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

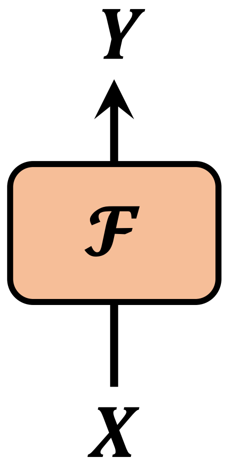
What is Paradigm?

- **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?

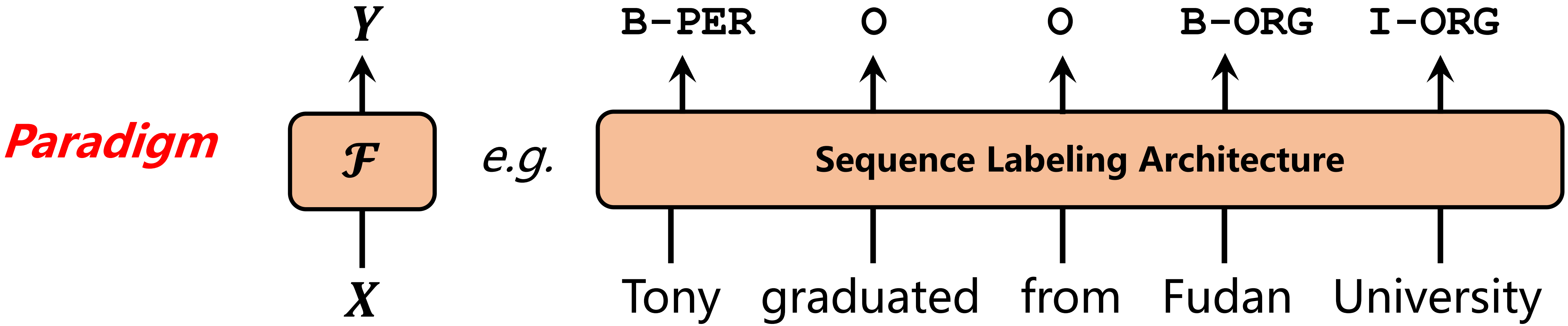
- **Definition from Wikipedia**
  - In science and philosophy, a **paradigm** is a distinct set of **concepts** or **thought patterns**, including theories, research methods, postulates, and standards for what constitutes legitimate contributions to a field.
- **Definition in the context of NLP**
  - *Paradigm is the general framework to model a class of tasks*

Paradigm



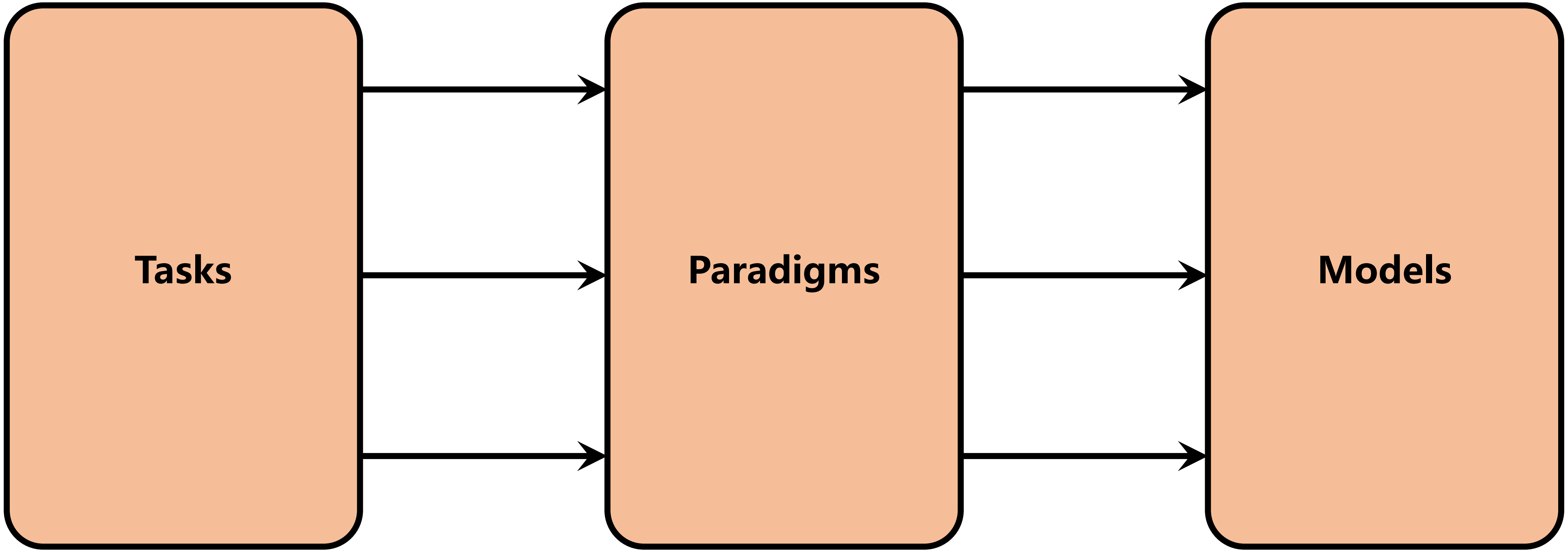
What is Paradigm?

- **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*



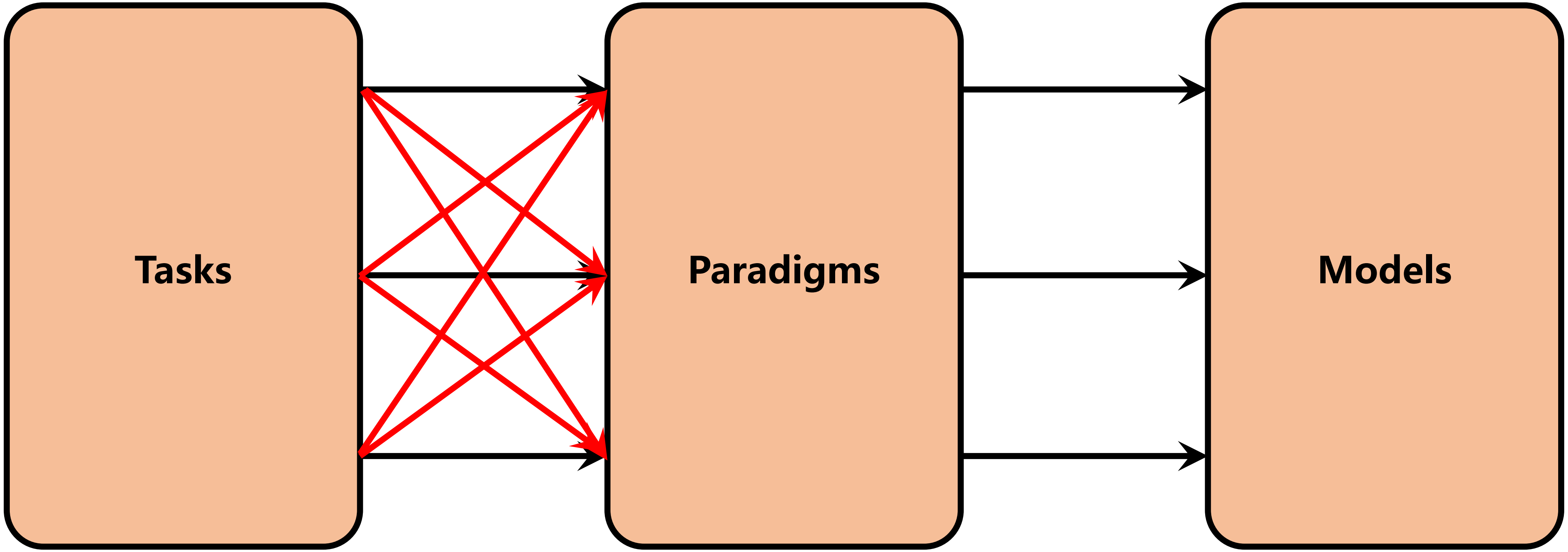
Paradigms, Tasks, and Models

- **A Rough Illustration**



Paradigms, Tasks, and Models

- **A Rough Illustration**

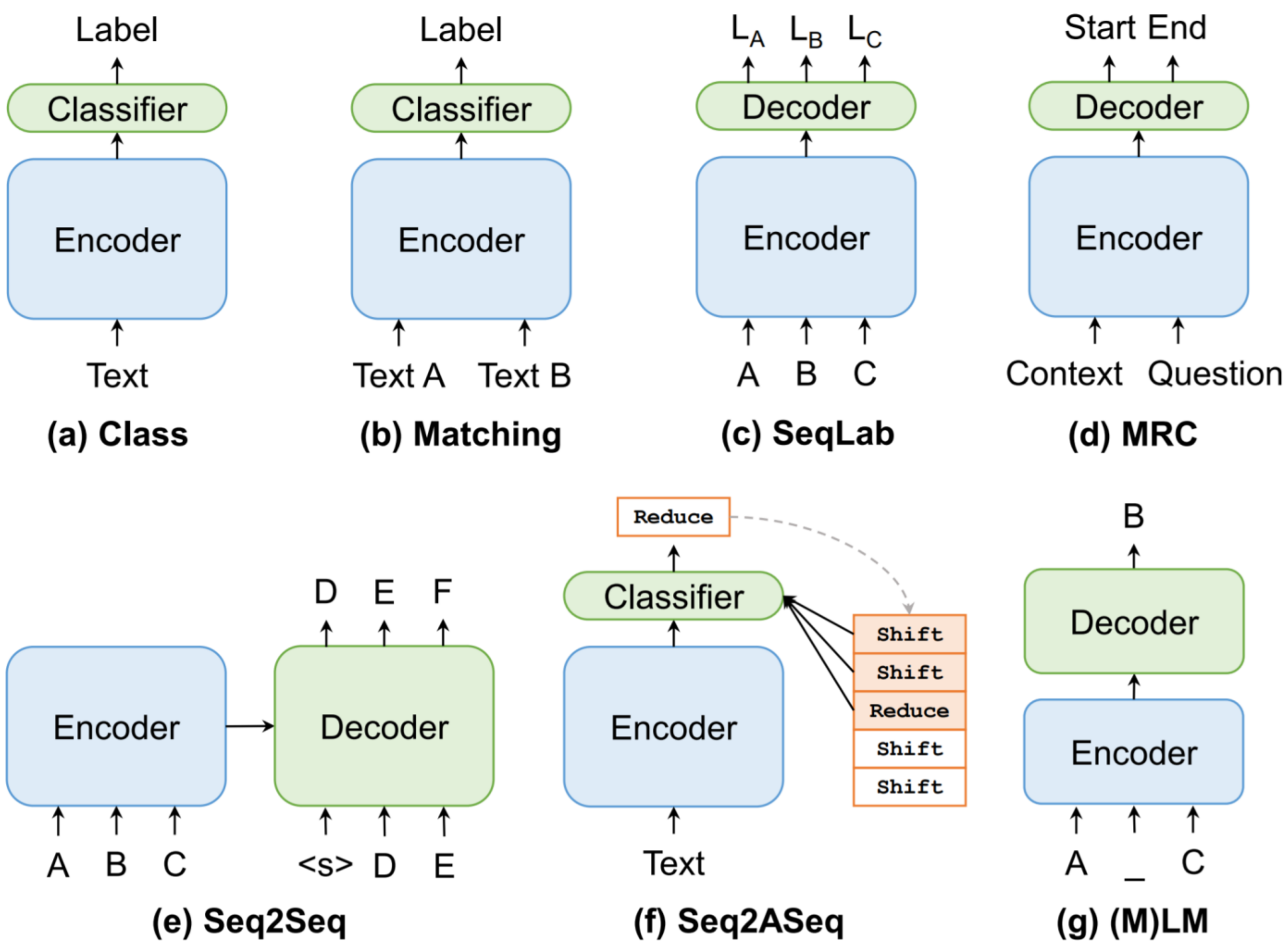


Outline

- 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

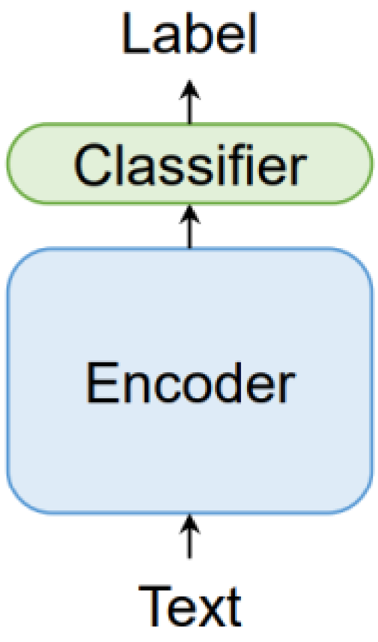


Classification (Class)

- **Paradigm**

$$\mathcal{Y} = \textsc{Cls}(\textsc{Enc}(\mathcal{X})).$$

- **Model**
  - $\textsc{Enc}(\cdot)$: CNN, RNN, Transformers…
  - $\textsc{Cls}(\cdot)$: (max/average/attention) pooling + MLP
- **Tasks**
  - Sentiment Analysis
  - Spam Detection
  - …

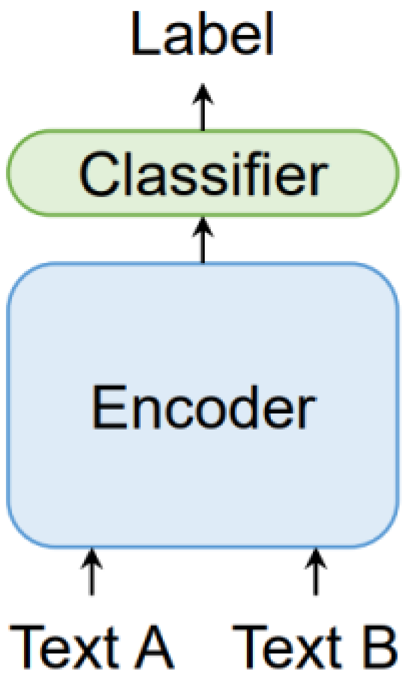


Matching

- **Paradigm**

$$\mathcal{Y} = \textsc{Cls}(\textsc{Enc}(\mathcal{X}_a, \mathcal{X}_b))$$

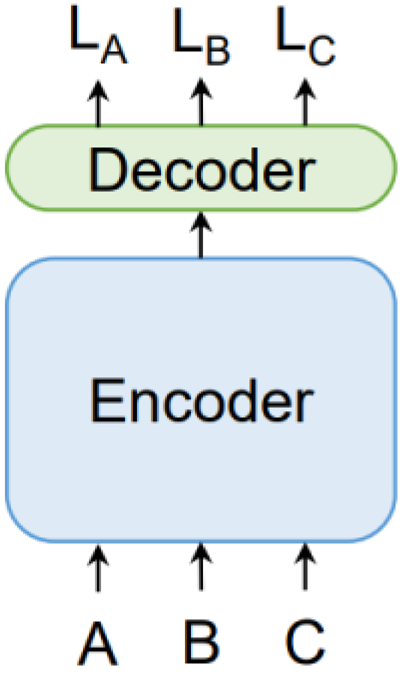
- **Model**
  - $\textsc{Enc}(\cdot)$: encode the two texts separately or jointly
  - $\textsc{Cls}(\cdot)$: capture the interaction, and then prediction
- **Tasks**
  - Natural Language Inference
  - Similarity Regression
  - ...



Sequence Labeling (SeqLab)

- **Paradigm**

$$y_1, \cdots, y_n = \textsc{Dec}(\textsc{Enc}(x_1, \cdots, x_n))$$



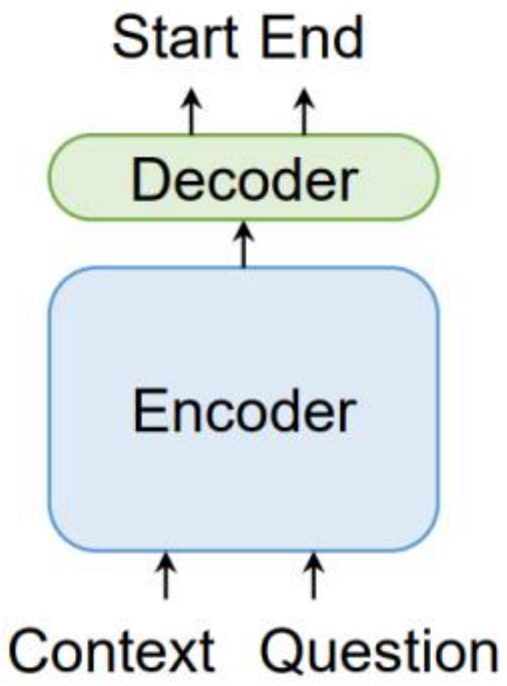
- **Model**
  - $\textsc{Enc}(\cdot)$: sequence model (RNN, Transformers…)
  - $\textsc{Dec}(\cdot)$: conditional random fields (CRF)
- **Tasks**
  - Named Entity Recognition (NER)
  - Part-Of-Speech Tagging
  - …

Machine Reading Comprehension (MRC)

- **Paradigm**

$$y_k \cdots y_{k+l} = \textsc{Dec}(\textsc{Enc}(\mathcal{X}_p, \mathcal{X}_q))$$

- **Model**
  - $\textsc{Enc}(\cdot)$: CNN, RNN, Transformers...
  - $\textsc{Dec}(\cdot)$: start/end position prediction
- **Tasks**
  - Machine Reading Comprehension

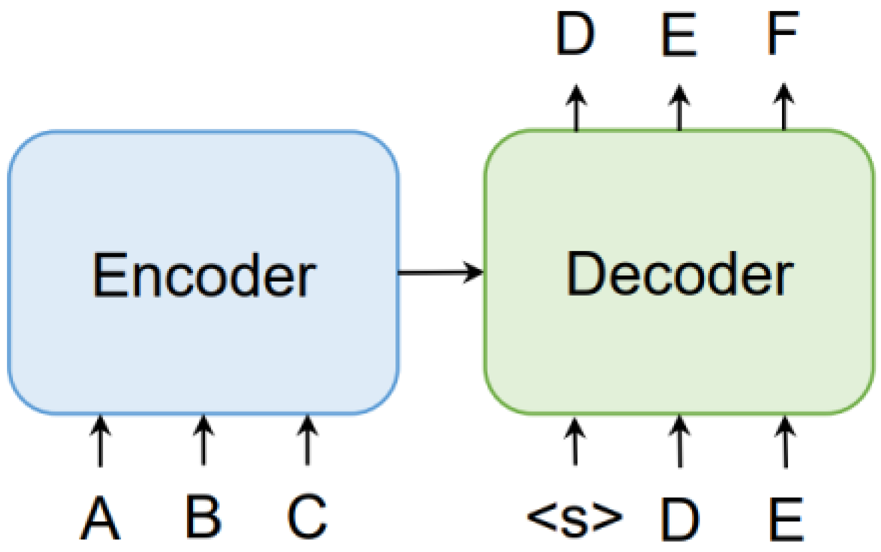


Sequence-to-Sequence (Seq2Seq)

- **Paradigm**

$$y_1, \cdots, y_m = \textsc{Dec}(\textsc{Enc}(x_1, \cdots, x_n))$$

- **Model**
  - $\textsc{Enc}(\cdot)$: CNN, RNN, Transformers…
  - $\textsc{Dec}(\cdot)$: CNN, RNN, Transformers…
- **Tasks**
  - Machine Translation
  - End-to-end dialogue system
  - …

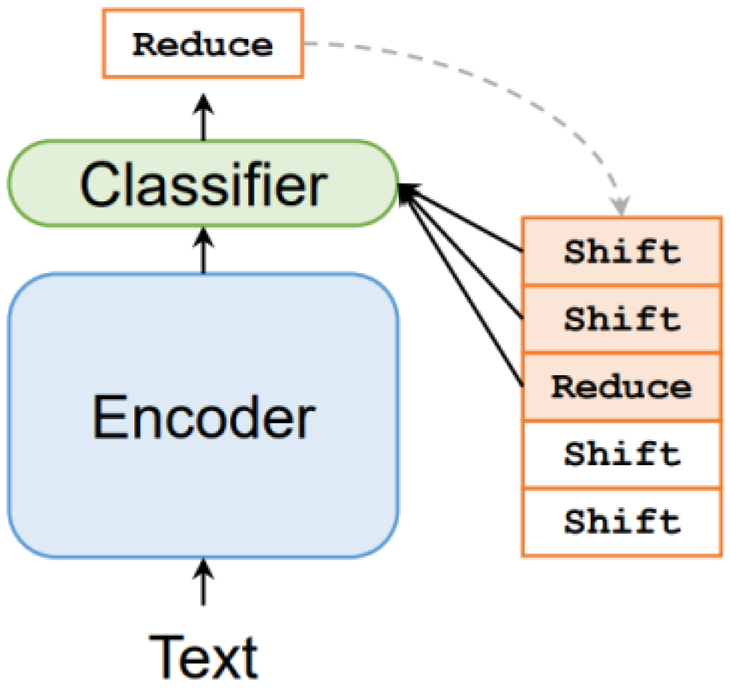


Sequence-to-Action-Sequence (Seq2ASeq)

- **Paradigm**

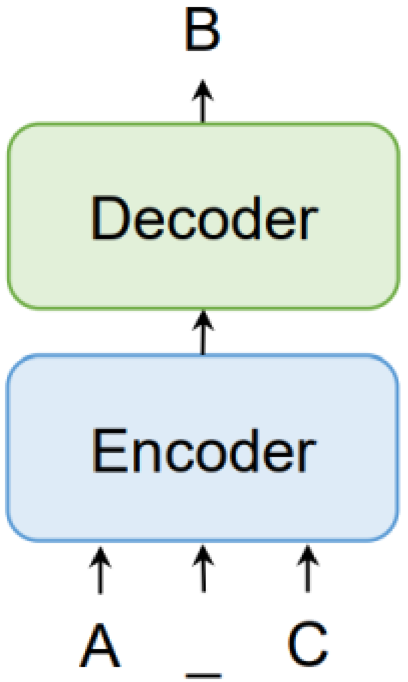
$$\mathcal{A} = \textsc{Cls}(\textsc{Enc}(\mathcal{X}), \mathcal{C})$$

- **Model**
  - $\textsc{Enc}(\cdot)$: CNN, RNN, Transformers…
  - $\textsc{Cls}(\cdot)$: predict an action conditioned on a configuration and the input text
- **Tasks**
  - Dependency Parsing
  - Constituency Parsing
  - …



(Masked) Language Model ((M)LM)

- **Paradigm**
  - LM: $x_k = \textsc{Dec}(x_1, \cdots, x_{k-1})$
  - MLM: $\bar{x} = \textsc{Dec}(\textsc{Enc}(\tilde{x}))$
- **Model**
  - $\textsc{Enc}(\cdot)$: CNN, RNN, Transformers…
  - $\textsc{Dec}(\cdot)$: simple classifier, or a auto-regressive decoder
- **Tasks**
  - Language Modeling
  - Masked Language Modeling
  - …



(g) (M)LM

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 Malfunkshun.* [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.

