

## Lecture 15: Integrating Knowledge in Language Models

국주현

# What does a language model know?

## EX) Bert large에서 상식 테스트

- iPod Touch is produced by Apple . (o)
- London Jazz Festival is located in London . (o)
- Dani Alves plays with Santos . (x) => Dani Alves는 바르셀로나에서 뛰었음
- Carl III used to communicate in German . (x) => 'Carl' 은 주로 스위스어에서 쓰이는 단어
- Ravens can fly . (o)

⇒ 언어모델은 그럴싸한(합리적인) 단어를 예측하지만, 사실이 아닌 단어들을 예측할 수 있음.

## Language model의 문제점

- 1) Unseen fact: 학습된 데이터셋 내에서만 문제 해결이 가능  
-> unseen fact(데이터셋 미포함)에 대해서는 사실을 예측하기 힘들
- 2) Rare fact: 사실을 학습하여도 특정 사실에 대한 데이터셋이 많지 않아 충분한 학습이 어려움
- 3) Model sensitivity: Model sensitivity 가 높으면 같은 의미의 문장이어도 다르게 받아들임  
"x was made in y" 와 "x was created in y" 는 의미상 같은 문장

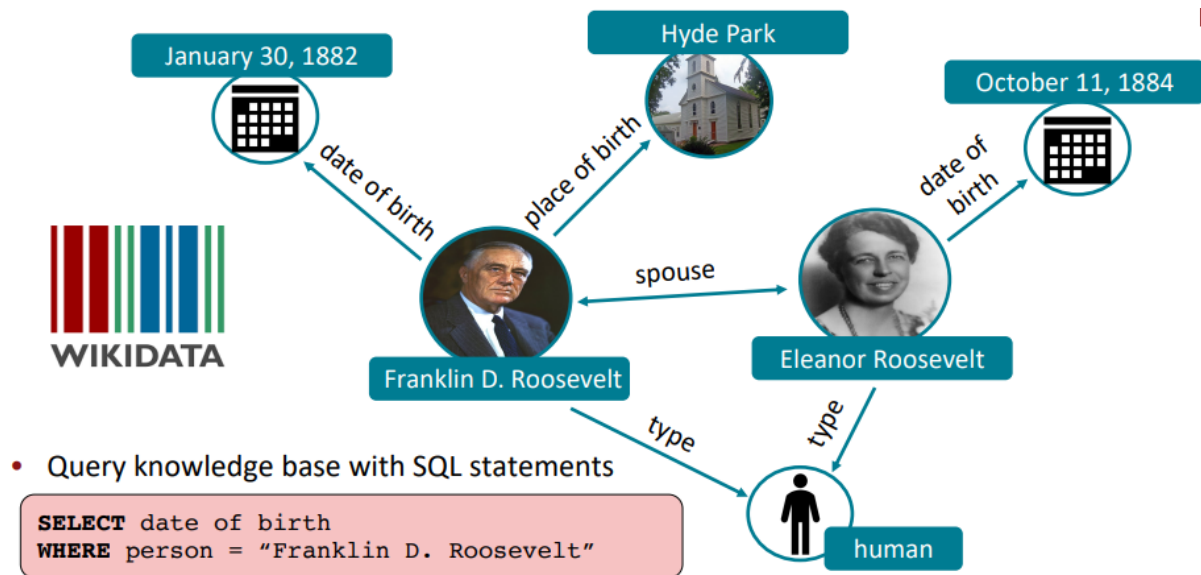
⇒ 따라서 언어모델은 학습한 지식을 온전히 기억할 수도 없고, external knowledge 가 있어야 더 정확한 사실을 예측할 수 있다!

# What does a language model know?

## Knowledge 기반 nlp task

### 전통적인 kg base 방법

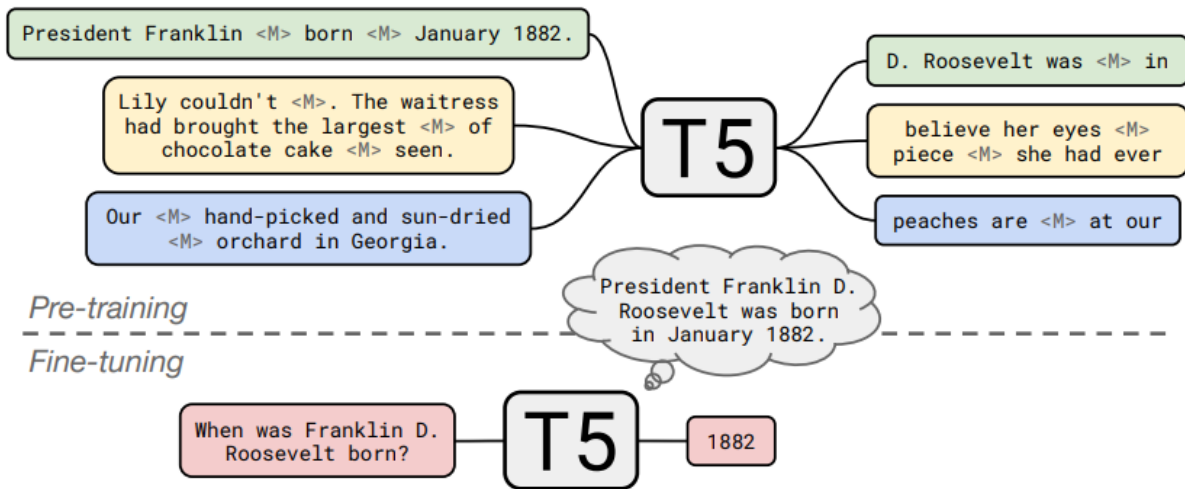
- Kg를 external knowledge로 활용하여 nlp 태스크 해결



