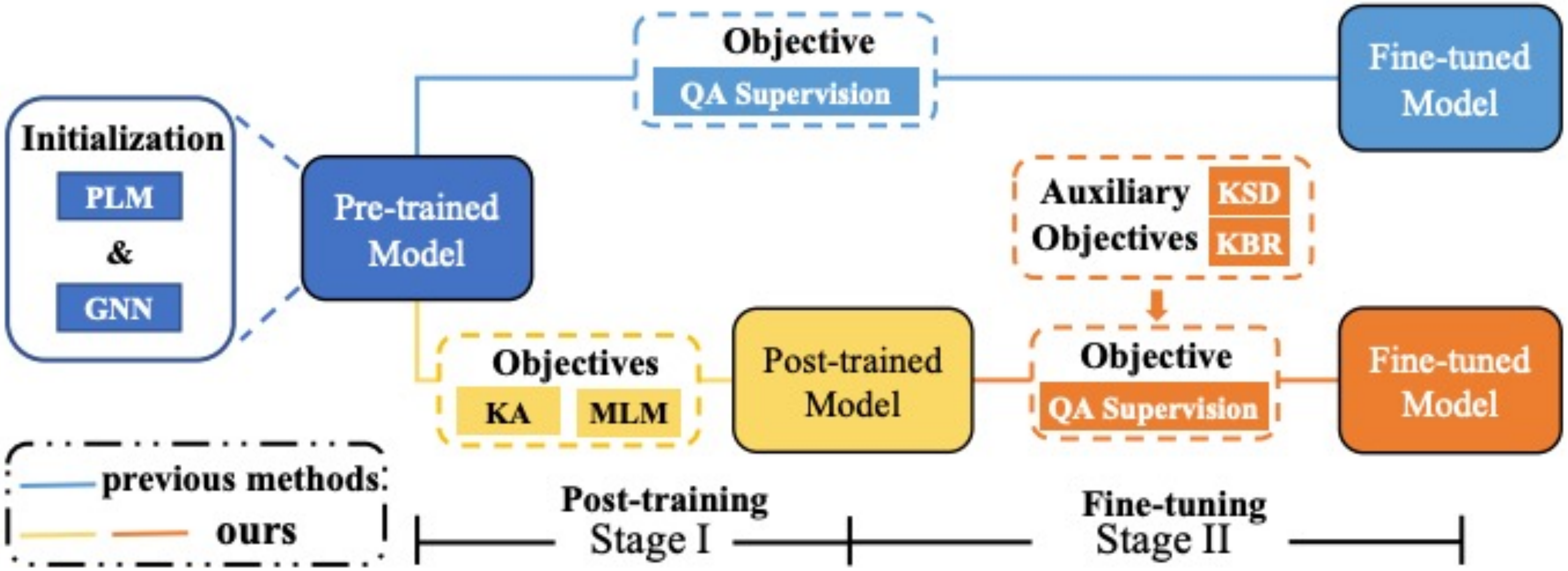
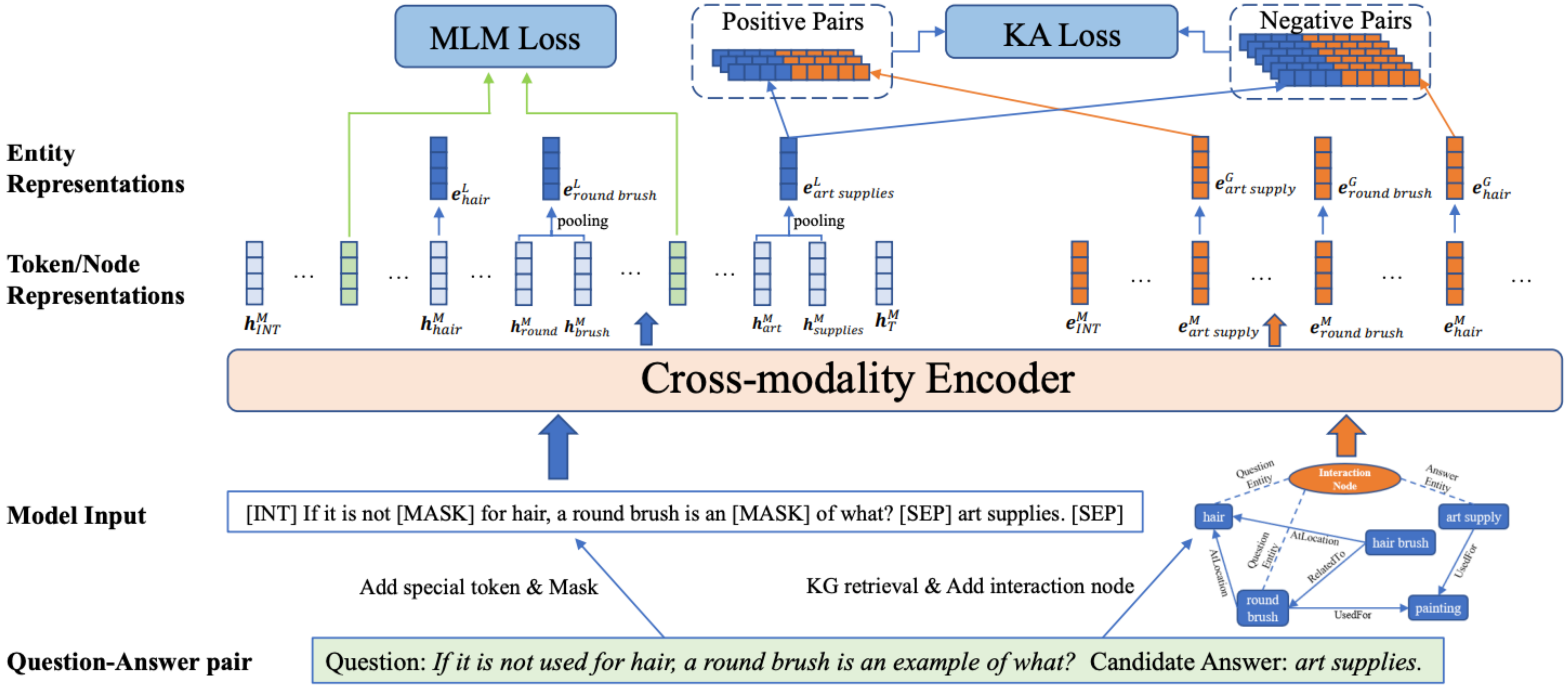