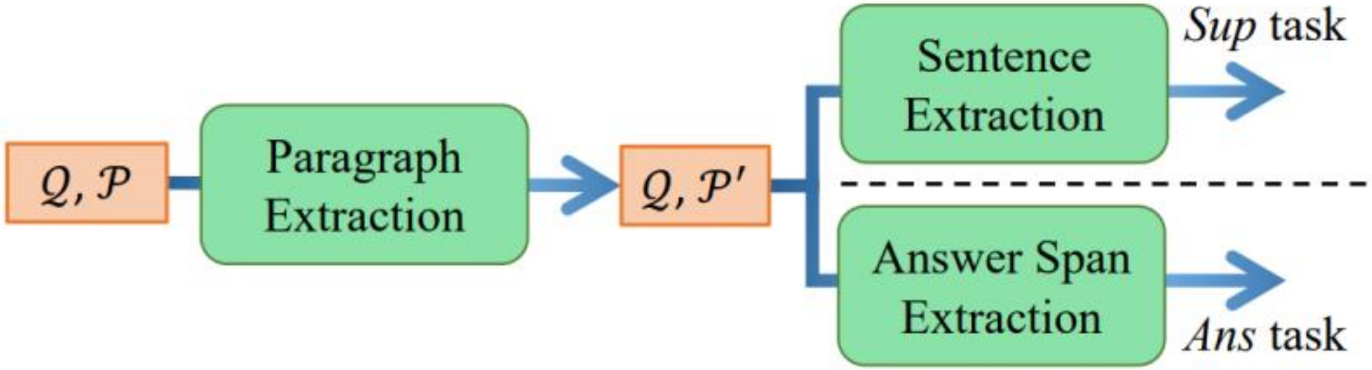
Paragraph B, Mother Love Bone:
[4] *Mother Love Bone was an American rock band that formed in Seattle, Washington in 1987.* [5] The band was active from 1987 to 1990. [6] *Frontman Andrew Wood's personality and compositions helped to catapult the group to the top of the burgeoning late 1980s/early 1990s 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: Malfunkshun
Supporting facts: 1, 2, 4, 6, 7

HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. EMNLP 2018

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 Malfunkshun.* [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.

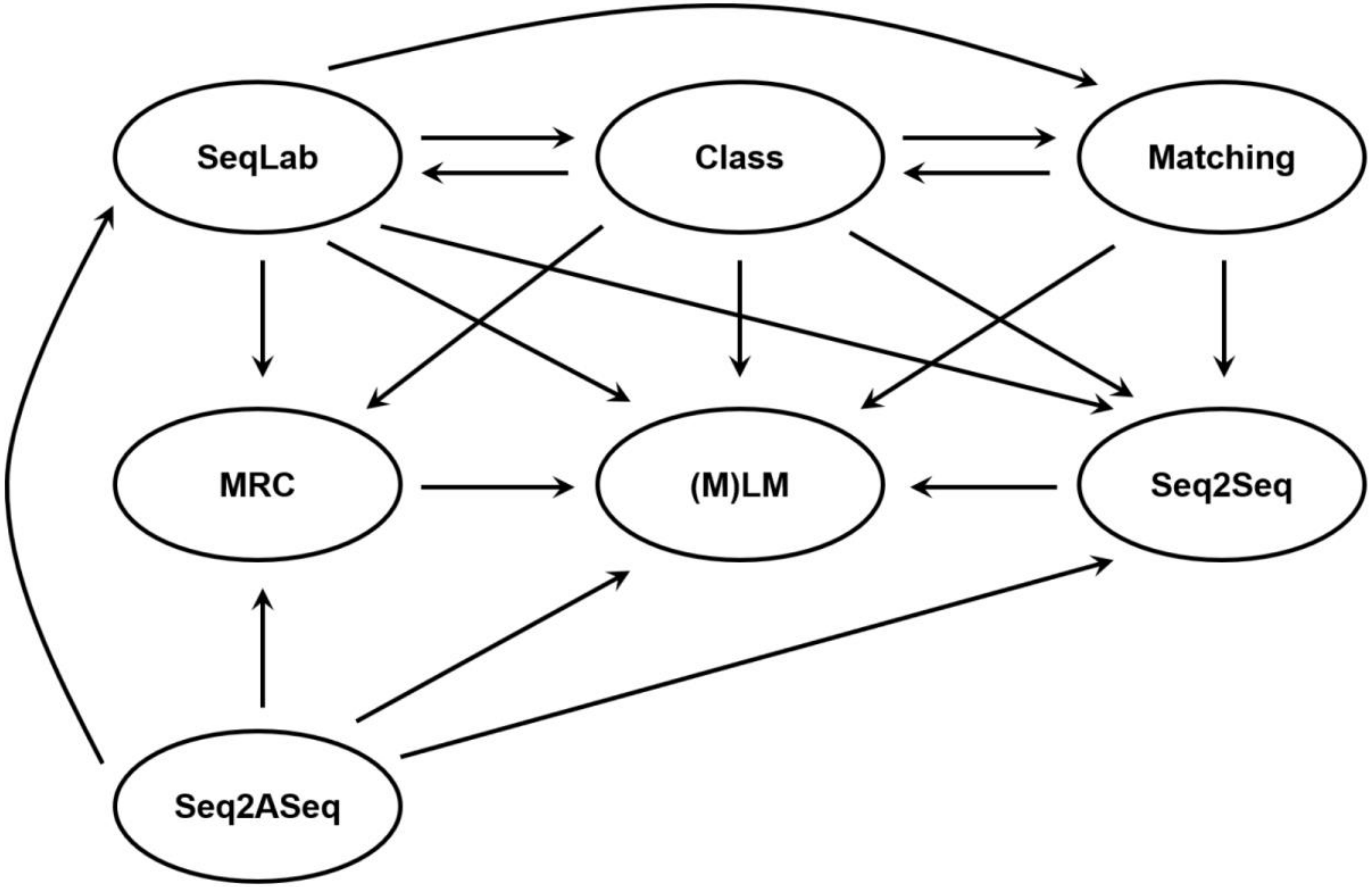
Q: What was the former band of the member of Mother Love Bone who died just before the release of "Apple"?
A: Malfunkshun
Supporting facts: 1, 2, 4, 6, 7

Graph-free Multi-hop Reading Comprehension: A Select-to-Guide Strategy. https://arxiv.org/abs/2107.11823

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

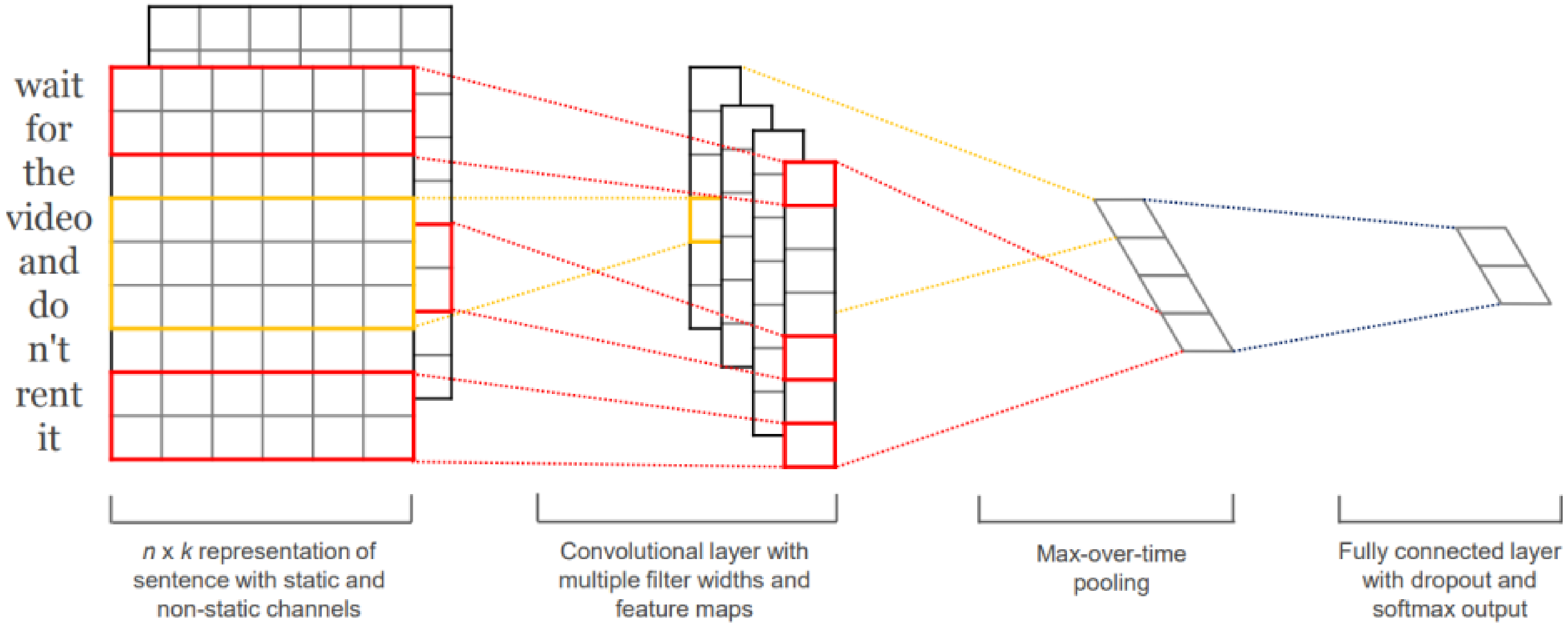
Table 1: Paradigms shift in natural language processing tasks. **TC**: text classification. **NLI**: natural language inference. **NER**: named entity recognition. **ABSA**: aspect-based sentiment analysis. **RE**: relation extraction. **Summ**: text summarization. **Parsing**: syntactic/semantic parsing. 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

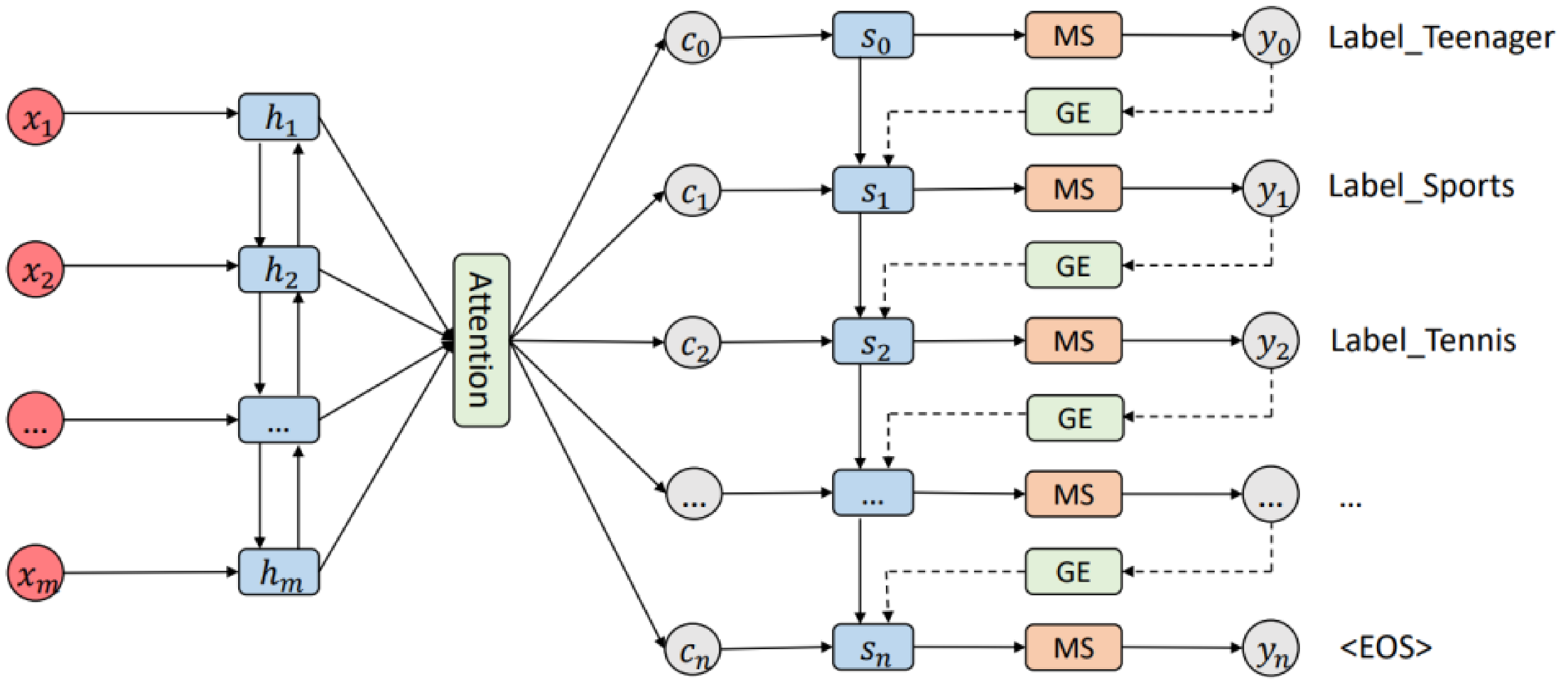
- **Traditional Paradigm: Class**
- **Shifted to…**
  - Seq2Seq
  - Matching
  - (M)LM



Convolutional Neural Networks for Sentence Classification. EMNLP 2014

Paradigm Shift in Text Classification

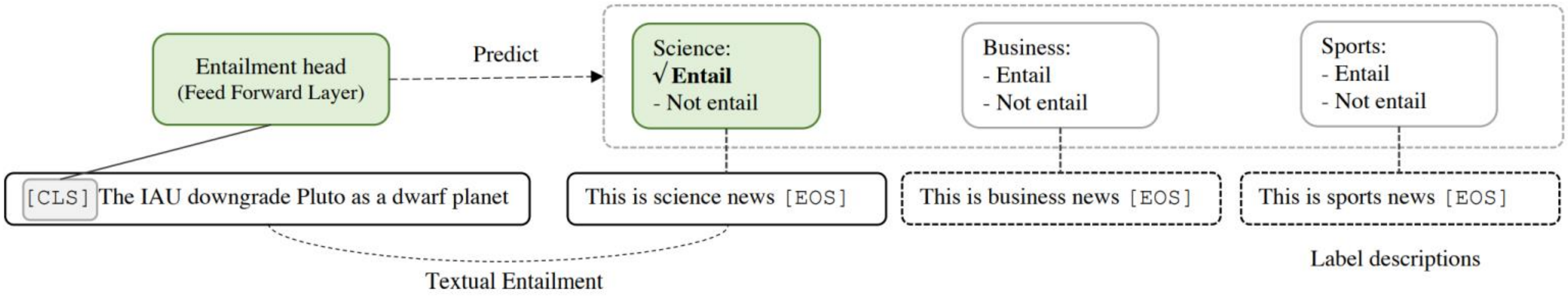
- **Traditional Paradigm:** Class
- **Shifted to…**
  - Seq2Seq
  - Matching
  - (M)LM



SGM: Sequence Generation Model for Multi-label Classification. COLING 2018

Paradigm Shift in Text Classification

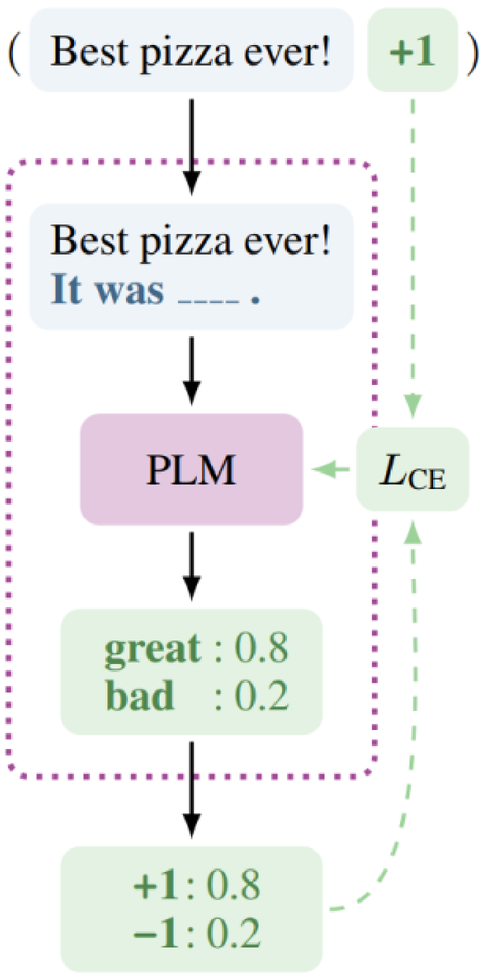
- **Traditional Paradigm:** Class
- **Shifted to…**
  - Seq2Seq
  - Matching
  - (M)LM



