

The image is a collage. On the left, there are several Transformers robots. A large, dark grey and red robot is in the background. In the foreground, there are blue and red robot parts, and a yellow and black robot. On the right, there is a close-up of Bert the Ernie puppet, who is yellow with a large orange nose, wide eyes, and a tuft of black hair. He is wearing a colorful striped shirt.

BERT

李宏毅

Hung-yi Lee

## 1-of-N Encoding

apple = [ 1 0 0 0 0 ]

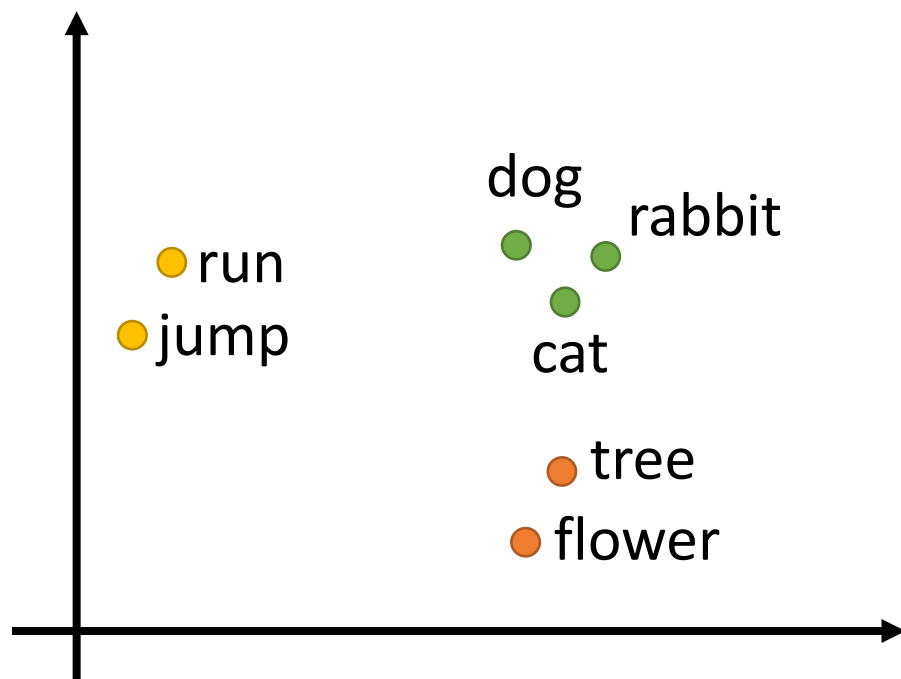
bag = [ 0 1 0 0 0 ]

cat = [ 0 0 1 0 0 ]

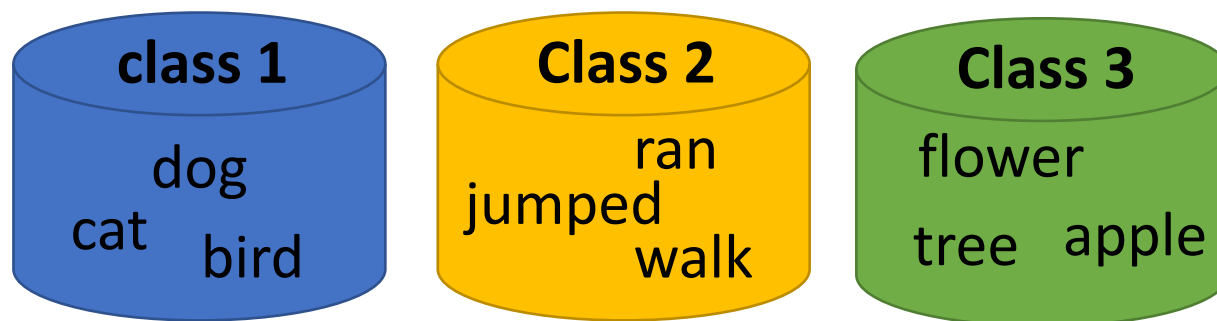
dog = [ 0 0 0 1 0 ]

elephant = [ 0 0 0 0 1 ]

## Word Embedding



## Word Class



# A word can have multiple senses.

Have you paid that money to the bank yet ?

