



# Paradigm Shift in NLP

Tianxiang Sun, Xiangyang Liu, Xipeng Qiu, Xuanjing Huang  
Fudan University

`txsun19@fudan.edu.cn`

11 Oct 2021



<https://arxiv.org/abs/2109.12575>



<https://txsun1997.github.io/nlp-paradigm-shift/>

# Outline

---

- Introduction
- The Seven Paradigms in NLP
- Paradigm Shift in NLP Tasks
- Potential Unified Paradigms
- Conclusion

# Outline

---

- **Introduction**
- **The Seven Paradigms in NLP**
- **Paradigm Shift in NLP Tasks**
- **Potential Unified Paradigms**
- **Conclusion**

# What is Paradigm?

---

- **Definition from Wikipedia**

- In science and philosophy, a **paradigm** is a distinct set of **concepts** or thought **patterns**, including theories, research methods, postulates, and standards for what constitutes legitimate contributions to a field.

- **Definition in the context of NLP**

- *Paradigm is the general framework to model a class of tasks*

# What is Paradigm?

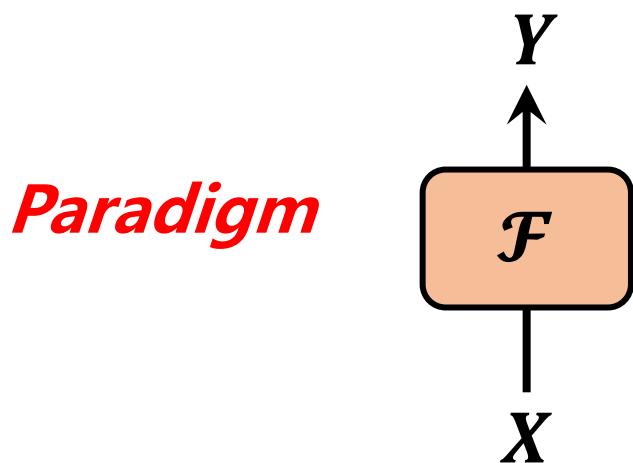
---

- **Definition from Wikipedia**

- In science and philosophy, a **paradigm** is a distinct set of **concepts** or thought **patterns**, including theories, research methods, postulates, and standards for what constitutes legitimate contributions to a field.

- **Definition in the context of NLP**

- *Paradigm is the general framework to model a class of tasks*



# What is Paradigm?

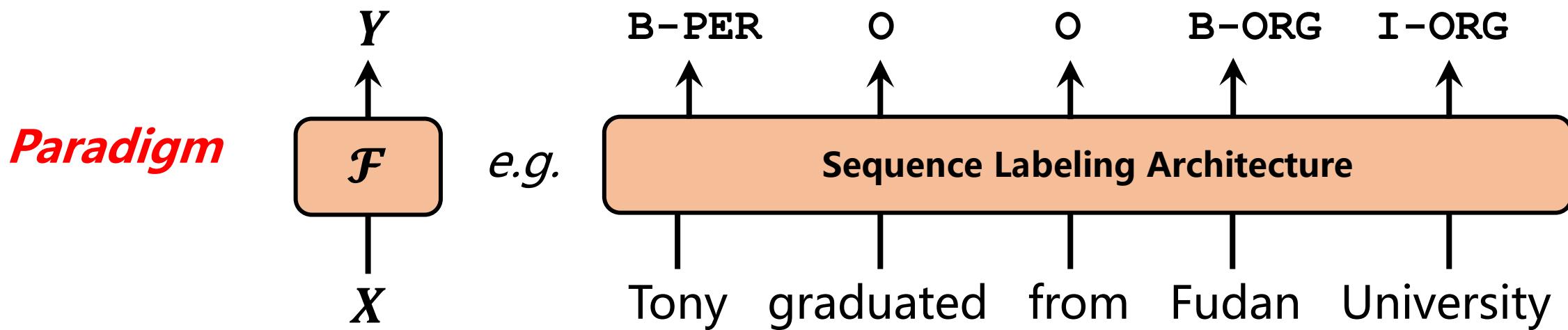
---

- **Definition from Wikipedia**

- In science and philosophy, a **paradigm** is a distinct set of **concepts** or thought **patterns**, including theories, research methods, postulates, and standards for what constitutes legitimate contributions to a field.

- **Definition in the context of NLP**

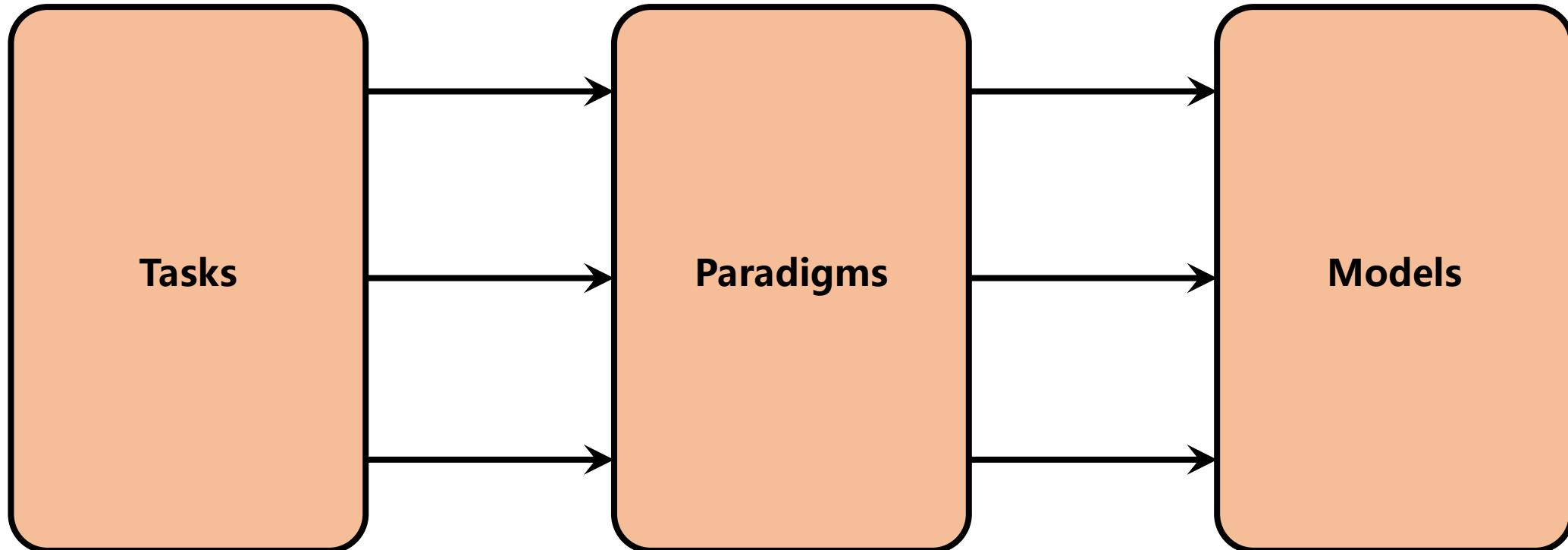
- *Paradigm is the general framework to model a class of tasks*



# Paradigms, Tasks, and Models

---

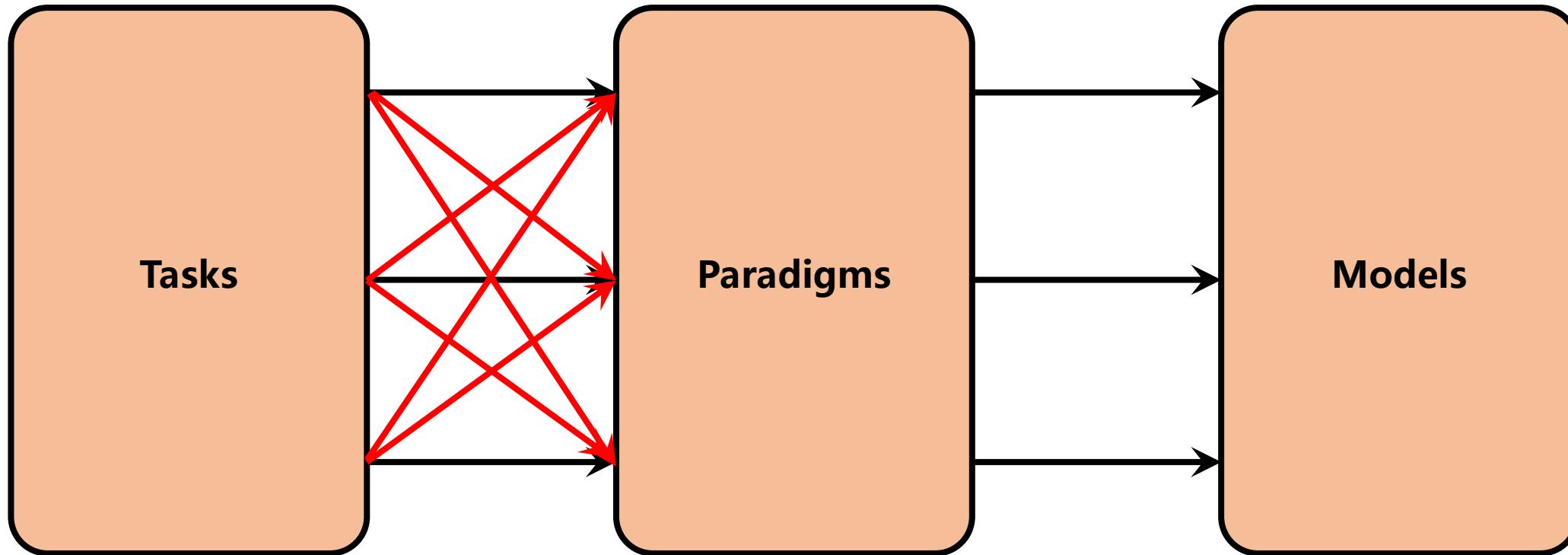
- A Rough Illustration



# Paradigms, Tasks, and Models

---

- A Rough Illustration



# Outline

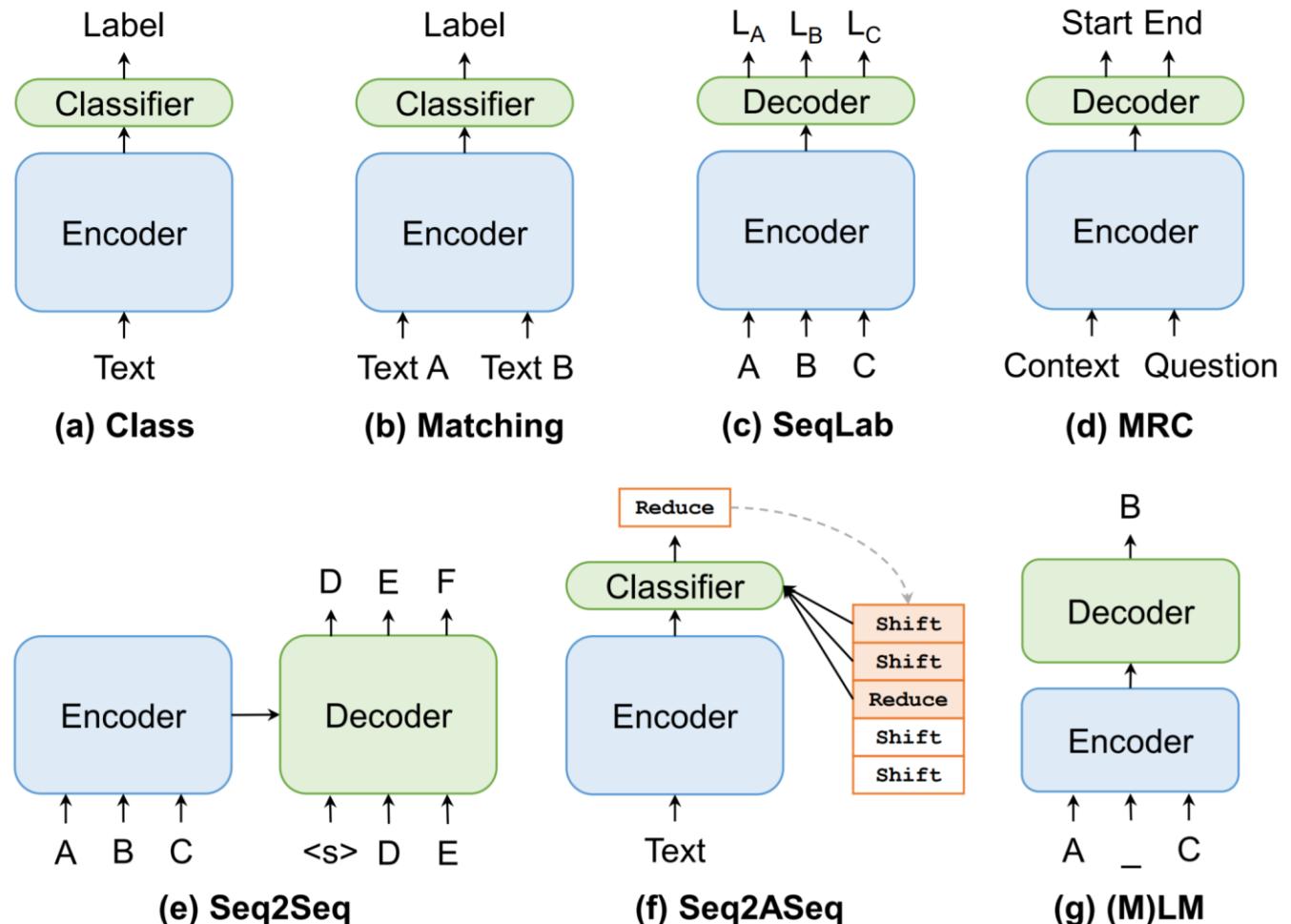
---

- Introduction
- **The Seven Paradigms in NLP**
- Paradigm Shift in NLP Tasks
- Potential Unified Paradigms
- Conclusion

# The Seven Paradigms in NLP

## • Seven Paradigms

- Class
- Matching
- SeqLab
- MRC
- Seq2Seq
- Seq2ASeq
- (M)LM



# Classification (**Class**)

---

- **Paradigm**

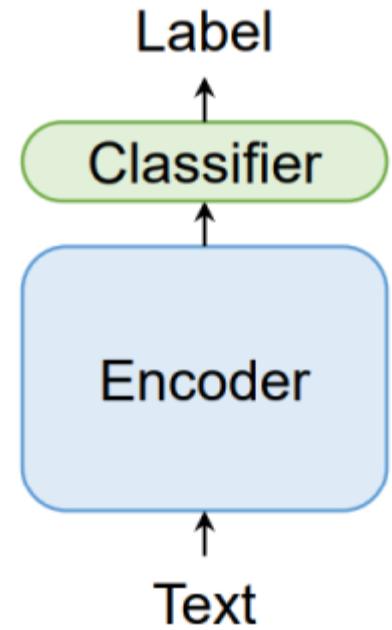
$$y = \text{CLS}(\text{ENC}(\mathcal{X})).$$

- **Model**

- $\text{ENC}(\cdot)$  : CNN, RNN, Transformers...
- $\text{CLS}(\cdot)$  : (max/average/attention) pooling + MLP

- **Tasks**

- Sentiment Analysis
- Spam Detection
- ...



# Matching

---

- **Paradigm**

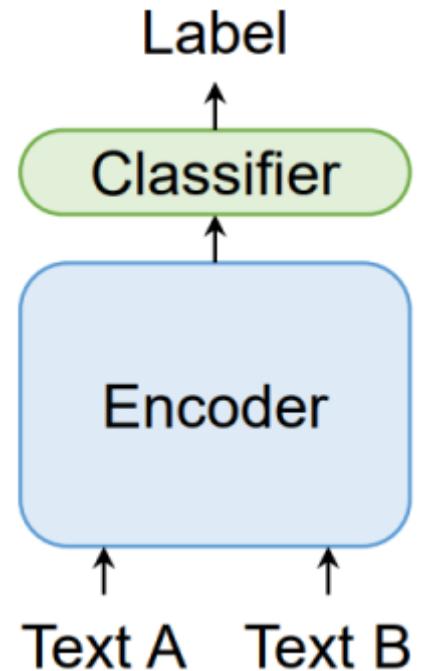
$$\mathcal{Y} = \text{CLS}(\text{ENC}(\mathcal{X}_a, \mathcal{X}_b))$$

- **Model**

- $\text{ENC}(\cdot)$  : encode the two texts separately or jointly
- $\text{CLS}(\cdot)$  : capture the interaction, and then prediction

- **Tasks**

- Natural Language Inference
- Similarity Regression
- ...



# Sequence Labeling (SeqLab)

- **Paradigm**

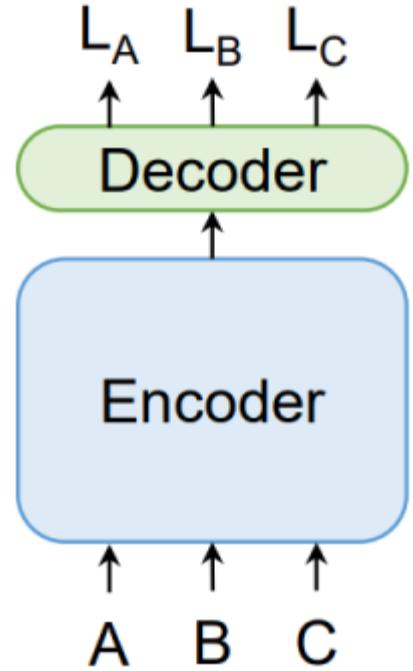
$$y_1, \dots, y_n = \text{DEC}(\text{ENC}(x_1, \dots, x_n))$$

- **Model**

- $\text{ENC}(\cdot)$  : sequence model (RNN, Transformers...)
- $\text{DEC}(\cdot)$  : conditional random fields (CRF)

- **Tasks**

- Named Entity Recognition (NER)
- Part-Of-Speech Tagging
- ...



# Machine Reading Comprehension (**MRC**)

---

- **Paradigm**

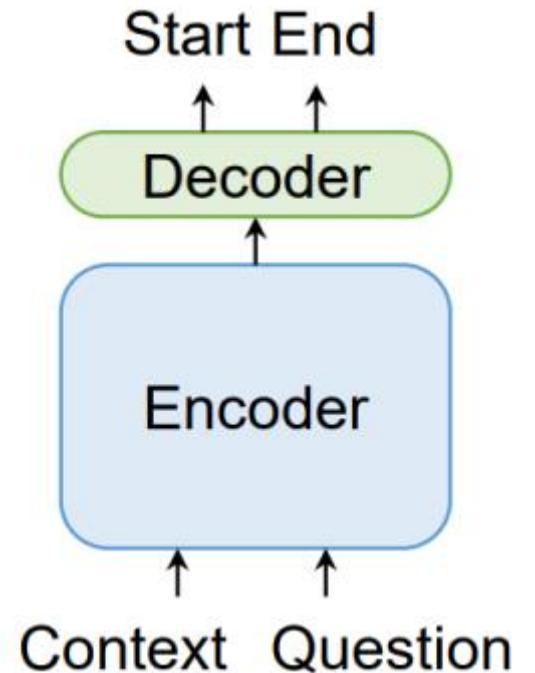
$$y_k \cdots y_{k+l} = \text{DEC}(\text{ENC}(\mathcal{X}_p, \mathcal{X}_q))$$

- **Model**

- $\text{ENC}(\cdot)$  : CNN, RNN, Transformers...
- $\text{DEC}(\cdot)$  : start/end position prediction

- **Tasks**

- Machine Reading Comprehension



# Sequence-to-Sequence (Seq2Seq)

- **Paradigm**

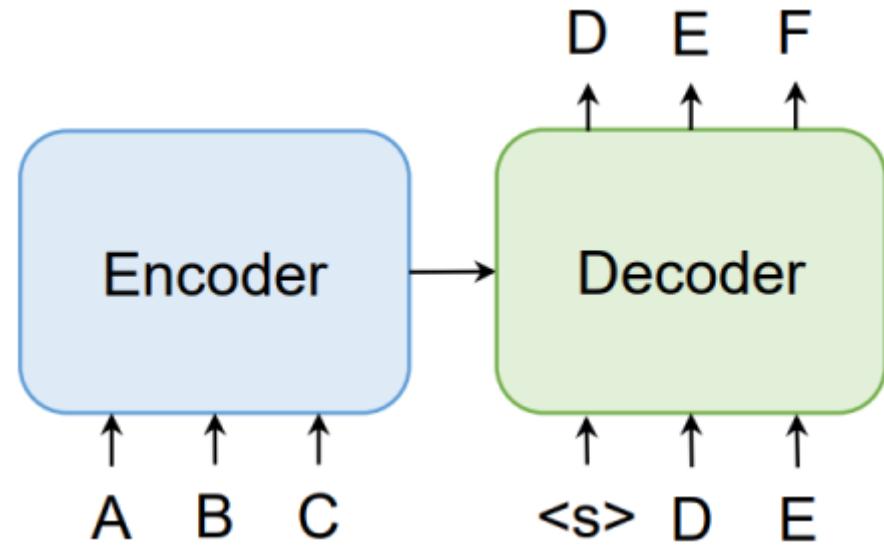
$$y_1, \dots, y_m = \text{DEC}(\text{ENC}(x_1, \dots, x_n))$$

- **Model**

- $\text{ENC}(\cdot)$  : CNN, RNN, Transformers...
- $\text{DEC}(\cdot)$  : CNN, RNN, Transformers...