Entailment as Few-Shot Learner. https://arxiv.org/abs/2104.14690

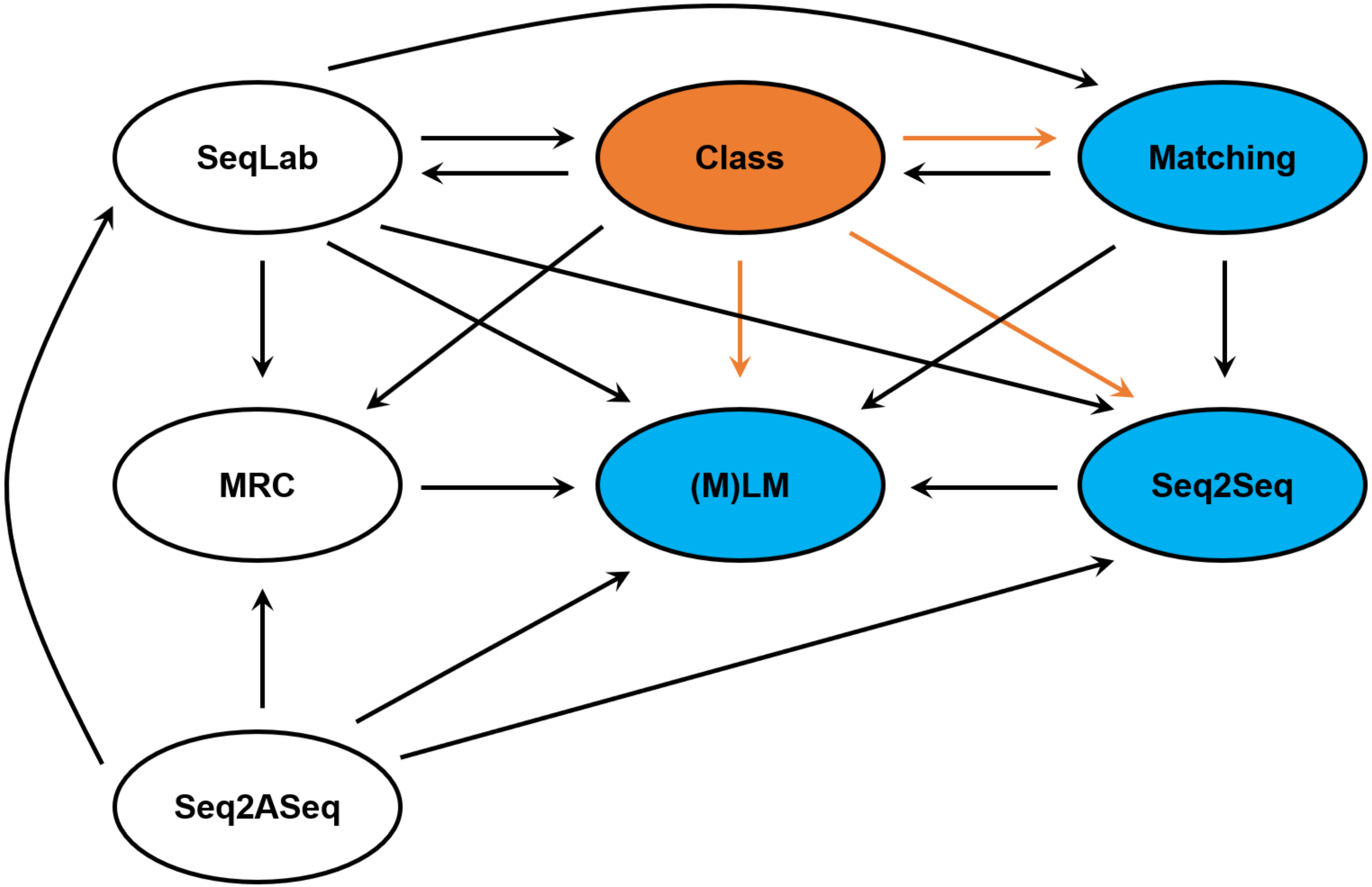
Paradigm Shift in Text Classification

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



Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. EACL 2021

Paradigm Shift in Text Classification

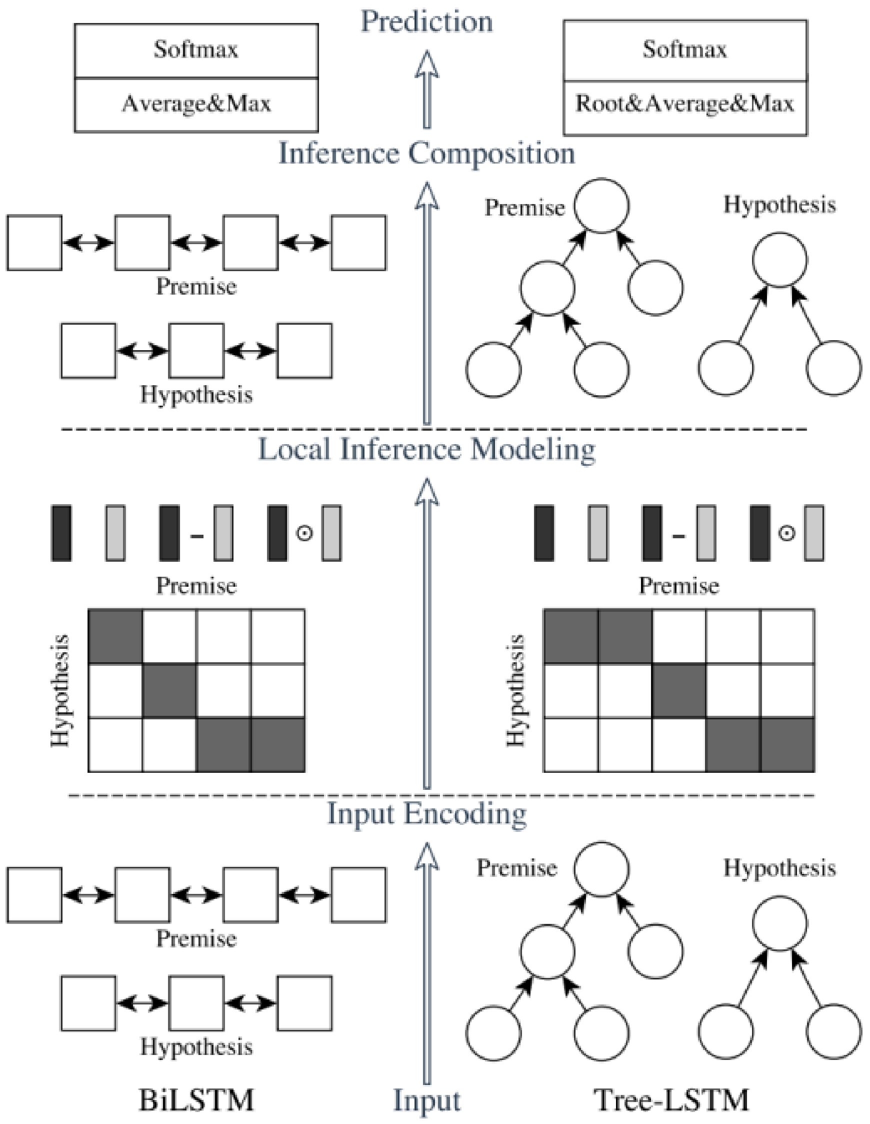


Paradigm Shift in NLI

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

Paradigm Shift in NLI

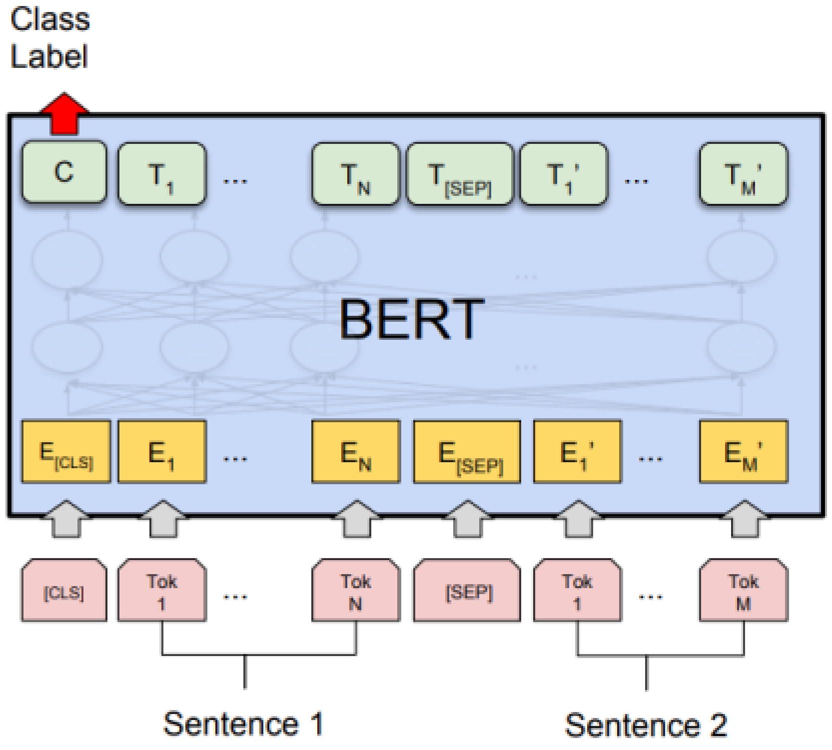
- **Traditional Paradigm: Matching**
- **Shifted to…**
  - Class
  - Seq2Seq
  - (M)LM



Enhanced LSTM for Natural Language Inference. ACL 2017

Paradigm Shift in NLI

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

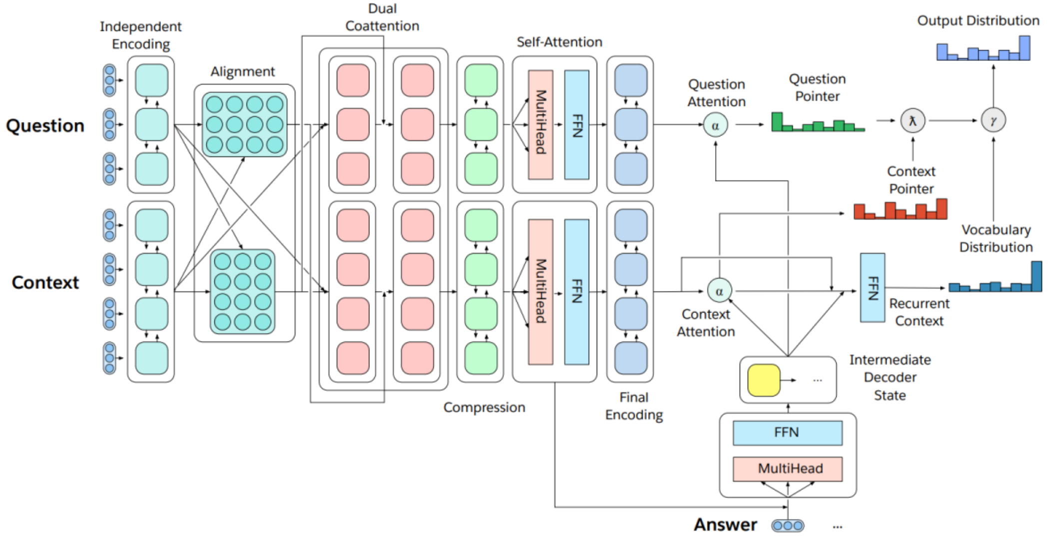


BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL 2019

Paradigm Shift in NLI

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

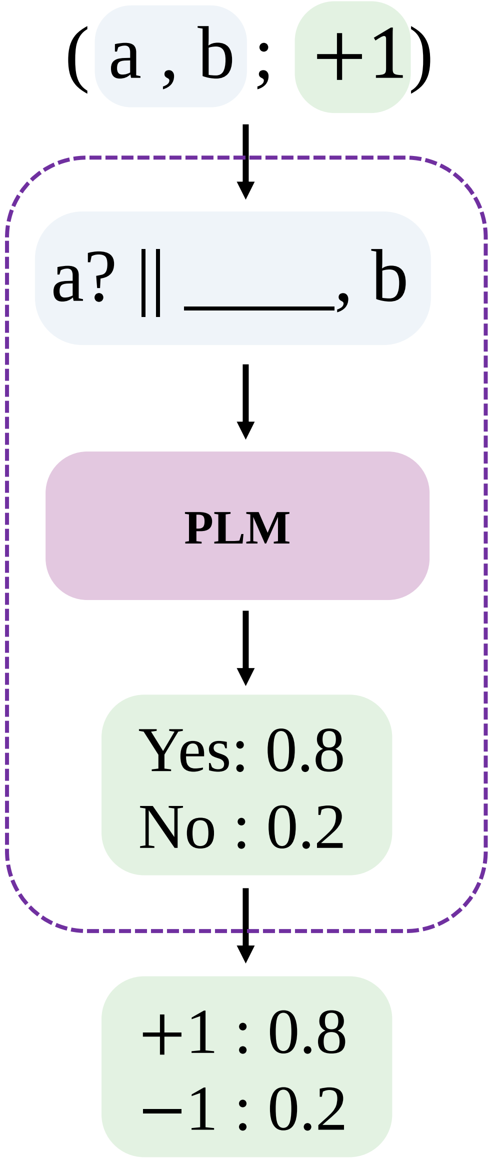
Question	Context	Answer
Hypothesis: Product and geography are what make cream skimming work. Entailment, neutral, or contradiction?	Premise: Conceptually cream skimming has two basic dimensions – product and geography.	Entailment



The Natural Language Decathlon: Multitask Learning as Question Answering. https://arxiv.org/abs/1806.08730

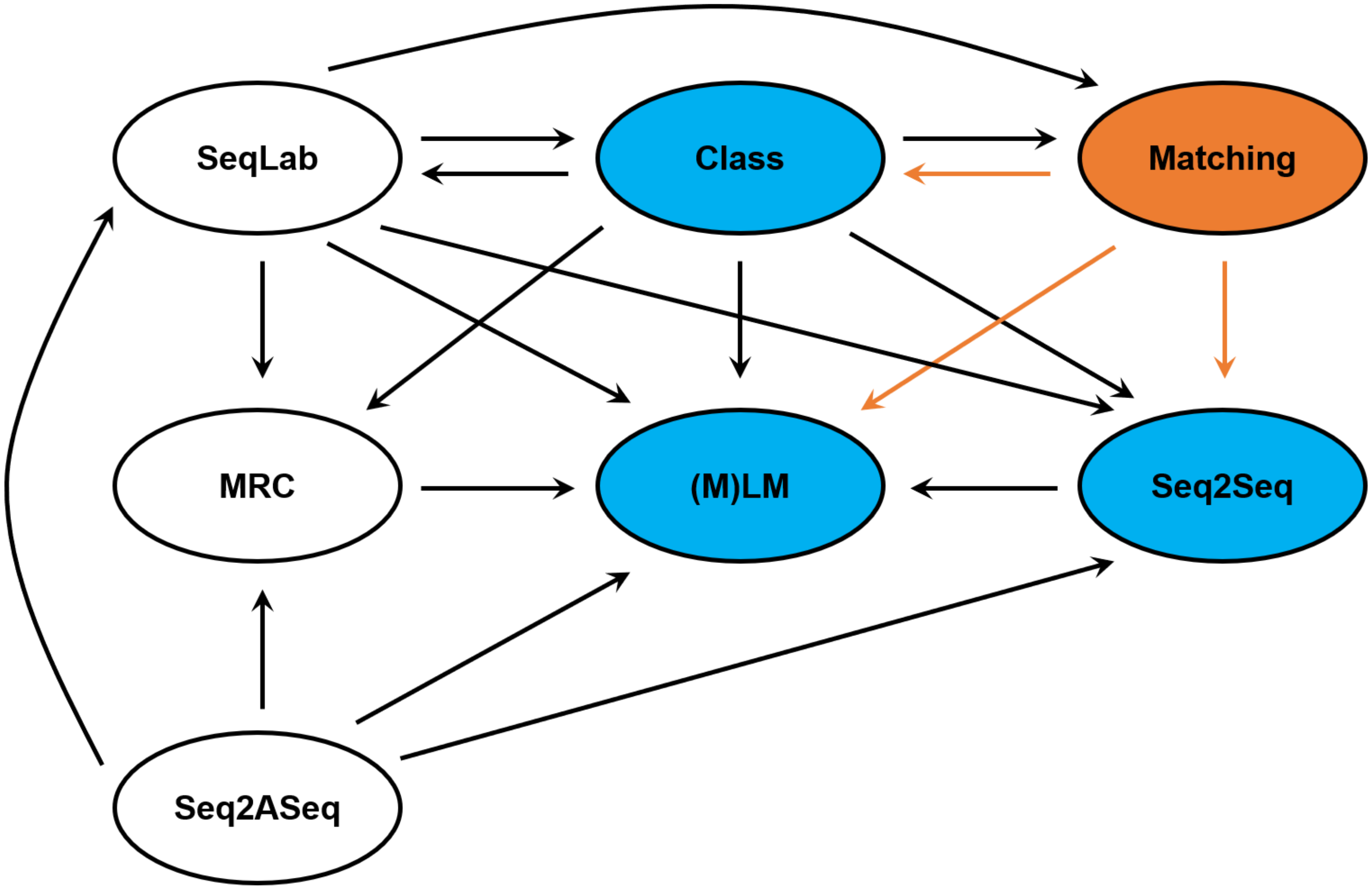
Paradigm Shift in NLI

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

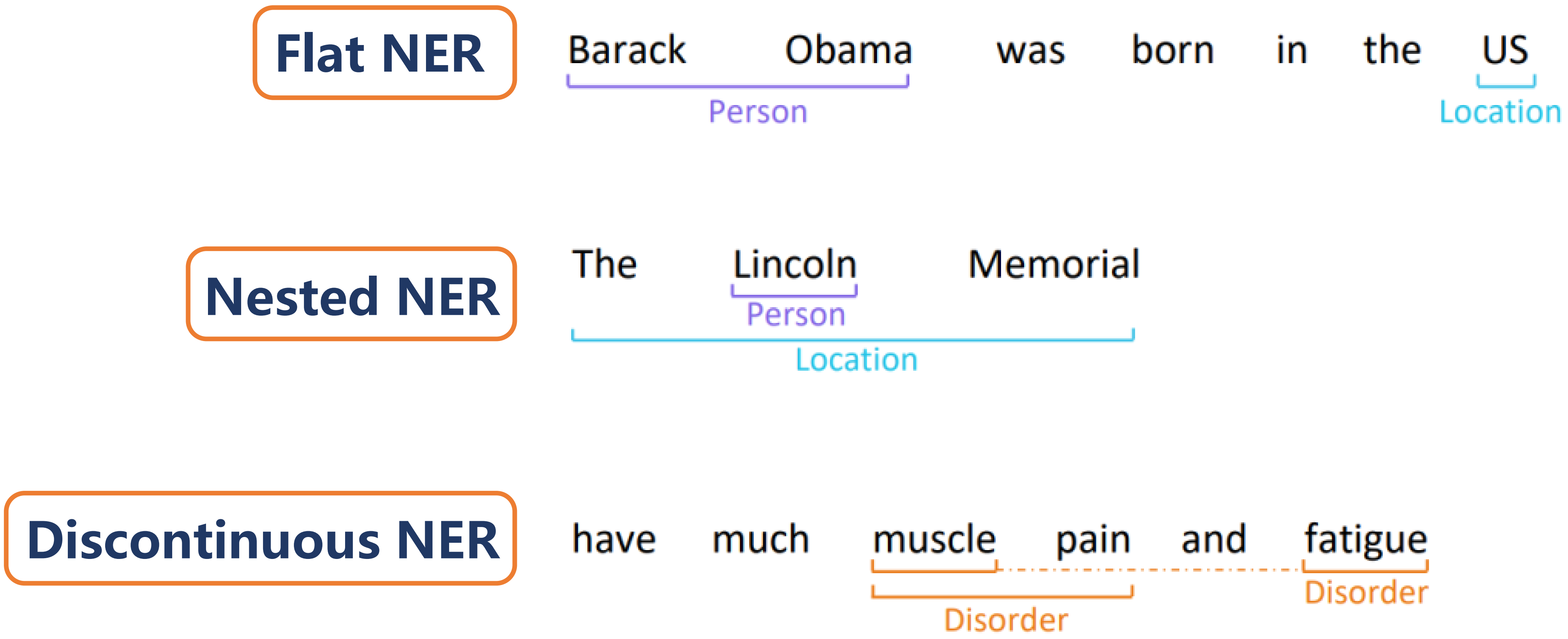


Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. EACL 2021

Paradigm Shift in NLI



Paradigm Shift in NER

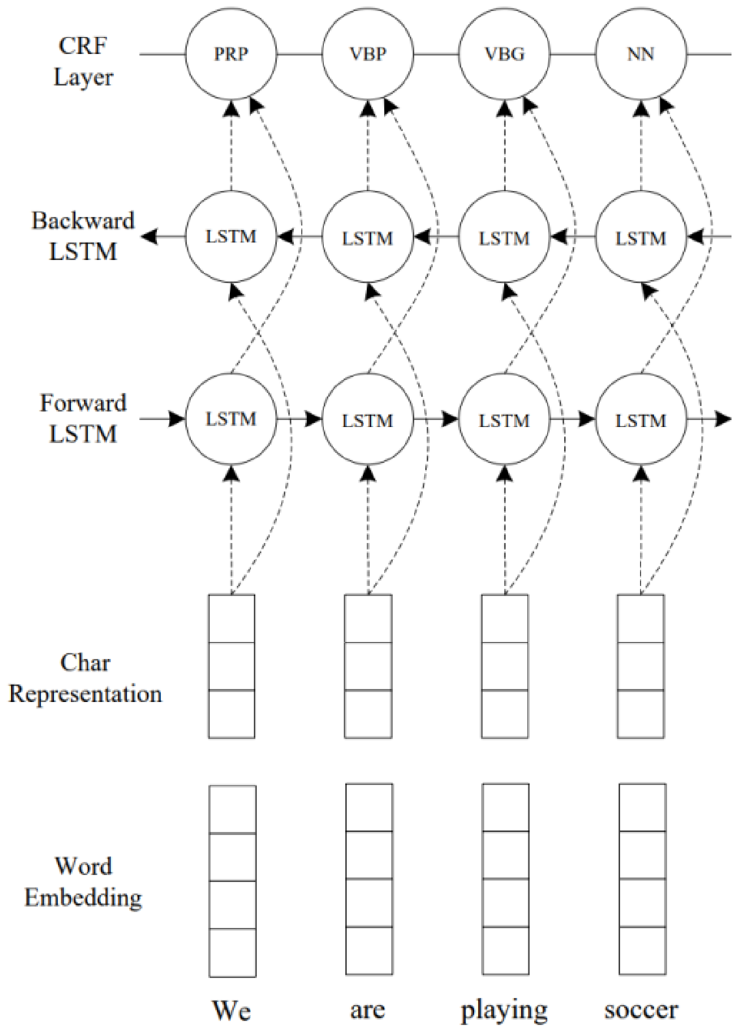


Paradigm Shift in NER

- **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

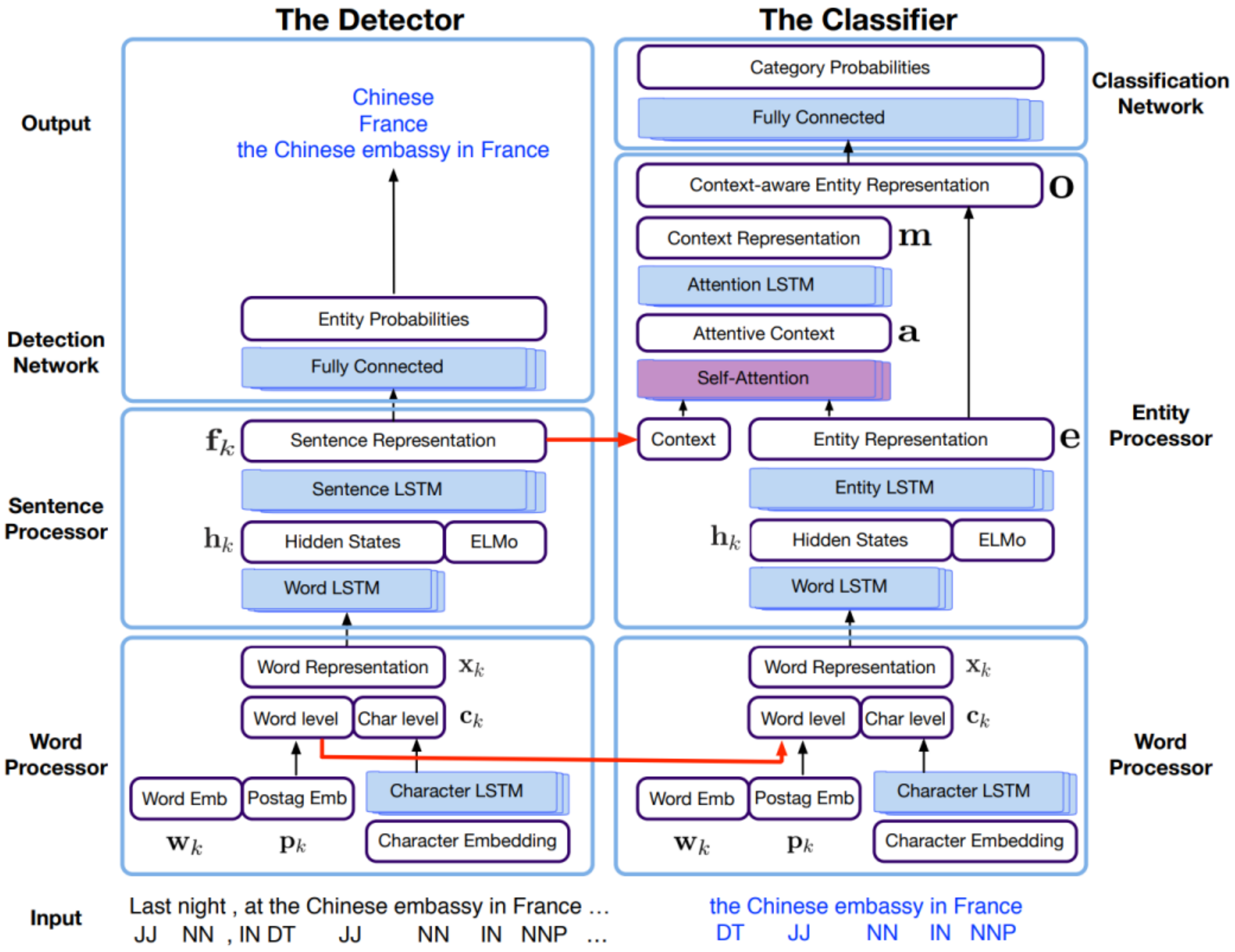
- **Traditional Paradigm:**
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Discontinuous NER)
- **Shifted to / Unified in…**
  - Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - Seq2Seq (All)