Sql 을 통해 데이터를 추출하거나 복잡한 nlp 파이프라인을 가진다는 단점 존재

### 언어모델 그 자체를 knowledge base로 활용!

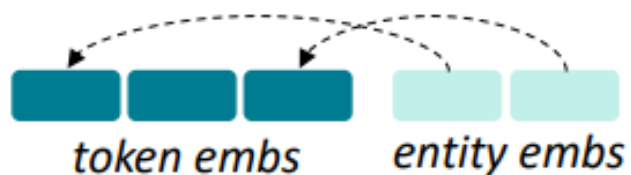
- 많은 data를 학습한 대형 언어모델 그 자체에서 knowledge 추출 및 활용



사전 학습된 언어모델에서 정보를 추출하므로, 정보 수정이 어렵다는 단점 존재

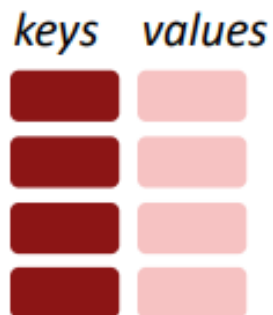
# Techniques to add knowledge to LMs

어떻게 knowledge를 lm에 적용시킬까?



## Add pretrained entity embeddings

- ERNIE
- KnowBERT



## Use an external memory

- KGLM
- kNN-LM



## Modify the training data

- WKLM
- ERNIE (another!), salient span masking

# Techniques to add knowledge to LMs

## Entity embeddings

- "U.S.A.", "United States of America" , "America" 는 서로 다른 임베딩을 가짐
- 위의 세 단어들을 모두 같은 임베딩 값을 적용시켜주면 어떨까?
  - ⇒ 이것이 entity embedding
- 엔티티 linking만 잘 된다면 entity embedding은 lm에 매우 유용할 수 있음
- What is entity linking?
  - 엔티티를 지식베이스에 연결하는 것 (문맥을 통해 워싱턴이 "사람 워싱턴"인 것을 파악)



# Techniques to add knowledge to LMs

## Entity embeddings

entity embedding은 워드 임베딩과 같지만, 지식그래프에서의 entity 사용을 위한 것!

$$\text{George Washington} = \begin{pmatrix} 0.111 \\ -0.345 \\ 0.876 \\ -0.201 \end{pmatrix}$$

다빈치  $\xrightarrow{\text{그리다}}$  모나리자

- Knowledge graph embedding methods (e.g., TransE)
  - 주로 트리플 관계; head entity 와 relation의 합이 tail entity가 되도록
- Word-entity co-occurrence methods (e.g., Wikipedia2Vec)
  - 엔티티가 주어지면, 주로 어떤 단어들이 해당 엔티티에 많이 등장하는가? 를 바탕으로 학습
- Transformer encodings of entity descriptions (e.g., BLINK)
  - 엔티티의 설명을 인코딩하여 학습하는 방법

## Add pretrained entity embeddings

어떻게 knowledge를 lm에 적용시킬까?

- 언어모델의 임베딩 공간에서 사전 훈련된 엔티티 임베딩을 어떻게 통합할까?  
⇒ Fusion layer를 통해 언어 context와 엔티티 정보를 통합

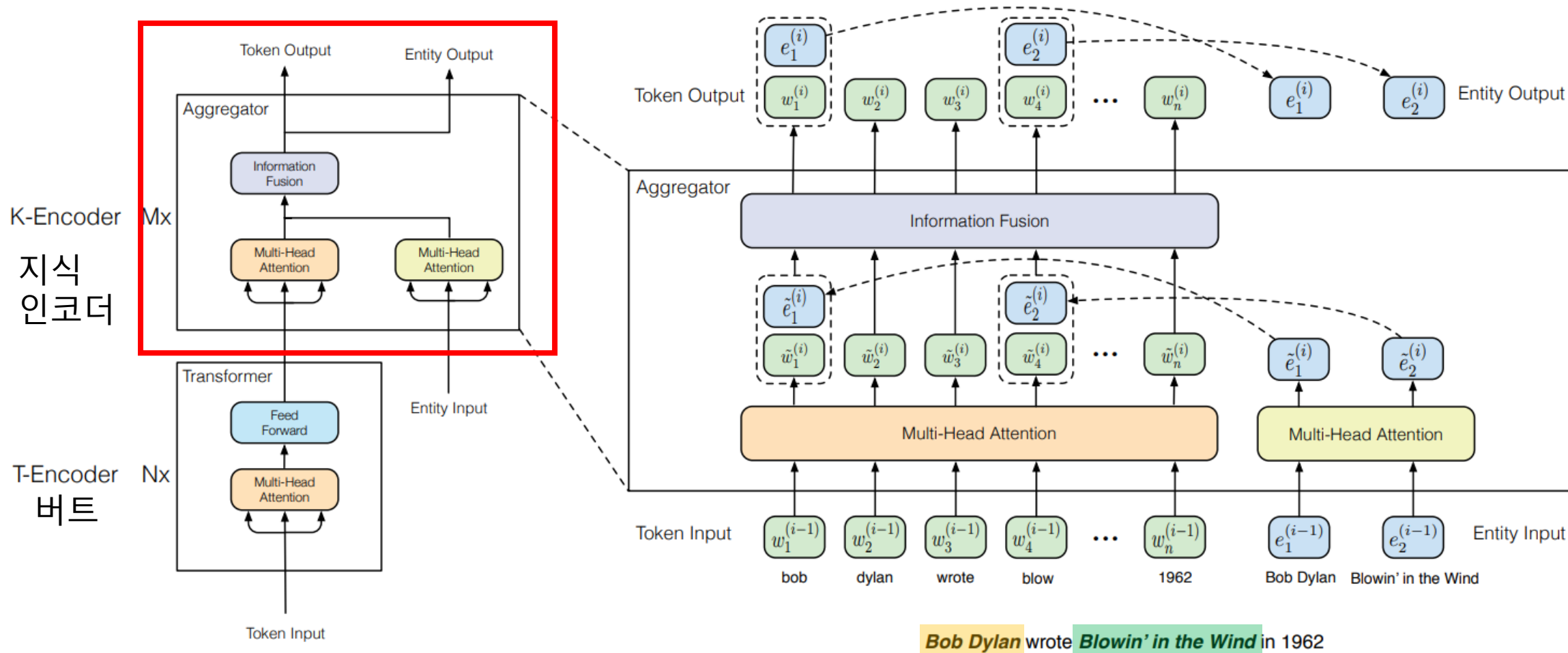
$$\mathbf{h}_j = F(\mathbf{W}_t \mathbf{w}_j + \mathbf{W}_e \mathbf{e}_k + \mathbf{b})$$