It is safest to deposit your money in the bank .

The victim was found lying dead on the river bank .

They stood on the river bank to fish.

The hospital has its own blood bank.

The third sense or not?

# More Examples



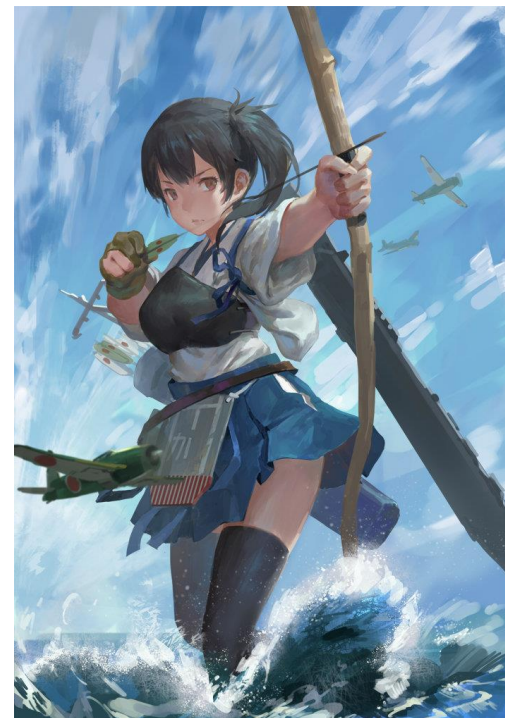
他是尼祿



她也是尼祿

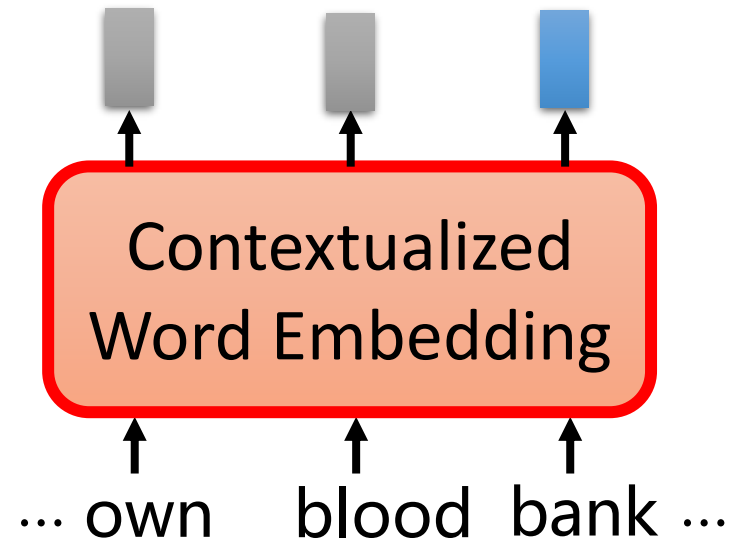
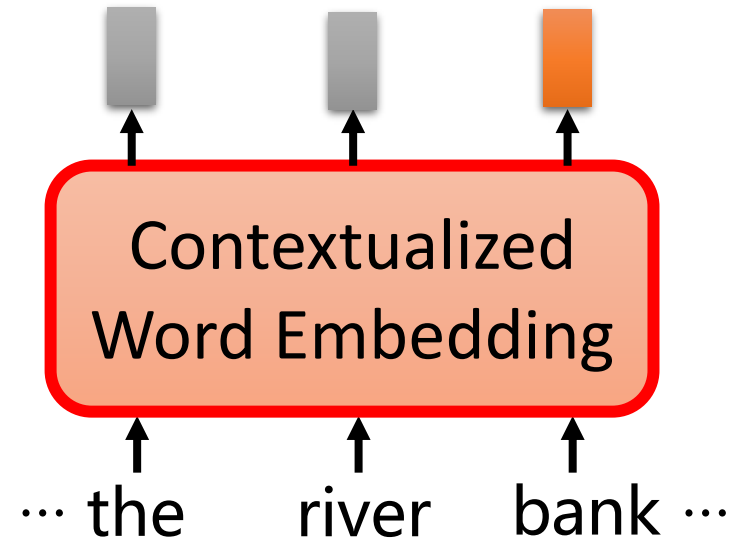
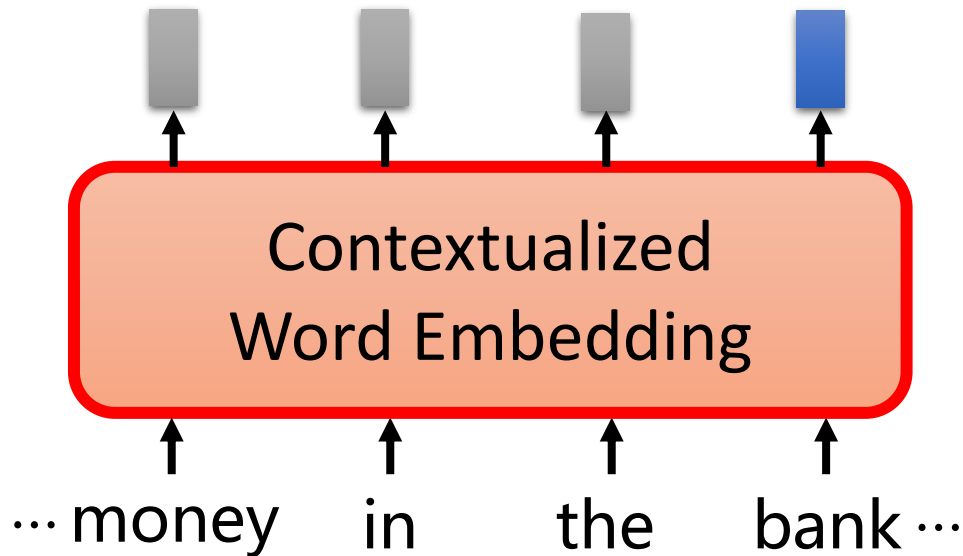