- **Tasks**

- Machine Translation
- End-to-end dialogue system
- ...



# Sequence-to-Action-Sequence (**Seq2ASeq**)

- **Paradigm**

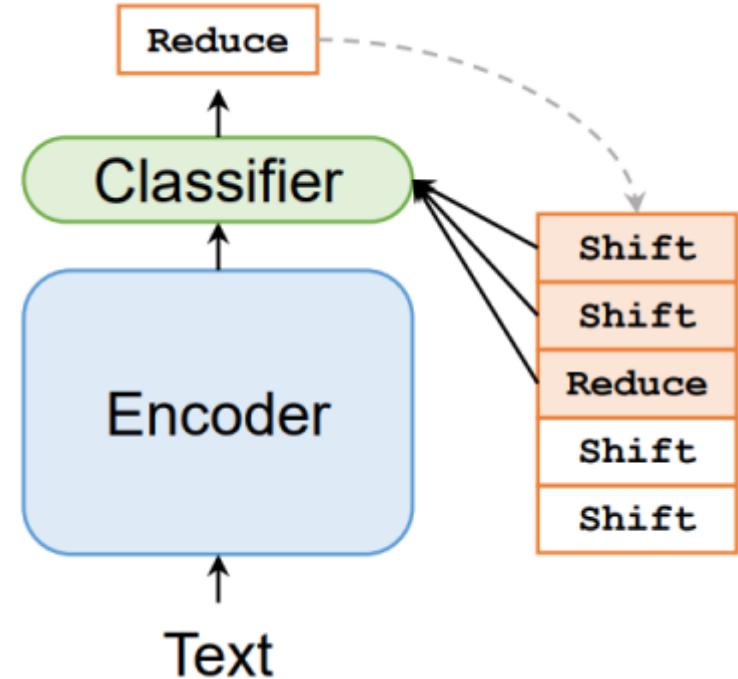
$$\mathcal{A} = \text{CLS}(\text{ENC}(\mathcal{X}), \mathcal{C})$$

- **Model**

- $\text{ENC}(\cdot)$  : CNN, RNN, Transformers...
- $\text{CLS}(\cdot)$  : predict an action conditioned on a configuration and the input text

- **Tasks**

- Dependency Parsing
- Constituency Parsing
- ...



# (Masked) Language Model ((M)LM)

- **Paradigm**

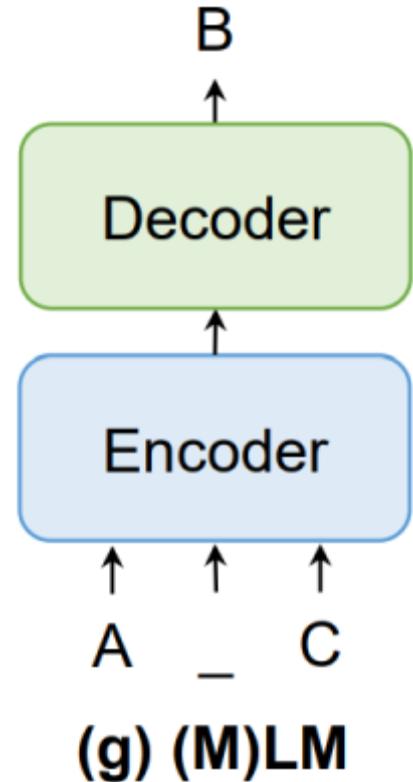
- LM:  $x_k = \text{DEC}(x_1, \dots, x_{k-1})$
- MLM:  $\bar{x} = \text{DEC}(\text{ENC}(\tilde{x}))$

- **Model**

- $\text{ENC}(\cdot)$  : CNN, RNN, Transformers...
- $\text{DEC}(\cdot)$  : simple classifier, or a auto-regressive decoder

- **Tasks**

- Language Modeling
- Masked Language Modeling
- ...



# Compound Paradigm

---

- Complicated NLP tasks can be solved by combining multiple fundamental paradigms
- An Example
  - HotpotQA

## Paragraph A, Return to Olympus:

[1] *Return to Olympus* is the only album by the alternative rock band *Mal funkshun*. [2] It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990. [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

## Paragraph B, Mother Love Bone:

[4] *Mother Love Bone* was an American rock band that formed in Seattle, Washington in 1987. [5] The band was active from 1987 to 1990. [6] Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene. [7] Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success. [8] The album was finally released a few months later.

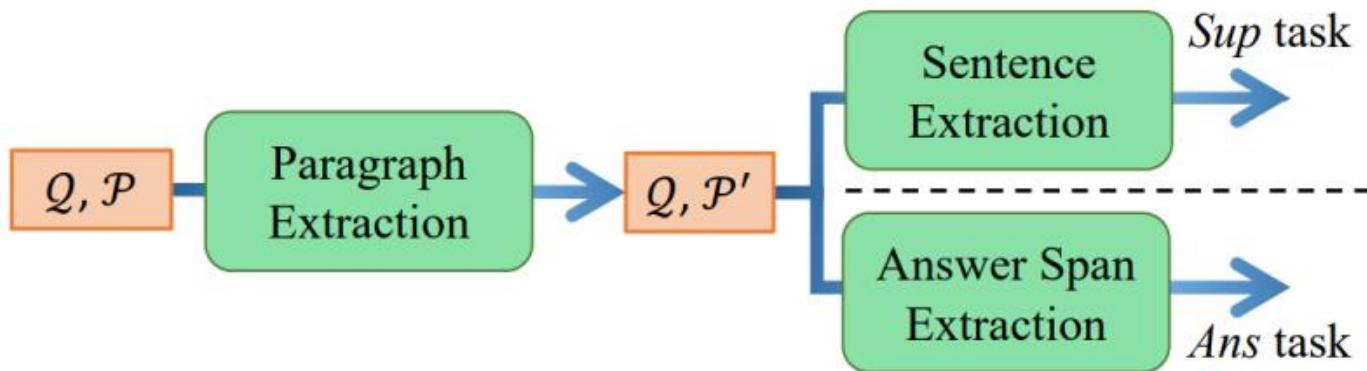
**Q:** What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?

**A:** Mal funkshun

**Supporting facts:** 1, 2, 4, 6, 7

# Compound Paradigm

- Complicated NLP tasks can be solved by combining multiple fundamental paradigms
- An Example
  - HotpotQA = Matching + MRC



## Paragraph A, Return to Olympus:

[1] *Return to Olympus* is the only album by the alternative rock band *Mal funkshun*. [2] It was released after the band had broken up and after lead singer Andrew Wood (later of Mother Love Bone) had died of a drug overdose in 1990. [3] Stone Gossard, of Pearl Jam, had compiled the songs and released the album on his label, Loosegroove Records.

## Paragraph B, Mother Love Bone:

[4] *Mother Love Bone* was an American rock band that formed in Seattle, Washington in 1987. [5] The band was active from 1987 to 1990. [6] Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s Seattle music scene. [7] Wood died only days before the scheduled release of the band's debut album, "Apple", thus ending the group's hopes of success. [8] The album was finally released a few months later.

**Q:** What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?

**A:** Mal funkshun

**Supporting facts:** 1, 2, 4, 6, 7

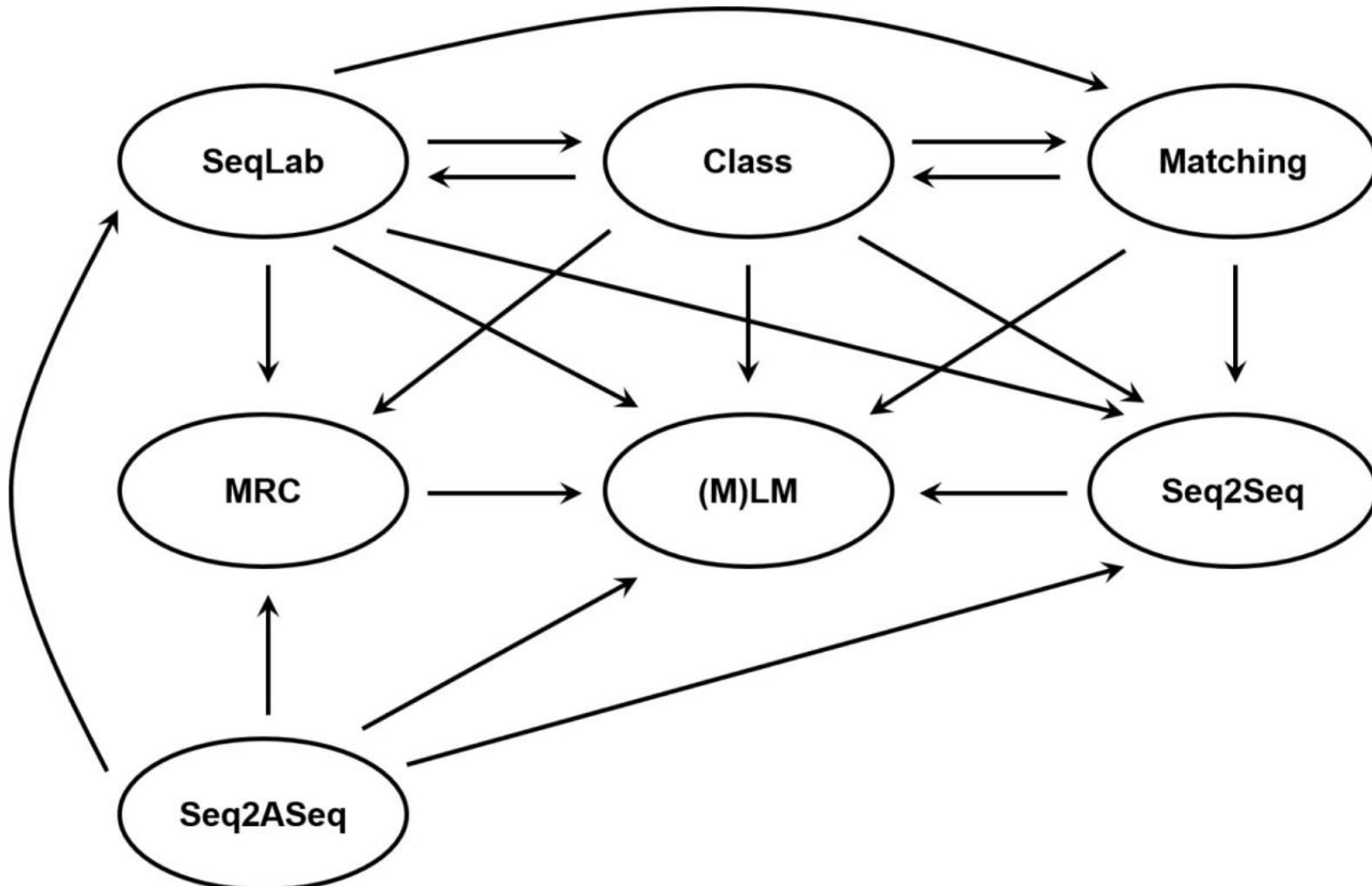
# Outline

---

- Introduction
- The Seven Paradigms in NLP
- **Paradigm Shift in NLP Tasks**
- Potential Unified Paradigms
- Conclusion

# Paradigm Shift in NLP

---



# Paradigm Shift in NLP

Task	Class	Matching	SeqLab	MRC	Seq2Seq	Seq2ASeq	(M) LM
<b>TC</b>	Input $\mathcal{X}$	$\mathcal{X}, \mathcal{L}$		$\mathcal{X}$ $y_1, \dots, y_m$ Yang et al. (2018a)	$f_{prompt}(\mathcal{X})$ $g(\mathcal{Y})$ Schick and Schütze (2021a)		
	Output $\mathcal{Y}$	$\mathcal{Y} \in \{0, 1\}$					
	Example Devlin et al. (2019)	Chai et al. (2020)					
<b>NLI</b>	Input $\mathcal{X}_a \oplus \mathcal{X}_b$	$\mathcal{X}_a, \mathcal{X}_b$		$f_{prompt}(\mathcal{X}_a, \mathcal{X}_b)$ $\mathcal{Y}$ McCann et al. (2018)	$f_{prompt}(\mathcal{X}_a, \mathcal{X}_b)$ $g(\mathcal{Y})$ Schick and Schütze (2021a)		
	Output $\mathcal{Y}$	$\mathcal{Y}$					
	Example Devlin et al. (2019)	Chen et al. (2017b)					
<b>NER</b>	Input $\mathcal{X}_{span}$	$x_1, \dots, x_n$ $y_1, \dots, y_n$ Ma and Hovy (2016)		$\mathcal{X}, \mathcal{Q}_y$ $\mathcal{X}_{span}$ Li et al. (2020)	$\mathcal{X}$ $(\mathcal{X}_{ent_i}, \mathcal{Y}_{ent_i})_{i=1}^m$ Yan et al. (2021b)	$(\mathcal{X}, \mathcal{C}_t)_{t=0}^{m-1}$ $\mathcal{A} = a_1, \dots, a_m$ Lample et al. (2016)	
	Output $\mathcal{Y}$						
	Example Fu et al. (2021)						
<b>ABSA</b>	Input $\mathcal{X}_{asp}$	$\mathcal{X}, \mathcal{S}_{aux}$		$\mathcal{X}, \mathcal{Q}_{asp}, \mathcal{Q}_{opin\&sent}$ $\mathcal{X}_{asp}, \mathcal{X}_{opin}, \mathcal{Y}_{sent}$ Mao et al. (2021)	$\mathcal{X}$ $(\mathcal{X}_{aspi}, \mathcal{X}_{opini}, \mathcal{Y}_{senti})_{i=1}^m$ Yan et al. (2021a)		$f_{prompt}(\mathcal{X})$ $g(\mathcal{Y})$ Li et al. (2021)
	Output $\mathcal{Y}$	$\mathcal{Y}$					
	Example Wang et al. (2016)	Sun et al. (2019)					
<b>RE</b>	Input $\mathcal{X}$			$\mathcal{X}, \mathcal{Q}_y$ $\mathcal{X}_{ent}$ Levy et al. (2017)	$\mathcal{X}$ $(\mathcal{Y}_i, \mathcal{X}_{subi}, \mathcal{X}_{obj})_{i=1}^m$ Zeng et al. (2018)		$f_{prompt}(\mathcal{X})$ $g(\mathcal{Y})$ Han et al. (2021)
	Output $\mathcal{Y}$						
	Example Zeng et al. (2014)						
<b>Summ</b>	Input $\mathcal{X}$	$(\mathcal{X}, \mathcal{S}_{candi})_{i=1}^n$ $\hat{\mathcal{S}}_{cand}$ Zhong et al. (2020)	$\mathcal{X}_1, \dots, \mathcal{X}_n$ $\mathcal{Y}_1, \dots, \mathcal{Y}_n \in \{0, 1\}^n$ Cheng and Lapata (2016)		$\mathcal{X}, \mathcal{Q}_{summ}$ $\mathcal{Y}$ McCann et al. (2018)		$\mathcal{X}$ , Keywords/Prompt $\mathcal{Y}$ Aghajanyan et al. (2021)
	Output $\mathcal{S}$						
	Example Zhong et al. (2020)						
<b>Parsing</b>	Input $\mathcal{X}$		$x_1, \dots, x_n$ $g(y_1, \dots, y_n)$ Strzyl et al. (2019)	$\mathcal{X}, \mathcal{Q}_{child}$ $\mathcal{X}_{parent}$ Gan et al. (2021)	$\mathcal{X}$ $g(y_1, \dots, y_m)$ Vinyals et al. (2015)	$(\mathcal{X}, \mathcal{C}_t)_{t=0}^{m-1}$ $\mathcal{A} = a_1, \dots, a_m$ Chen and Manning (2014)	$(\mathcal{X}, \mathcal{Y}_i)_{i=1}^k$ $\hat{\mathcal{Y}}$ Choe and Charniak (2016)
	Output $\mathcal{Y}$						
	Example Strzyl et al. (2019)						

Table 1: Paradigms shift in natural language processing tasks. **TC**: text classification. **NLI**: natural language inference. **NER**: named entity recognition. **ABSA**: aspect-based sentiment analysis. **RE**: relation extraction. **Summ**: text summarization. **Parsing**: syntactic/semantic parsing.  $f$  and  $g$  indicate pre-processing and post-processing, respectively. In (M) LM,  $f(\cdot)$  is usually implemented as a template and  $g(\cdot)$  is a verbalizer. In parsing tasks,  $g(\cdot)$  is a function that reconstructs the structured representation (*e.g.* dependency tree) from the output sequence.  $\mathcal{L}$  means label description.  $\oplus$  means concatenation.  $\mathcal{X}_{asp}, \mathcal{X}_{opin}, \mathcal{Y}_{sent}$  mean aspect, opinion, and sentiment, respectively.  $\mathcal{S}_{aux}$  means auxiliary sentence.  $\mathcal{X}_{sub}, \mathcal{X}_{obj}$  stand for subject entity and object entity, respectively.  $\mathcal{S}_{cand}$  means candidate summary.  $\mathcal{C}_t$  is configuration  $t$  and  $\mathcal{A}$  is a sequence of actions. More details can be found in Section 3.

# Paradigm Shift in Text Classification

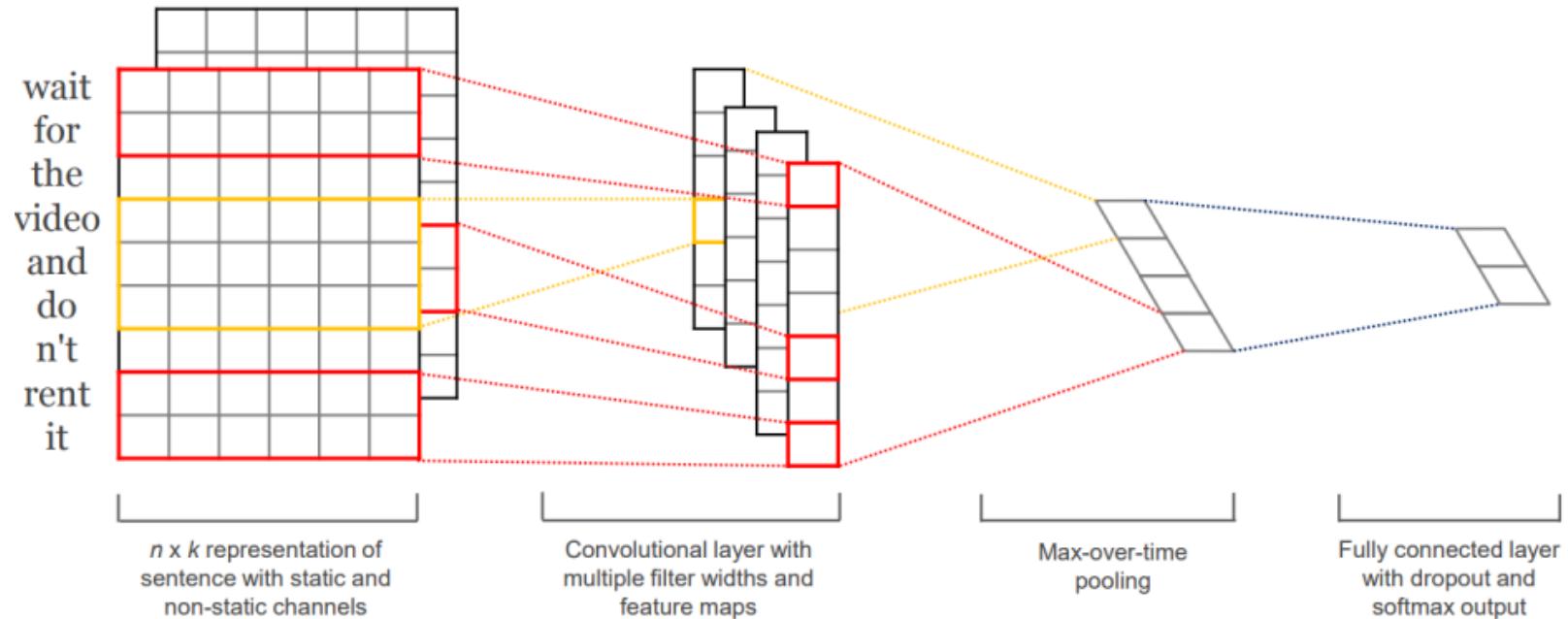
---

- Traditional Paradigm: **Class**
- Shifted to...
  - Seq2Seq
  - Matching
  - (M)LM

# Paradigm Shift in Text Classification

- Traditional Paradigm: **Class**
- Shifted to...