Word j의 임베딩    entity j의 임베딩

# Add pretrained entity embeddings

## 1. ERNIE: Enhanced Language Representation with Informative Entities(칭화대)



버트는 Bob Dylan의 정보를 모름 -> 기존의 학습된 entity embedding(transE)을 가져와서 정보 통합!  
(이때, 두개 이상의 토큰은 첫번째 토큰에 정보 Aggregation!)



# Add pretrained entity embeddings

## 1. ERNIE: Enhanced Language Representation with Informative Entities

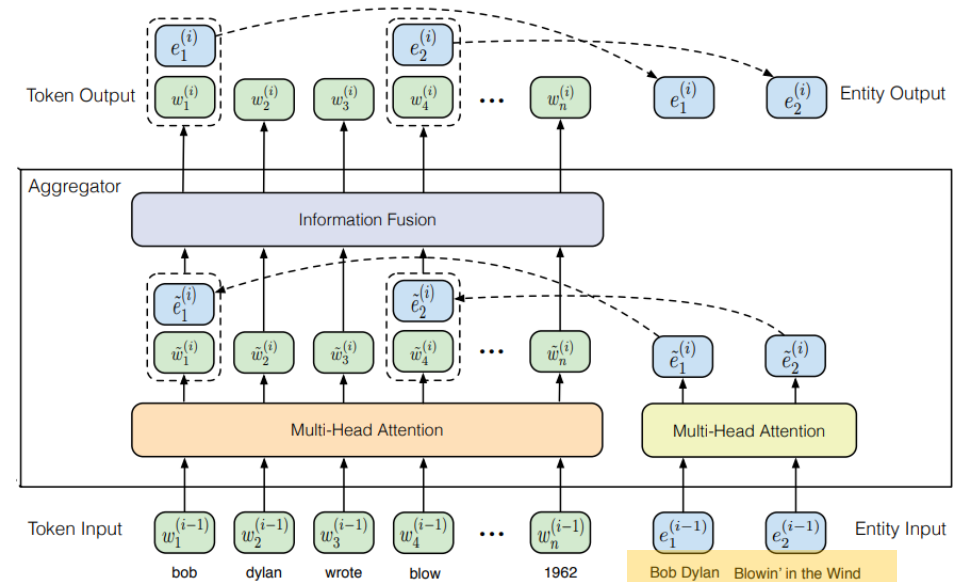
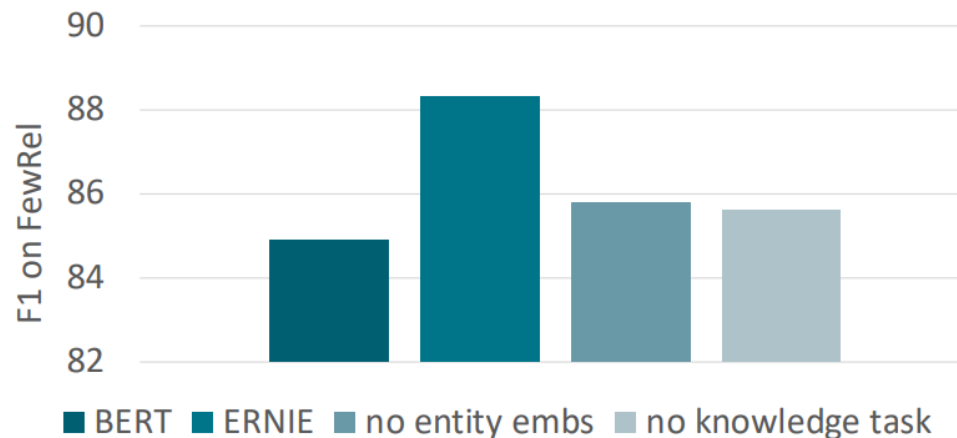
- Pretrain task
  - Masked language model and next sentence prediction
  - Knowledge pretraining task (dEA1; denoising entity auto-encoder):

$$p(e_j | w_i) = \frac{\exp(\mathbf{W}\mathbf{w}_i \cdot \mathbf{e}_j)}{\sum_{k=1}^m \exp(\mathbf{W}\mathbf{w}_i \cdot \mathbf{e}_k)}$$