這是  
加賀號護衛艦



這也是加賀  
號護衛艦

# Contextualized Word Embedding



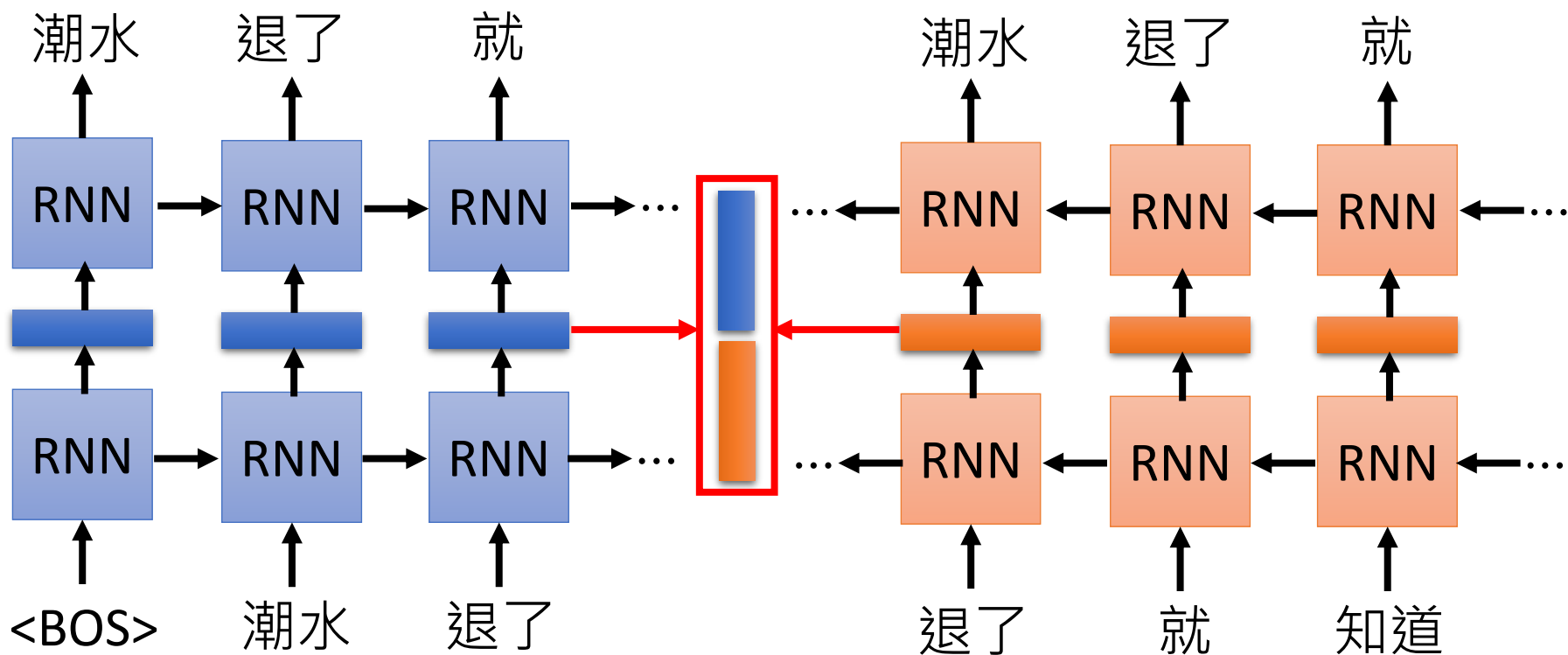
# Embeddings from Language Model (ELMO)