- Seq2Seq
- Matching
- (M)LM

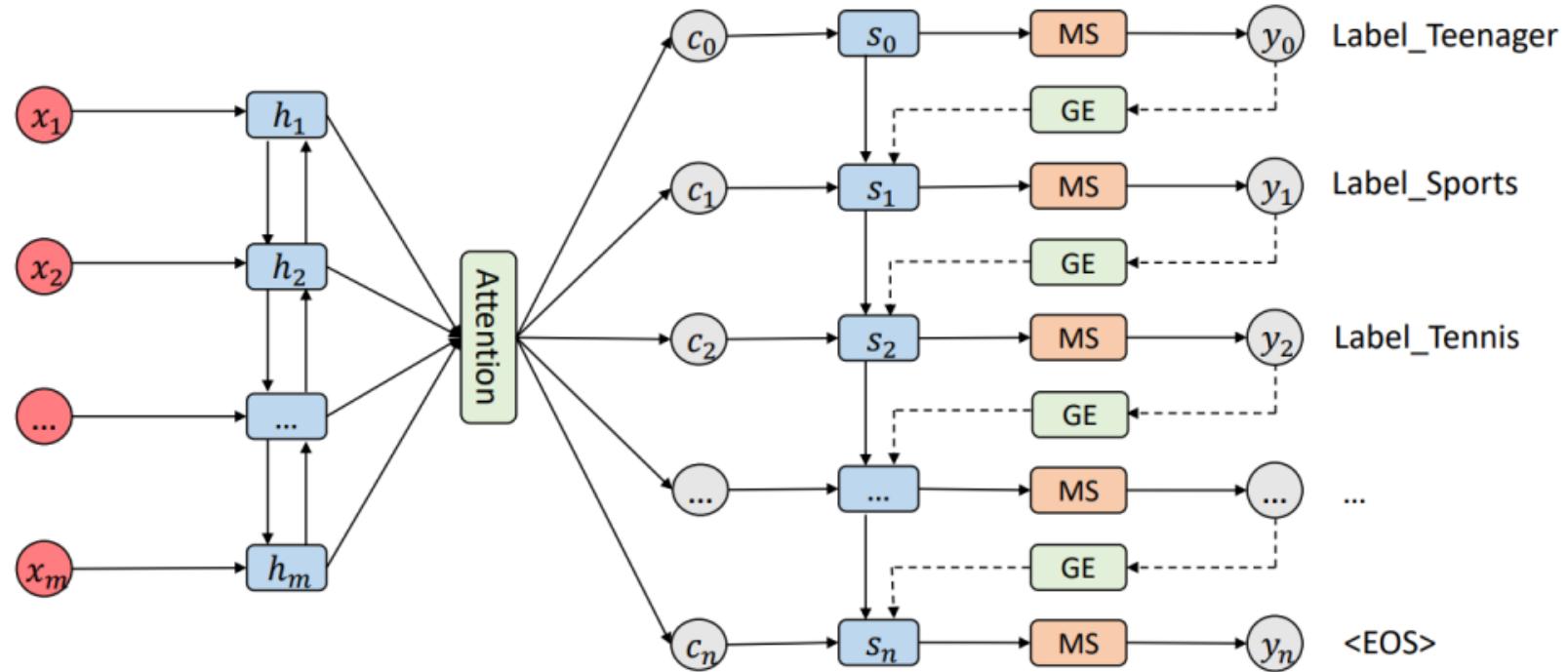


# Paradigm Shift in Text Classification

- Traditional Paradigm: Class
- Shifted to...

- Seq2Seq

- Matching
- (M)LM

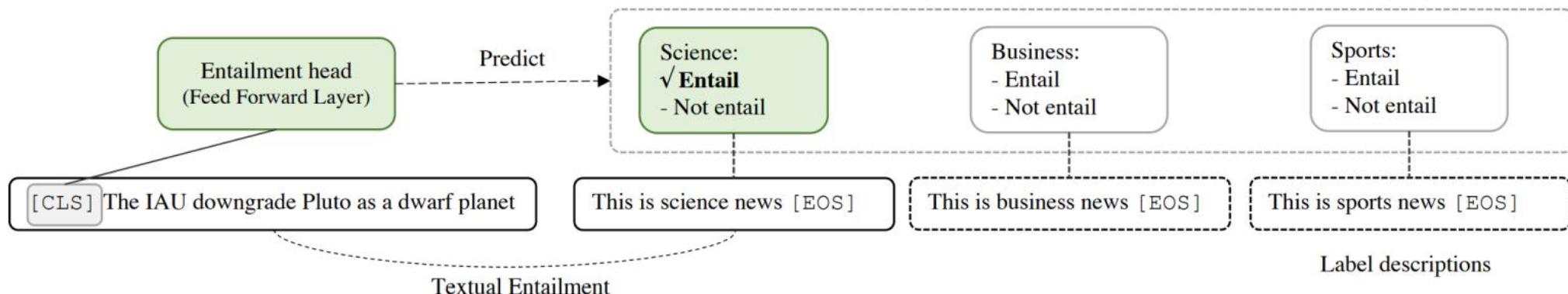


# Paradigm Shift in Text Classification

- Traditional Paradigm: Class

- Shifted to...

- Seq2Seq
- Matching
- (M)LM

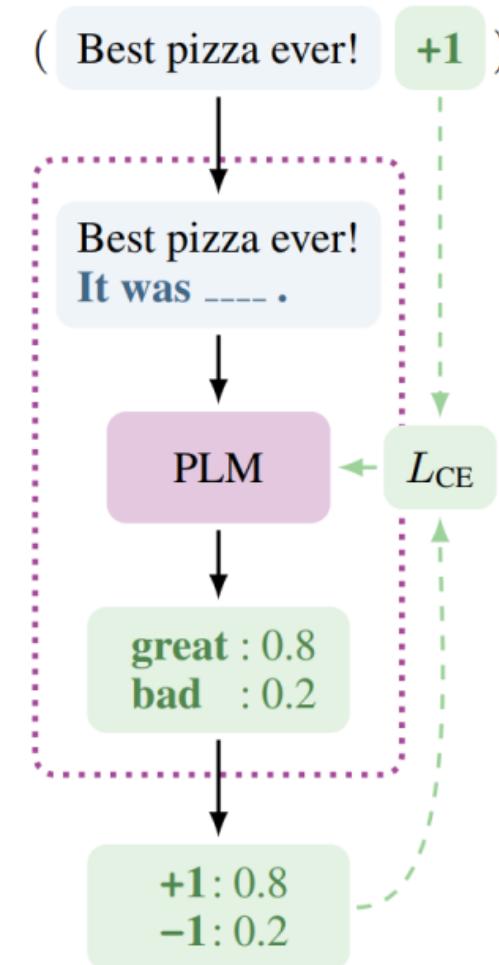


# Paradigm Shift in Text Classification

- Traditional Paradigm: Class

- Shifted to...

- Seq2Seq
- Matching
- (M)LM



# Paradigm Shift in NLI

---

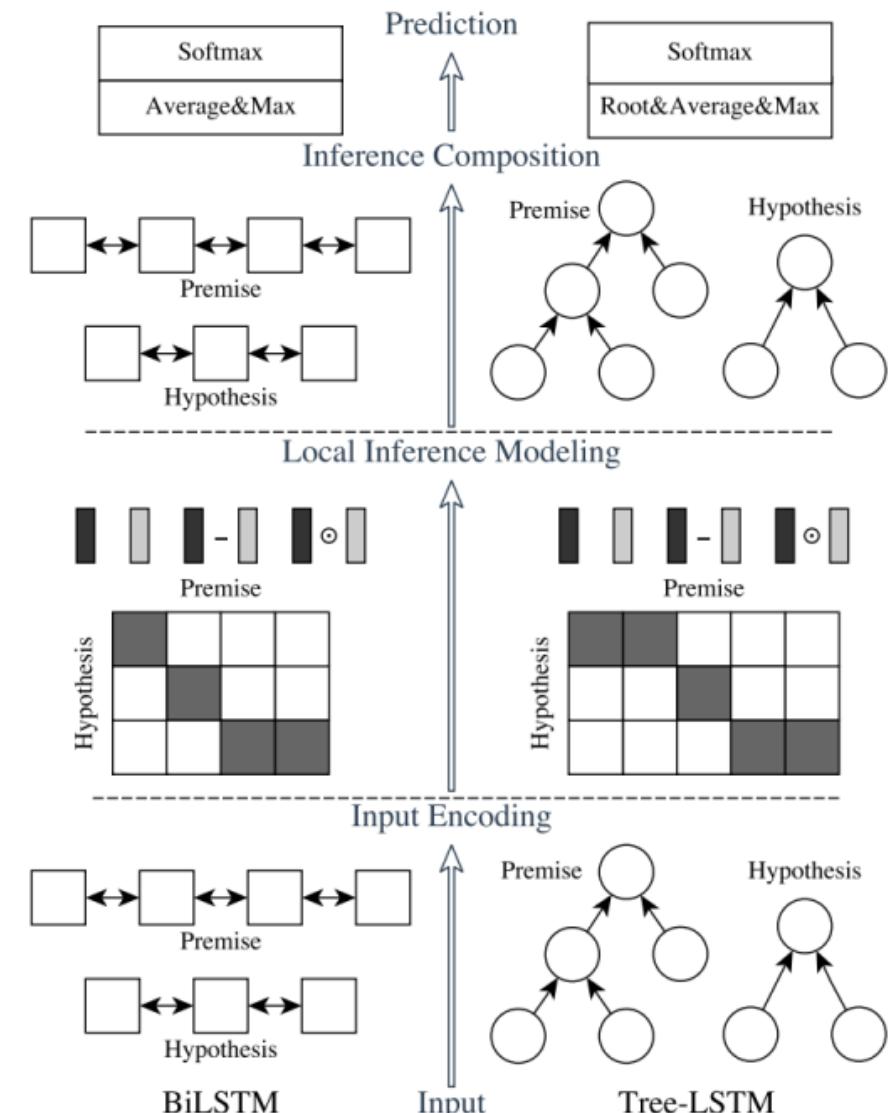
- Traditional Paradigm: **Matching**
- Shifted to...
  - Class
  - Seq2Seq
  - (M)LM

# Paradigm Shift in NLI

- Traditional Paradigm: **Matching**

- Shifted to...

- Class
- Seq2Seq
- (M)LM



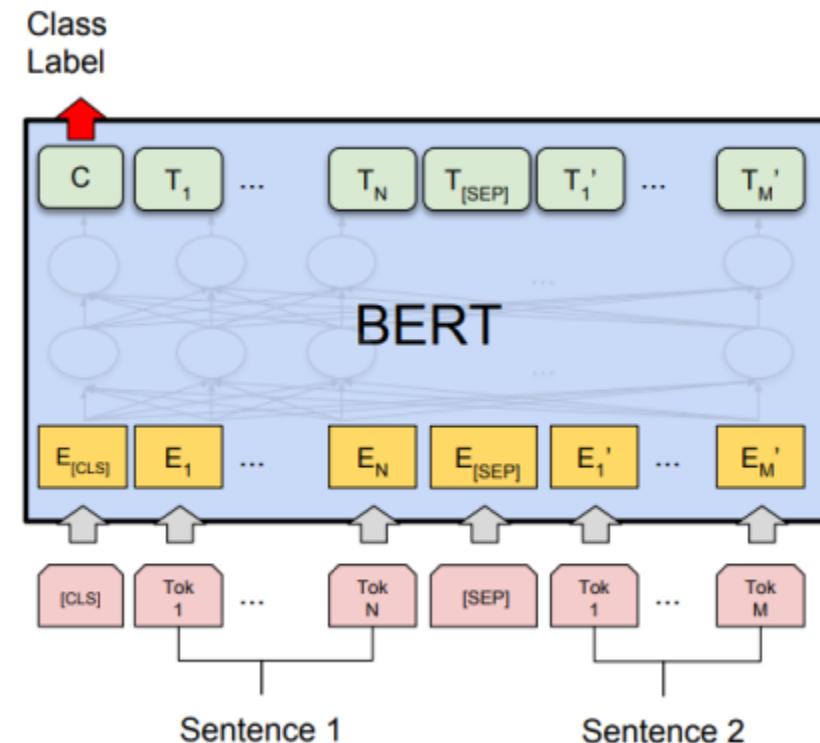
# Paradigm Shift in NLI

- Traditional Paradigm: Matching

- Shifted to...

- Class

- Seq2Seq
  - (M)LM



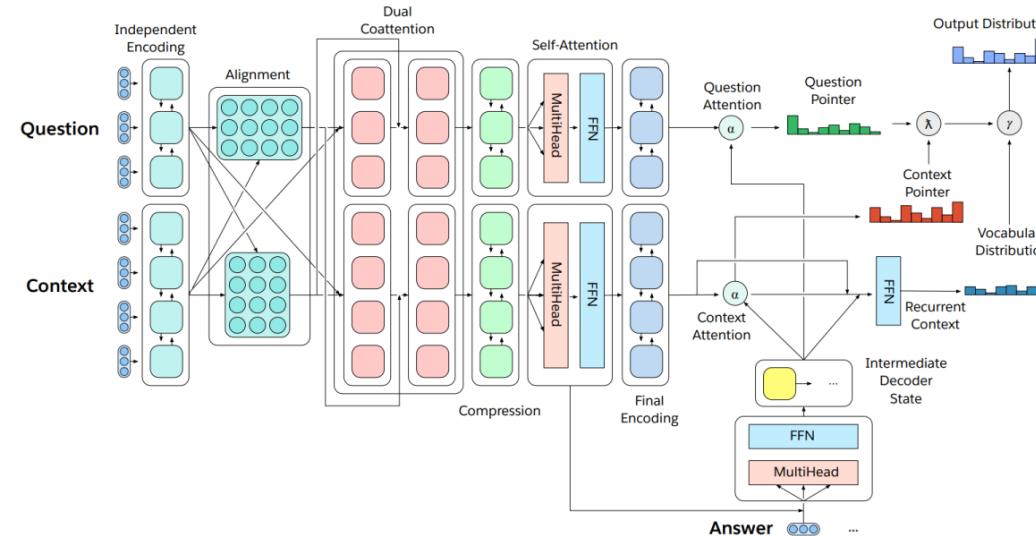
# Paradigm Shift in NLI

- Traditional Paradigm: Matching

- Shifted to...

- Class
- Seq2Seq
- (M)LM

Question	Context	Answer
<p>Hypothesis: Product and geography are what make cream skimming work. <b>Entailment</b>, neutral, or contradiction?</p>	<p>Premise: Conceptually cream skimming has two basic dimensions – product and geography.</p>	<p><b>Entailment</b></p>

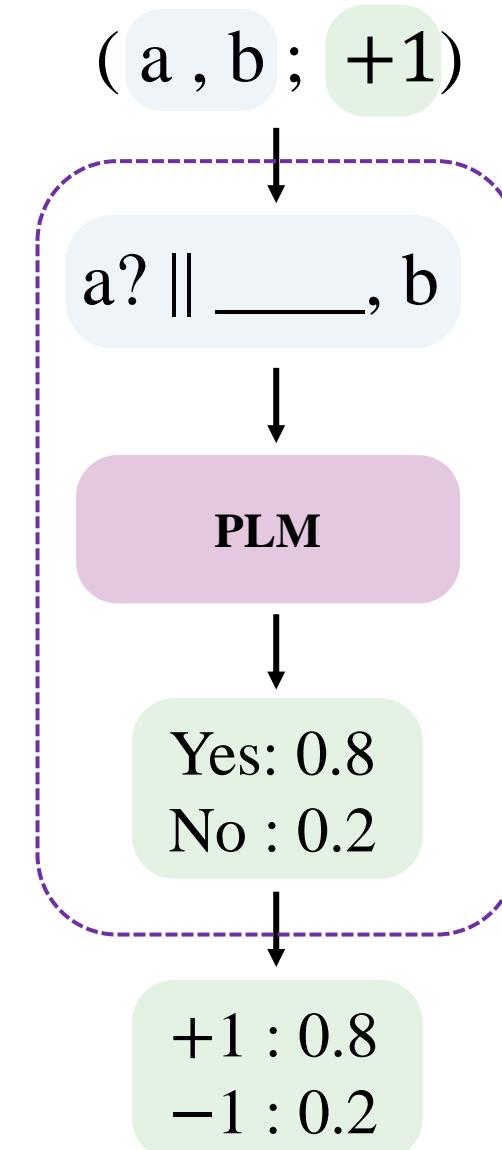


# Paradigm Shift in NLI

- Traditional Paradigm: Matching

- Shifted to...

- Class
- Seq2Seq
- (M)LM



# Paradigm Shift in NER

---

## Flat NER

Barack      Obama  
                Person

## Nested NER

The      Lincoln      Memorial  
                Person  
                Location

## Discontinuous NER

have      much      muscle      pain      and      fatigue  
  Disorder

# Paradigm Shift in NER

---

- **Traditional Paradigm:**

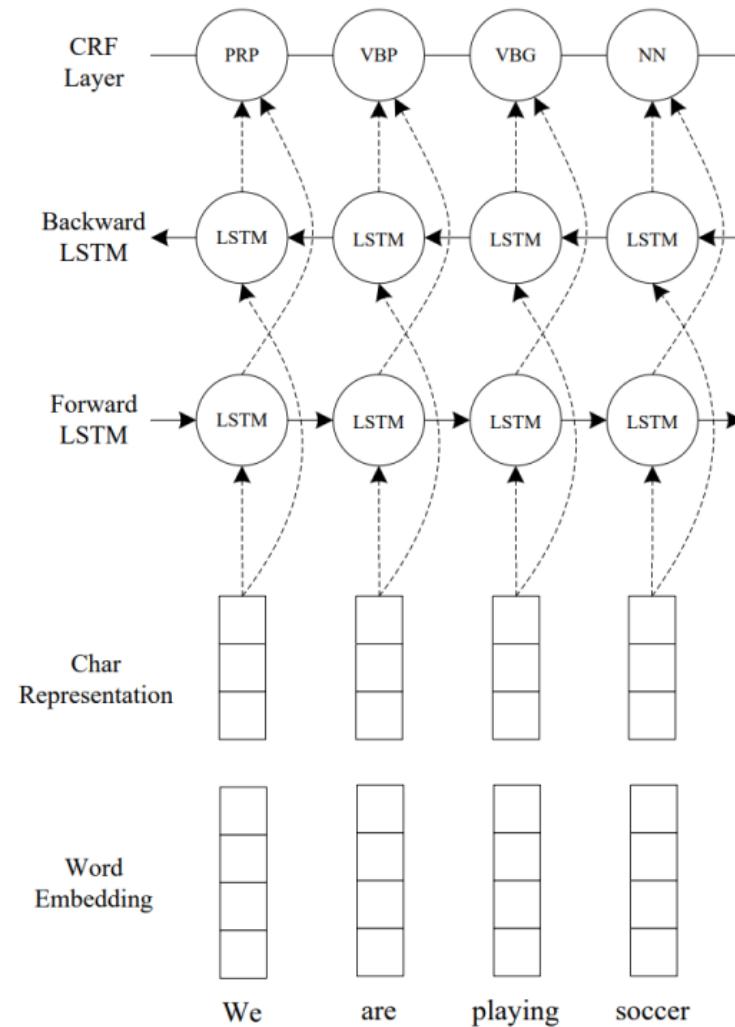
- SeqLab (Flat NER)
- Class (Nested NER)
- Seq2ASeq (Discontinuous NER)

- **Shifted to / Unified in...**

- Class (Flat&Nested NER)
- MRC (Flat&Nested NER)
- Seq2Seq (All)

# Paradigm Shift in NER

- **Traditional Paradigm:**
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Discontinuous NER)
- **Shifted to / Unified in...**
  - Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - Seq2Seq (All)