Token-entity alignment를 masking

- ✓ 5% : 기존 entity를 다른 random entity로 대체
- ✓ 15% : token-entity alignment를 마스킹.
- ✓ 나머지 : unchanged.

$$\mathcal{L}_{ERNIE} = \mathcal{L}_{MLM} + \mathcal{L}_{NSP} + \mathcal{L}_{dEA}$$



Bob Dylan wrote *Blowin' in the Wind* in 1962

## Add pretrained entity embeddings

### 2. Jointly learn to link entities with KnowBERT

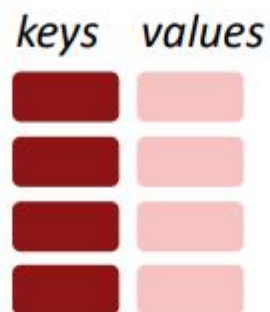
- 버트의 확장으로 entity linker를 사전훈련(bert에 kb와 entity 연결 학습이 추가됨)

$$\mathcal{L}_{KnowBERT} = \mathcal{L}_{NSP} + \mathcal{L}_{MLM} + \mathcal{L}_{EL}$$

Predict over set of hard candidates (not just those in sentence)

- EL을 학습하면서 knowledge 인코딩을 더 잘 학습하게 됨  
(다운스트림 태스크에서 earnie보다 더 우수한 성능)
- Knowbert도 earnie처럼 fusion layer 를 학습

## Use an external memory



### Add pretrained entity embeddings

- ERNIE
- KnowBERT

### Use an external memory

- KGLM
- kNN-LM

### Modify the training data

- WKLM
- ERNIE (another!), salient span masking

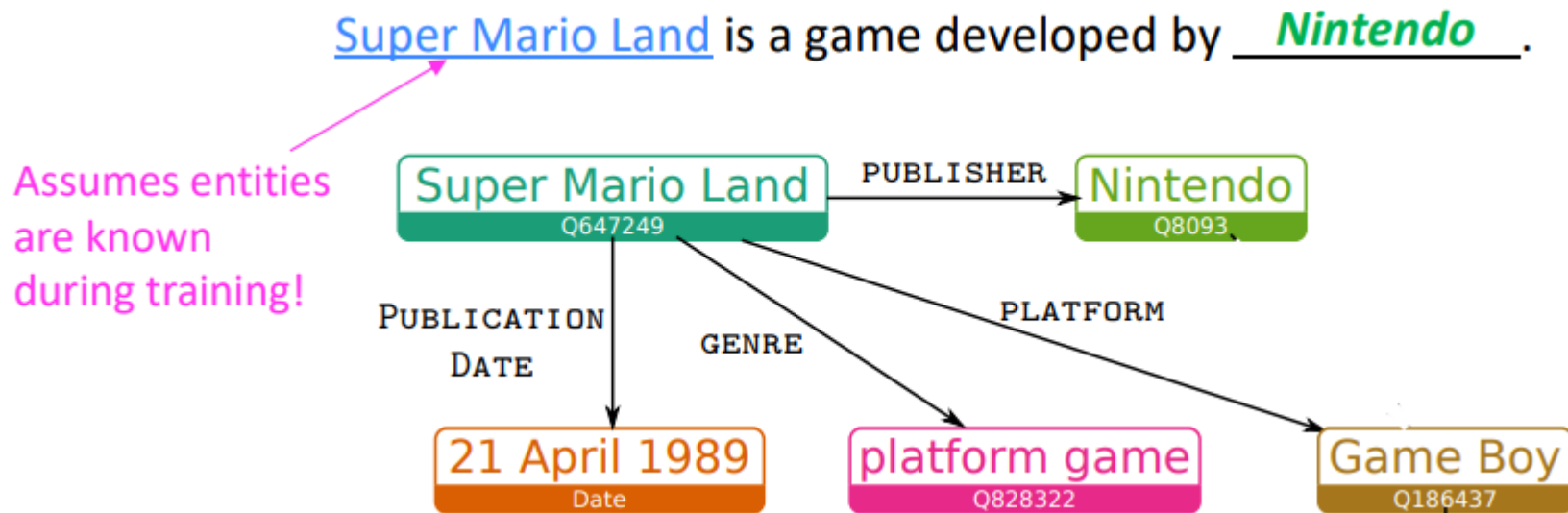
## Use an external memory

### 1. KGLM

- Lstm 구조를 사용(문장의 이전 시퀀스, 이전 entity가 주어지면 그 다음 시퀀스와 entity 예측)

$$P(x^{(t+1)}, \varepsilon^{(t+1)} | x^{(t)}, \dots, x^{(1)}, \varepsilon^{(t)}, \dots, \varepsilon^{(1)})$$