<https://arxiv.org/abs/1802.05365>



- RNN-based language models (trained from lots of sentences)

e.g. given “潮水 退了 就 知道 誰 沒穿 褲子”

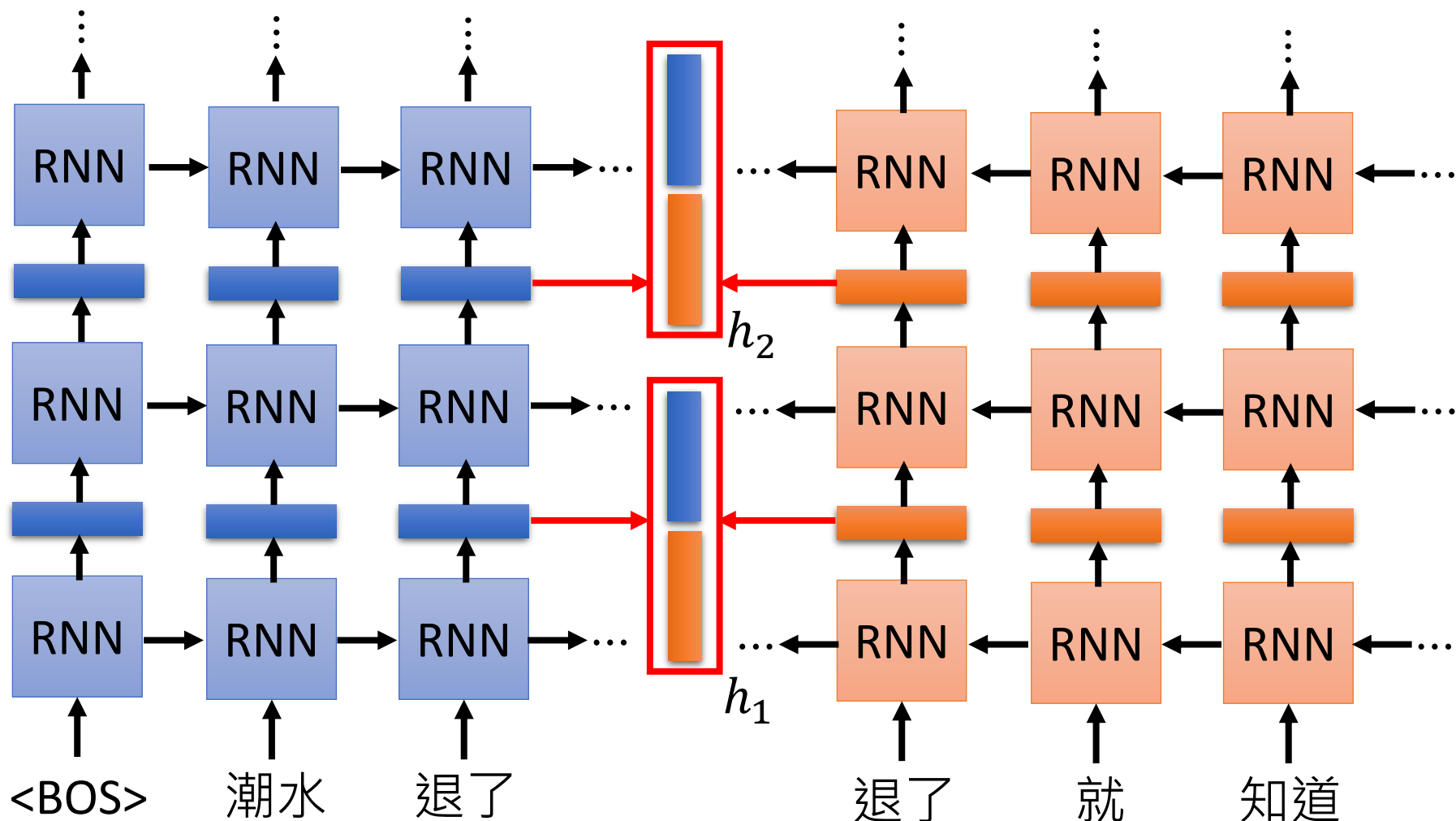




# ELMO

Each layer in deep LSTM can generate a latent representation.

Which one should we use???