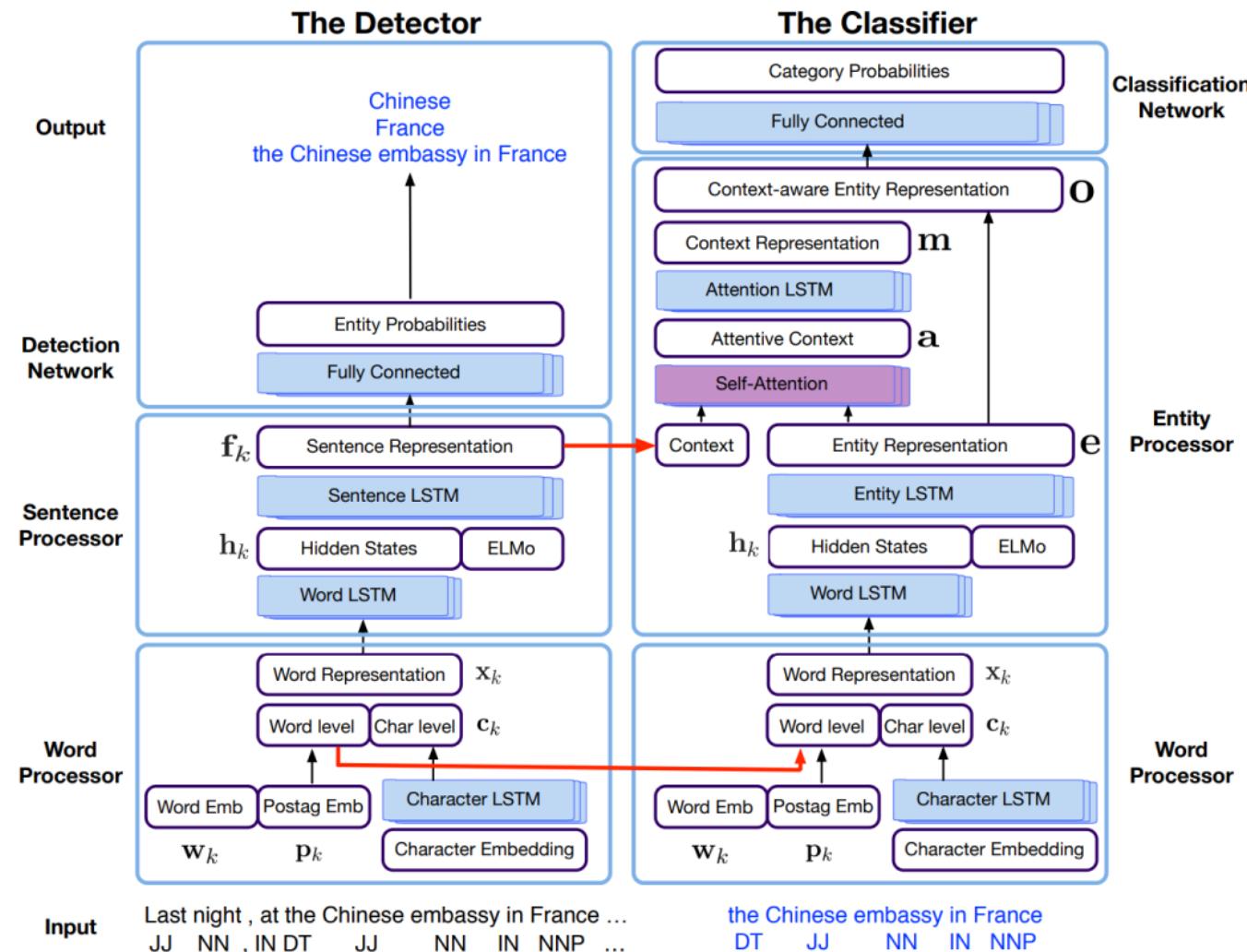
# Paradigm Shift in NER

- **Traditional Paradigm:**

- SeqLab (Flat NER)
- **Class** (Nested NER)
- Seq2ASeq (Discontinuous NER)

- **Shifted to / Unified in...**

- Class (Flat&Nested NER)
- MRC (Flat&Nested NER)
- Seq2Seq (All)



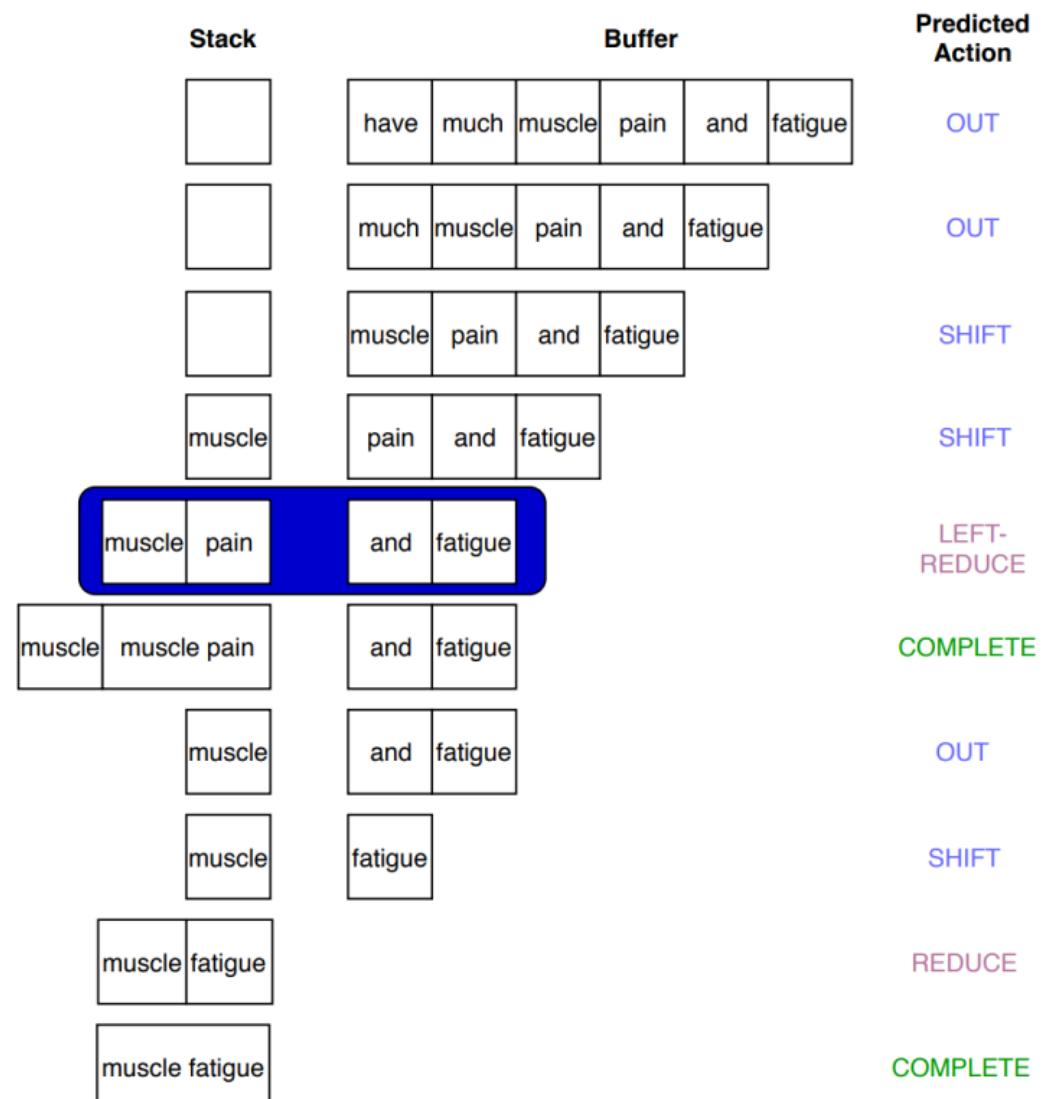
# Paradigm Shift in NER

- **Traditional Paradigm:**

- SeqLab (Flat NER)
- Class (Nested NER)
- **Seq2ASeq** (Discontinuous NER)

- **Shifted to / Unified in...**

- Class (Flat&Nested NER)
- MRC (Flat&Nested NER)
- Seq2Seq (All)



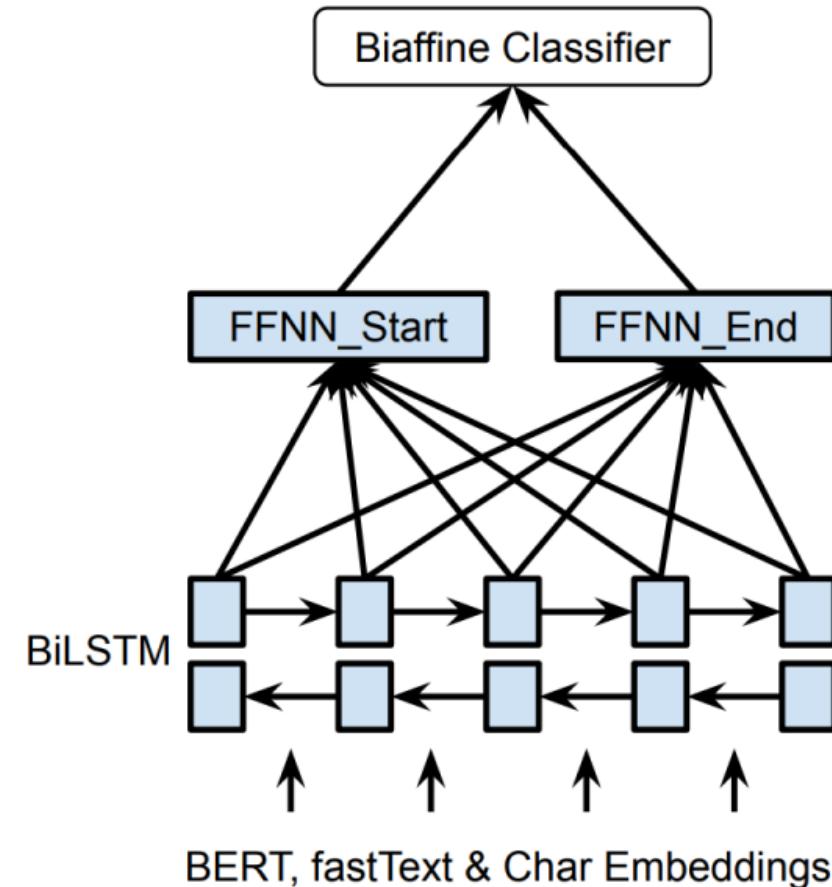
# Paradigm Shift in NER

- **Traditional Paradigm:**

- SeqLab (Flat NER)
- Class (Nested NER)
- Seq2ASeq (Discontinuous NER)

- **Shifted to / Unified in...**

- **Class** (Flat&Nested NER)
- MRC (Flat&Nested NER)
- Seq2Seq (All)



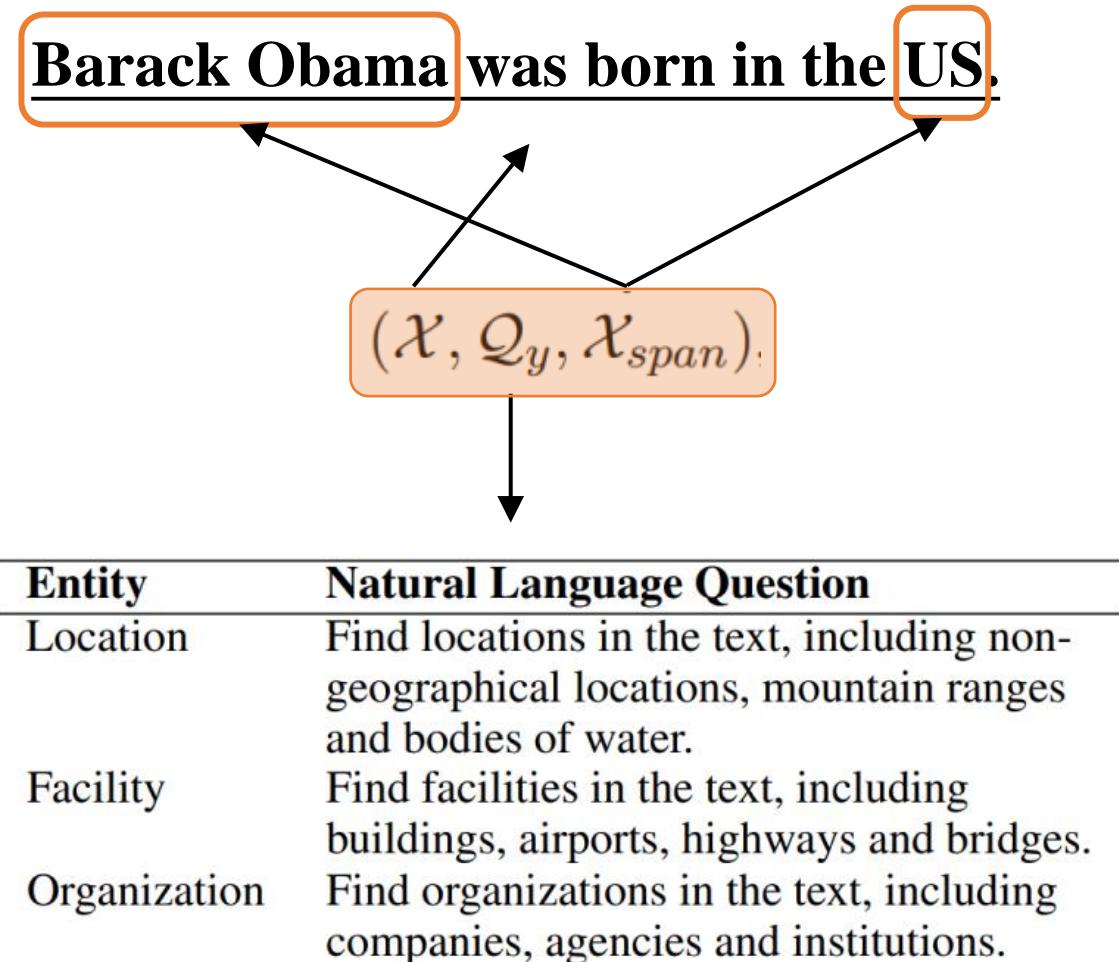
# Paradigm Shift in NER

- **Traditional Paradigm:**

- SeqLab (Flat NER)
- Class (Nested NER)
- Seq2ASeq (Discontinuous NER)

- **Shifted to / Unified in...**

- Class (Flat&Nested NER)
- **MRC (Flat&Nested NER)**
- Seq2Seq (All)



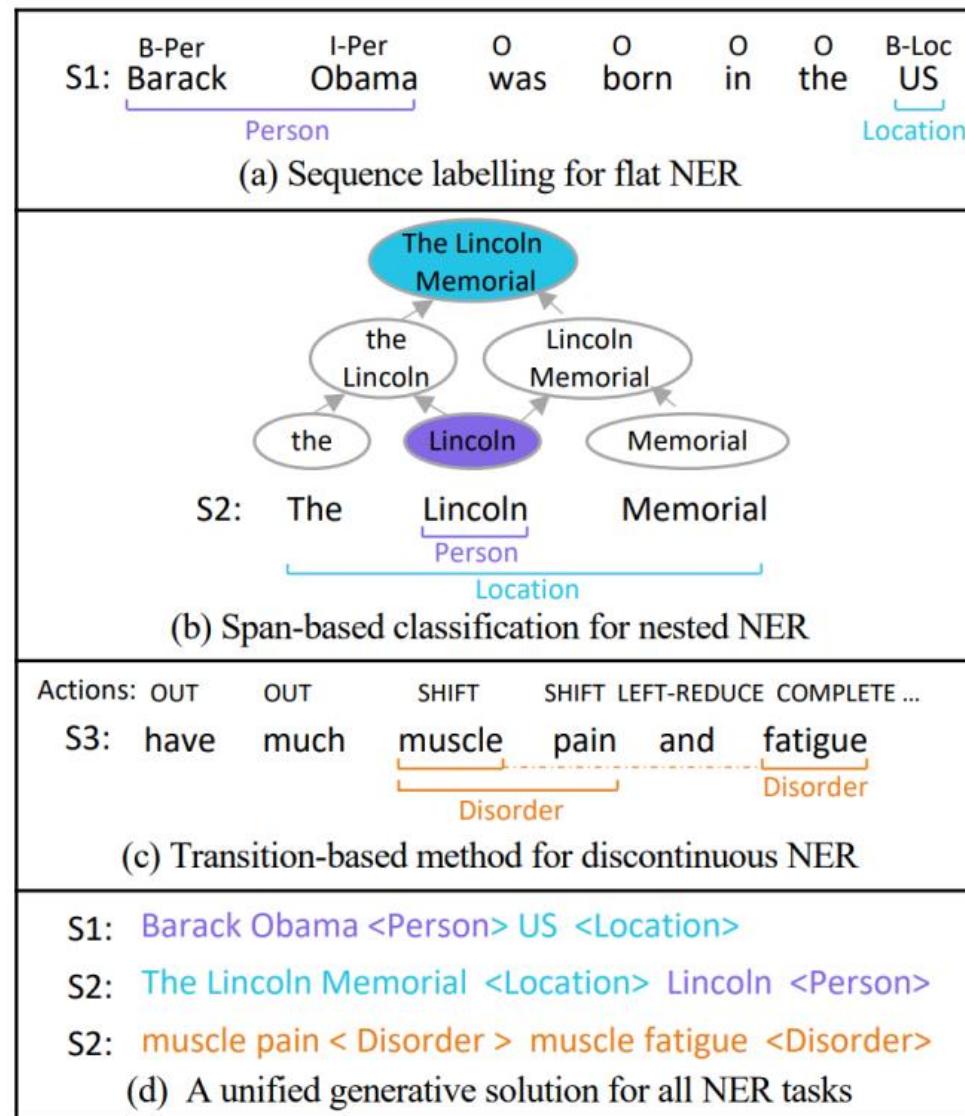
# Paradigm Shift in NER

- **Traditional Paradigm:**

- SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Discontinuous NER)

- Shifted to / Unified in...

- Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - Seq2Seq (All)



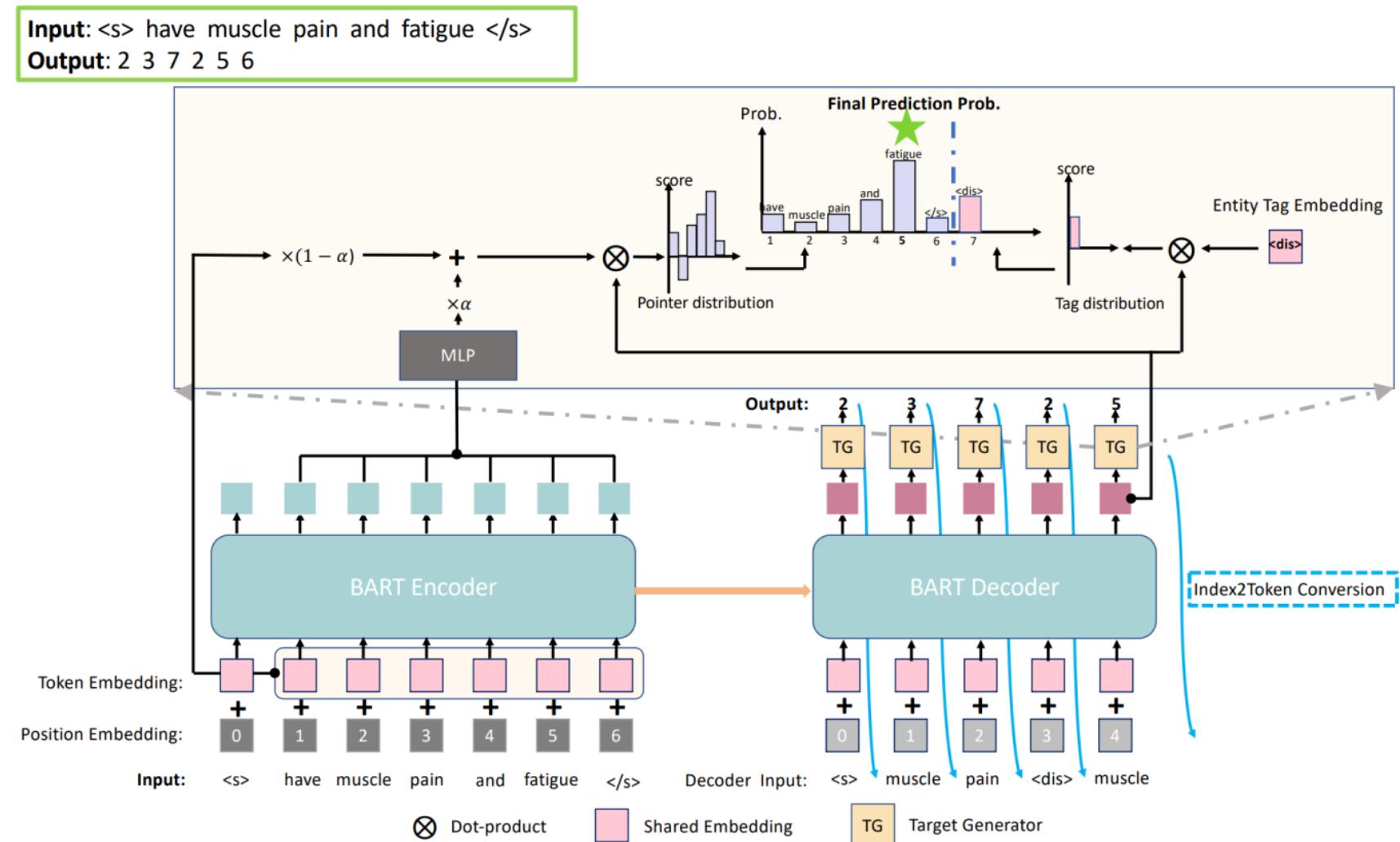
# Paradigm Shift in NER

- **Traditional Paradigm**

- SeqLab (Flat NER)
- Class (Nested NER)
- Seq2ASeq (Discontinuous)

- **Shifted to / Unified**

- Class (Flat&Nested)
- MRC (Flat&Nested)
- **Seq2Seq (All)**



# Paradigm Shift in ABSA

S: The **drinks** are always ***well made*** and **wine selection** is ***fairly priced***.