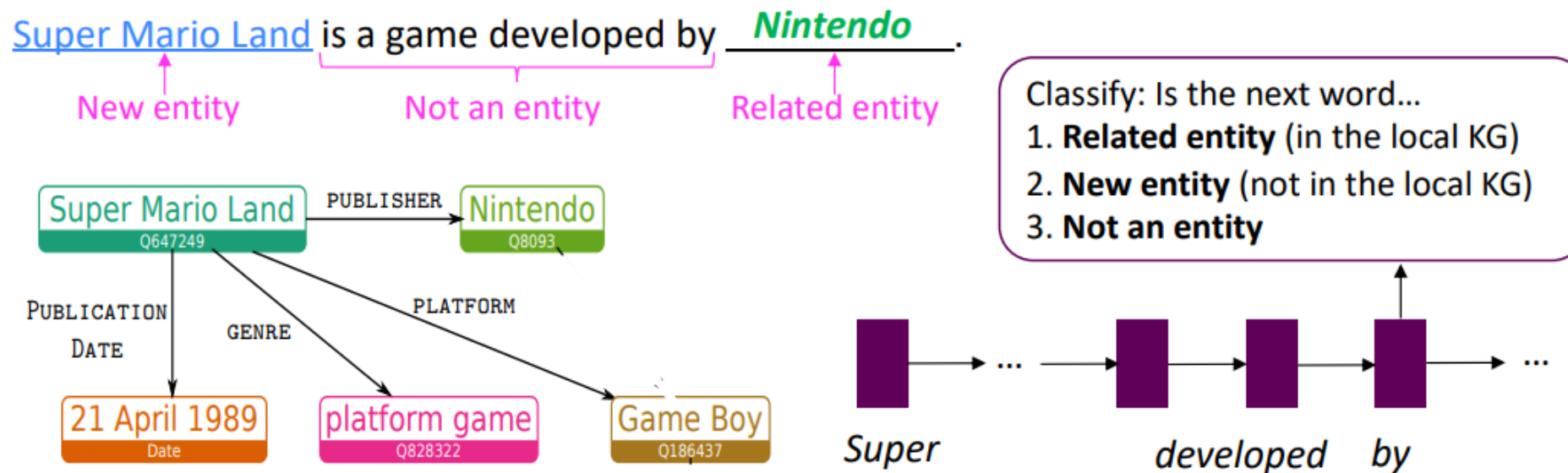
#### 1) KG에서 시퀀스와 관련된 Subgraph 탐색



# Use an external memory

## 1. KGLM

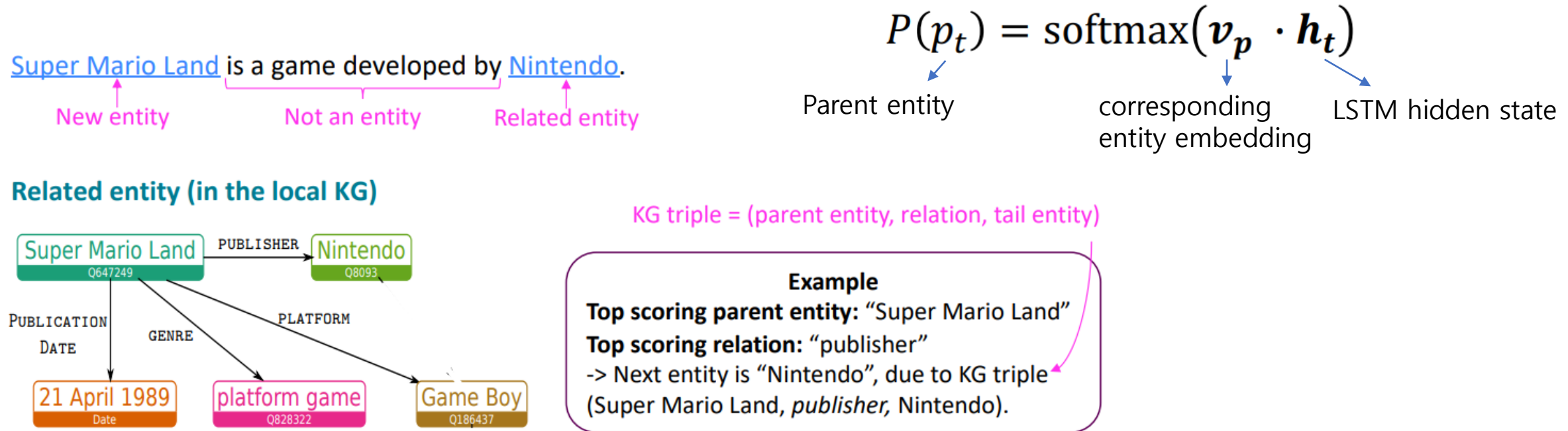
### 2) LSTM과 Subgraph를 활용하여 word의 유형 예측



# Use an external memory

## 1. KGLM

3) LSTM hidden state 와 pretrained entity embeddings를 활용하여 sub kg에서 top-scoring entity 를 찾기



- standard vocabulary와 더불어 top-scoring entity에 대한 entity aliases(별칭; ex: nintendo company등등) 를 포함한 확장된 어휘집합에서 예측을 수행

## Use an external memory

### 1. KGLM

#### 4) New entity (not in the subgraph KG)

Super Mario Land is a game developed by Nintendo.

The diagram illustrates the classification of entities in the sentence "Super Mario Land is a game developed by Nintendo." The words "Super Mario Land" and "Nintendo" are underlined in blue. Below "Super Mario Land" is a pink arrow pointing up to the text "New entity". Below "Nintendo" is a pink arrow pointing up to the text "Related entity". A pink bracket spans the words "is a game developed by" with the text "Not an entity" centered below it.