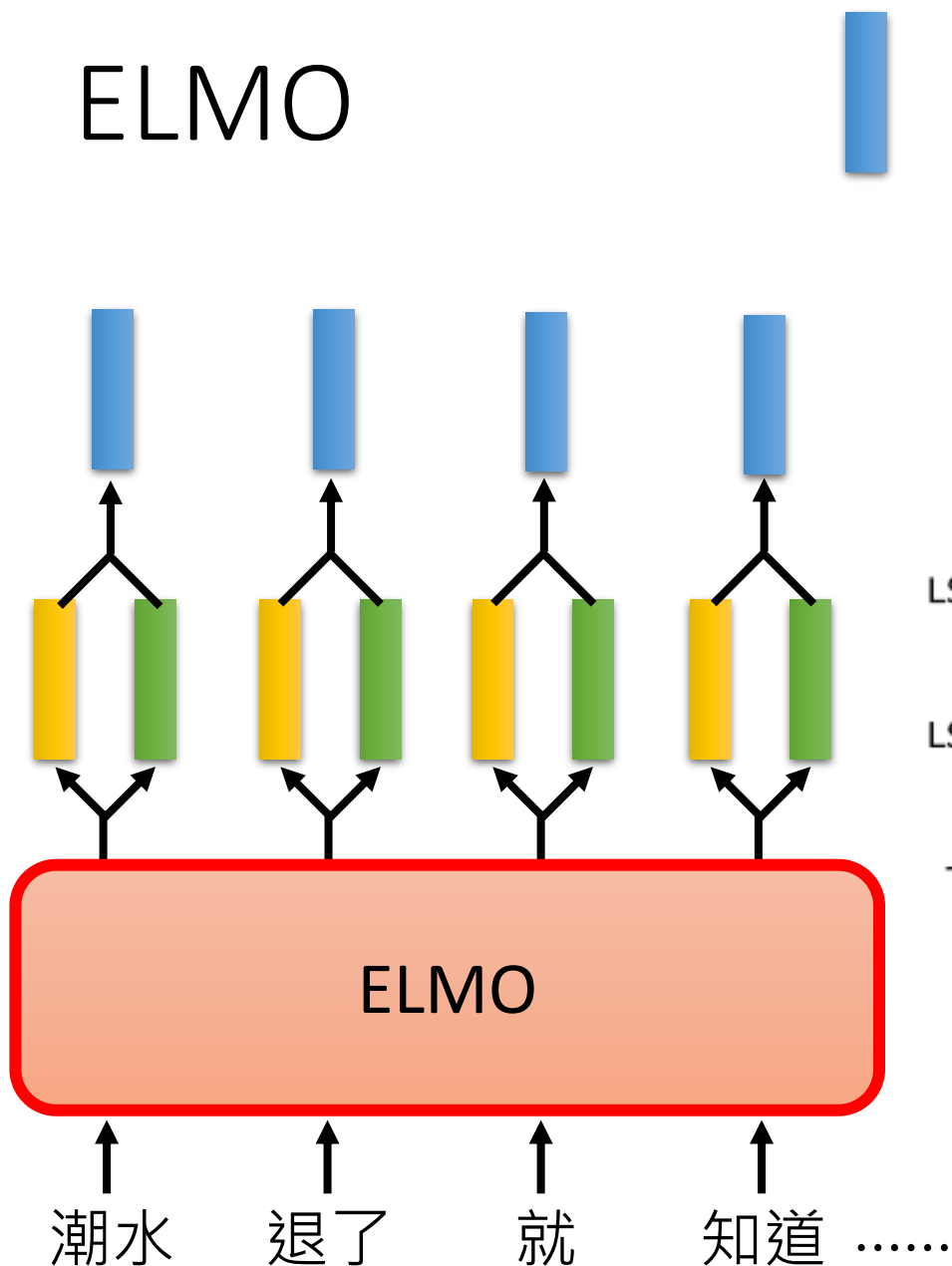




我全都要

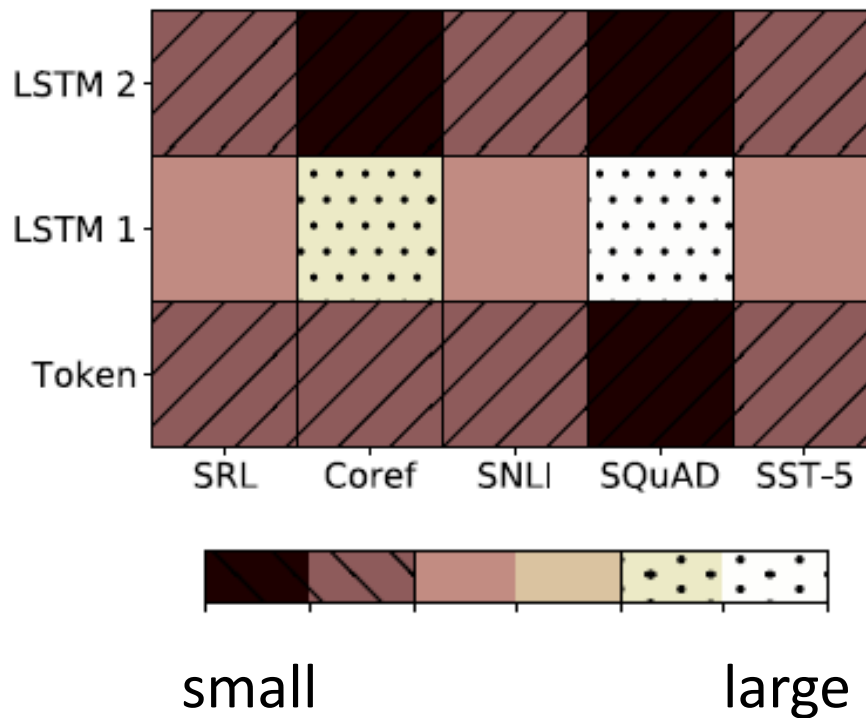


# ELMO



$$\text{Blue