Initialization
PLM
&
GNN
Pre-trained Model
Objective
QA Supervision
Fine-tuned Model
Auxiliary Objectives
KSD
KBR
Objectives
KA
MLM
Post-trained Model
Objective
QA Supervision
Fine-tuned Model
previous methods:
ours
Post-training
Stage I
Fine-tuning
Stage II

Figure 3



$$\{\tilde{\mathbf{h}}_{int}^{l}, \mathbf{h}_1^{l}, .., \mathbf{h}_T^{l}\} = \text{LM-Enc}\left(\{\mathbf{h}_{int}^{l-1}, \mathbf{h}_1^{l-1}, .., \mathbf{h}_T^{l-1}\}\right) \quad (1)$$

$$\{\tilde{\mathbf{e}}_{int}^{l}, \mathbf{e}_1^{l}, ..., \mathbf{e}_J^{l}\} = \text{GNN}\left(\{\mathbf{e}_{int}^{l-1}, \mathbf{e}_1^{l-1}, ..., \mathbf{e}_J^{l-1}\}\right) \quad (2)$$

Subgraph 생성 방법

- MHGRN 방법론을 따라 subgraph 생성을 하고, QA-GNN의 방법을 따라 노드 개수(하이퍼 파라미터) 에 맞춰서 subgraph 완성
- MHGRN 방법론
  1) Question 문장과 Answer 문장을 tokenize(NLTK)를 시키고 tokenize화된 단어가 KG(conceptnet)에 있는지 살펴본다.
  2) Question entity 와 Answer entity사이의 2-hop 연결된 모든 엔티티들을 가져온다.
- QA-GNN 방법론
  1) MHGRN방법론으로부터 만들어진 graph의 각 entity를 Question과 Answer를 연결한 QA-context와 concat([QA-context;text(entity)])해서 LM에 넣어서 relevance score를 계산한다.
  2) Top-노드개수의 relevance score를 기반으로 pruning 진행한다.
  3) QA-context 를 엔티티로 변환하여 pruning된 subgraph에 추가한다.