- LSTM hidden state 와 pretrained entity embeddings 를 활용하여 전체 kg에서 top-scoring entity 를 찾기
- top-scoring entity => new entity
- Next word는 entity aliases(별칭; ex: nintendo company등등)를 포함한 확장된 어휘집합에서 예측 수행

#### 5) Not an entity

- 그냥 simple lstm
- Next word는 일반적인 어휘집합에서 예측 수행

### • 성능

- Fact completion task에서 GPT-2 와 AWD-LSTM보다 우수함
- GPT-2는 일반적인 토큰 예측, KGLM은 더 구체적인 토큰 예측 가능

## Use an external memory

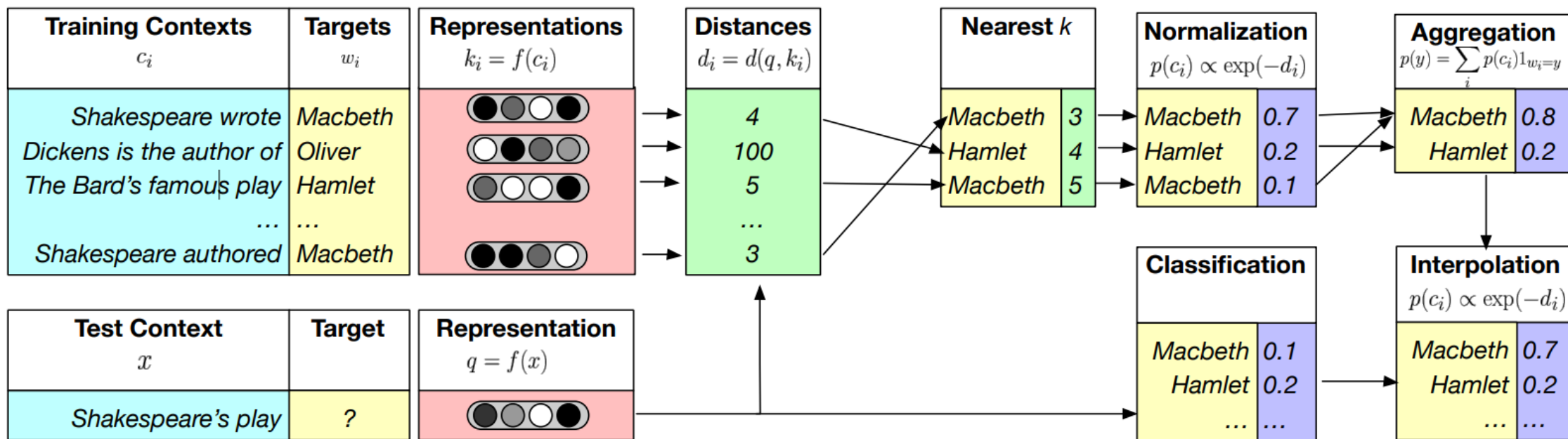
### 2. More recent takes: Nearest Neighbor Language Models (kNN-LM)

- KNN, 거리에 기반하여 유사한 text sequence 들을 탐색 후, 단어 예측진행

*“Dickens is the author of \_\_\_\_\_”*  $\approx$  *“Dickens wrote \_\_\_\_\_”*

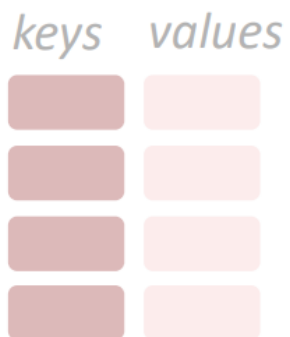
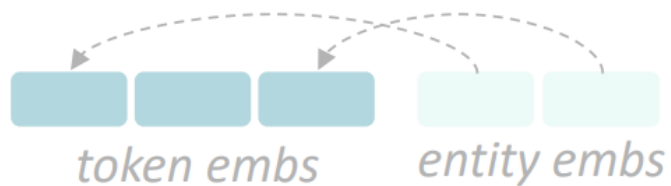
- Knn 확률값과 lm 확률값을 결합하여 사용

$$P(y|x) = \lambda P_{kNN}(y|x) + (1 - \lambda)P_{LM}(y|x)$$





## Modify the training data



## Add pretrained entity embeddings

- ERNIE
- KnowBERT

## Use an external memory

- KGLM
- kNN-LM

## Modify the training data

- WKLM
- ERNIE (another!), salient span masking

## Modify the training data

### 1. Pretrained Encyclopedia: Weakly Supervised KnowledgePretrained Language Model (WKLM)

- false knowledge 생성

true knowledge인 기존 데이터의 일부분을 같은 type의 다른 entity로 변경해서 생성

*True knowledge statement:*

J.K. Rowling is the author of Harry Potter.



*Negative knowledge statement:*

J.R.R. Tolkien is the author of Harry Potter.