**a<sub>1</sub>**                    ***o<sub>1</sub>***                    **a<sub>2</sub>**                    ***o<sub>2</sub>***

**s<sub>1</sub> Positive**                    **s<sub>2</sub> Positive**

Subtask	Input	Output	Task Type
<b>Aspect Term Extraction(AE)</b>	S	<b>a<sub>1</sub>, a<sub>2</sub></b>	Extraction
<b>Opinion Term Extraction(OE)</b>	S	<b><i>o<sub>1</sub>, o<sub>2</sub></i></b>	Extraction
<b>Aspect-level Sentiment Classification(ALSC)</b>	<b>S + a<sub>1</sub></b>	<b>s<sub>1</sub></b>	Classification
	<b>S + a<sub>2</sub></b>	<b>s<sub>2</sub></b>	
<b>Aspect-oriented Opinion Extraction(AOE)</b>	<b>S + a<sub>1</sub></b>	<b><i>o<sub>1</sub></i></b>	Extraction
	<b>S + a<sub>2</sub></b>	<b><i>o<sub>2</sub></i></b>	
<b>Aspect Term Extraction and Sentiment Classification(AESC)</b>	S	<b>(a<sub>1</sub>, s<sub>1</sub>), (a<sub>2</sub>, s<sub>2</sub>)</b>	Extraction & Classification
<b>Pair Extraction(Pair)</b>	S	<b>(a<sub>1</sub>, o<sub>1</sub>), (a<sub>2</sub>, o<sub>2</sub>)</b>	Extraction
<b>Triplet Extraction(Triplet)</b>	S	<b>(a<sub>1</sub>, o<sub>1</sub>, s<sub>1</sub>), (a<sub>2</sub>, o<sub>2</sub>, s<sub>2</sub>)</b>	Extraction & Classification

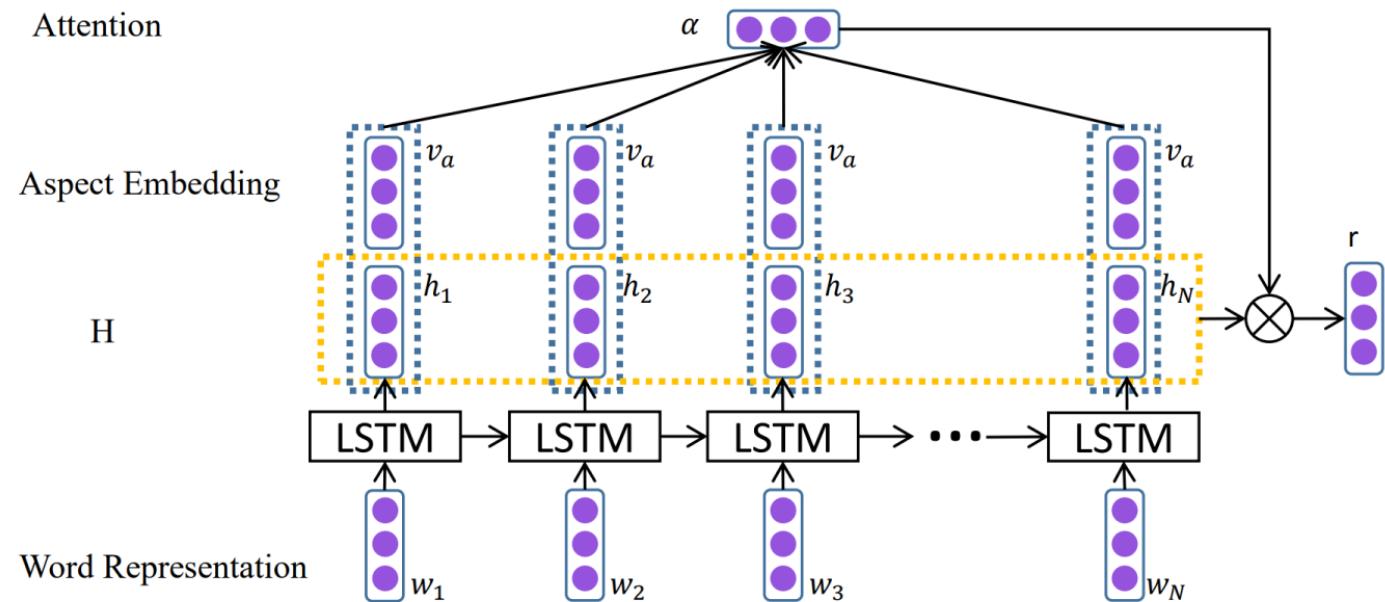
# Paradigm Shift in ABSA

---

- **Traditional Paradigm:**
  - SeqLab (AE, OE, AOE, ...)
  - Class (ALSC...)
- **Shifted to / Unified in...**
  - Matching (ALSC)
  - MRC (All)
  - Seq2Seq (All)
  - (M)LM (All)

# Paradigm Shift in ABSA

- Traditional Paradigm:
  - SeqLab (AE, OE, AOE, ...)
  - Class (ALSC...)
- Shifted to / Unified in...
  - Matching (ALSC)
  - MRC (All)
  - Seq2Seq (All)
  - (M)LM (All)



# Paradigm Shift in ABSA

- **Traditional Paradigm:**

- SeqLab (AE, OE, AOE, ...)
- Class (ALSC...)

- **Shifted to / Unified in...**

- **Matching** (ALSC)
- MRC (All)
- Seq2Seq (All)
- (M)LM (All)

X: LOC1 is often considered the coolest area of London.  
Aspect: *Safety*



**QA-M**

What do you think of the *safety* of LOC1? [X]

**NLI-M**

LOC1- *safety*. [X]

**QA-B**

The polarity of the aspect *safety* of LOC1 is positive. [X]

**NLI-B**

LOC1- *safety* - positive. [X]

# Paradigm Shift in ABSA

- **Traditional Paradigm:**

- SeqLab (AE, OE, AOE, ...)
- Class (ALSC...)

- **Shifted to / Unified in...**

- Matching (ALSC)
- **MRC** (All)
- Seq2Seq (All)
- (M)LM (All)

**Original training example:**

- **input text:** The **ambience** was **nice** , but **service** was **not so great**.
- **annotations:** (**ambience**, **nice**, **positive**), (**service**, **no so great**, **negative**)



**Converted training example 1:**

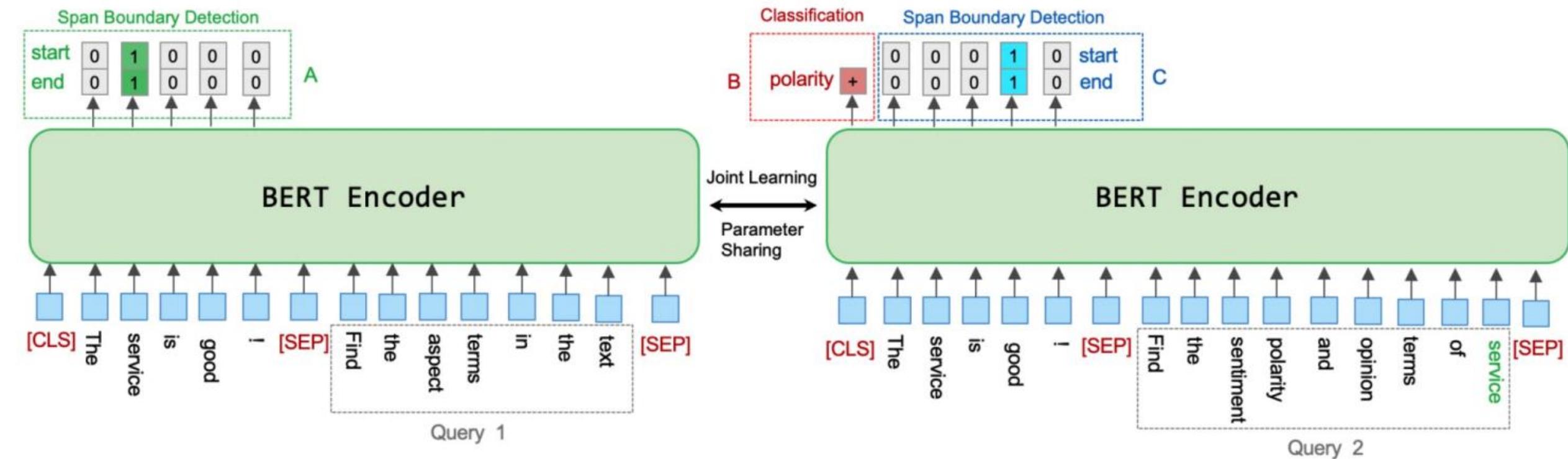
- **query-1:** Find the *aspect terms* in the text.
- **answer-1:** **ambience, service**
- **query-2:** Find the *sentiment polarity* and *opinion terms* for **ambience** in the text.
- **answer-2:** (**nice, positive**)

**Converted training example 2:**

- **query-1:** Find the *aspect terms* in the text.
- **answer-1:** **ambience, service**
- **query-2:** Find the *sentiment polarity* and *opinion terms* for **service** in the text.
- **answer-2:** (**not so great, negative**)

# Paradigm Shift in ABSA

- Traditional Paradigm:



# Paradigm Shift in ABSA

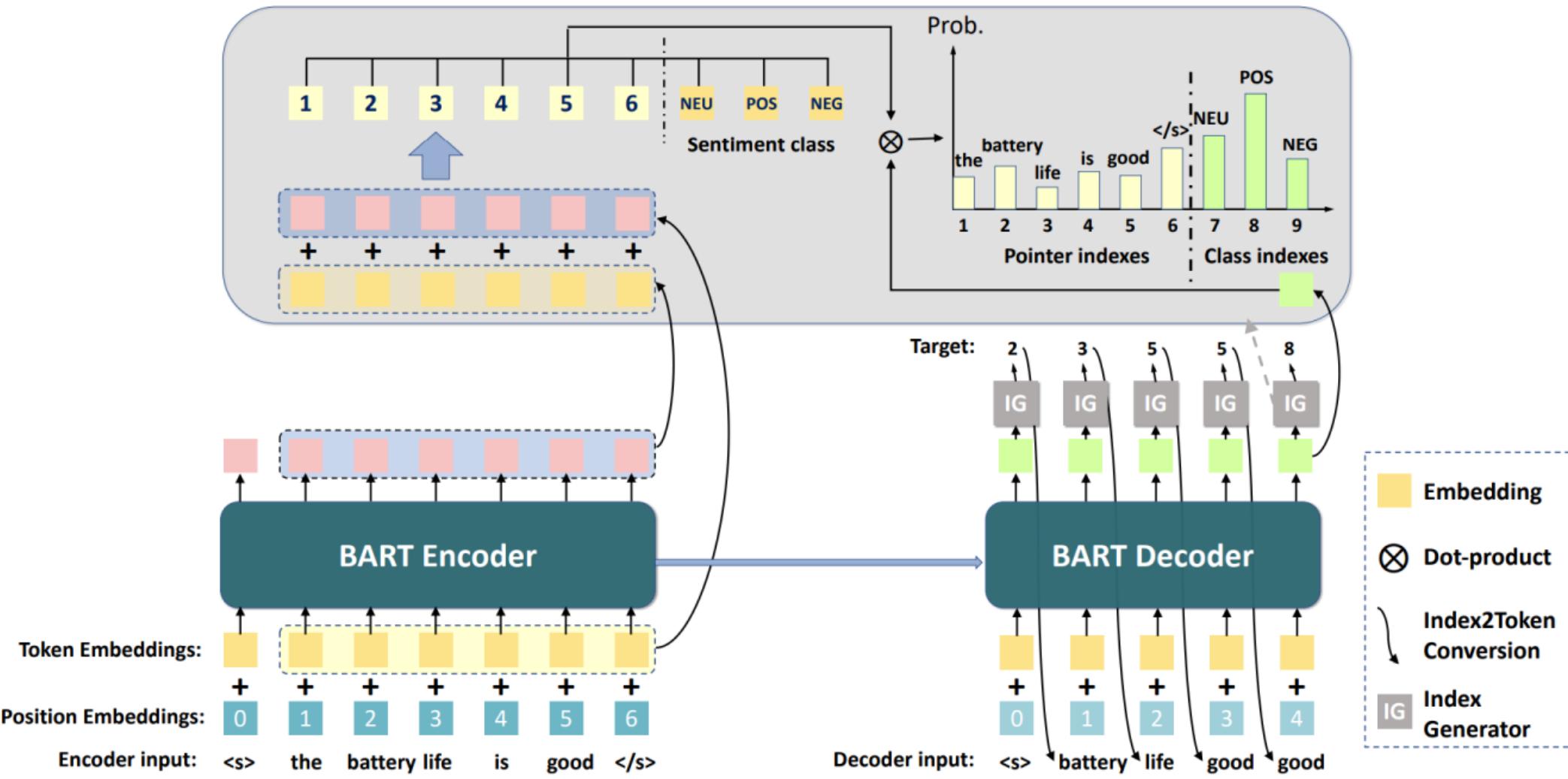
- Traditional Paradigm:
  - SeqLab (AE, OE, AOE, ...)
  - Class (ALSC...)
- Shifted to / Unified in...
  - Matching (ALSC)
  - MRC (All)
  - **Seq2Seq** (All)
  - (M)LM (All)

Token: The wine list is interesting and has good values, but the service is dreadful

Position index: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

Subtask	Target Sequence
<b>AE</b>	1, 2, 12, 12, </s>
<b>OE</b>	4, 4, 7, 8, 14, 14, </s>
<b>ALSC</b>	<u>1</u> , <u>2</u> , POS, </s>
	<u>12</u> , <u>12</u> , POS, </s>
<b>AOE</b>	<u>1</u> , <u>2</u> , 4, 4, 7, 8, </s>
	<u>12</u> , <u>12</u> , 14, 14, </s>
<b>AESC</b>	1, 2, POS, 12, 12, NEG, </s>
<b>Pair</b>	1, 2, 4, 4, 1, 2, 7, 8, 12, 12, 14, 14, </s>
<b>Triplet</b>	1, 2, 4, 4, POS, 1, 2, 7, 8, POS, 12, 12, 14, 14, POS, </s>

# Paradigm Shift in ABSA



# Paradigm Shift in ABSA

- **Traditional Paradigm:**
  - SeqLab (AE, OE, AOE, ...)
  - Class (ALSC...)
- **Shifted to / Unified in...**
  - Matching (ALSC)
  - MRC (All)
  - Seq2Seq (All)
  - **(M)LM (All)**

Consistency  
prompt

Polarity  
prompt

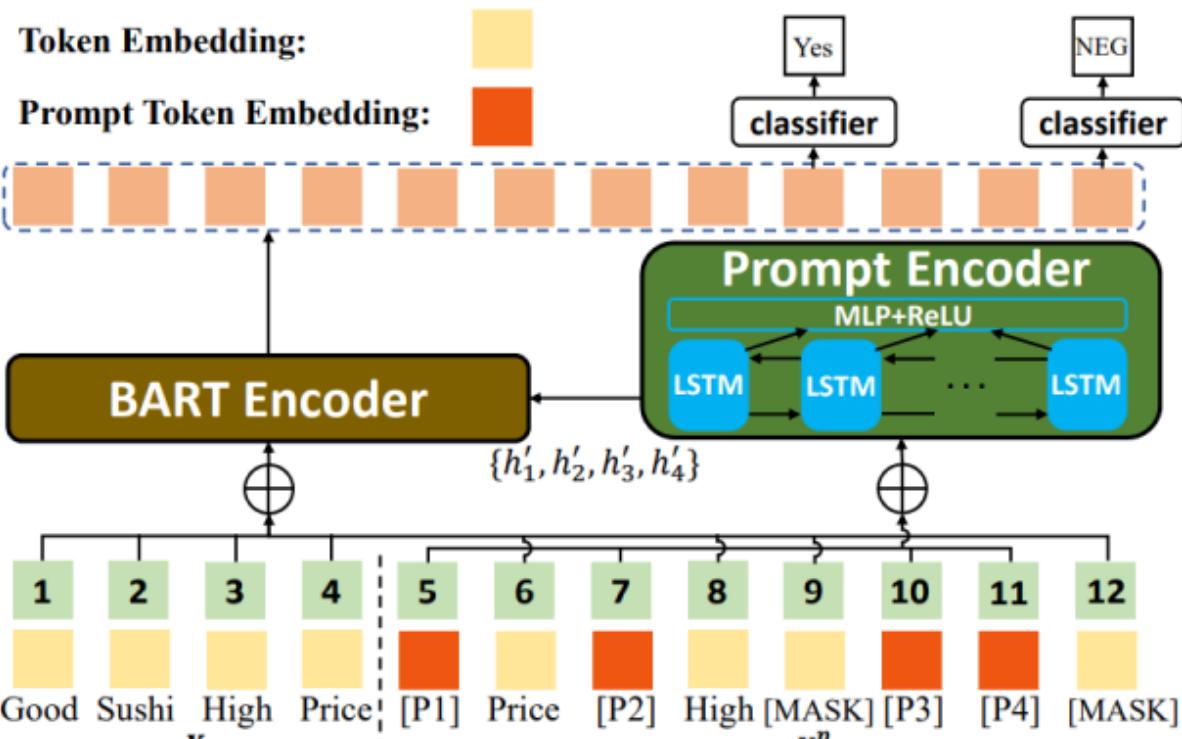
The owners are great fun and the  
beer selection is worth staying for.



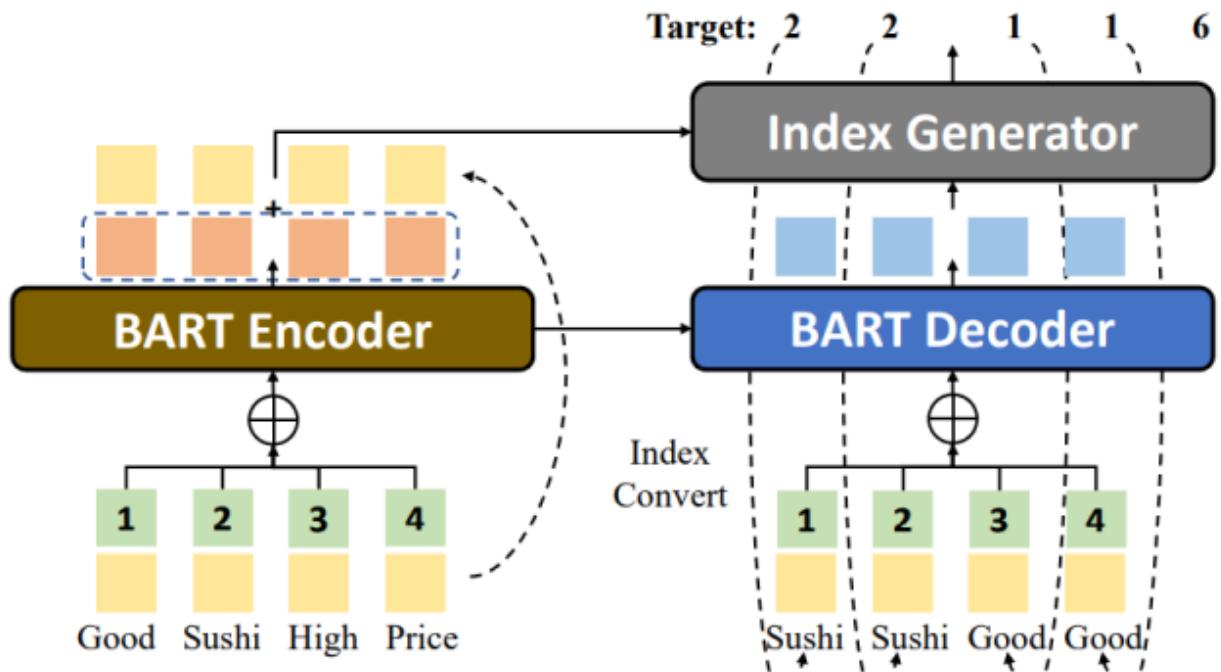
The owners are great fun? [MASK] .

This is [MASK] .