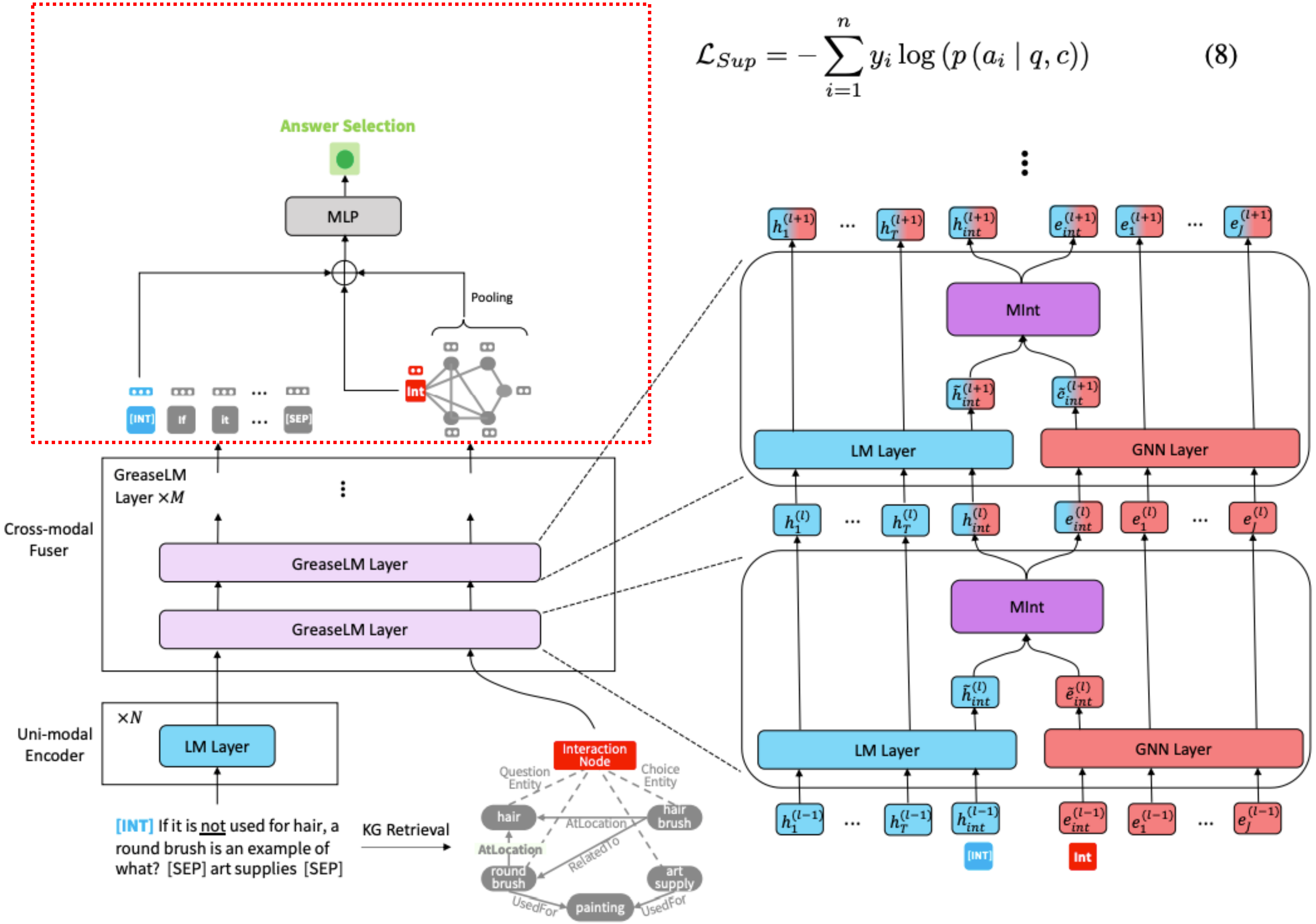
Q. If it’s not used for hair a round brush is an example of what?

A. hair brush B. ladies bathroom C. **art supplies** D. shower E. hair salon

"qc": ["example", "round", "round_brush", "use", "used"], "ac": ["art", "art_supplies", "art_supply", "supplies", "supply"]