- 생성된 false knowledge 에 대해서 replace가 되었는지 안되었는지 학습 (참인지 거짓인지 구별하게끔 학습)

True/False knowledge  
판별을 위한 loss

$$\mathcal{L}_{entRep} = \mathbb{I}_{e \in \mathcal{E}^+} \log P(e | C) + (1 - \mathbb{I}_{e \in \mathcal{E}^+}) \log(1 - P(e | C))$$



true entity mention(0 or 1)



context

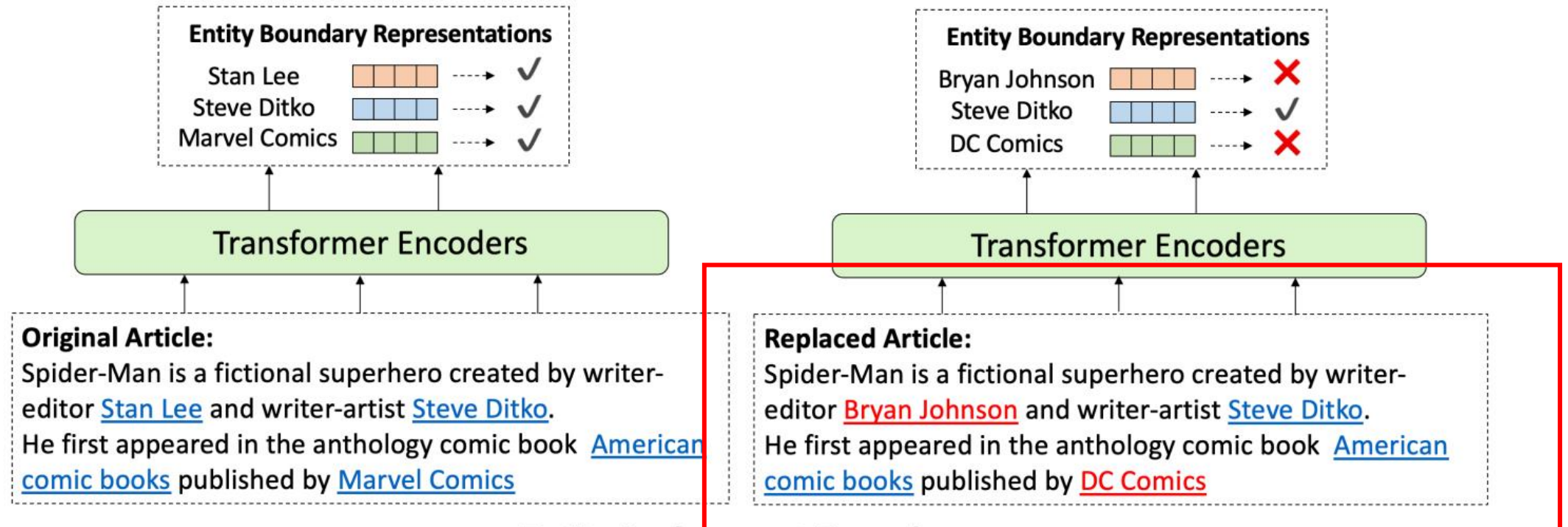
최종 loss

$$\mathcal{L}_{WKLM} = \mathcal{L}_{MLM} + \mathcal{L}_{entRep}$$

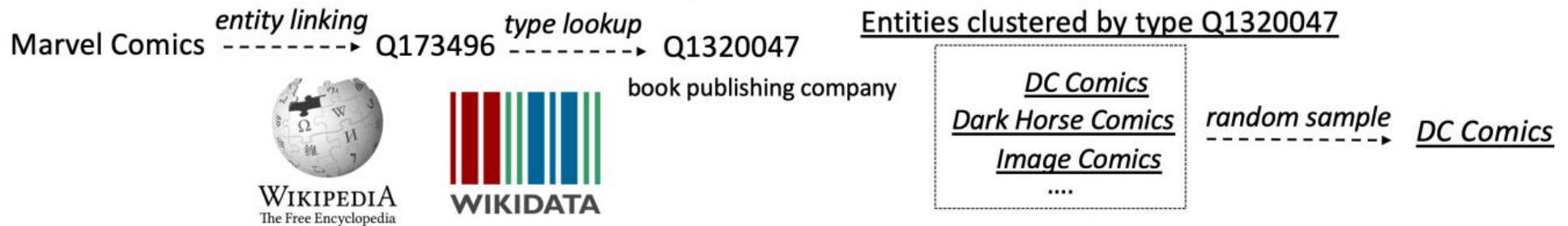
# Modify the training data

## 1. Pretrained Encyclopedia: Weakly Supervised KnowledgePretrained Language Model (WKLM)

- 각 엔티티 타입이랑 같은 타입의 엔티티를 검색하고 랜덤으로 교체해서 손상된 data 생성



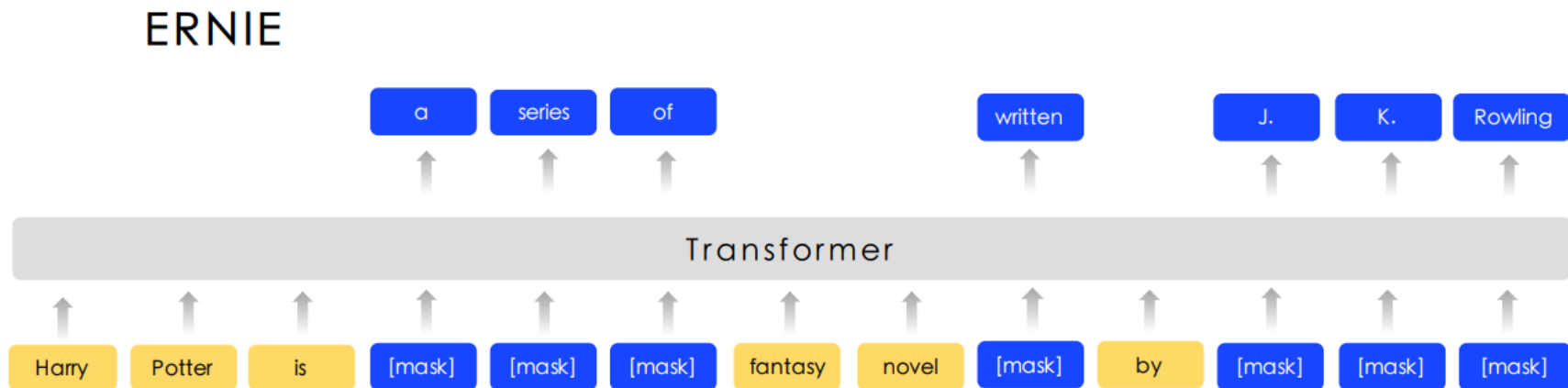
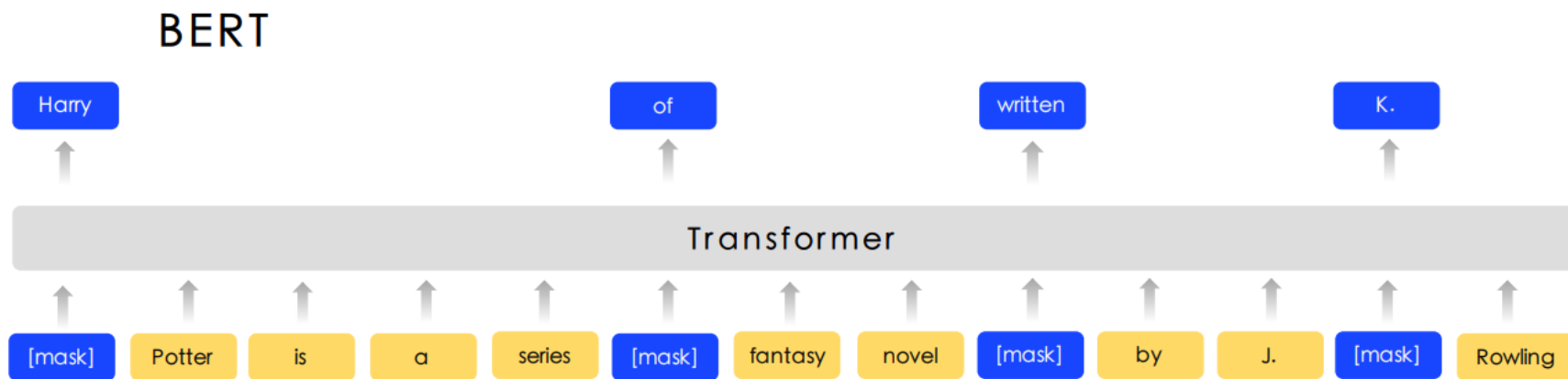
### Entity Replacement Procedure



## Modify the training data

### 1. ERNIE1: Enhanced Representation through Knowledge Integration(바이두)

- 별도의 Entity Embedding 사용하지 않고, masking기법을 통해 knowledge 정보 주입.
- 일련된 같은 구문 마스킹 진행( JK Rowling)



# Evaluating knowledge in LMs

## 1. LAMA

- 기성 LM이 얼마나 relational knowledge를 반영하고 있는지 확인하는 downstream task
- KG triples와 question-answer pair에서 cloze statements를 생성