End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. ACL 2016

Paradigm Shift in NER

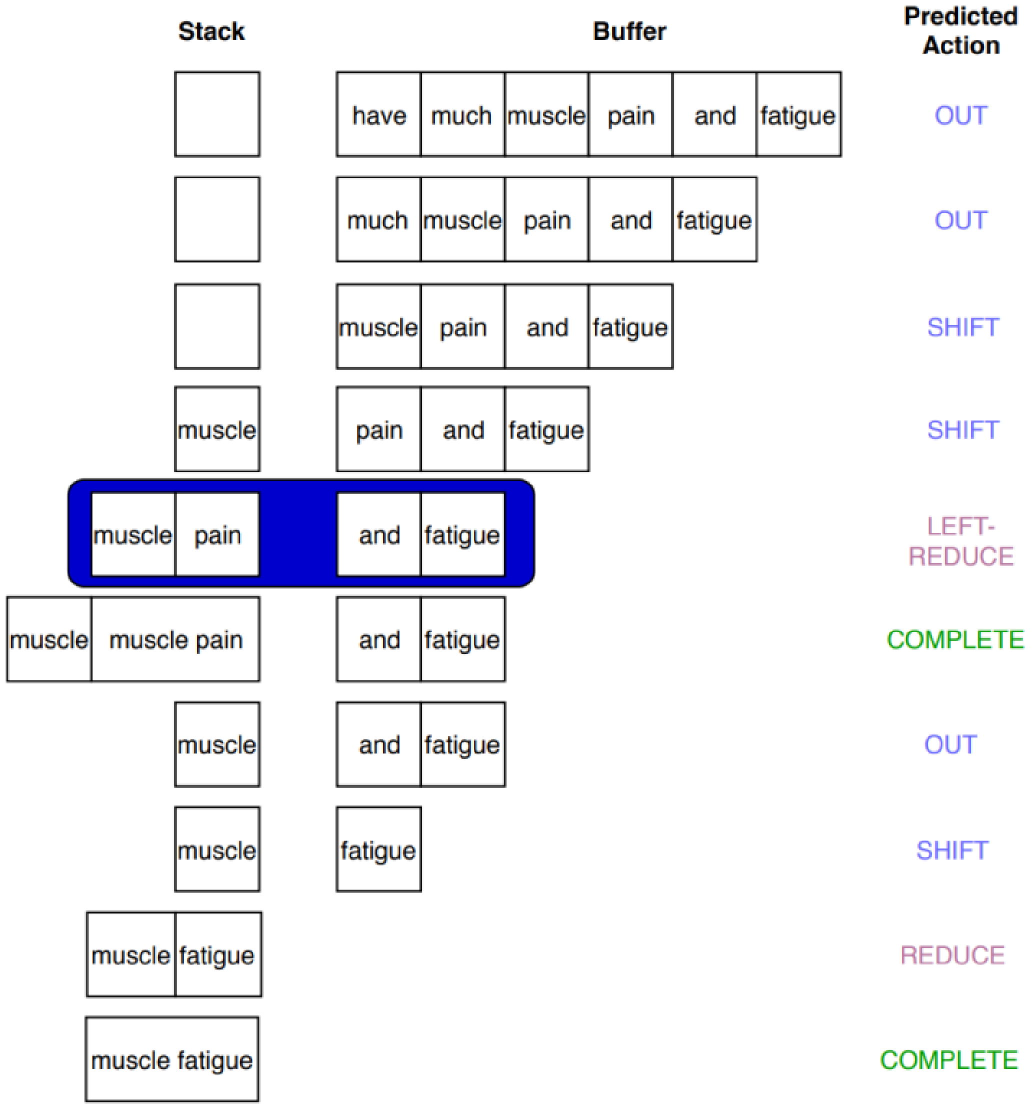
- **Traditional Paradigm:**
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Discontinuous NER)
- **Shifted to / Unified in…**
  - Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - Seq2Seq (All)



Multi-Grained Named Entity Recognition. ACL 2019

Paradigm Shift in NER

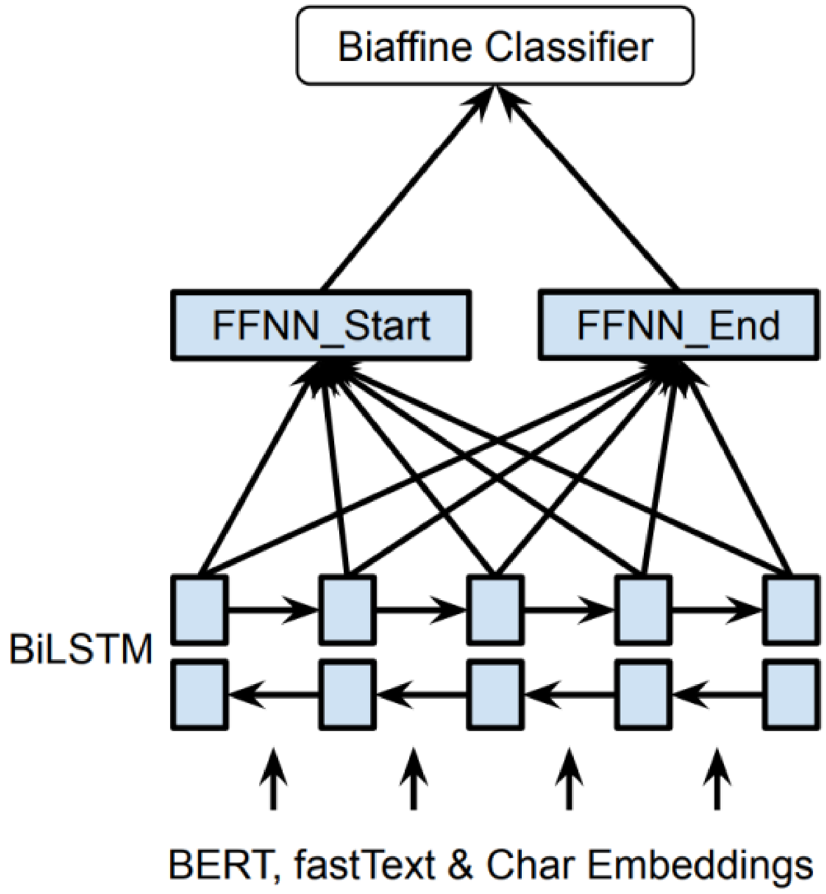
- **Traditional Paradigm:**
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Discontinuous NER)
- **Shifted to / Unified in...**
  - Class (Flat&Nested NER)
  - MRC (Flat&Nested NER)
  - Seq2Seq (All)



An Effective Transition-based Model for Discontinuous NER. ACL 2020

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)



Named Entity Recognition as Dependency Parsing. ACL 2020

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)

Matrix (l×l×c) Labeling:

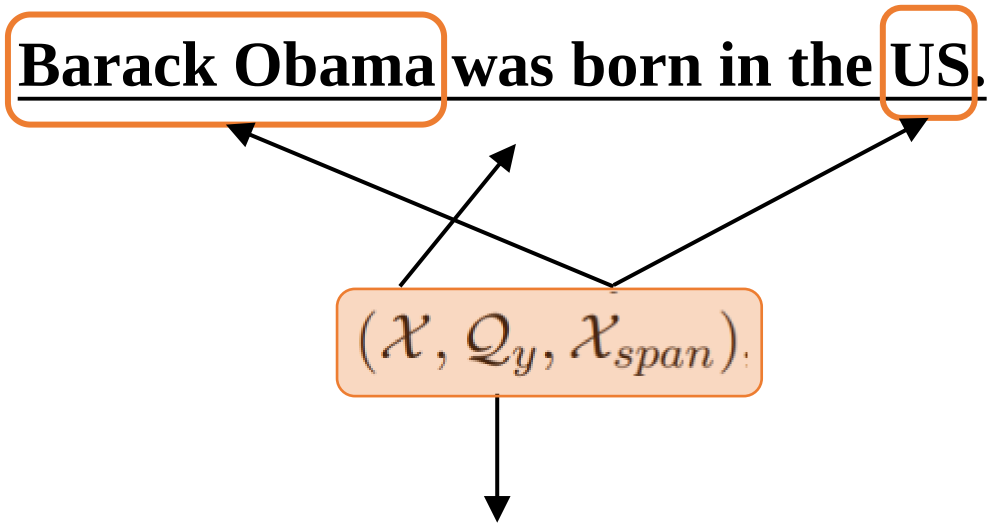
	The	Lincoln	Memorial
The	0	0	2
Lincoln	-1	1	0
Memorial	-1	-1	0

$$
\begin{aligned}
h_s(i) &= \mathrm{FFNN}_s(x_{s_i}) \\
h_e(i) &= \mathrm{FFNN}_e(x_{e_i}) \\
r_m(i) &= h_s(i)^\top \mathrm{U}_m h_e(i) \\
&\quad + W_m(h_s(i) \oplus h_e(i)) + b_m
\end{aligned}
$$

Named Entity Recognition as Dependency Parsing. ACL 2020

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)

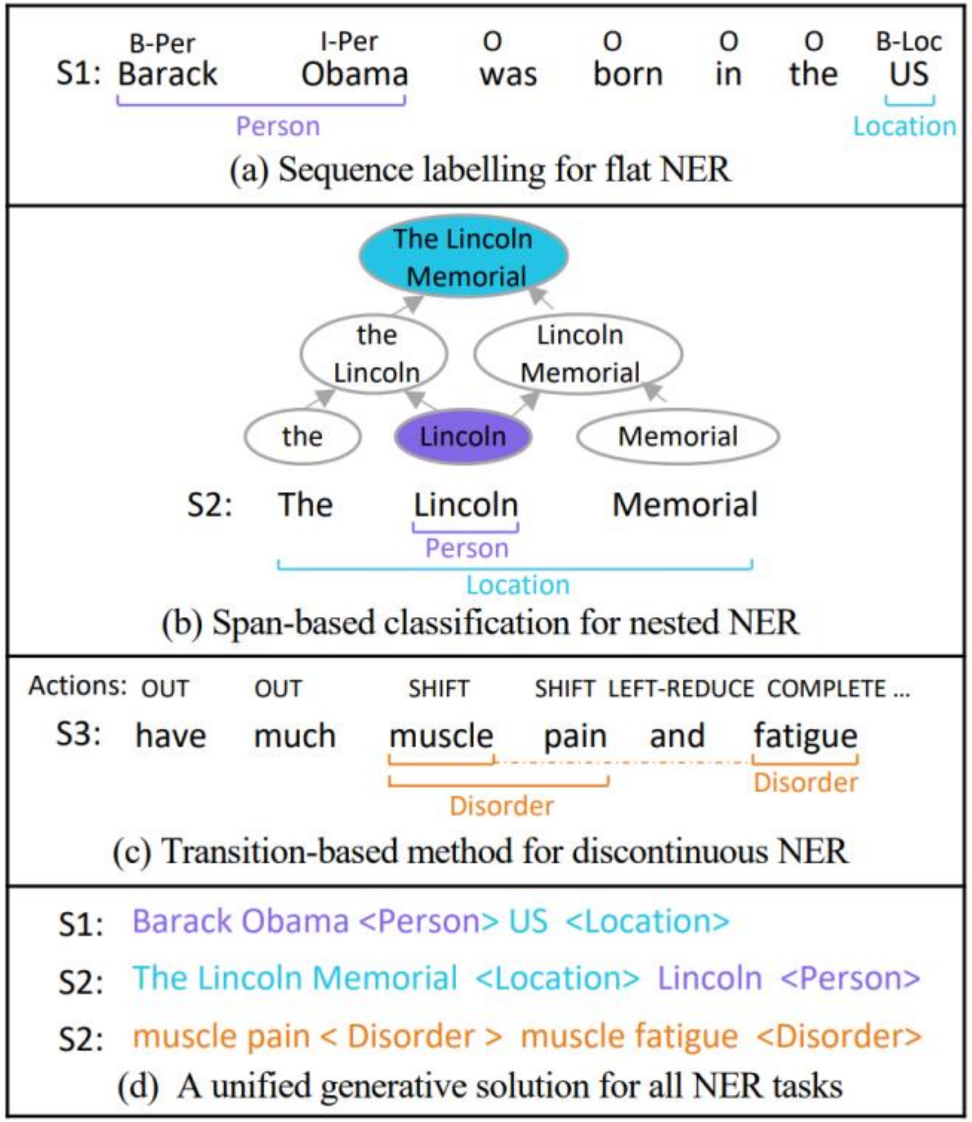


Entity	Natural Language Question
Location	Find locations in the text, including non-geographical locations, mountain ranges and bodies of water.
Facility	Find facilities in the text, including buildings, airports, highways and bridges.
Organization	Find organizations in the text, including companies, agencies and institutions.

A Unified MRC Framework for Named Entity Recognition. ACL 2020

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)

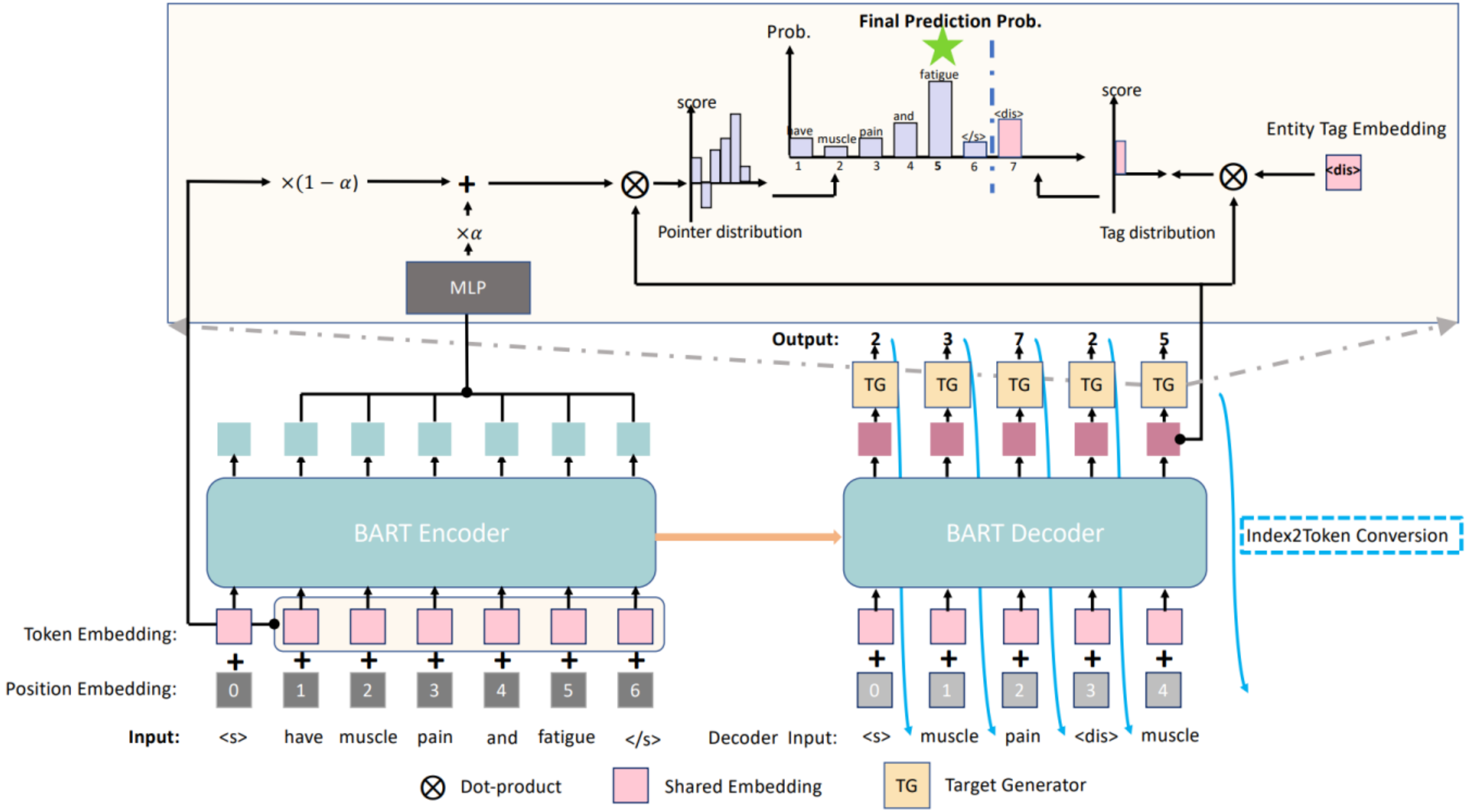


A Unified Generative Framework for Various NER Subtasks. ACL 2021

Paradigm Shift in NER

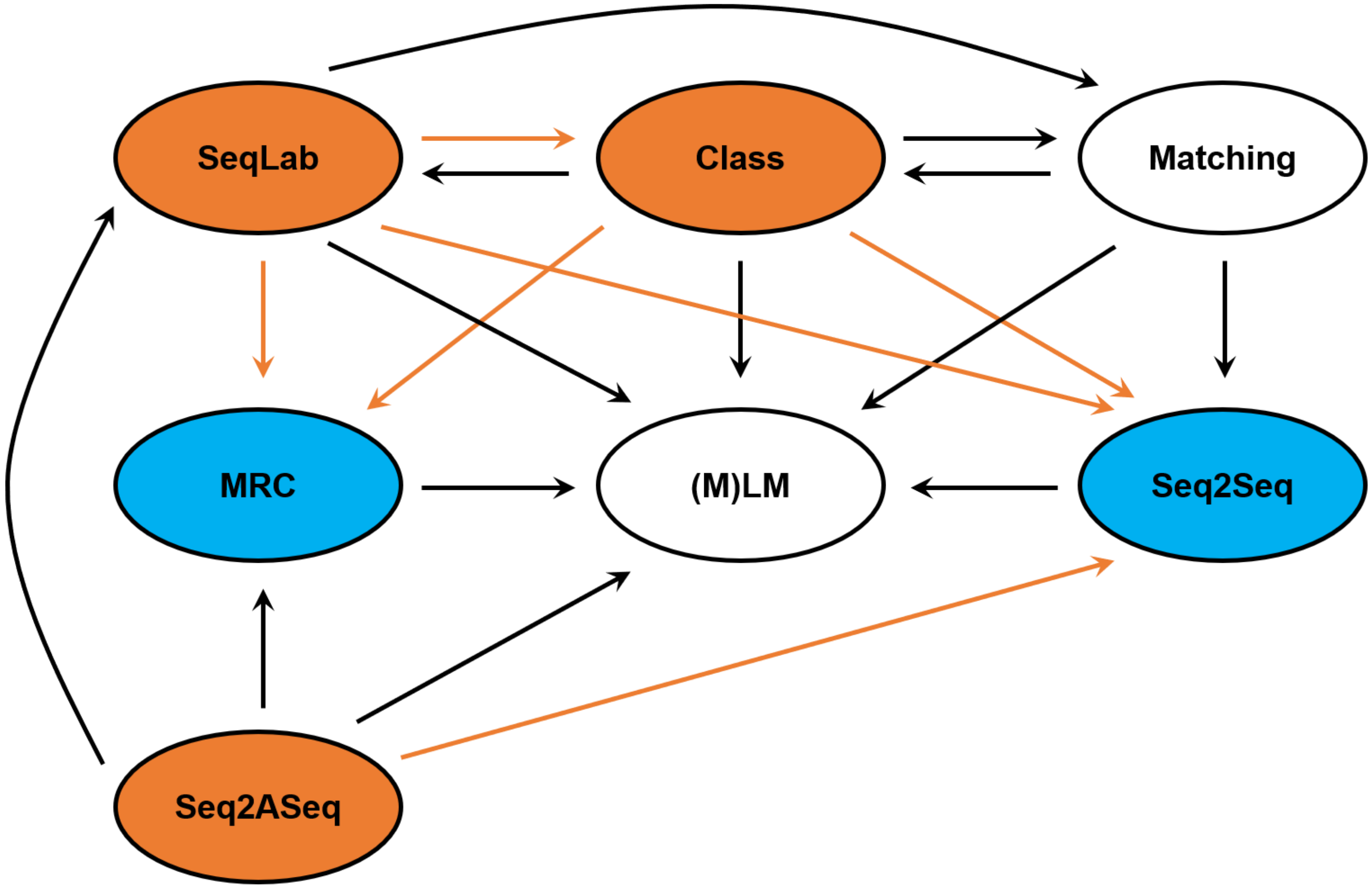
- **Traditional Paradig**
  - SeqLab (Flat NER)
  - Class (Nested NER)
  - Seq2ASeq (Disconti
- **Shifted to / Unifie**
  - Class (Flat&Nested
  - MRC (Flat&Nested
  - Seq2Seq (All)

Input: <s> have muscle pain and fatigue </s>
Output: 2 3 7 2 5 6



A Unified Generative Framework for Various NER Subtasks. ACL 2021

Paradigm Shift in NER



Paradigm Shift in ABSA

Subtask	Input	Output	Task Type
Aspect Term Extraction(*AE*)	S	a_1, a_2	Extraction
Opinion Term Extraction(*OE*)	S	o_1, o_2	Extraction
Aspect-level Sentiment Classification(*ALSC*)	S + a_1	s_1	Classification
	S + a_2	s_2	
Aspect-oriented Opinion Extraction(*AOE*)	S + a_1	o_1	Extraction
	S + a_2	o_2	
Aspect Term Extraction and Sentiment Classification(*AESC*)	S	(a_1, s_1), (a_2, s_2)	Extraction & Classification
Pair Extraction(*Pair*)	S	(a_1, o_1), (a_2, o_2)	Extraction
Triplet Extraction(*Triplet*)	S	(a_1, o_1, s_1), (a_2, o_2, s_2)	Extraction & Classification

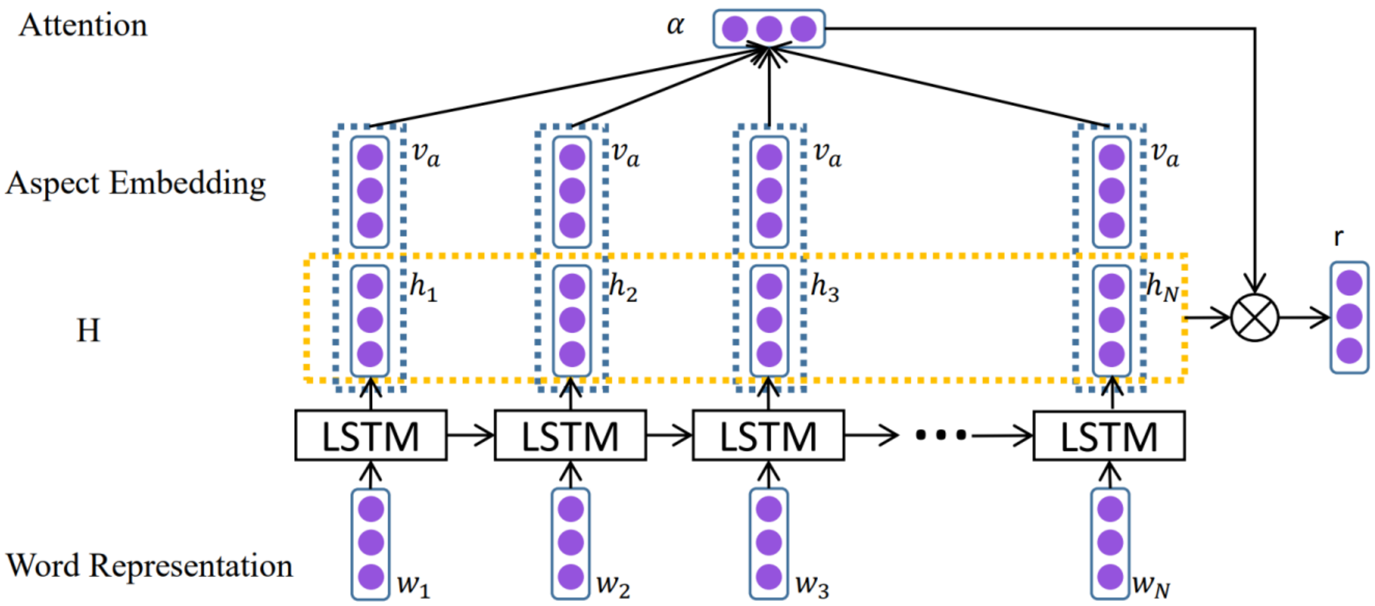
A Unified Generative Framework for Aspect-Based Sentiment Analysis. ACL 2021