Figure GREASELM

QA Supervision



$$p(a_i \mid q, c) \propto \exp\left(\mathrm{MLP}\left(\mathbf{h}_{int}^{M}, \mathbf{e}_{int}^{M}, \mathbf{g}\right)\right) \qquad (7)$$

$$\mathcal{L}_{Sup} = -\sum_{i=1}^{n} y_i \log\left(p\left(a_i \mid q, c\right)\right) \qquad (8)$$

$$\{\tilde{\mathbf{h}}_{int}^{l}, \mathbf{h}_1^{l}, .., \mathbf{h}_T^{l}\} = \text{LM-Enc}\left(\{\mathbf{h}_{int}^{l-1}, \mathbf{h}_1^{l-1}, .., \mathbf{h}_T^{l-1}\}\right) \qquad (1)$$

$$\{\tilde{\mathbf{e}}_{int}^{l}, \mathbf{e}_1^{l}, ..., \mathbf{e}_J^{l}\} = \mathrm{GNN}\left(\{\mathbf{e}_{int}^{l-1}, \mathbf{e}_1^{l-1}, ..., \mathbf{e}_J^{l-1}\}\right) \qquad (2)$$

$$[\mathbf{h}_{int}^{l}; \mathbf{e}_{int}^{l}] = \mathrm{MLP}\left(\left[\tilde{\mathbf{h}}_{int}^{l}; \tilde{\mathbf{e}}_{int}^{l}\right]\right) \qquad (3)$$

Figure 4

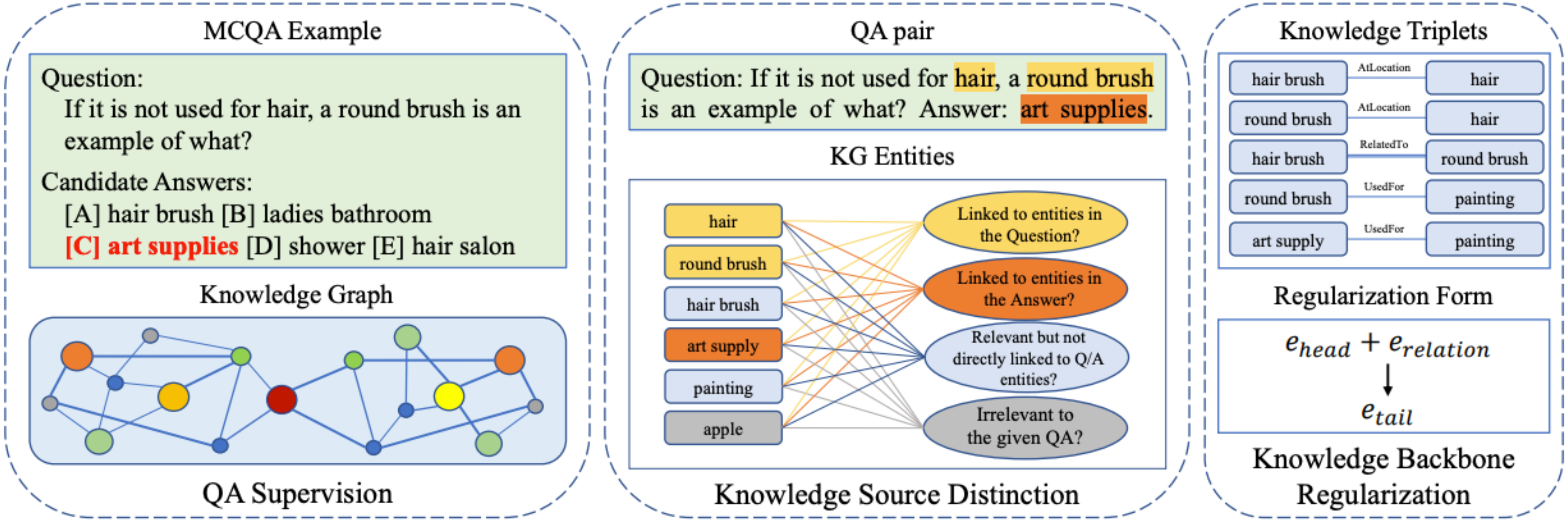


Figure 4: An overview of the knowledge-aware fine-tuning objectives.