- relation extraction (RE) 과 question answering systems 태스크로 모델이 가진 지식 평가
- 한계점
  - 사전학습된 데이터에 따라 모델이 가진 지식이 상이할 수 있음
  - 모델이 lama에서 성능이 높을수록 왜 성능이 더 높은지 이유가 확실치 않음
  - Lm은 phrasing 에 민감(lama에서는 relation 으로 정의된 템플릿이 하나 -> 어떤 모델은 잘 이해하지 못할 수 있음)

## 2. A More Challenging Probe: LAMA-UnHelpful Names (LAMA-UHN)

- relational knowledge(entity)가 있어야 답변 가능한 data만 남김
- BERT는 겉보기(sequence의 표면적 형태; 예시에서는 이름)에 의존하는 경향 발견
- Lama와 비교시 lama-uhn은 8% 성능 저하됨을 알 수 있음

Native language of  
French-speaking actors  
according to BERT

Person Name	BERT
Jean Marais	French
Daniel Ceccaldi	Italian
Orane Demazis	Albanian
Sylvia Lopez	Spanish
Annick Alane	English

## Evaluating knowledge in LMs

### 3. Developing better prompts to query knowledge in LMs

- LM 은 Query(prompts)에 매우 민감함
- LM이 사실을 알 수 있지만, prompts로 인해 lama와 같은 태스크에서 좋은 성능을 내지 못할 수 있음
- 예시) 사전학습에서 학습을 한 corpus여도 query가 corpus와 매우 다르기때문에, 모델은 답하기 힘들

Example: “The birth place of Barack Obama is Honolulu, Hawaii” (pretraining corpus) versus “Barack Obama was born in \_\_\_\_\_” (query)

- 종속성 구문분석과 의역을 활용해 다양한 prompts 생성
- 앙상블을 통해 prompts의 다양성을 높임 => 언어모델은 사전학습시 봤던 context 를 볼 확률이 높아짐  
=> 언어모델이 가진 지식을 더 잘 추출할 수 있음