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)



Attention-based LSTM for Aspect-level Sentiment Classification. EMNLP 2016

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]**

Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. NAACL 2019

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)

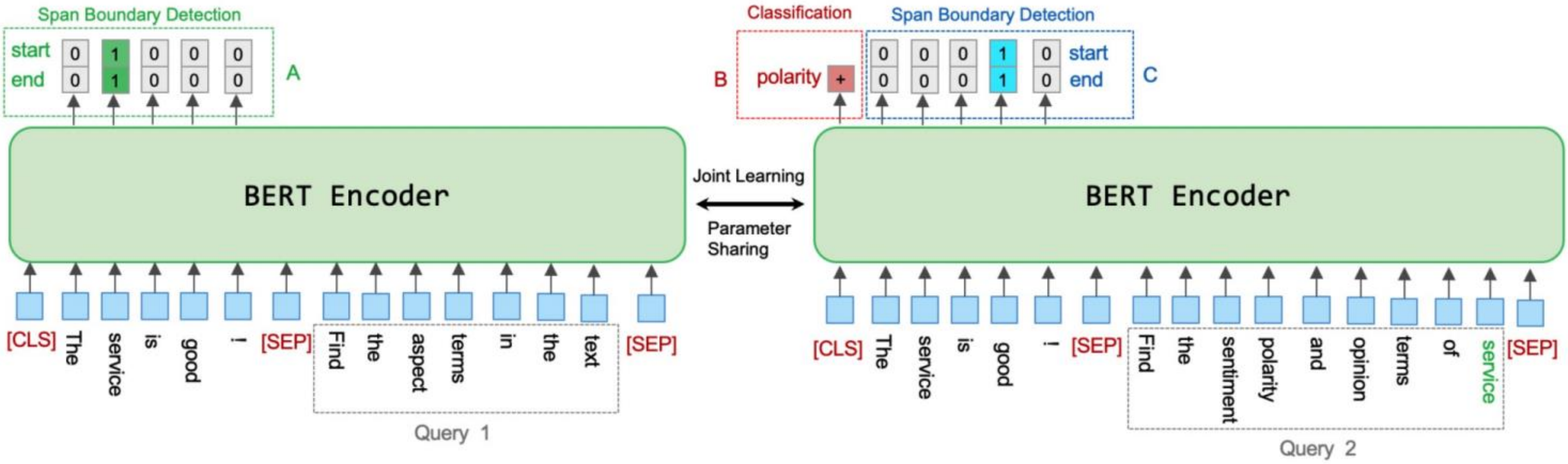
Converted training example 1:
- **query-1**: Find the *aspect terms* in the text.
- **answer-1**: ambience, service
- **query-2**: Find the *sentiment polarity* and *opinion terms* for ambience in the text.
- **answer-2**: (nice, positive)

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)

A Joint Training Dual-MRC Framework for Aspect Based Sentiment Analysis. AAAI 2021

Paradigm Shift in ABSA

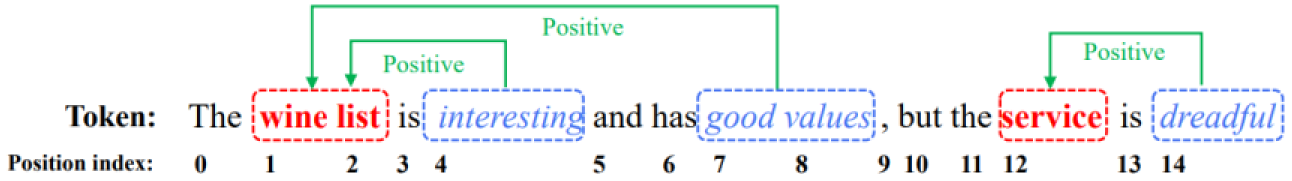
- **Traditional Paradigm:**



A Joint Training Dual-MRC Framework for Aspect Based Sentiment Analysis. AAAI 2021

Paradigm Shift in ABSA

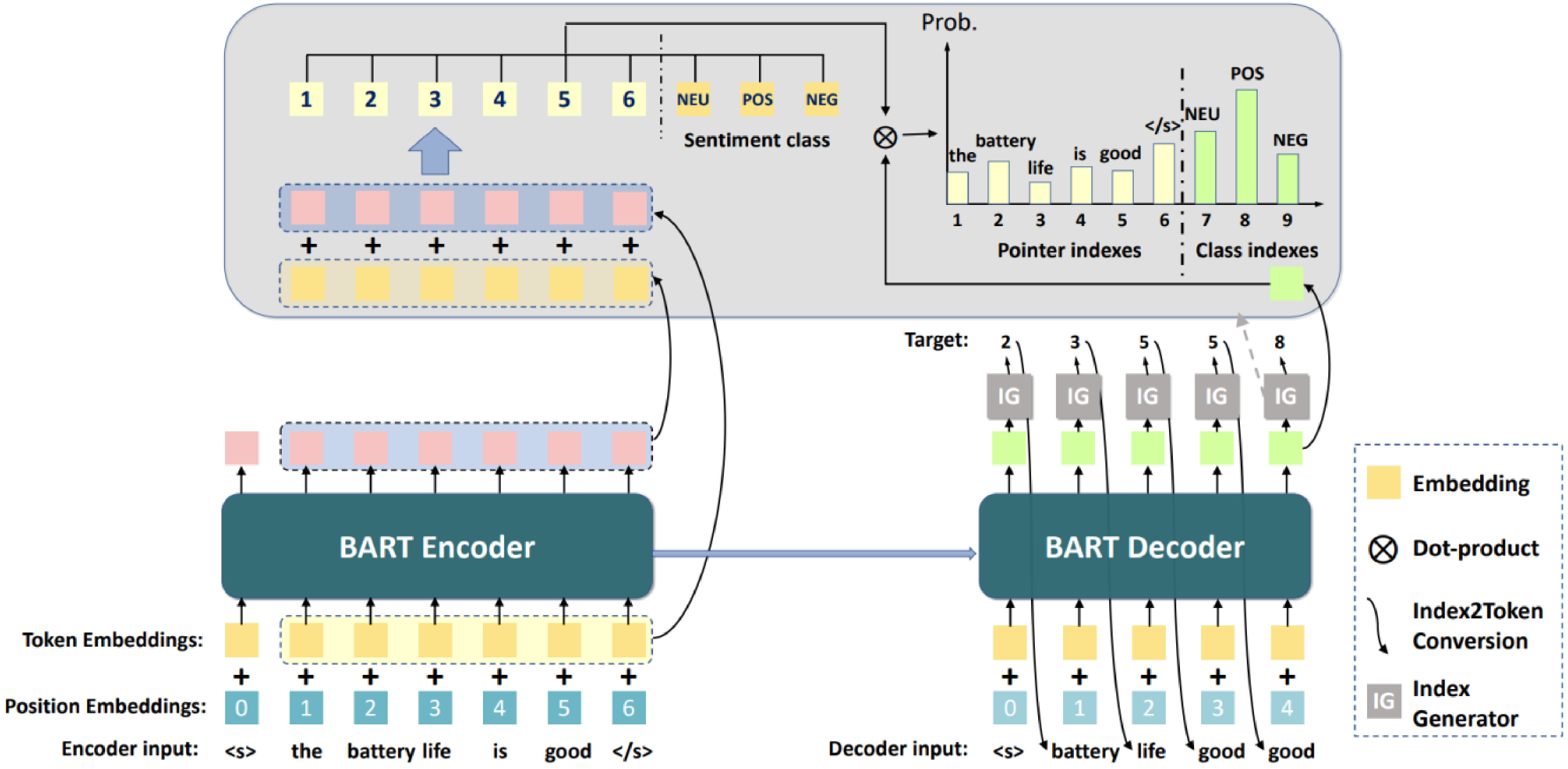
- **Traditional Paradigm:**
  - SeqLab (AE, OE, AOE, ...)
  - Class (ALSC...)
- **Shifted to / Unified in...**
  - Matching (ALSC)
  - MRC (All)
  - Seq2Seq (All)
  - (M)LM (All)



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

A Unified Generative Framework for Aspect-Based Sentiment Analysis. ACL 2021

Paradigm Shift in ABSA



A Unified Generative Framework for Aspect-Based Sentiment Analysis. ACL 2021

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)

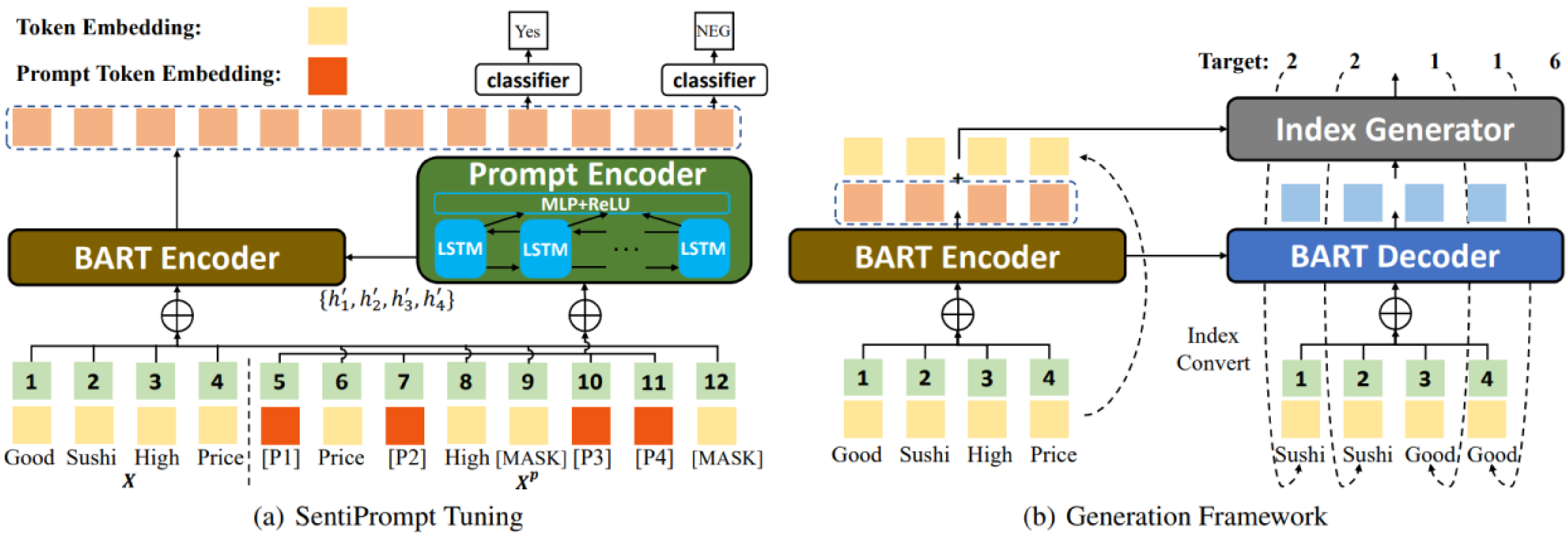
The owners are great fun and the beer selection is worth staying for.

↓

Consistency prompt	**The owners are great fun? [MASK] .**
Polarity prompt	**This is [MASK] .**

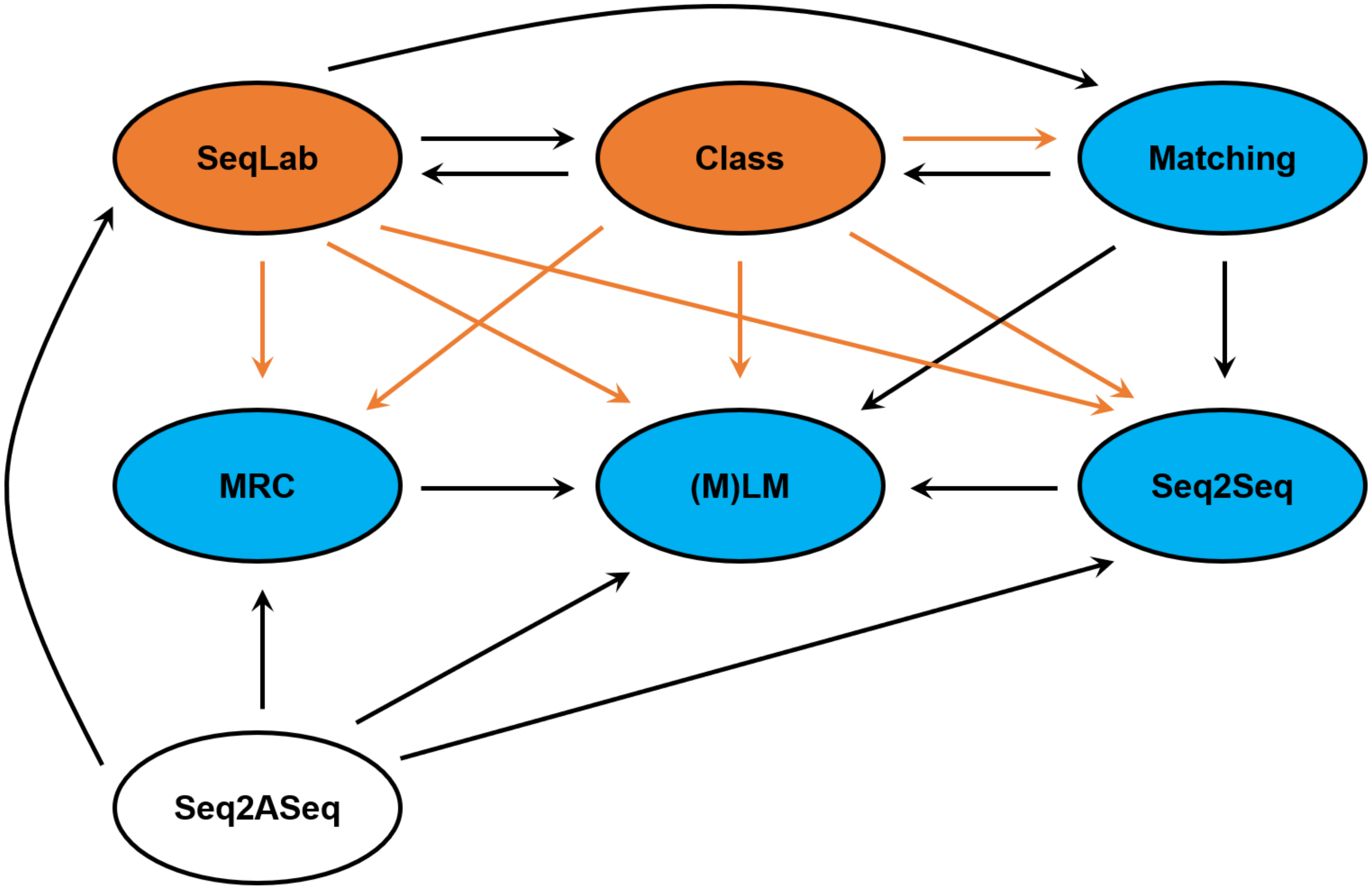
SentiPrompt: Sentiment Knowledge Enhanced Prompt-Tuning for Aspect-Based Sentiment Analysis. https://arxiv.org/abs/2109.08306

Paradigm Shift in ABSA



SentiPrompt: Sentiment Knowledge Enhanced Prompt-Tuning for Aspect-Based Sentiment Analysis. https://arxiv.org/abs/2109.08306

Paradigm Shift in ABSA

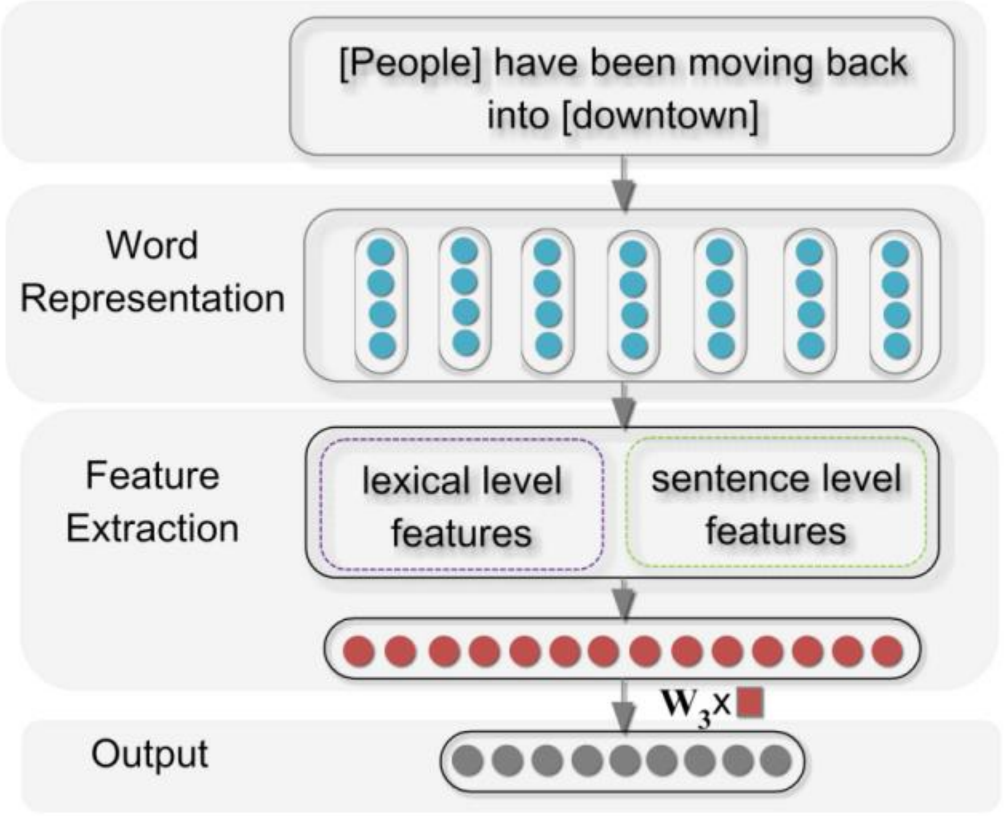


Paradigm Shift in Relation Extraction

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

Paradigm Shift in Relation Extraction

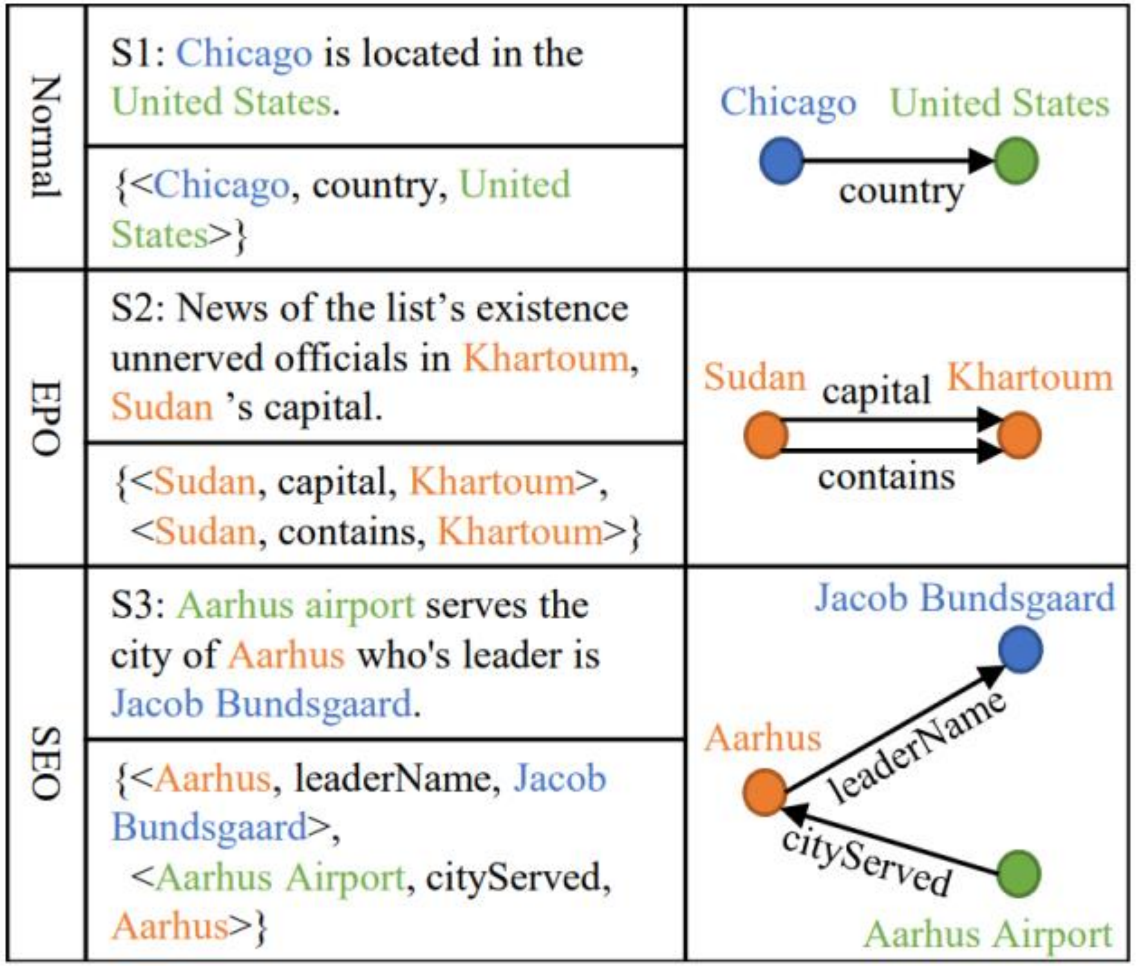
- **Traditional Paradigm:**
  - SeqLab (entity extraction)
  - Class (relation classification)
- **Shifted to / Unified in…**
  - Seq2Seq
  - MRC
  - (M)LM