# Paradigm Shift in ABSA



(a) SentiPrompt Tuning



(b) Generation Framework

# Paradigm Shift in Relation Extraction

---

- **Traditional Paradigm:**
  - SeqLab (entity extraction)
  - Class (relation classification)
- **Shifted to / Unified in...**
  - Seq2Seq
  - MRC
  - (M)LM

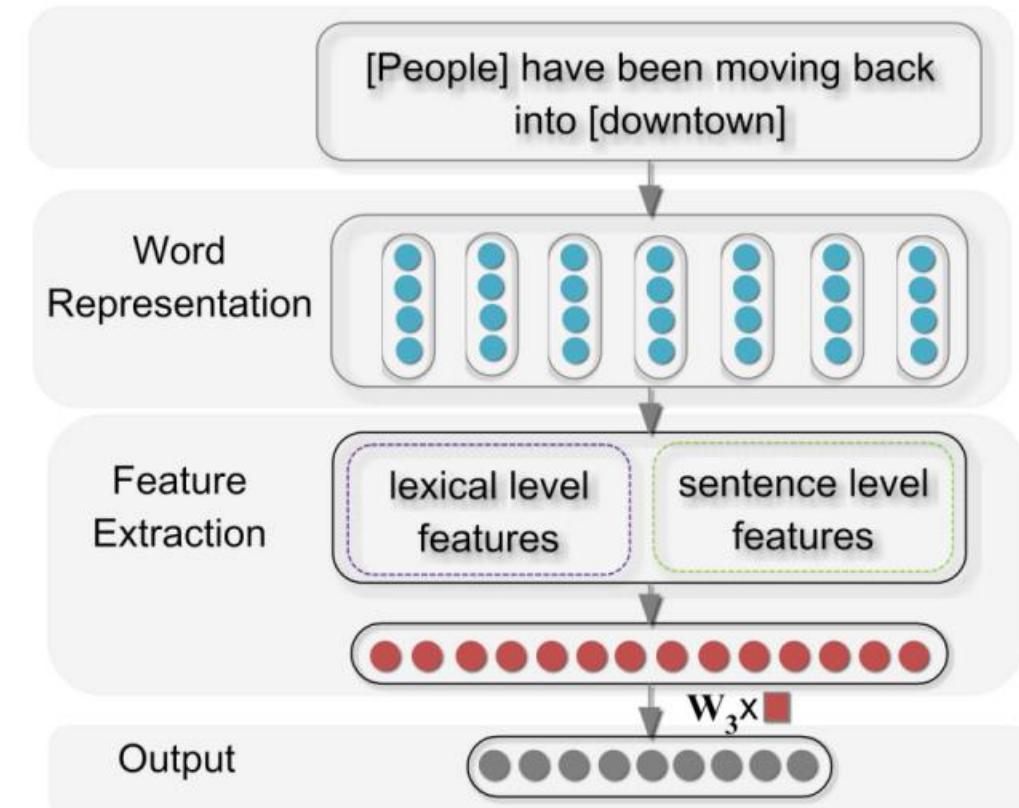
# Paradigm Shift in Relation Extraction

- **Traditional Paradigm:**

- SeqLab (entity extraction)
- **Class** (relation classification)

- **Shifted to / Unified in...**

- Seq2Seq
- MRC
- (M)LM



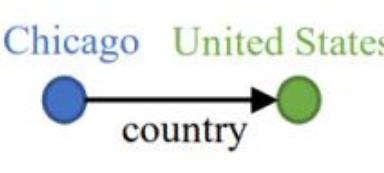
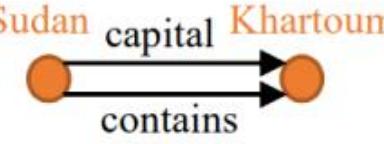
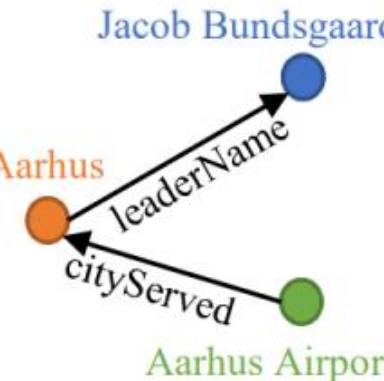
# Paradigm Shift in Relation Extraction

- **Traditional Paradigm:**

- SeqLab (entity extraction)
- Class (relation classification)

- **Shifted to / Unified in...**

- Seq2Seq
- MRC
- (M)LM

Normal	S1: Chicago is located in the United States.  {<Chicago, country, United States>}	
EPO	S2: News of the list's existence unnerved officials in Khartoum, Sudan 's capital.  {<Sudan, capital, Khartoum>, <Sudan, contains, Khartoum>}	
SEO	S3: Aarhus airport serves the city of Aarhus who's leader is Jacob Bundsgaard.  {<Aarhus, leaderName, Jacob Bundsgaard>, <Aarhus Airport, cityServed, Aarhus>}	

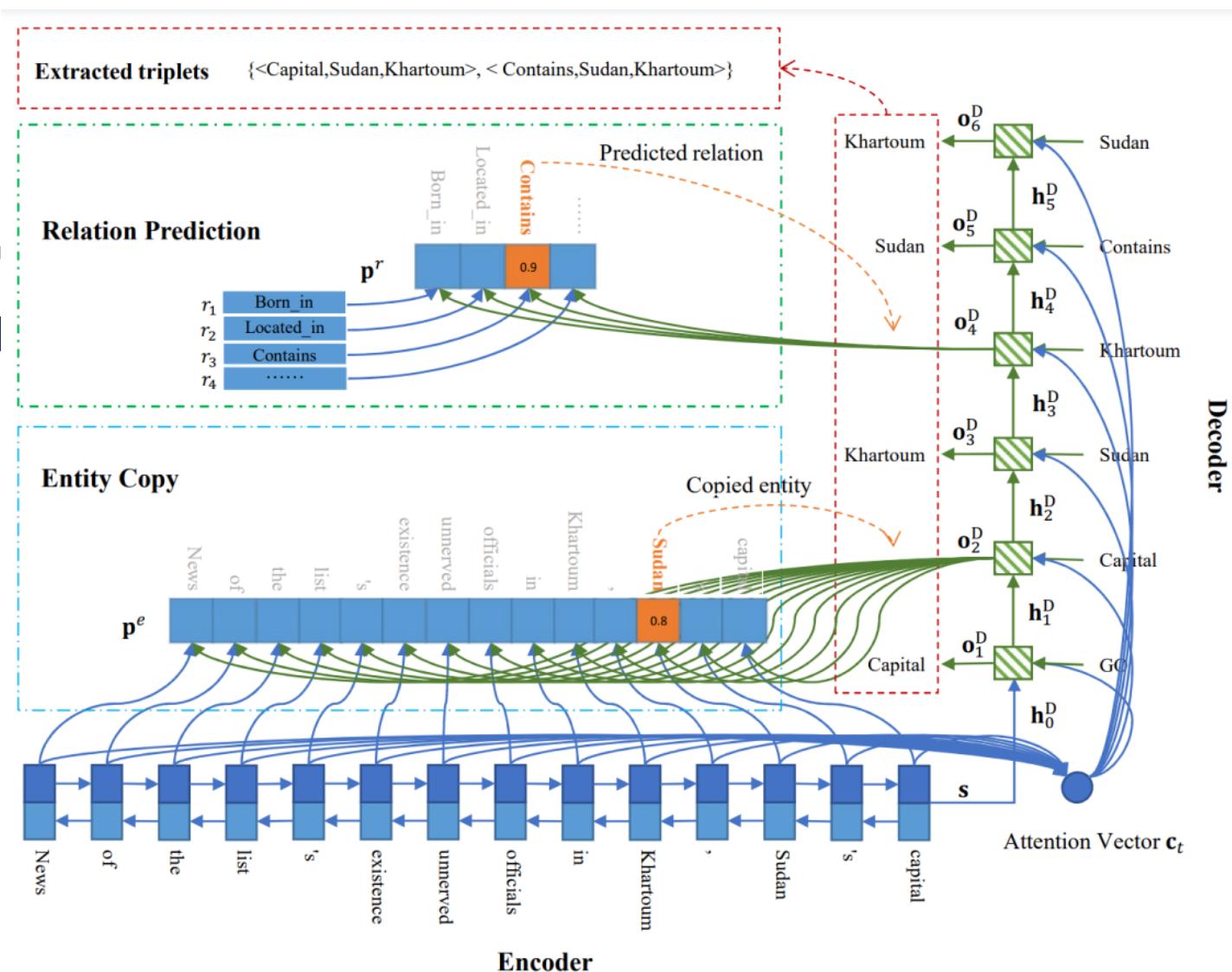
# Paradigm Shift in Relation Extraction

## • Traditional Paradigm

- SeqLab (entity extraction)
- Class (relation classification)

## • Shifted to / Unified into

- Seq2Seq
- MRC
- (M)LM



# Paradigm Shift in Relation Extraction

---

- **Traditional Paradigm:**

- SeqLab (entity extraction)
- Class (relation classification)

- **Shifted to / Unified in...**

- Seq2Seq
- MRC (entity prediction)
- (M)LM

Relation	Question	Sentence & Answers
<i>educated_at</i>	What is <b>Albert Einstein</b> 's alma mater?	<b>Albert Einstein</b> was awarded a PhD by the <u>University of Zürich</u> , with his dissertation titled...
<i>occupation</i>	What did <b>Steve Jobs</b> do for a living?	<b>Steve Jobs</b> was an American <u>businessman</u> , <u>inventor</u> , and <u>industrial designer</u> .
<i>spouse</i>	Who is <b>Angela Merkel</b> married to?	<b>Angela Merkel</b> 's second and current husband is quantum chemist and professor <b>Joachim Sauer</b> , who has largely...

# Paradigm Shift in Relation Extraction

---

- **Traditional Paradigm:**

- SeqLab (entity extraction)
- Class (relation classification)

- **Shifted to / Unified in...**

- Seq2Seq
- MRC (triplet extraction)
- (M)LM

Formulate RESUME  
dataset as Multi-turn QA:

---

Q1 Person:	who is mentioned in the text?	A: $e_1$
Q2 Company:	which companies did $e_1$ work for?	A: $e_2$
Q3 Position:	what was $e_1$ 's position in $e_2$ ?	A: $e_3$
Q4 Time:	During which period did $e_1$ work for $e_2$ as $e_3$	A: $e_4$

---

# Paradigm Shift in Relation Extraction

---

- **Traditional Paradigm:**

- SeqLab (entity extraction)
- Class (relation classification)

- **Shifted to / Unified in...**

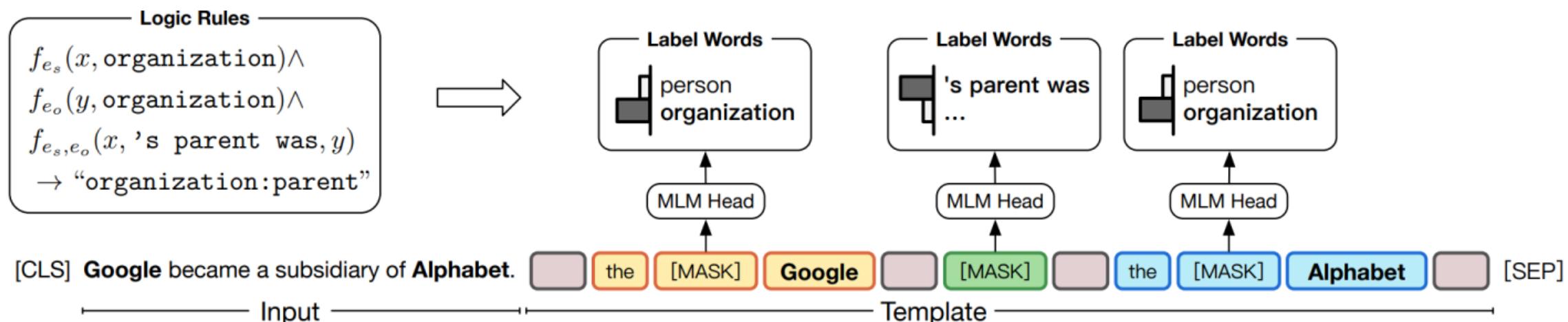
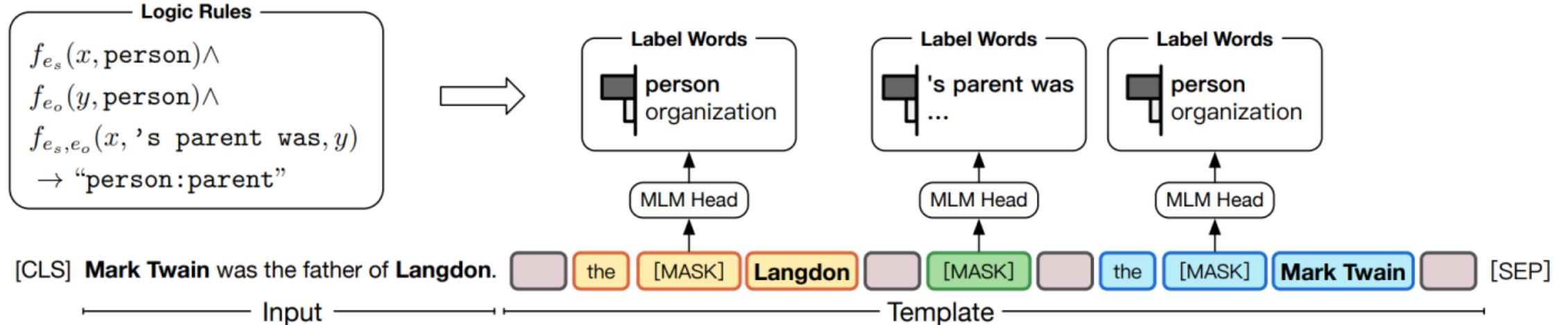
- Seq2Seq
- MRC (triplet extraction)
- (M)LM

*Mark Twain was the father of Langdon.*



[p] the person *Langdon* [p] 's parent was  
[p] the person *Mark Twain* [p].

# Paradigm Shift in Relation Extraction



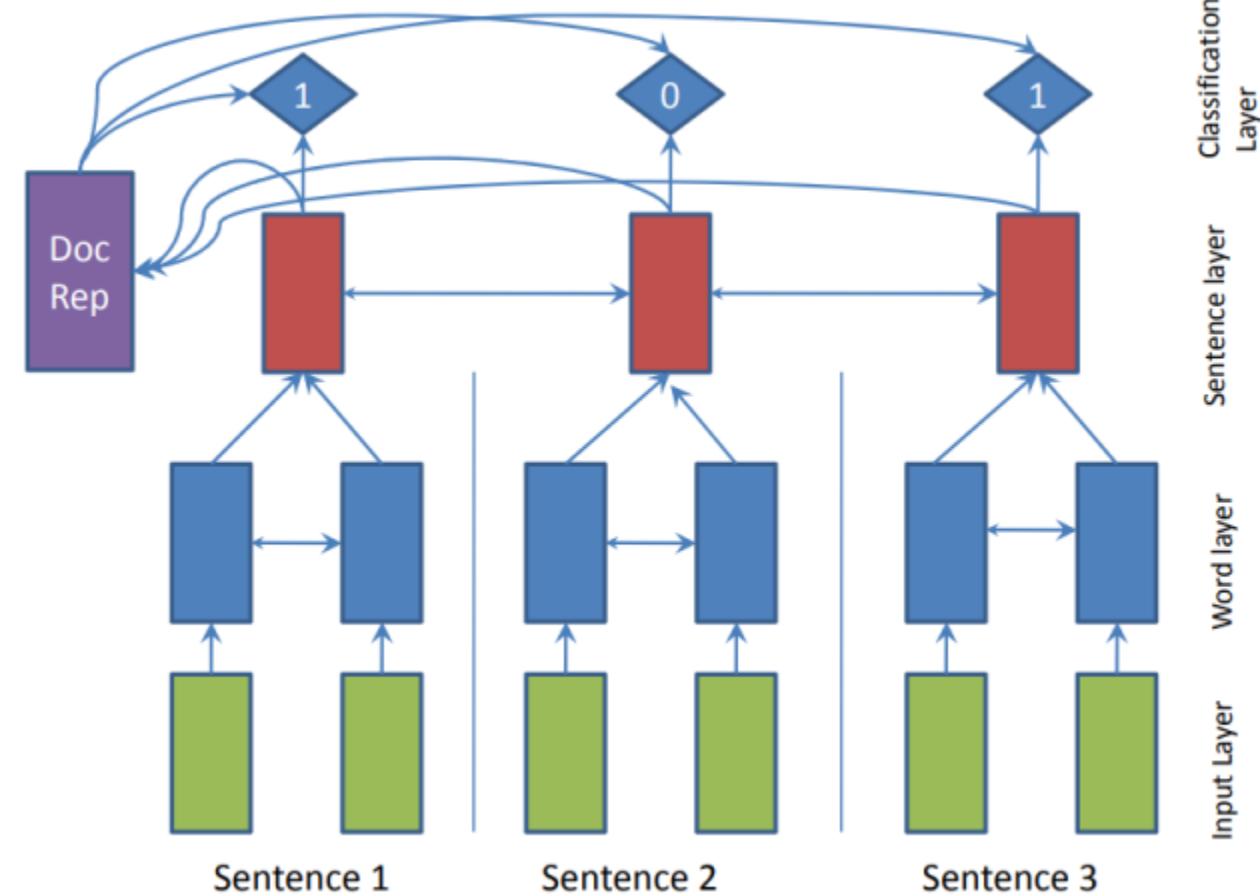
# Paradigm Shift in Text Summarization

---

- **Traditional Paradigm:**
  - SeqLab (extractive)
  - Seq2Seq (abstractive)
- **Shifted to / Unified in...**
  - Matching (extractive)
  - (M)LM (abstractive)

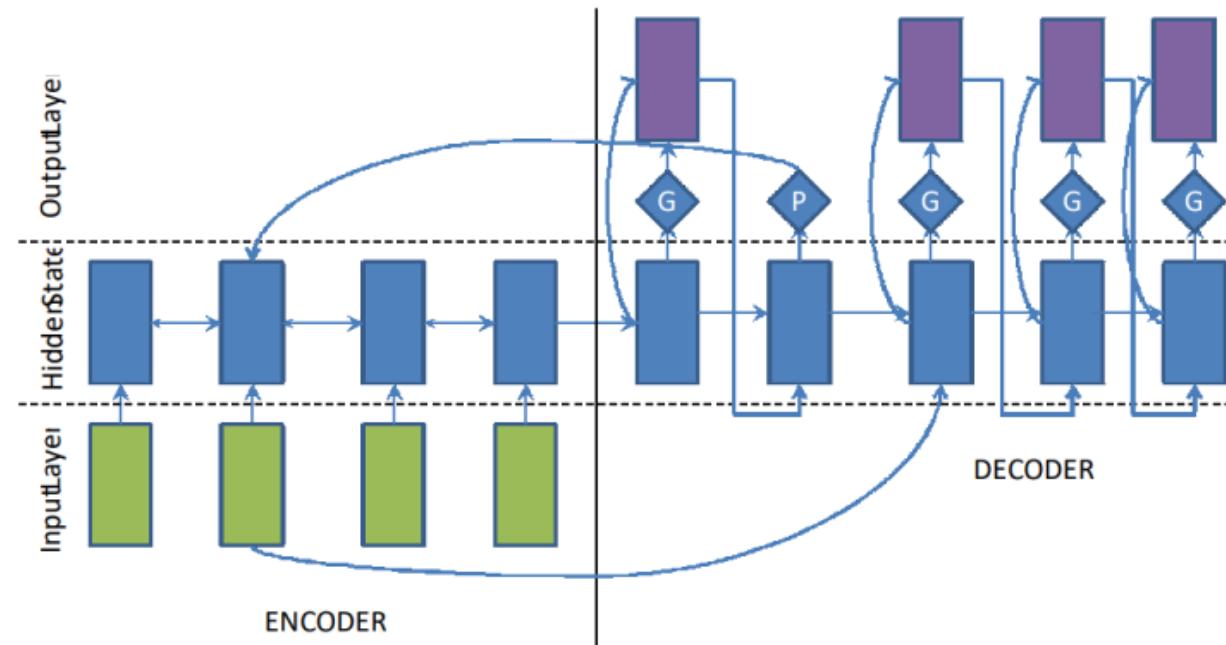
# Paradigm Shift in Text Summarization

- **Traditional Paradigm:**
  - SeqLab (extractive)
  - Seq2Seq (abstractive)
- **Shifted to / Unified in...**
  - Matching (extractive)
  - (M)LM (abstractive)



# Paradigm Shift in Text Summarization

- Traditional Paradigm:
  - SeqLab (extractive)
  - Seq2Seq (abstractive)
- Shifted to / Unified in...
  - Matching (extractive)
  - (M)LM (abstractive)