Figure 5

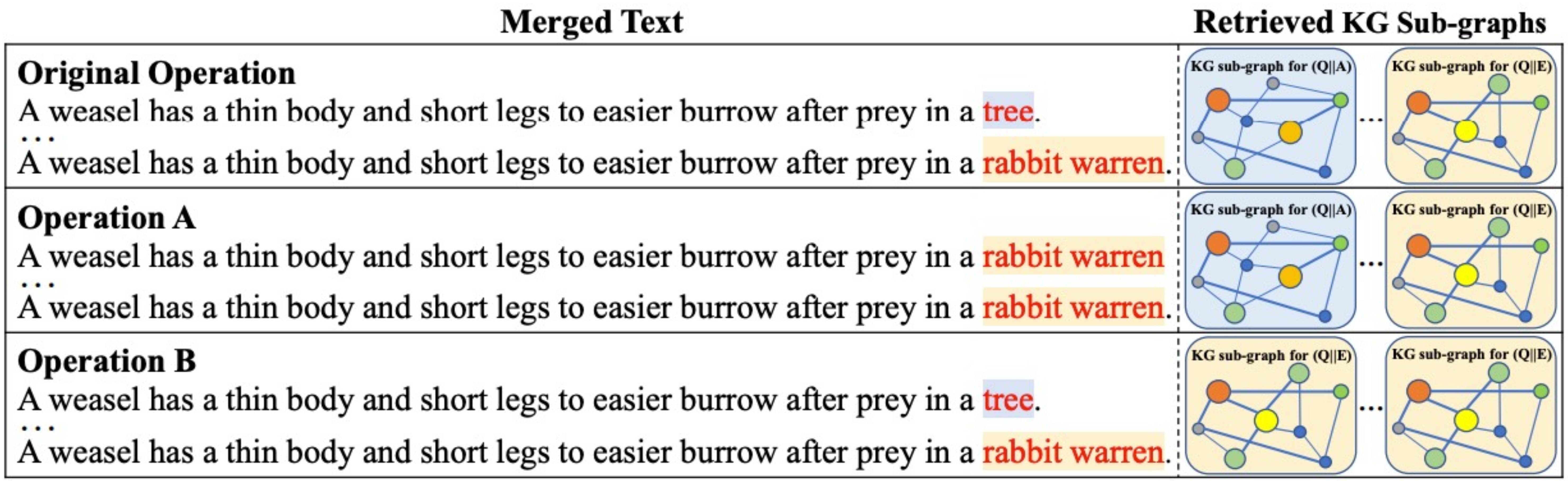


Figure 5: Operations A and B are used for quantitative analysis (section). The unprocessed MCQA example is shown in Figure 1. $Q||A$ denotes the text obtained by merging the question Q and the candidate answer A.

Model	test-reason	test-param
GreaseLM (Zhang et al. 2022)	73.4	69.0
+ post-training	73.9	70.9
+ knowledge-aware fine-tuning	74.2	71.2
+ **FiTs (Ours)**	**74.5**	**71.6**

Table 3: The test-reason set evaluates models' joint reasoning ability, while the test-param set measures models' parametric knowledge.

Figure 5

Model	IHtest-Acc (%)
RoBERTa-Large (w/o KG)♣	68.7
RGCN (Schlichtkrull et al. 2018)♣	68.4
KagNet (Lin et al. 2019)♣	69.0
MHGRN (Feng et al. 2020)♣	71.1
QA-GNN (Yasunaga et al. 2021)♣	73.4
GreaseLM (Zhang et al. 2022)♣	74.2
GreaseLM (Our implementation)	73.6
+ **FiTs** (**Ours**)	**76.2**

Table 1: Performance comparison on CommonsenseQA in-house split. We report the in-house Test (IHtest) accuracy using the data split of Lin et al. (2019), because the official test is hidden. ♣: results from Zhang et al. (2022); all other results are reproduced by ourselves.

Model	Test-Acc (%)
AristoRoBERTa (no KG)♣	78.4
RGCN (Schlichtkrull et al. 2018)♣	74.6
MHGRN (Feng et al. 2020)♣	80.6
QA-GNN (Yasunaga et al. 2021)♣	82.8
GreaseLM (Zhang et al. 2022)♣	84.8
GreaseLM (Our implementation)	84.2
+ **FiTs** (**Ours**)	**86.0**

Table 2: Test accuracy comparison on OpenBookQA. ♣: results from Zhang et al. (2022); all other results are reproduced by ourselves.

Model	test-reason	test-param
GreaseLM (Zhang et al. 2022)	73.4	69.0
+ post-training	73.9	70.9
+ knowledge-aware fine-tuning	74.2	71.2
+ **FiTs** (**Ours**)	**74.5**	**71.6**

Table 3: The test-reason set evaluates models' joint reasoning ability, while the test-param set measures models' parametric knowledge.

감사합니다

발표 경청해 주셔서 감사합니다

정지원 성균관대학교 인공지능학과 석사 과정
jwjw9603@g.skku.edu