Relation Classification via Convolutional Deep Neural Network. COLING 2014

Paradigm Shift in Relation Extraction

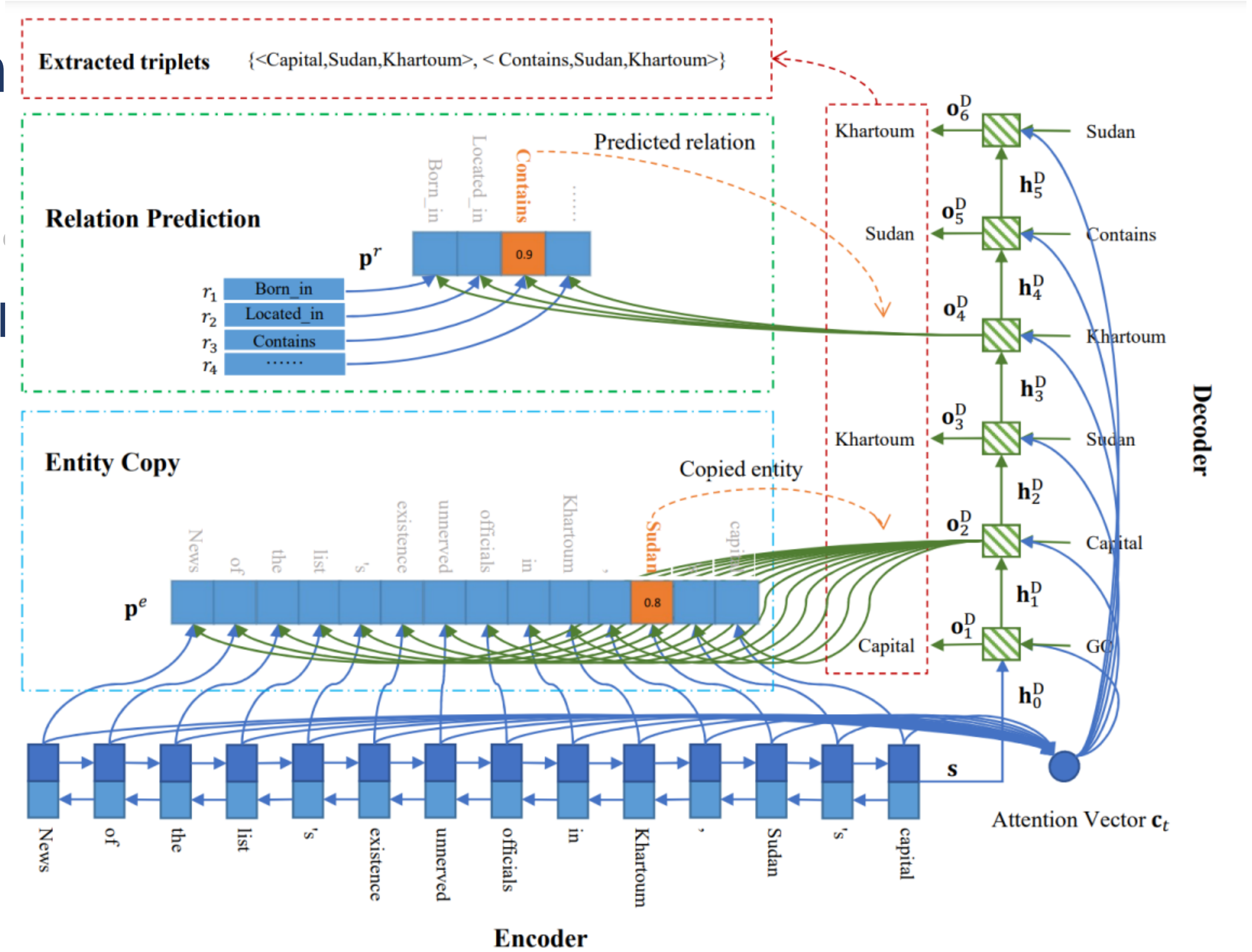
- **Traditional Paradigm:**
  - SeqLab (entity extraction)
  - Class (relation classification)
- **Shifted to / Unified in…**
  - Seq2Seq
  - MRC
  - (M)LM



Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism. ACL 2018

Paradigm Shift in Relation Extraction

- **Traditional Paradigm**
  - SeqLab (entity extracti
  - Class (relation classific
- **Shifted to / Unified i**
  - Seq2Seq
  - MRC
  - (M)LM



Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism. ACL 2018

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 Template
$educated\_at(x, y)$	Where did x graduate from? In which university did x study? What is x's alma mater?
$occupation(x, y)$	What did x do for a living? What is x's job? What is the profession of x?
$spouse(x, y)$	Who is x's spouse? Who did x marry? Who is x married to?

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

Zero-Shot Relation Extraction via Reading Comprehension. CoNLL 2017

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

Entity-Relation Extraction as Multi-Turn Question Answering. ACL 2019

Paradigm Shift in Relation Extraction

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

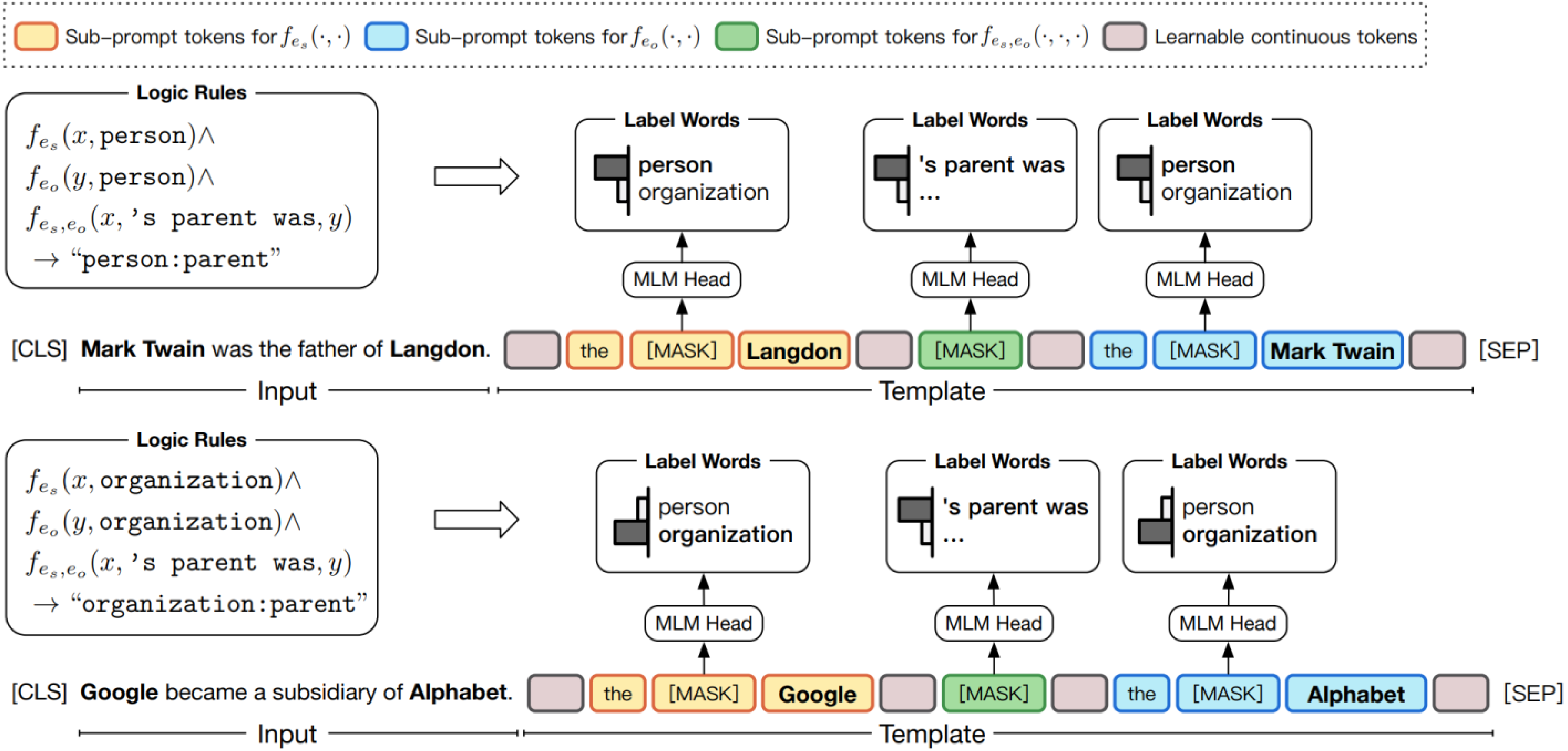
***Mark Twain* was the father of *Langdon*.**

↓

[p] the <u>person</u> *Langdon* [p] <u>'s parent was</u> [p] the <u>person</u> *Mark Twain* [p].

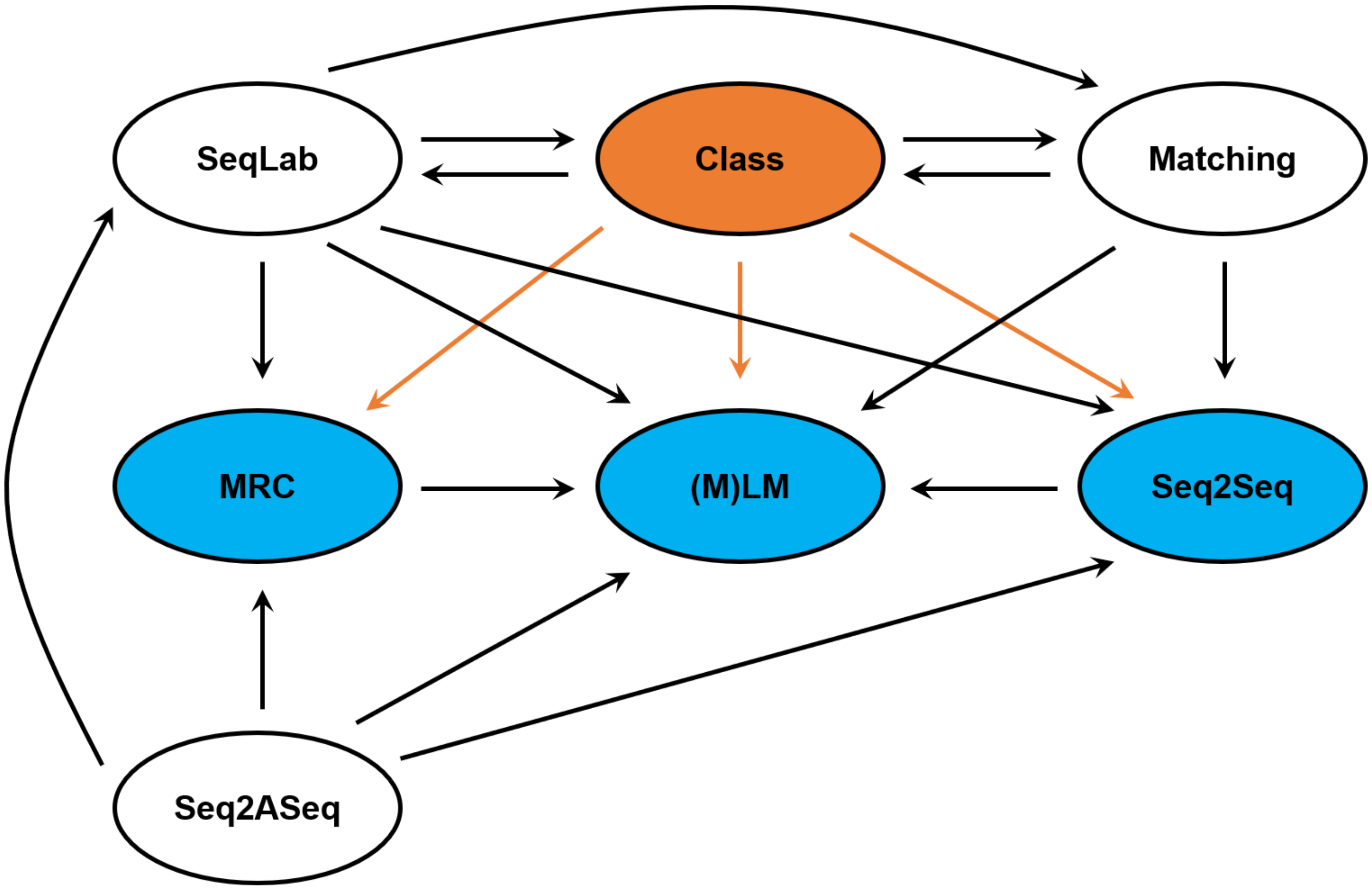
PTR: Prompt Tuning with Rules for Text Classification. https://arxiv.org/abs/2105.11259

Paradigm Shift in Relation Extraction



PTR: Prompt Tuning with Rules for Text Classification. https://arxiv.org/abs/2105.11259

Paradigm Shift in Relation Extraction

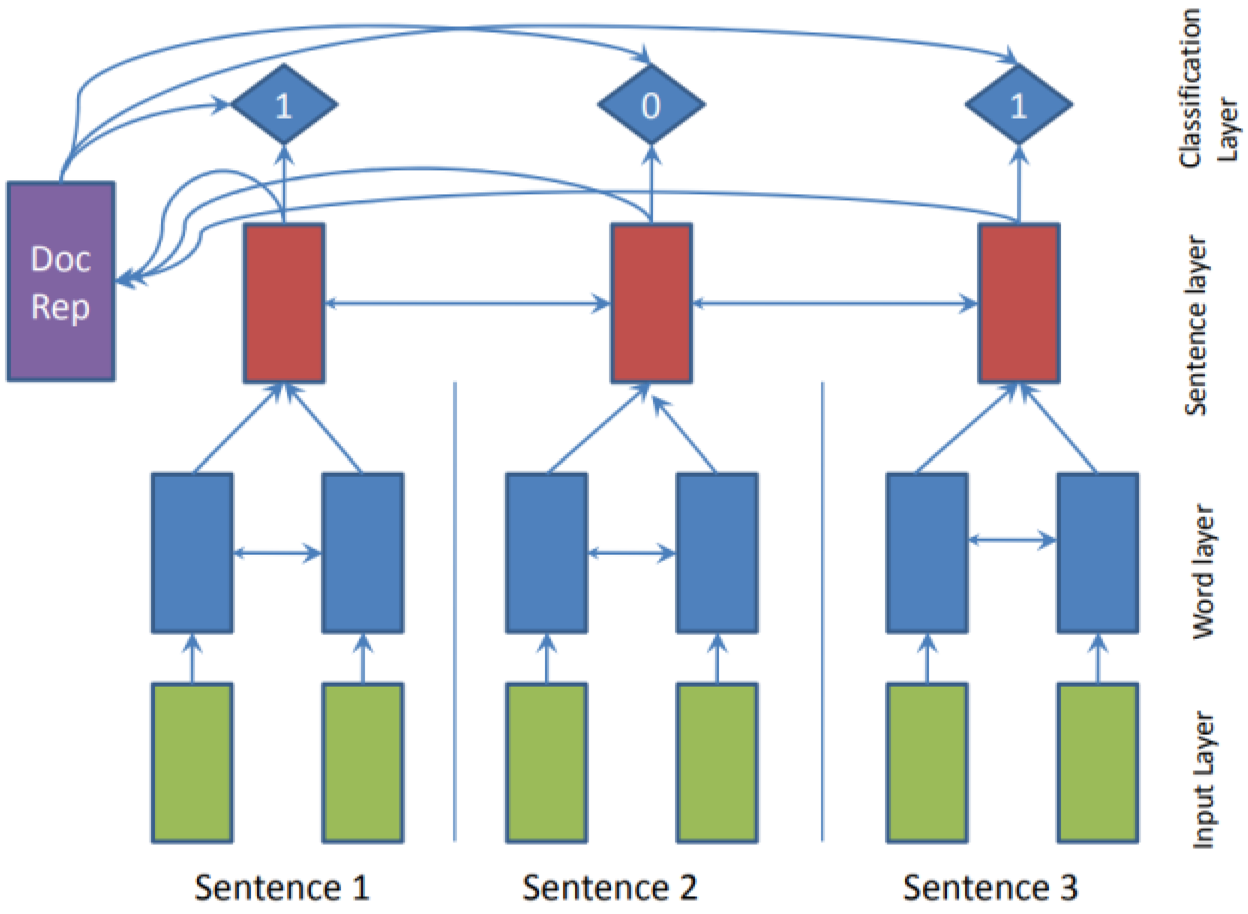


Paradigm Shift in Text Summarization

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

Paradigm Shift in Text Summarization

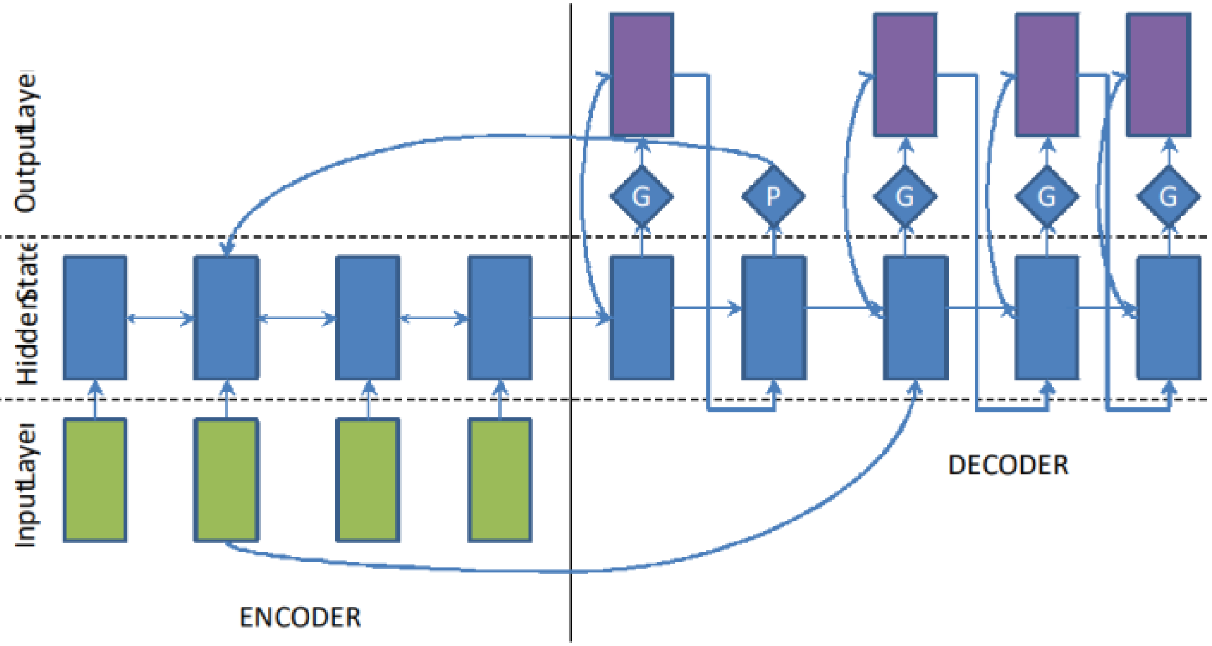
- **Traditional Paradigm:**
  - SeqLab (extractive)
  - Seq2Seq (abstractive)
- **Shifted to / Unified in…**
  - Matching (extractive)
  - (M)LM (abstractive)



SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents. AAAI 2017

Paradigm Shift in Text Summarization

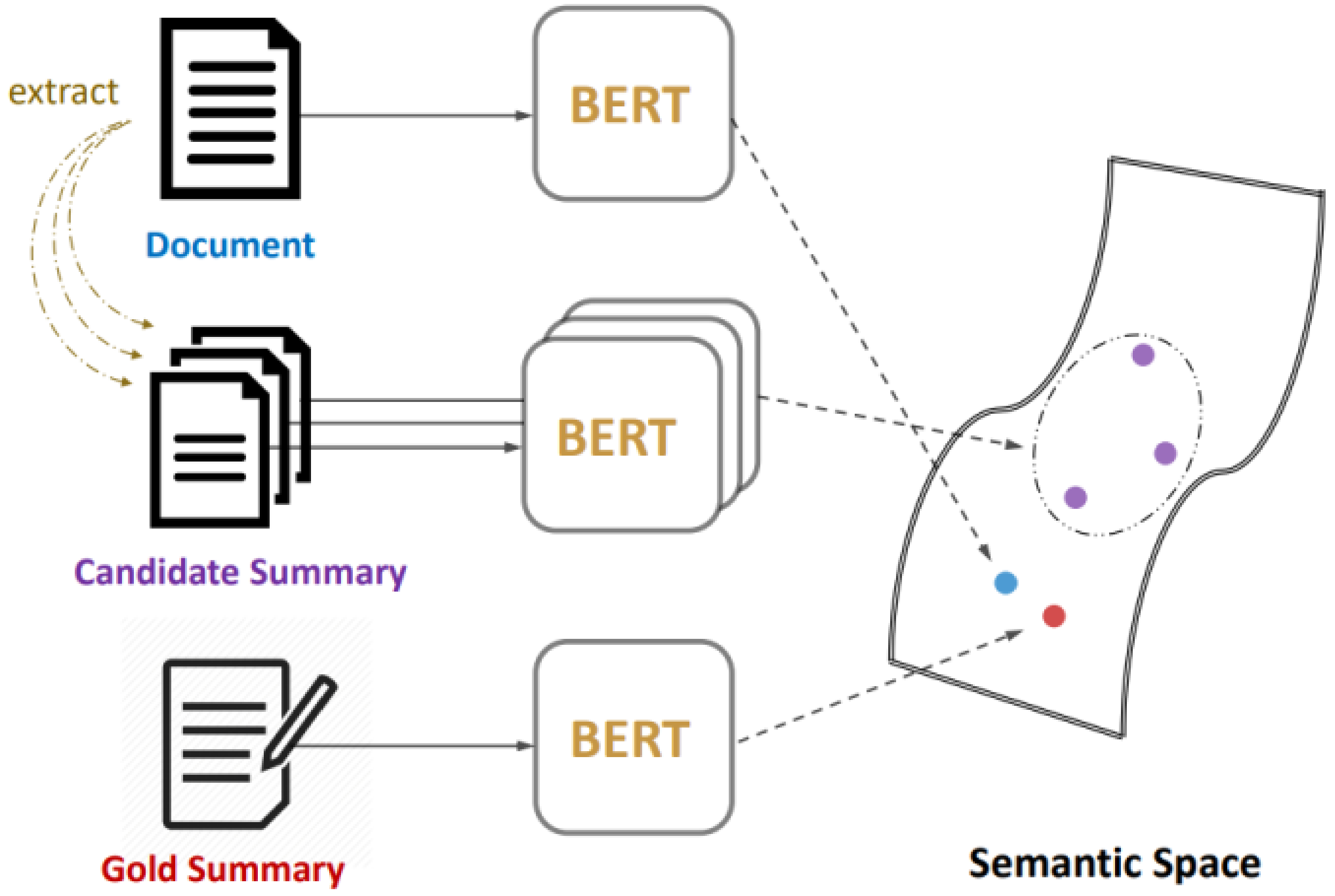
- **Traditional Paradigm:**
  - SeqLab (extractive)
  - Seq2Seq (abstractive)
- **Shifted to / Unified in…**
  - Matching (extractive)
  - (M)LM (abstractive)



Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. CoNLL 2016

Paradigm Shift in Text Summarization

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



Extractive Summarization as Text Matching. ACL 2020

Paradigm Shift in Text Summarization

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

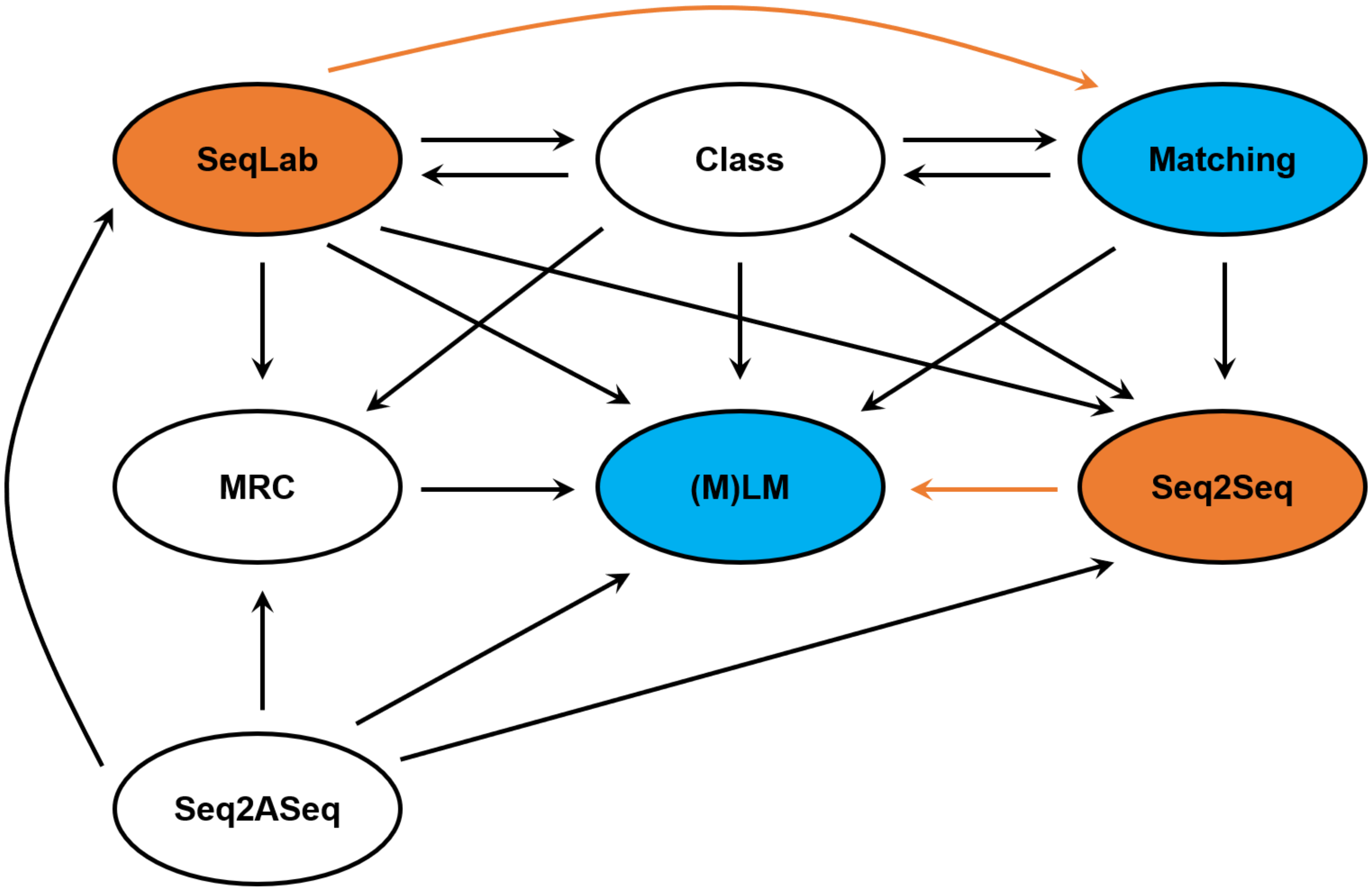
```
<!DOCTYPE html>
<html>
    <title> <mask>12 </title>
    <body>
        ˜ south korea on monday announced sweeping
            tax reforms , including income and
            corporate tax cuts to boost growth by
            stimulating sluggish private
            consumption and business investment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

↓

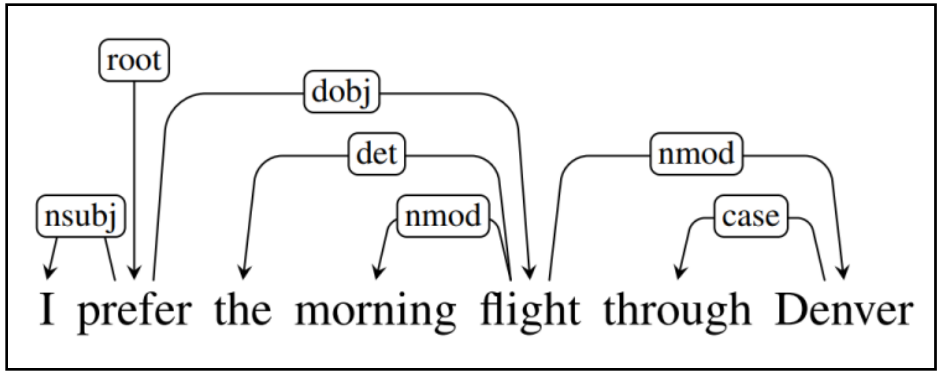
```
<!DOCTYPE html>
<html>
    <title> ˜ South Korea Announces Tax Reforms To
          Boost Economic Growth ˜ </title>
    <body>
        ˜ south korea on monday announced sweeping
            tax reforms...
    </body>
</html>
```

HTLM: Hyper-Text Pre-Training and Prompting of Language Models. https://arxiv.org/abs/2107.06955

Paradigm Shift in Text Summarization



Paradigm Shift in Parsing

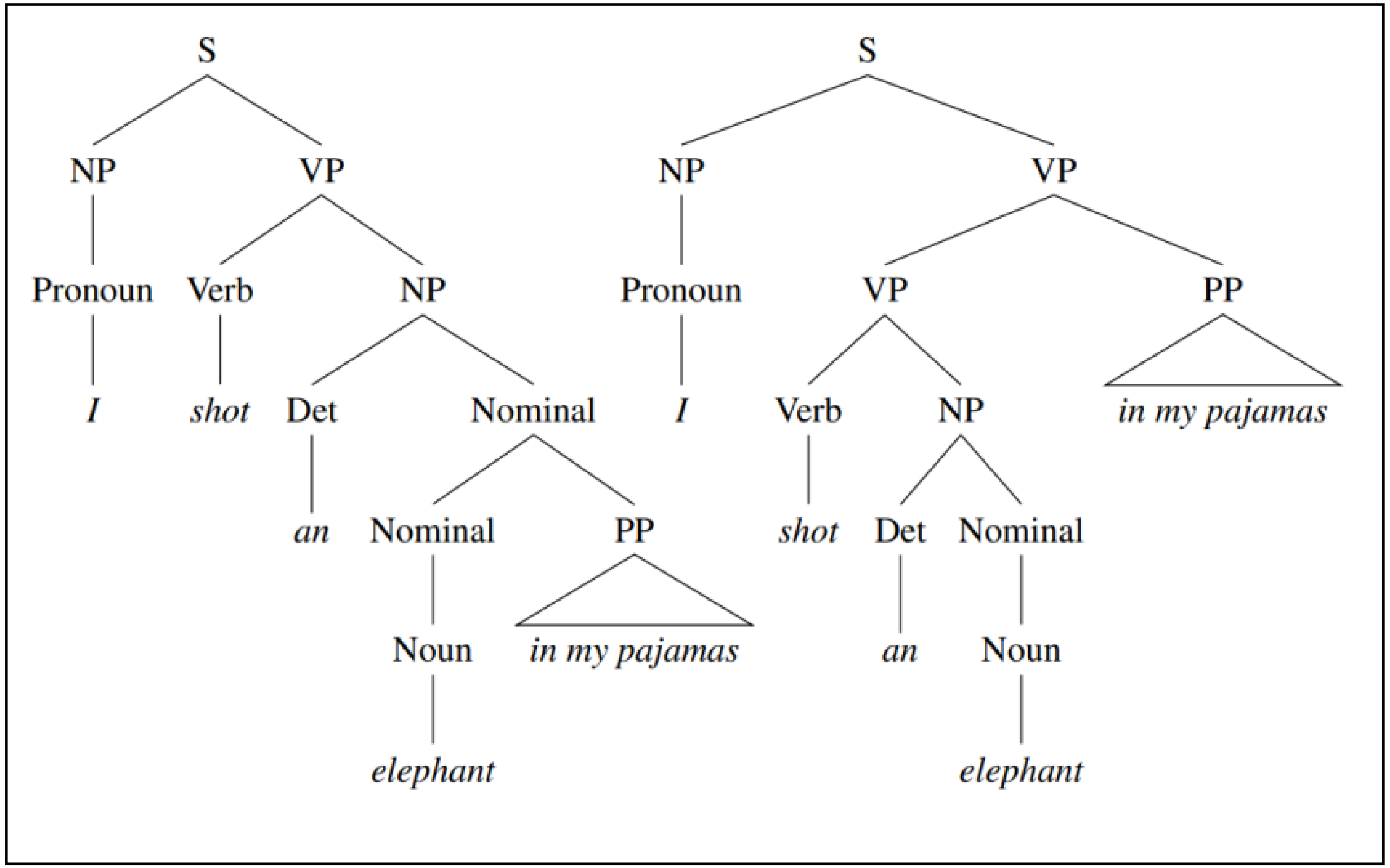


Dependency Parsing

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;
```

Semantic Parsing



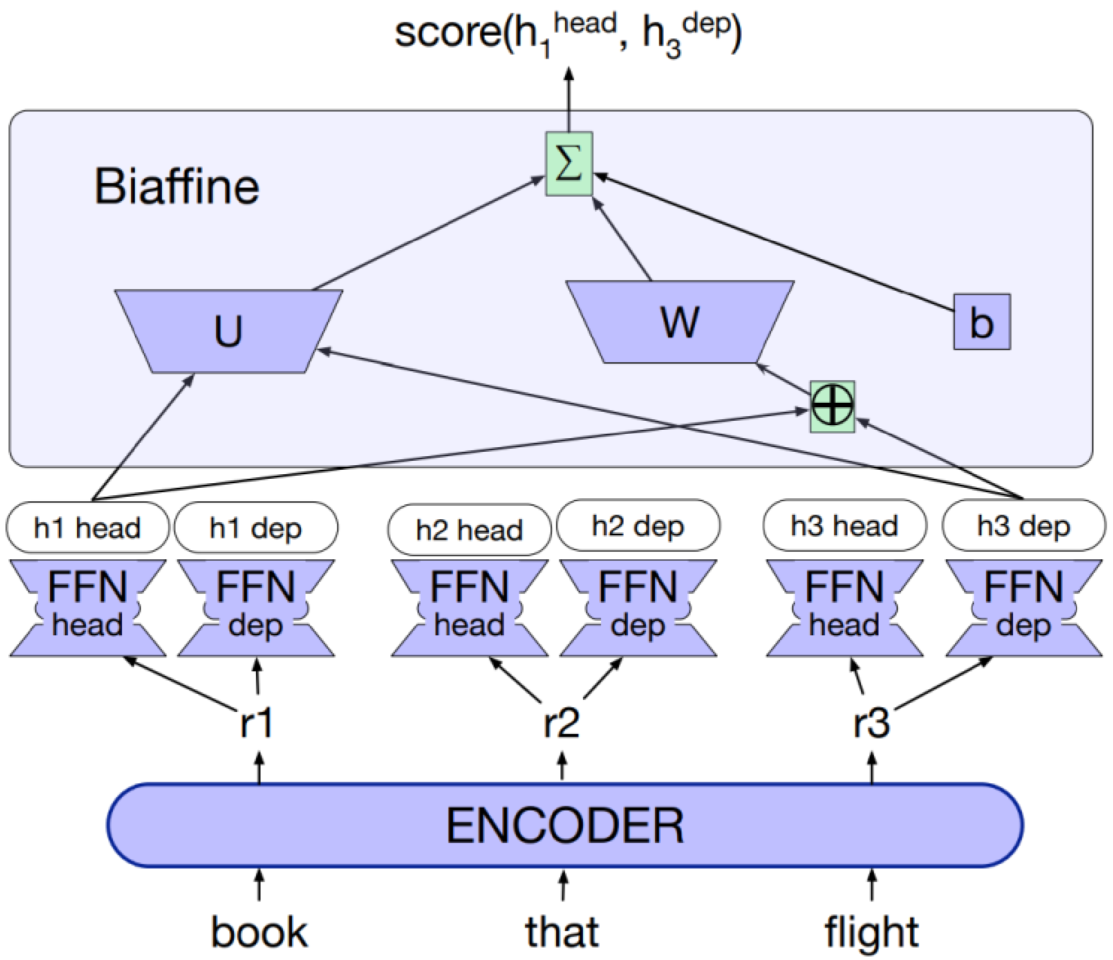
Constituency Parsing

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

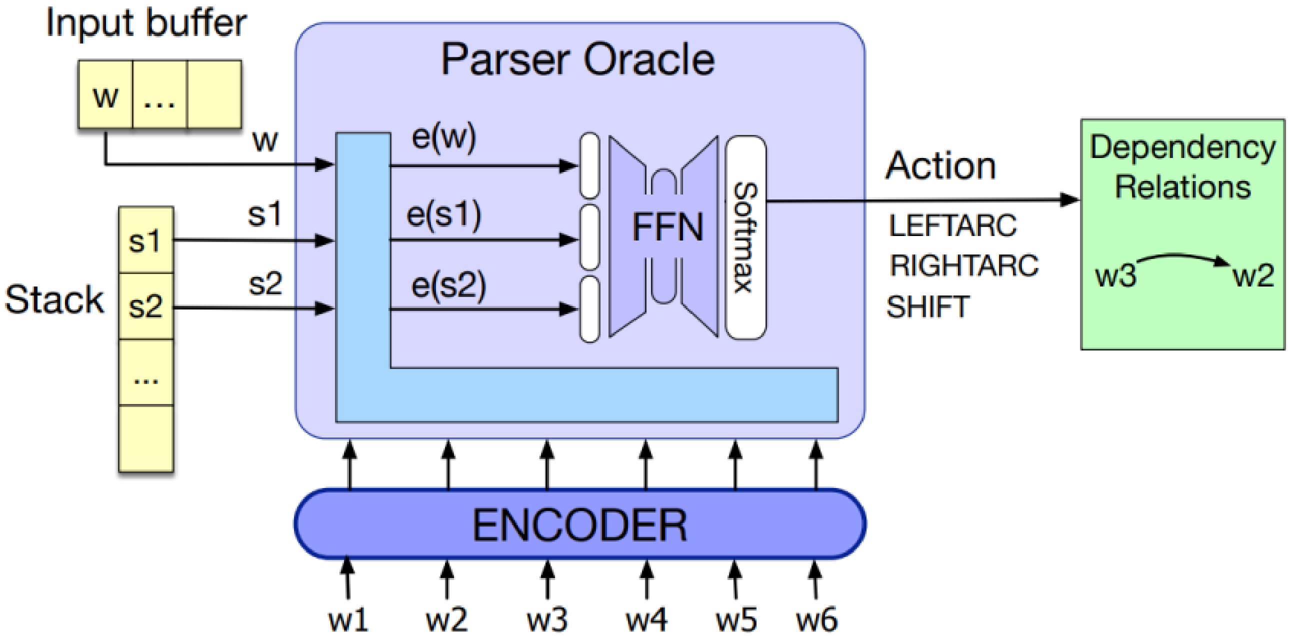
- **Traditional Paradigm:**
  - Class (graph-based)
  - Seq2ASeq (transition-based)
- **Shifted to / Unified in…**
  - SeqLab
  - Seq2Seq
  - (M)LM
  - MRC



https://web.stanford.edu/~jurafsky/slp3/14.pdf

Paradigm Shift in Parsing

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

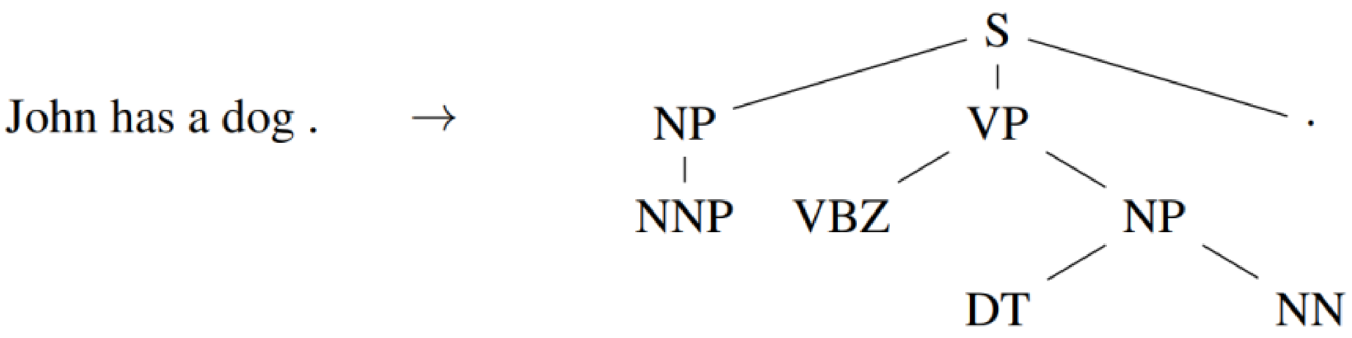


https://web.stanford.edu/~jurafsky/slp3/14.pdf

Paradigm Shift in Parsing

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

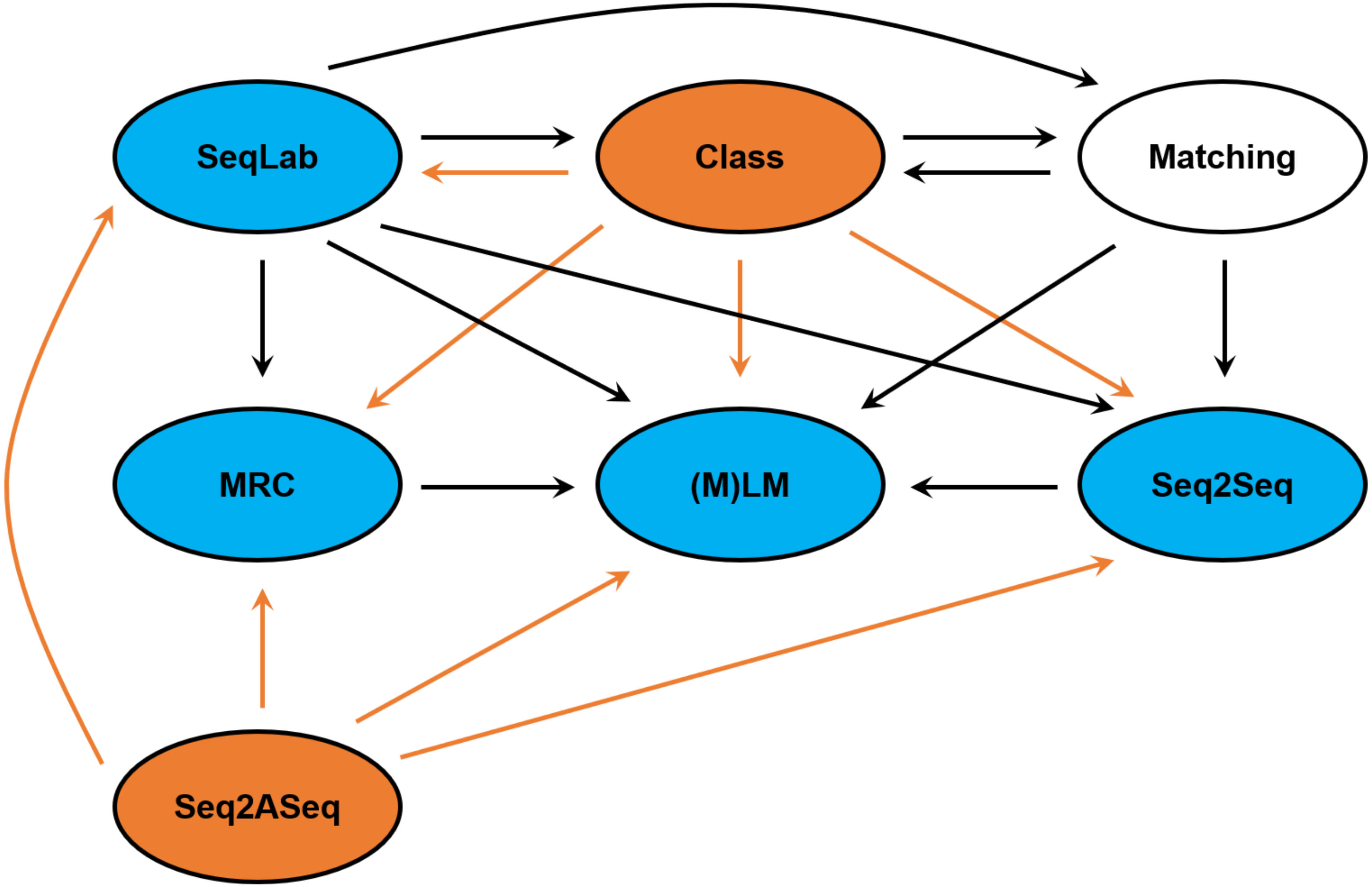
Linearize a parsing tree:



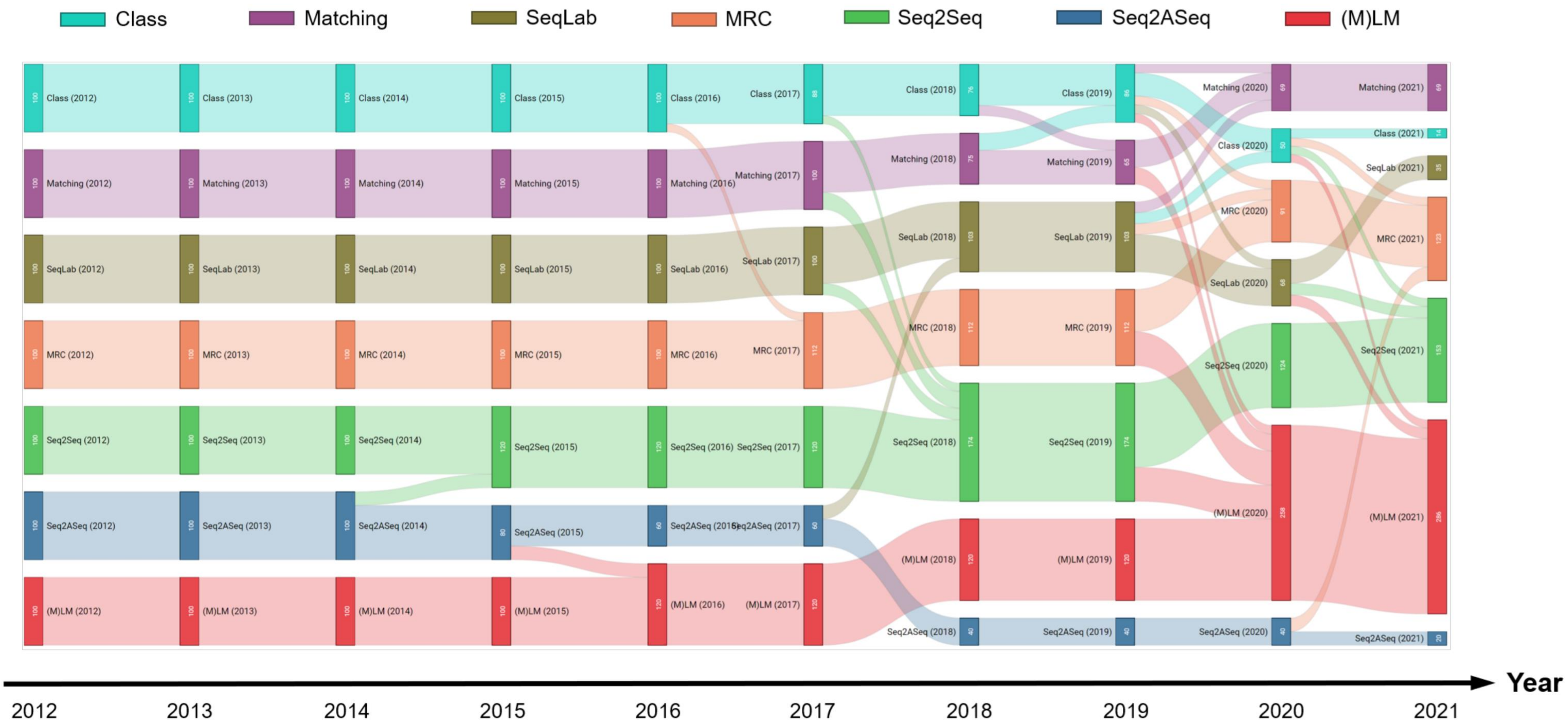
John has a dog . → (S (NP NNP $)_{NP}$ (VP VBZ (NP DT NN $)_{NP}$ $)_{VP}$. $)_{S}$

Grammar as a Foreign Language. NIPS 2015

Paradigm Shift in Parsing



Trends of Paradigm Shift



Online version: https://txsun1997.github.io/nlp-paradigm-shift/sankey.html

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

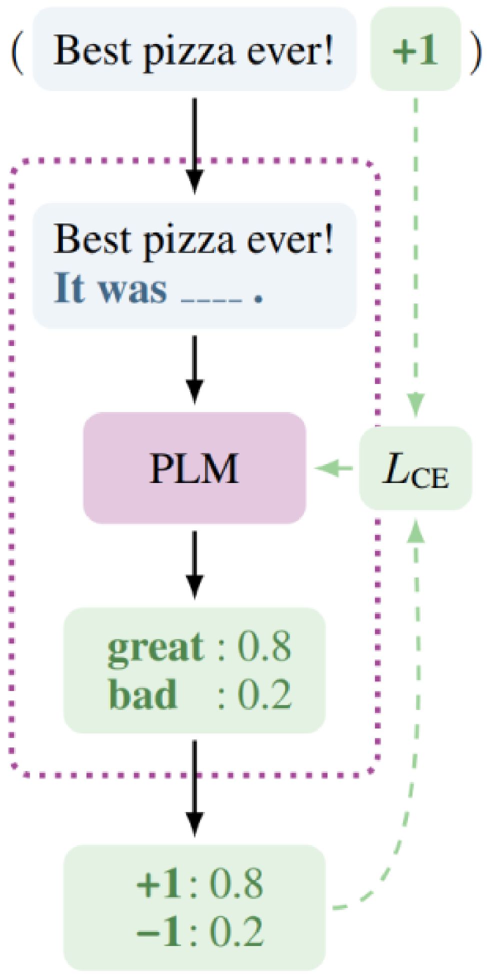
Why Unified Paradigm?

- **Data Efficiency**
  - Task-specific models usually required large-scale annotated data, while unified models can achieve considerable performance with much less data
- **Generalization**
  - Unified models can easily generalize to unseen tasks
- **Convenience**
  - Unified models are easier and cheaper to deploy and serve. They are born to be commercial black-box APIs

Potential Unified Paradigms

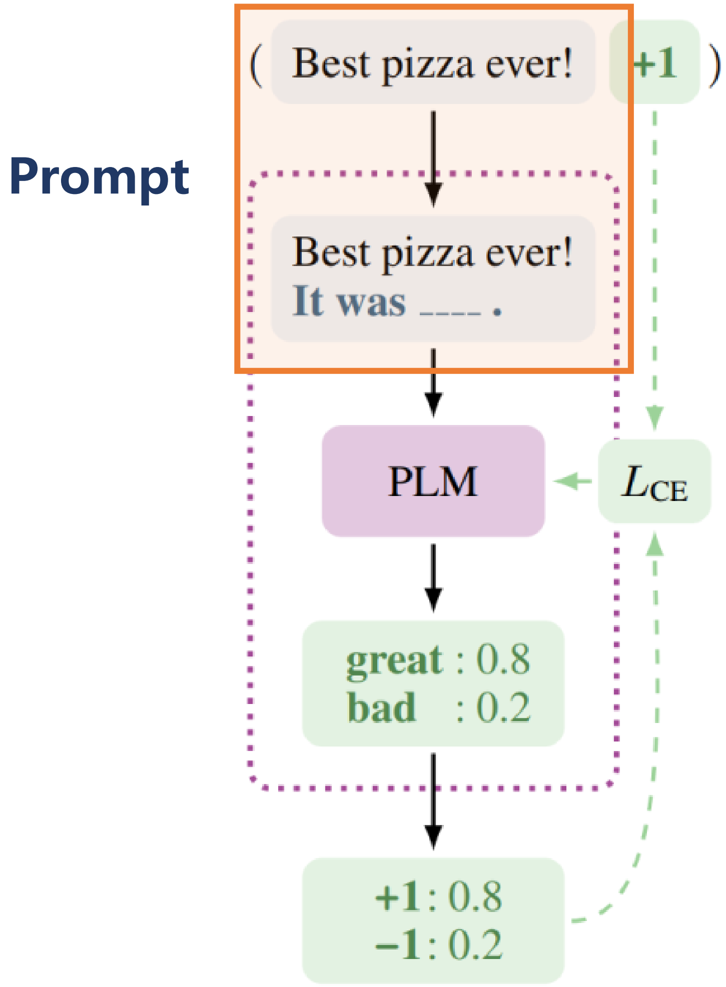
- (M)LM
- Matching
- MRC
- Seq2Seq

(M)LM



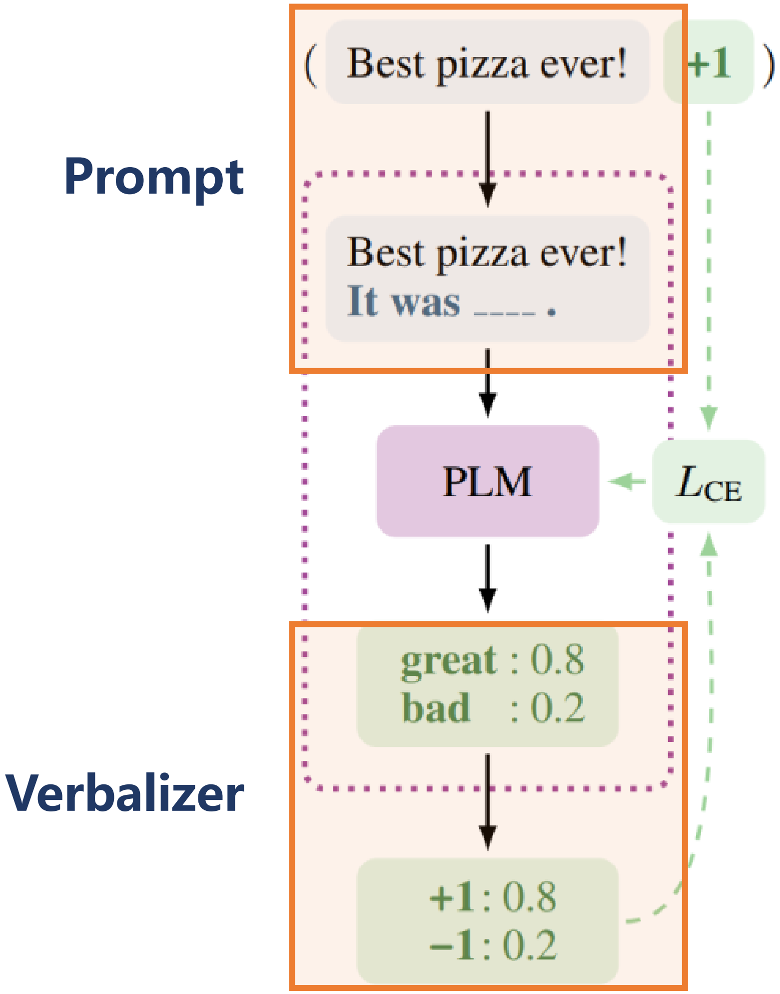
Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. EACL 2021

(M)LM



Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. EACL 2021

(M)LM



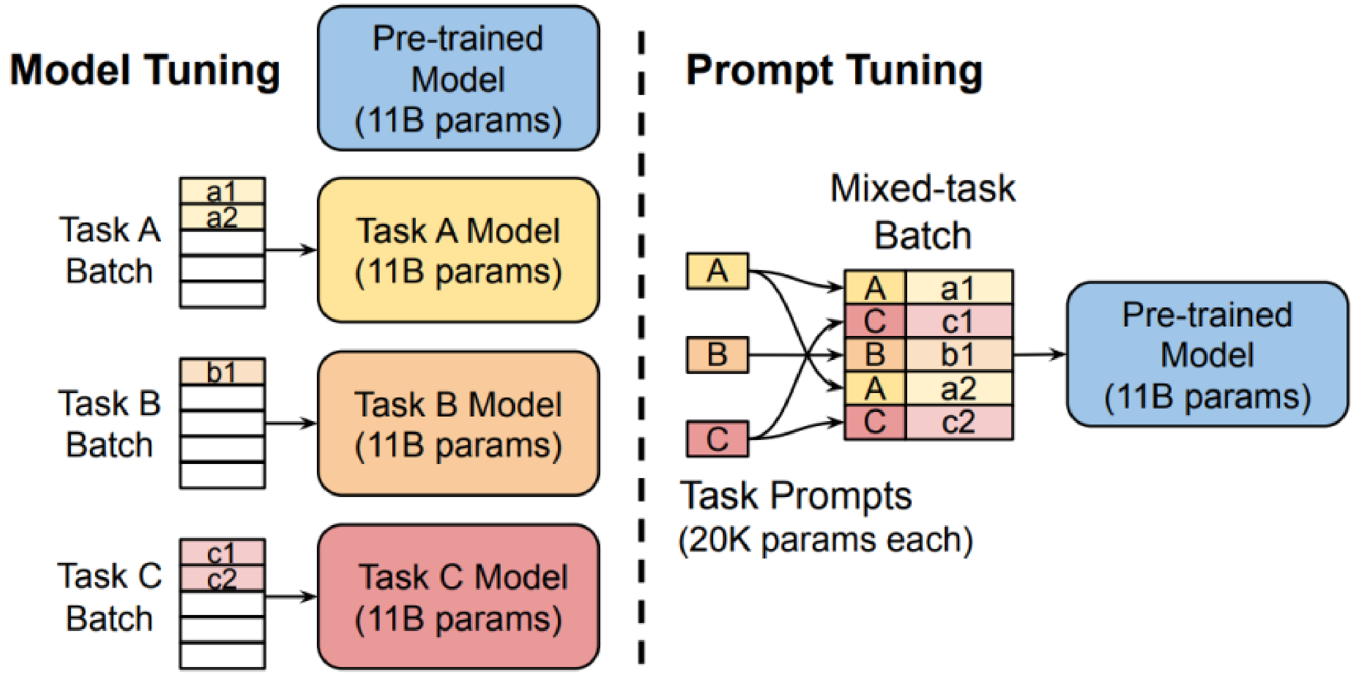
Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference. EACL 2021

(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

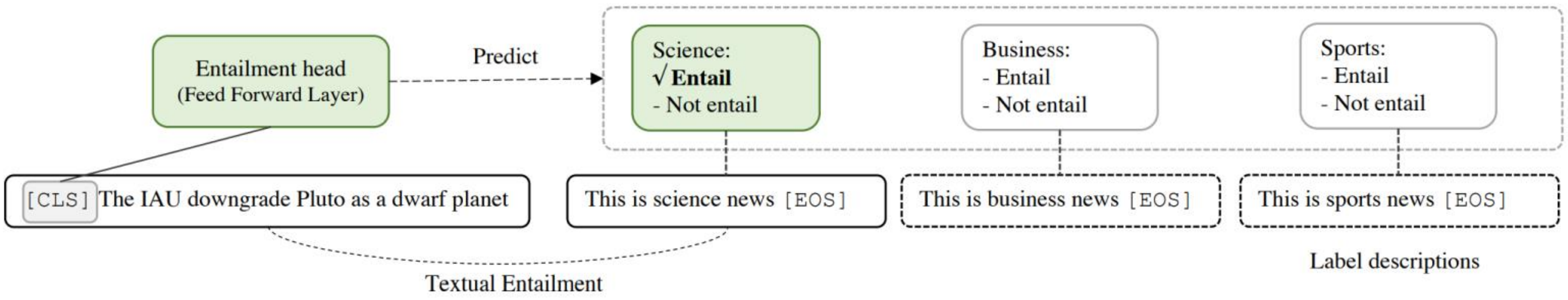
(M)LM

- **Parameter-Efficient Tuning**
  - Only tuning prompts can match the performance of fine-tuning
  - Mixed-task inference



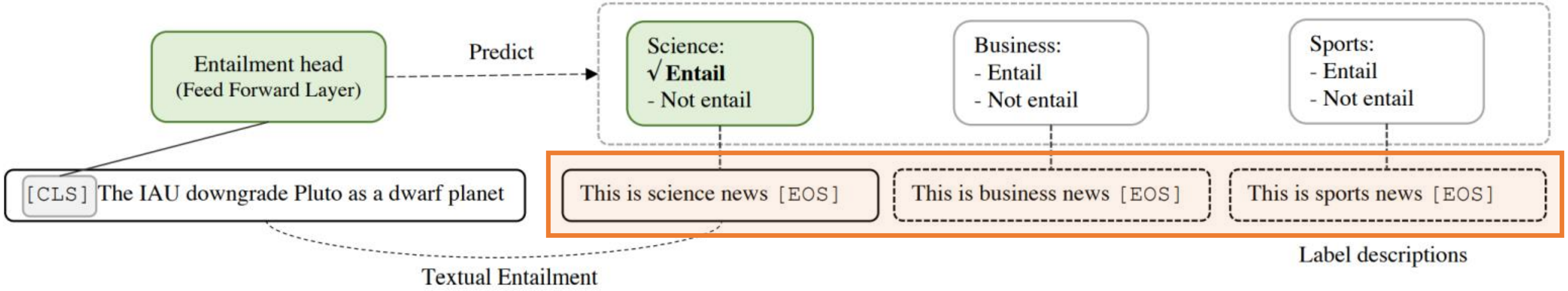
The Power of Scale for Parameter-Efficient Prompt Tuning. https://arxiv.org/abs/2104.08691

Matching



Entailment as Few-Shot Learner. https://arxiv.org/abs/2104.14690

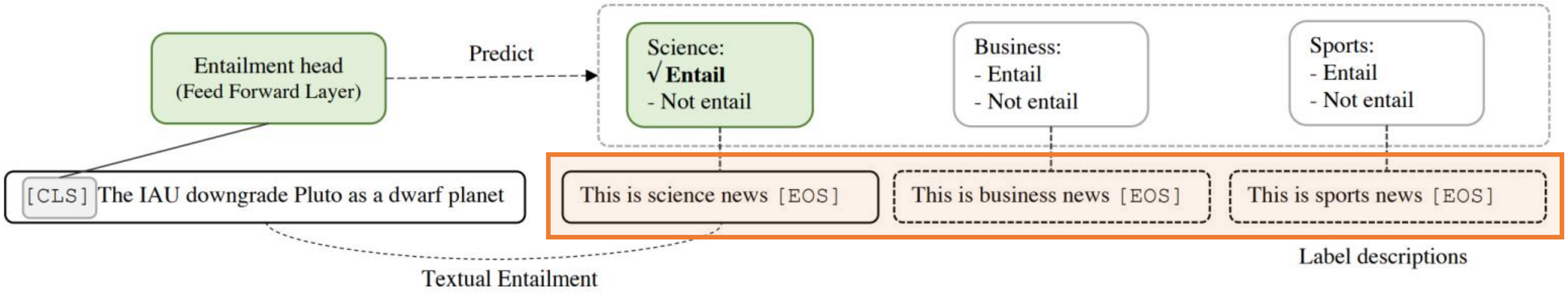
Matching



Label Description

Entailment as Few-Shot Learner. https://arxiv.org/abs/2104.14690

Matching

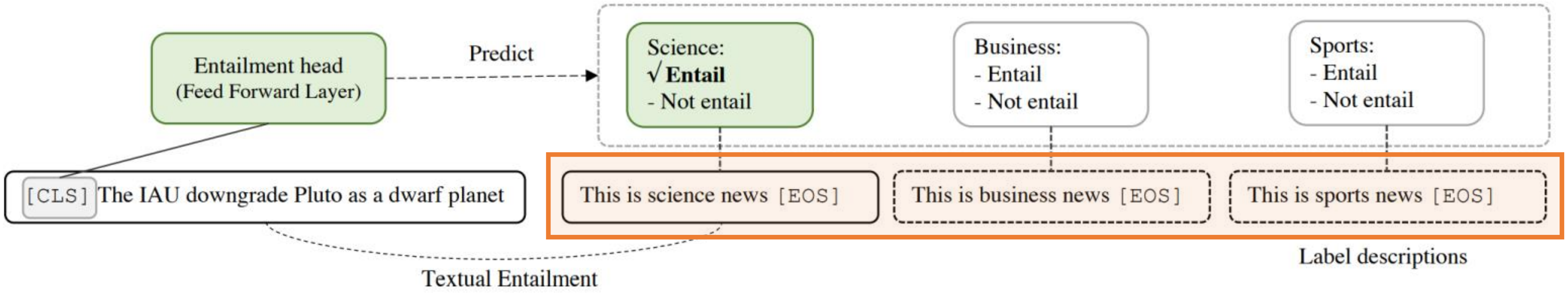


Label Description

- **Label Description**
  - Manually designed (can be the same as prompt)
  - Generated by reinforcement learning (Chai et al.)

Entailment as Few-Shot Learner. https://arxiv.org/abs/2104.14690

Matching



Label Description

- **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

Entailment as Few-Shot Learner. https://arxiv.org/abs/2104.14690

(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

MRC

- **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
What is a major importance of Southern California in relation to California and the US?	...Southern California is a major economic center for the state of California and the US....	major economic center
What is the translation from English to German?	Most of the planet is ocean water.	Der Großteil der Erde ist Meerwasser
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...
Hypothesis: Product and geography are what make cream skimming work. Entailment, neutral, or contradiction?	Premise: Conceptually cream skimming has two basic dimensions – product and geography.	Entailment
Is this sentence positive or negative?	A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.	positive

Question	Context	Answer
What has something experienced?	Areas of the Baltic that have experienced eutrophication.	eutrophication
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 Bernie Wrightson.	Bernie Wrightson
What is the change in dialogue state?	Are there any Eritrean restaurants in town?	food: Eritrean
What is the translation from English to SQL?	The table has column names... Tell me what the notes are for South Australia	SELECT notes from table WHERE 'Current Slogan' = 'South Australia'
Who had given help? Susan or Joan?	Joan made sure to thank Susan for all the help she had given.	Susan

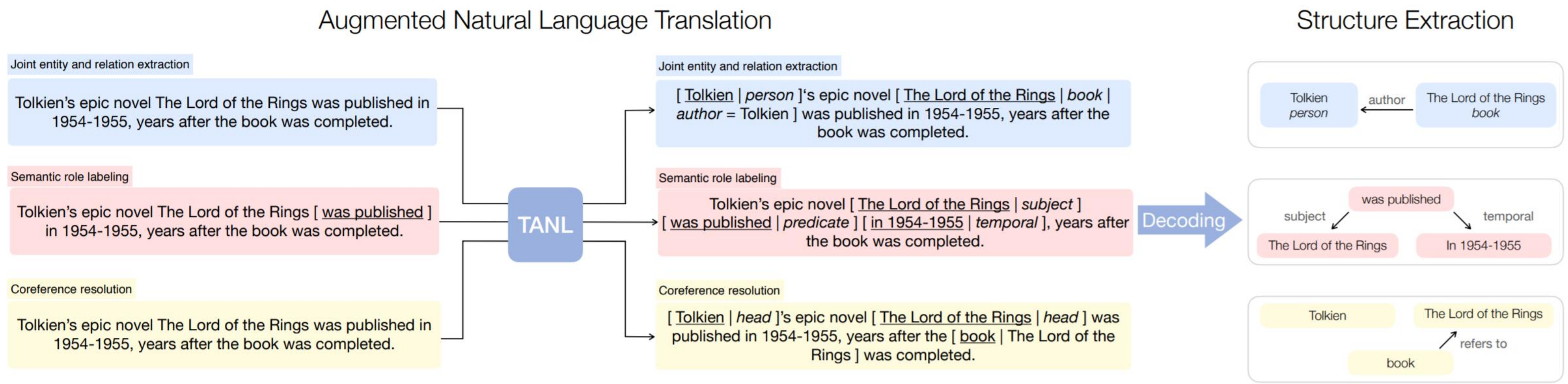
The Natural Language Decathlon: Multitask Learning as Question Answering. https://arxiv.org/abs/1806.08730

MRC

- **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?**
  - All NLP tasks as open-domain QA?
  - Dense Passage Retriever (DPR) may help (REALM, RAG, …)
  - Question $\xrightarrow[retrieval]{}$ Context $\xrightarrow[MRC]{}$ Answer

Seq2Seq

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



Structured prediction as translation between augmented natural languages. ICLR 2021

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

https://arxiv.org/abs/2109.12575

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