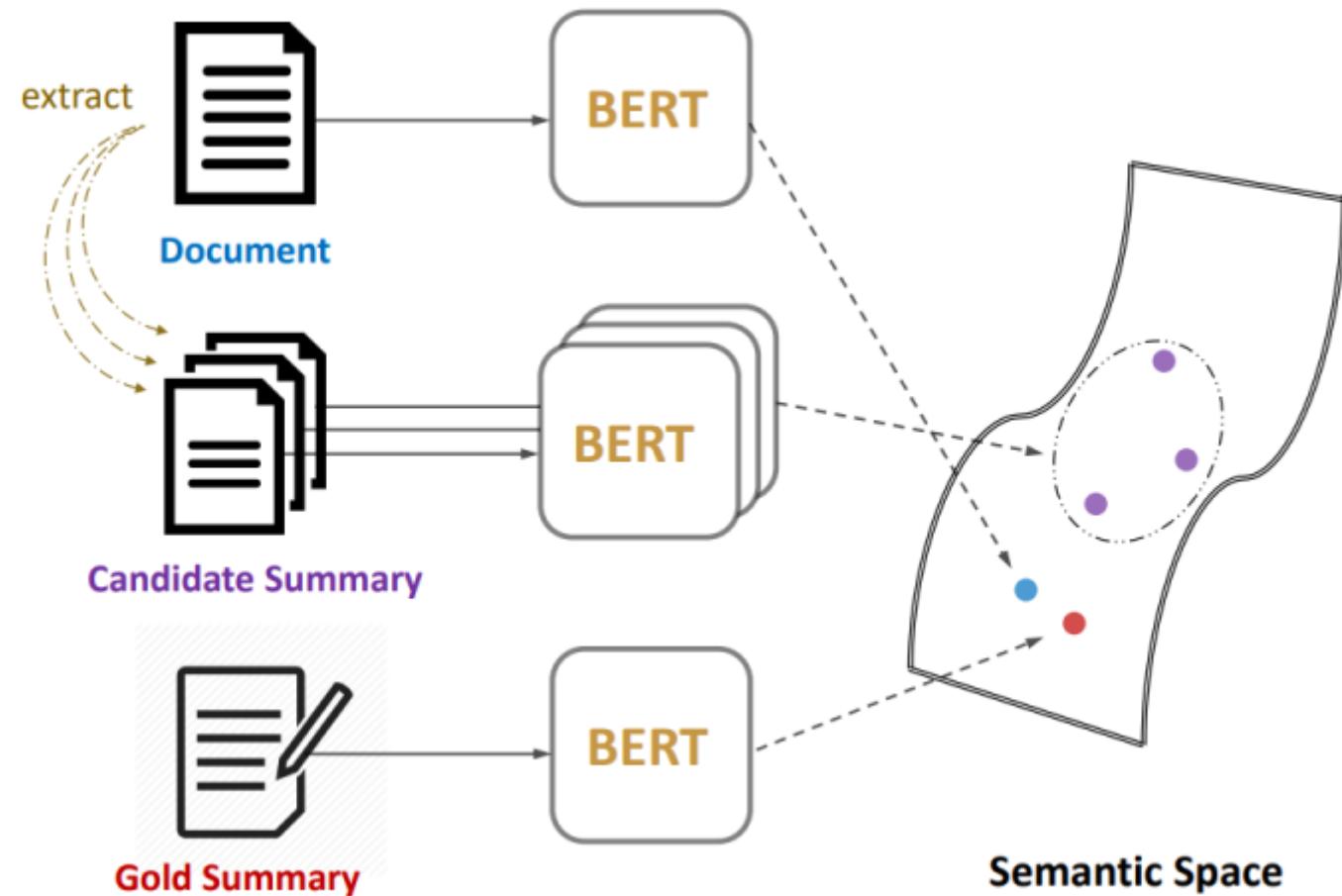
# Paradigm Shift in Text Summarization

- **Traditional Paradigm:**

- SeqLab (extractive)
- Seq2Seq (abstractive)

- **Shifted to / Unified in...**

- **Matching** (extractive)
- (M)LM (abstractive)



# Paradigm Shift in Text Summarization

- **Traditional Paradigm:**

- SeqLab (extractive)
- Seq2Seq (abstractive)

- **Shifted to / Unified in...**

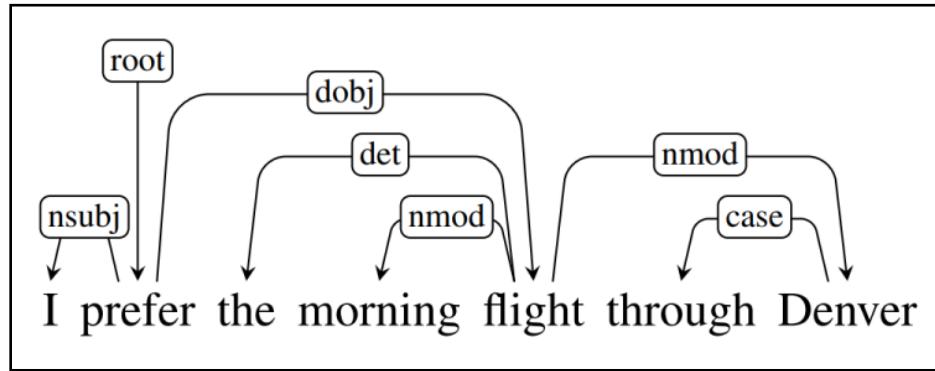
- Matching (extractive)
- **(M)LM** (abstractive)

```
<!DOCTYPE html>
<html>
  <title> <mask>12 </title>
  <body>
    ~ south korea on monday announced sweeping
      tax reforms , including income and
      corporate tax cuts to boost growth by
      stimulating sluggish private
      consumption and business investment .
  </body>
</html>
```

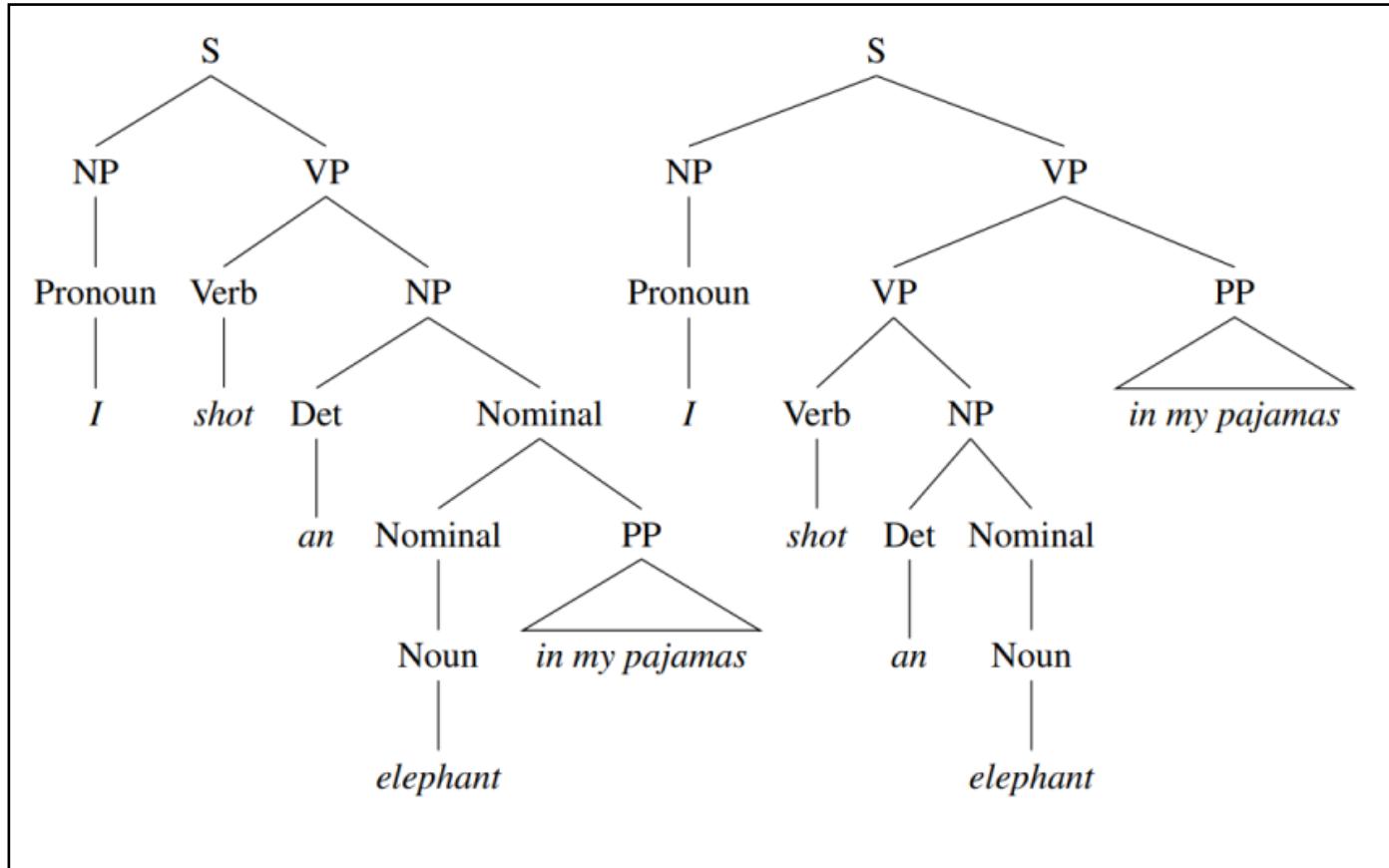


```
<!DOCTYPE html>
<html>
  <title> ~ South Korea Announces Tax Reforms To
      Boost Economic Growth ~ </title>
  <body>
    ~ south korea on monday announced sweeping
      tax reforms...
  </body>
</html>
```

# Paradigm Shift in Parsing



Dependency Parsing



Semantic Parsing

*which country had the highest carbon emissions last year*

```
SELECT      country.name
FROM        country, co2_emissions
WHERE       country.id = co2_emissions.country_id
AND         co2_emissions.year = 2014
ORDER BY    co2_emissions.volume DESC
LIMIT       1;
```

Constituency Parsing

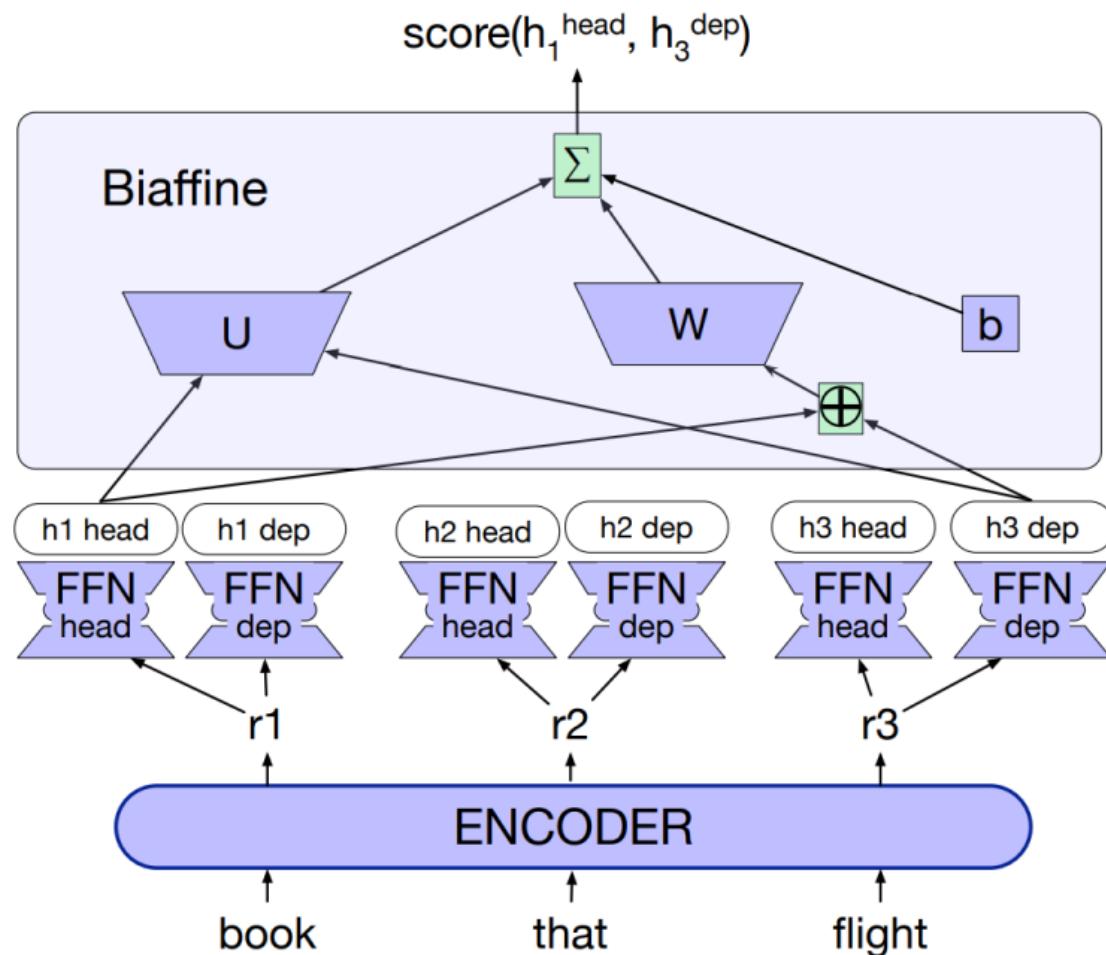
# Paradigm Shift in Parsing

---

- **Traditional Paradigm:**
  - Class (graph-based)
  - Seq2ASeq (transition-based)
- **Shifted to / Unified in...**
  - SeqLab
  - Seq2Seq
  - (M)LM
  - MRC

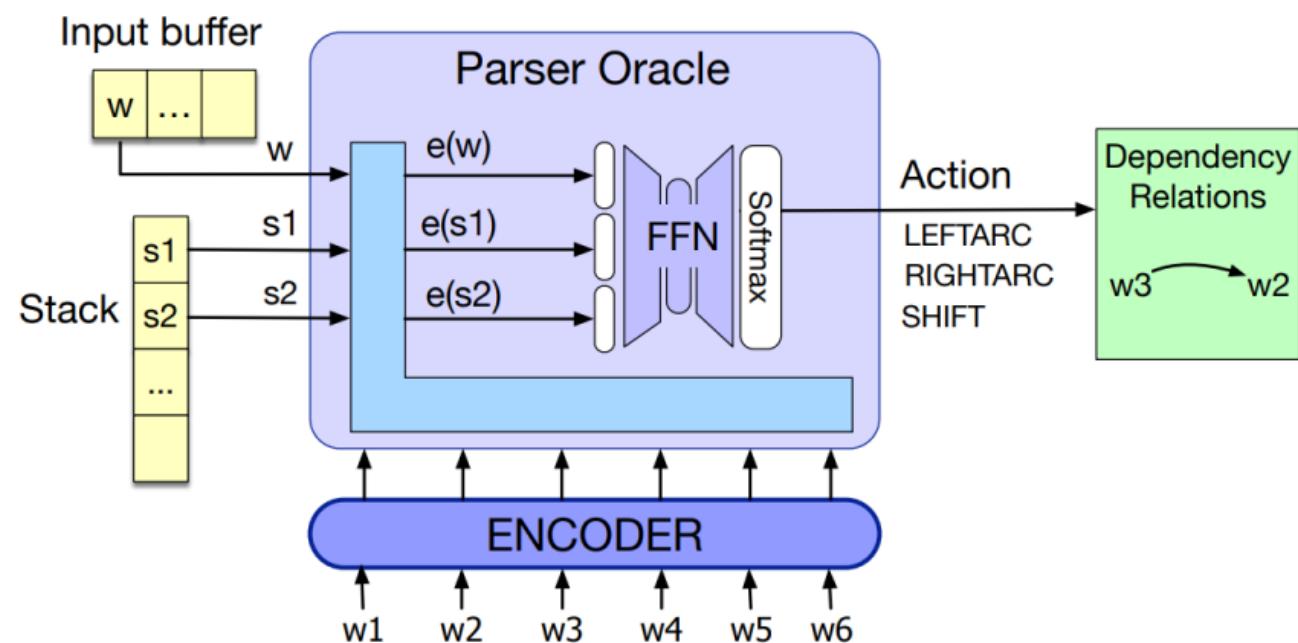
# Paradigm Shift in Parsing

- Traditional Paradigm:
  - Class (graph-based)
  - Seq2ASeq (transition-based)
- Shifted to / Unified in...
  - SeqLab
  - Seq2Seq
  - (M)LM
  - MRC



# Paradigm Shift in Parsing

- Traditional Paradigm:
  - Class (graph-based)
  - Seq2ASeq (transition-based)
- Shifted to / Unified in...
  - SeqLab
  - Seq2Seq
  - (M)LM
  - MRC



# Paradigm Shift in Parsing

---

- **Traditional Paradigm:**

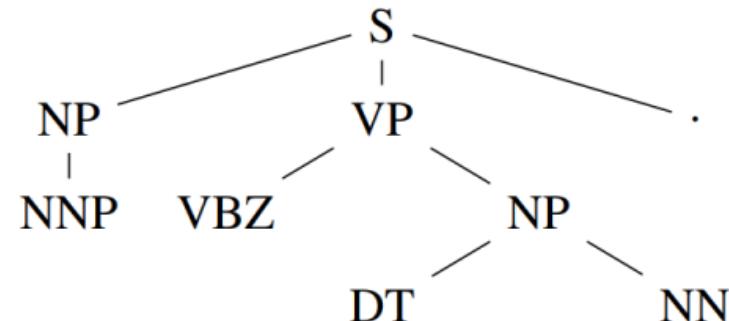
- Class (graph-based)
- Seq2ASeq (transition-based)

- **Shifted to / Unified in...**

- SeqLab
- Seq2Seq
- (M)LM
- MRC

John has a dog . →

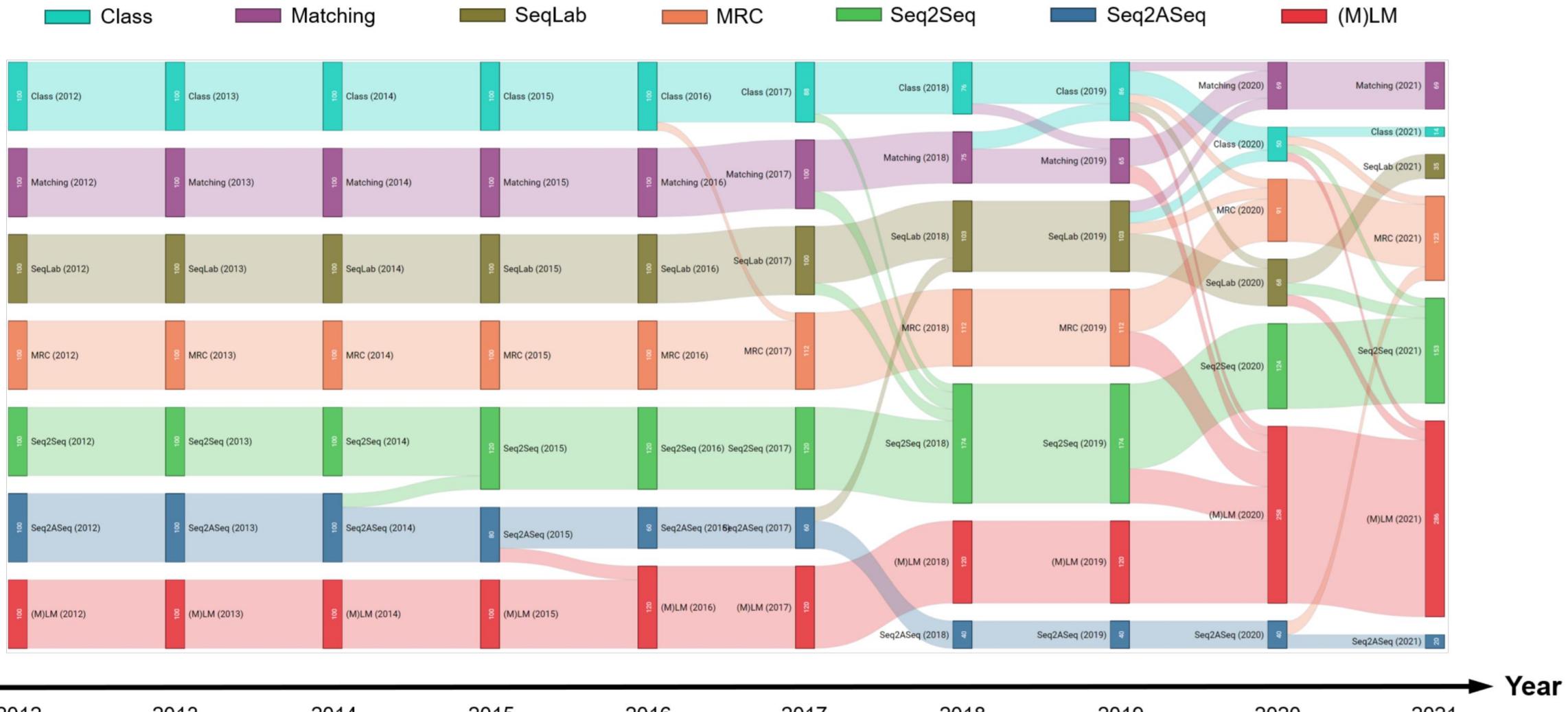
Linearize a parsing tree:



John has a dog . →

(S (NP NNP )<sub>NP</sub> (VP VBZ (NP DT NN )<sub>NP</sub> )<sub>VP</sub> . )<sub>S</sub>

# Trends of Paradigm Shift



# Trends of Paradigm Shift

---

- **More General and Flexible Paradigms are Dominating**
  - Traditional: Class, SeqLab, Seq2ASeq
  - General: Matching, MRC, Seq2Seq, (M)LM
- **The Impact of Pre-trained LMs**
  - Formulate a NLP task as one that PLMs are good at!

# Outline

---

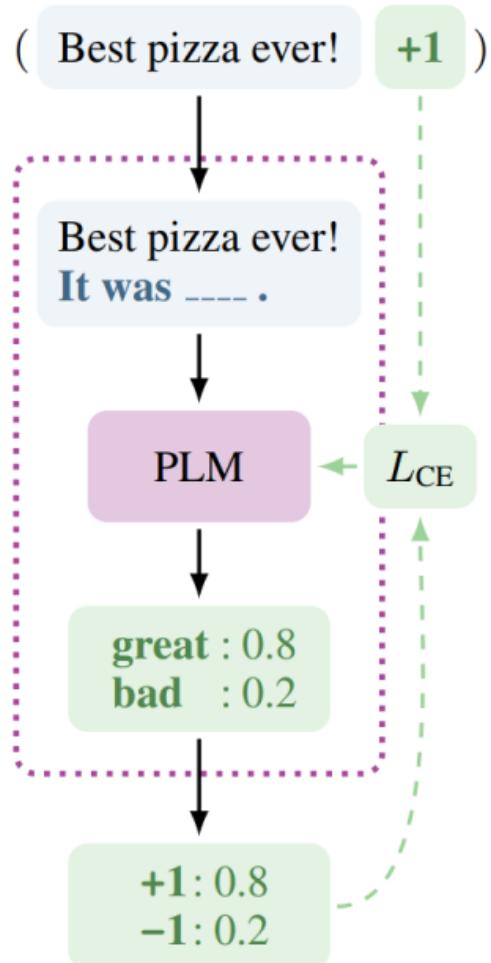
- Introduction
- The Seven Paradigms in NLP
- Paradigm Shift in NLP Tasks
- **Potential Unified Paradigms**
- Conclusion

# Potential Unified Paradigms

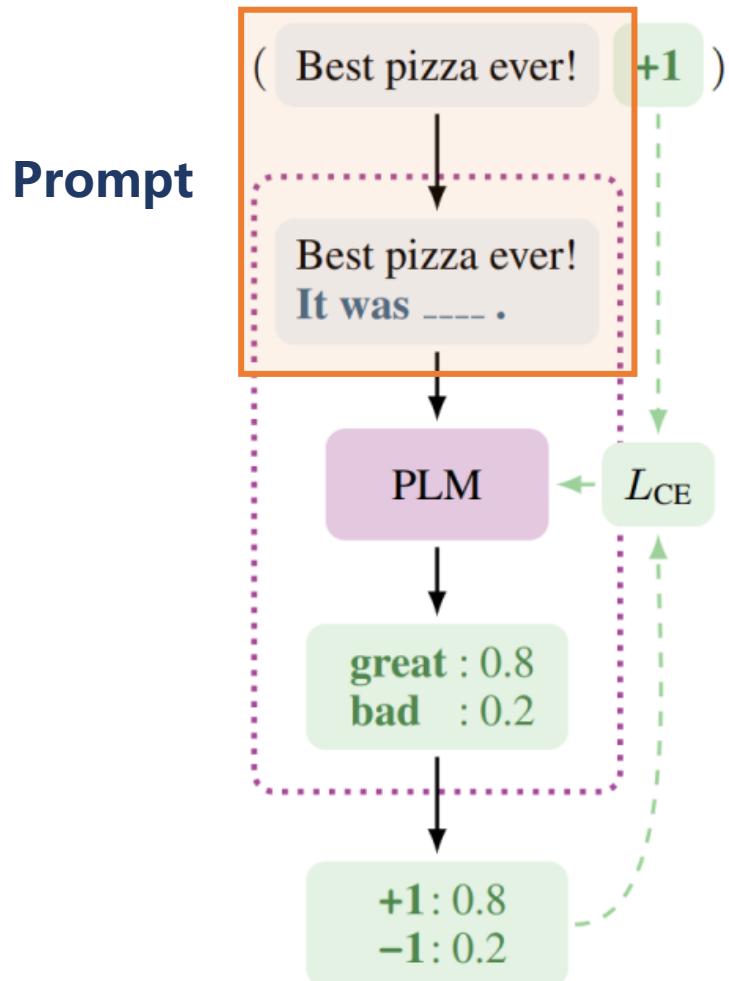
---

- (M)LM
- Matching
- MRC
- Seq2Seq

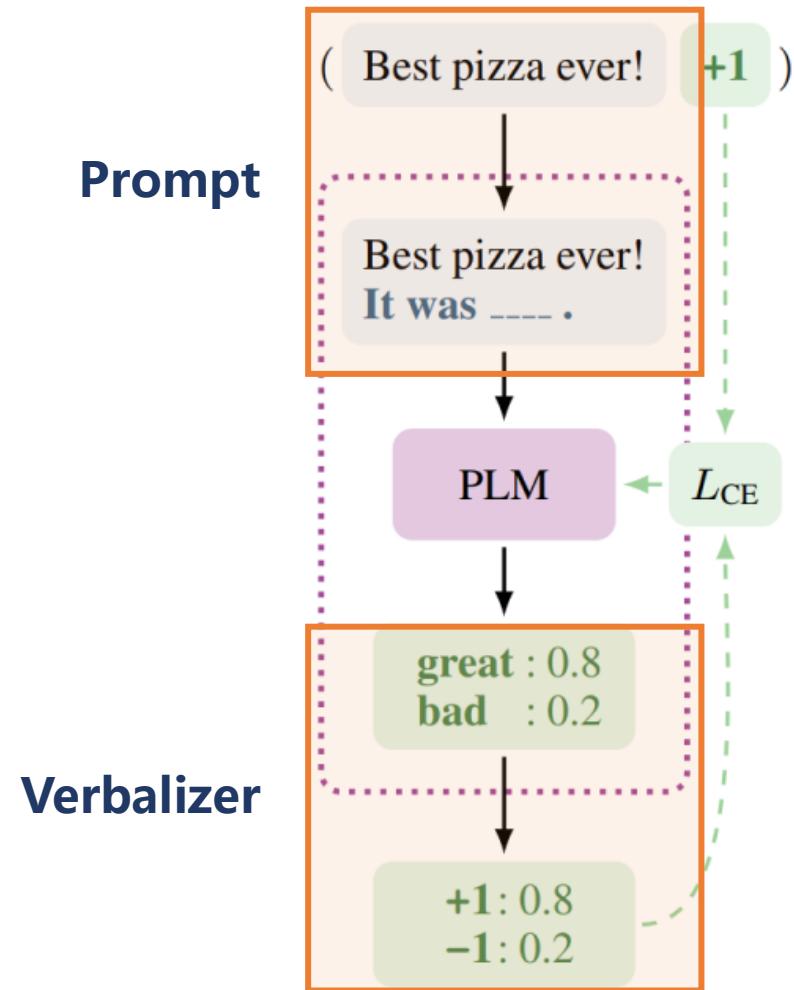
# (M)LM



# (M)LM



# (M)LM



# (M)LM

---

- **Prompt**

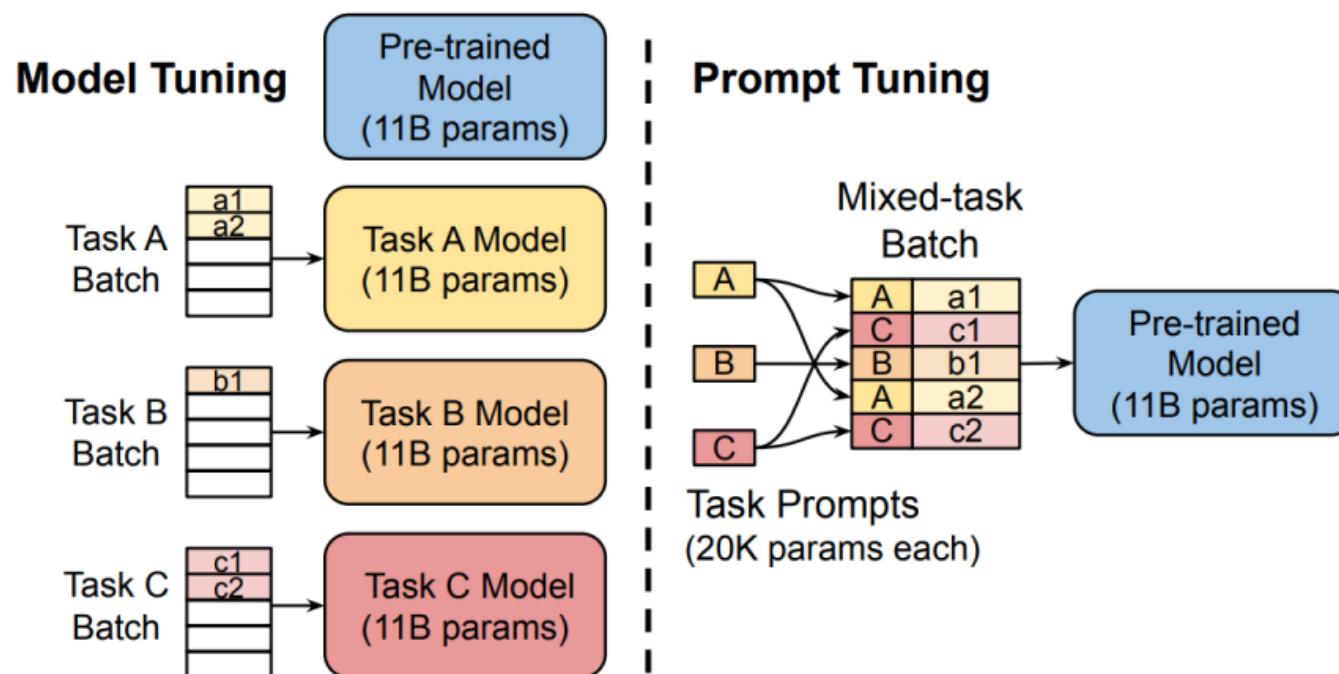
- Manually designed
- Mined from corpora
- Generated by paraphrasing
- Generated by another PLM
- Learned by gradient search/descent

- **Verbalizer**

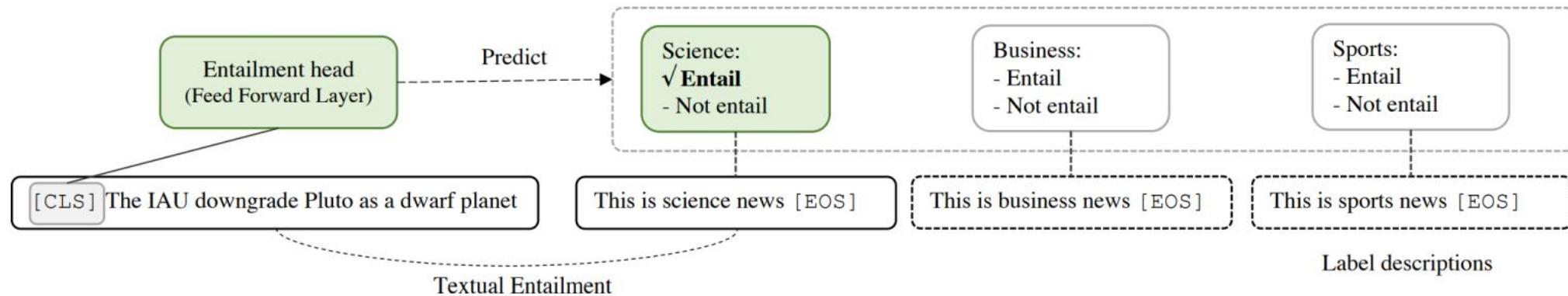
- Manually designed
- Automatically searched
- Constructed and refined with KB

- **Parameter-Efficient Tuning**

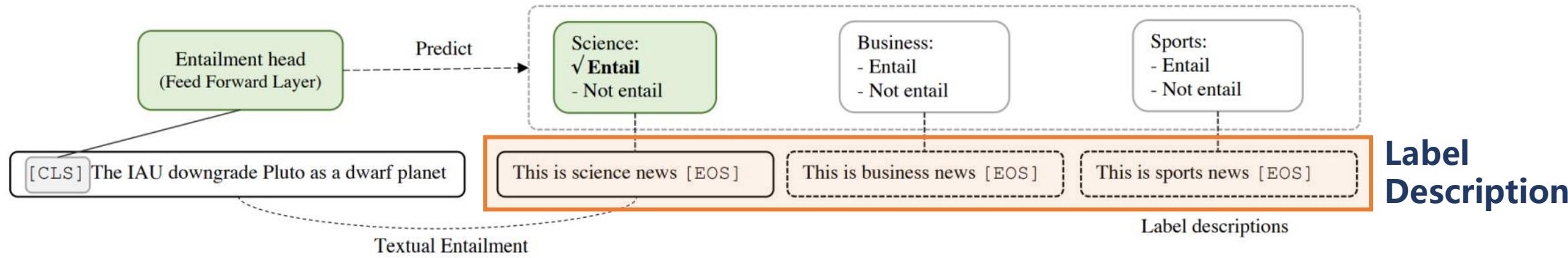
- Only tuning prompts can match the performance of fine-tuning
- Mixed-task inference



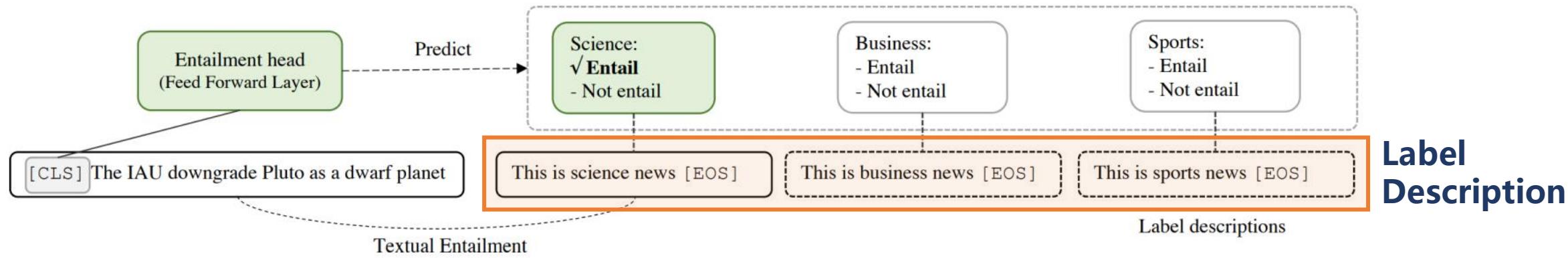
# Matching



# Matching



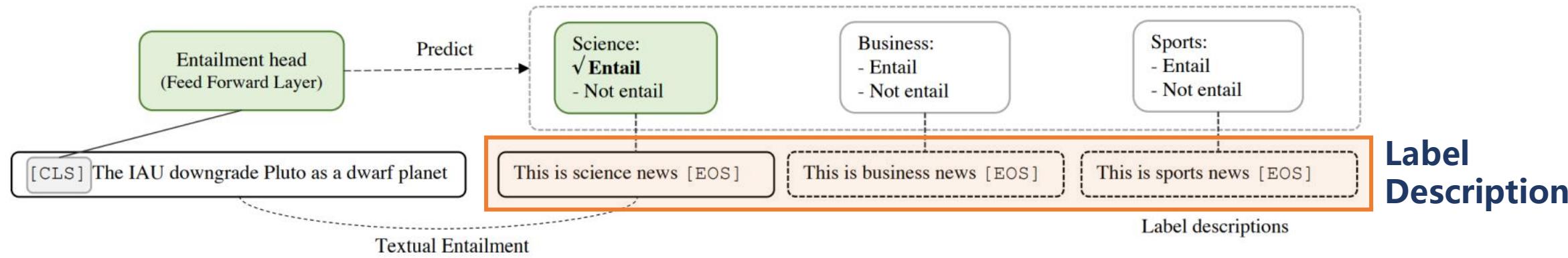
# Matching



- **Label Description**

- Manually designed (can be the same as prompt)
- Generated by reinforcement learning ([Chai et al.](#))

# Matching



- **Label Description**

- Manually designed (can be the same as prompt)
- Generated by reinforcement learning ([Chai et al.](#))

- **The Entailment Model**

- Fine-tuning a PLM on MNLI

# (M)LM or Matching?

---

- **(M)LM**

- [MASK] -> MLM head, instead of randomly initialized classifier
- Require modifications of input (prompt) and output (verbalizer)
- Pre-trained LMs can be directly used (even zero-shot)
- Compatible with generation tasks

- **Matching**

- [CLS] -> MNLI/NSP head, instead of randomly initialized classifier
- Only label descriptions are required (less engineering!)
- Contrastive learning can be applied
- Suffer from domain adaption (due to the requirement of supervised data)
- Only support NLU tasks

## • A Highly General Paradigm

- A task can be solved as a MRC one as long as its input can be formulated as [context, question, answer].

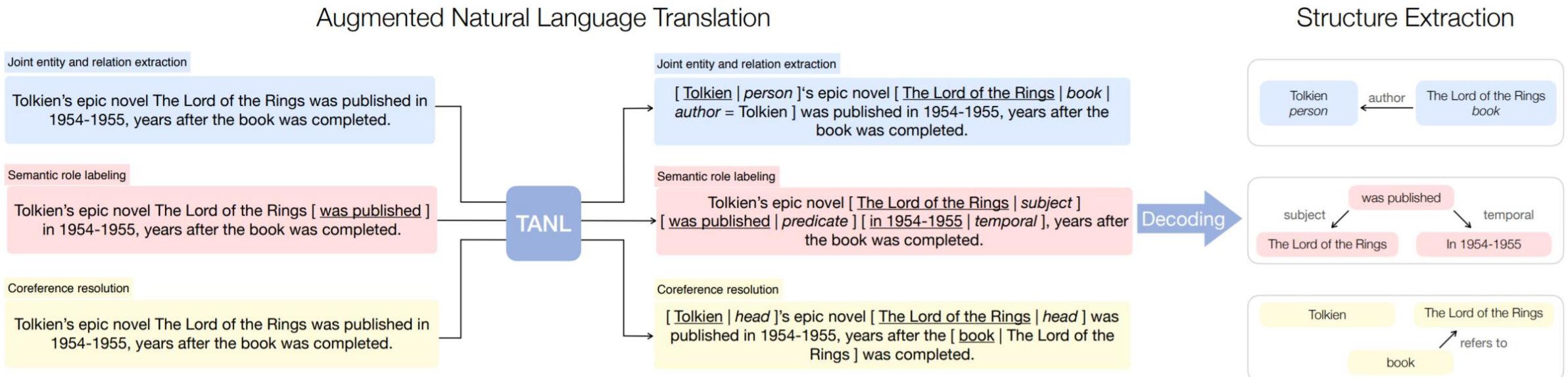
### Examples

Question	Context	Answer	Question	Context	Answer
What is a major importance of Southern California in relation to California and the US?	...Southern California is a <b>major economic center</b> for the state of California and the US....	major economic center	What has something experienced?	Areas of the Baltic that have experienced <b>eutrophication</b> .	eutrophication
What is the translation from English to German?	Most of the planet is ocean water.	Der Großteil der Erde ist Meerwasser	Who is the illustrator of Cycle of the Werewolf?	Cycle of the Werewolf is a short novel by Stephen King, featuring illustrations by comic book artist <b>Bernie Wrightson</b> .	Bernie Wrightson
What is the summary?	Harry Potter star Daniel Radcliffe gains access to a reported £320 million fortune...	Harry Potter star Daniel Radcliffe gets £320M fortune...	What is the change in dialogue state?	Are there any Eritrean restaurants in town?	food: Eritrean
Hypothesis: Product and geography are what make cream skimming work. <b>Entailment</b> , neutral, or contradiction?	Premise: Conceptually cream skimming has two basic dimensions – product and geography.	Entailment	What is the translation from English to SQL?	The <b>table</b> has column names... Tell me what the <b>notes</b> are for <b>South Australia</b>	SELECT notes from table WHERE 'Current Slogan' = 'South Australia'
Is this sentence <b>positive</b> or negative?	A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.	positive	Who had given help? <b>Susan</b> or <b>Joan</b> ?	Joan made sure to thank Susan for all the help she had given.	Susan

- **A Highly General Paradigm**
  - A task can be solved as a MRC one as long as its input can be formulated as [context, question, answer].
- **MRC has been applied to many tasks...**
  - entity-relation extraction, coreference resolution, entity linking, dependency parsing, dialog state tracking, event extraction, aspect-based sentiment analysis...
- **How to Utilize the Power of Pre-Training?**

# Seq2Seq

- A Highly General and Flexible Paradigm
  - Suitable for complicated tasks (e.g. structured prediction, discontinuous NER, triplet extraction, etc.)



# Seq2Seq

---

- **A Highly General and Flexible Paradigm**
  - Suitable for complicated tasks (e.g. structured prediction, discontinuous NER, triplet extraction, etc.)
- **Powered by Pre-training**
  - MASS, BART, T5...
- **Compatible with (M)LM and MRC**
- **However...**
  - High Latency at Inference Time (Non-autoregressive? Early exiting?)

# Outline

---

- Introduction
- The Seven Paradigms in NLP
- Paradigm Shift in NLP Tasks
- Potential Unified Paradigms
- Conclusion

# Conclusion

---

- **(M)LM, aka prompt-based tuning, is exploding in popularity...**
  - Does the power come from the pre-trained MLM head?
  - What if the classification head can be replaced with the NSP head, entailment head, or other classification/generation heads?
  - What if pre-training can also boost other paradigms?
- **More attention is needed on other promising paradigms**
  - **Matching:** less engineering, benefit from supervised data and contrastive learning
  - **MRC:** general, interpretable
  - **Seq2Seq:** compatibility, flexible to handle very complicated tasks



---

# Thank You!

Any question or suggestion is welcome!

[txsun19@fudan.edu.cn](mailto:txsun19@fudan.edu.cn)



<https://arxiv.org/abs/2109.12575>



<https://txsun1997.github.io/nlp-paradigm-shift/>