Bar} = \alpha_1 \text{Yellow Bar} + \alpha_2 \text{Green Bar}$$

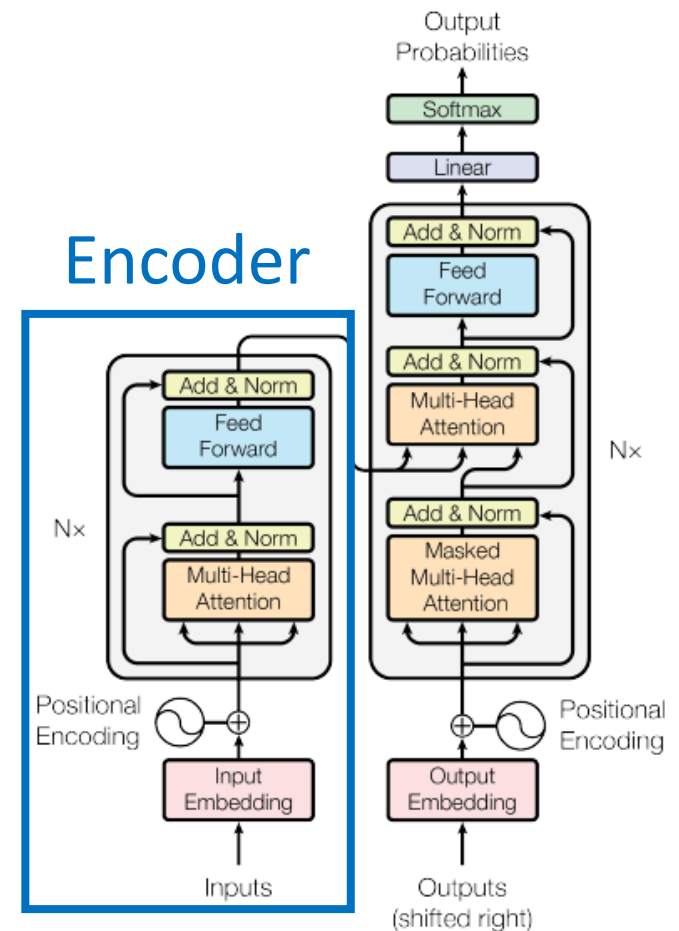
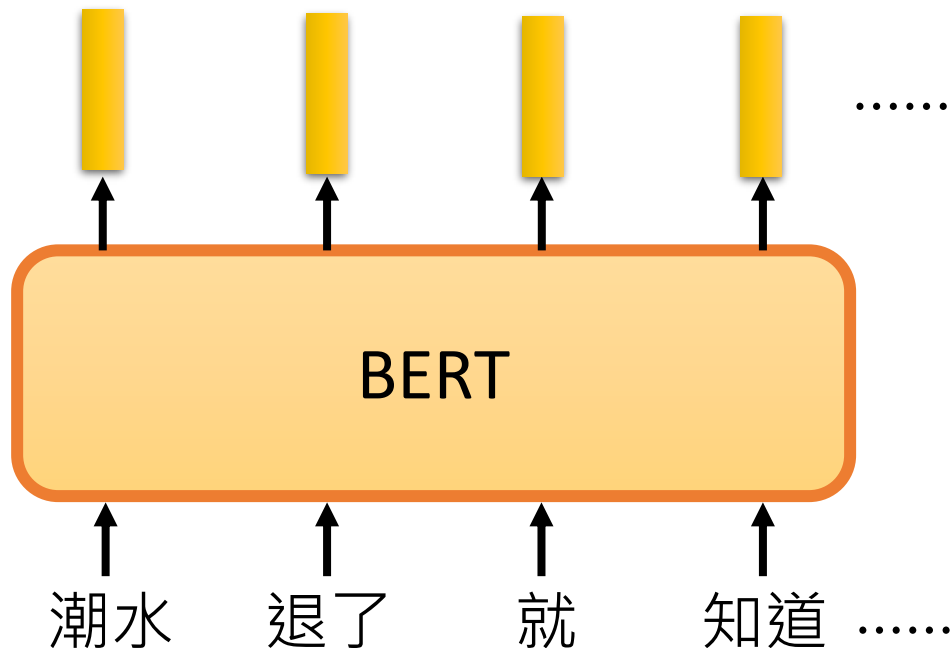
Learned with the downstream tasks



# Bidirectional Encoder Representations from Transformers (BERT)

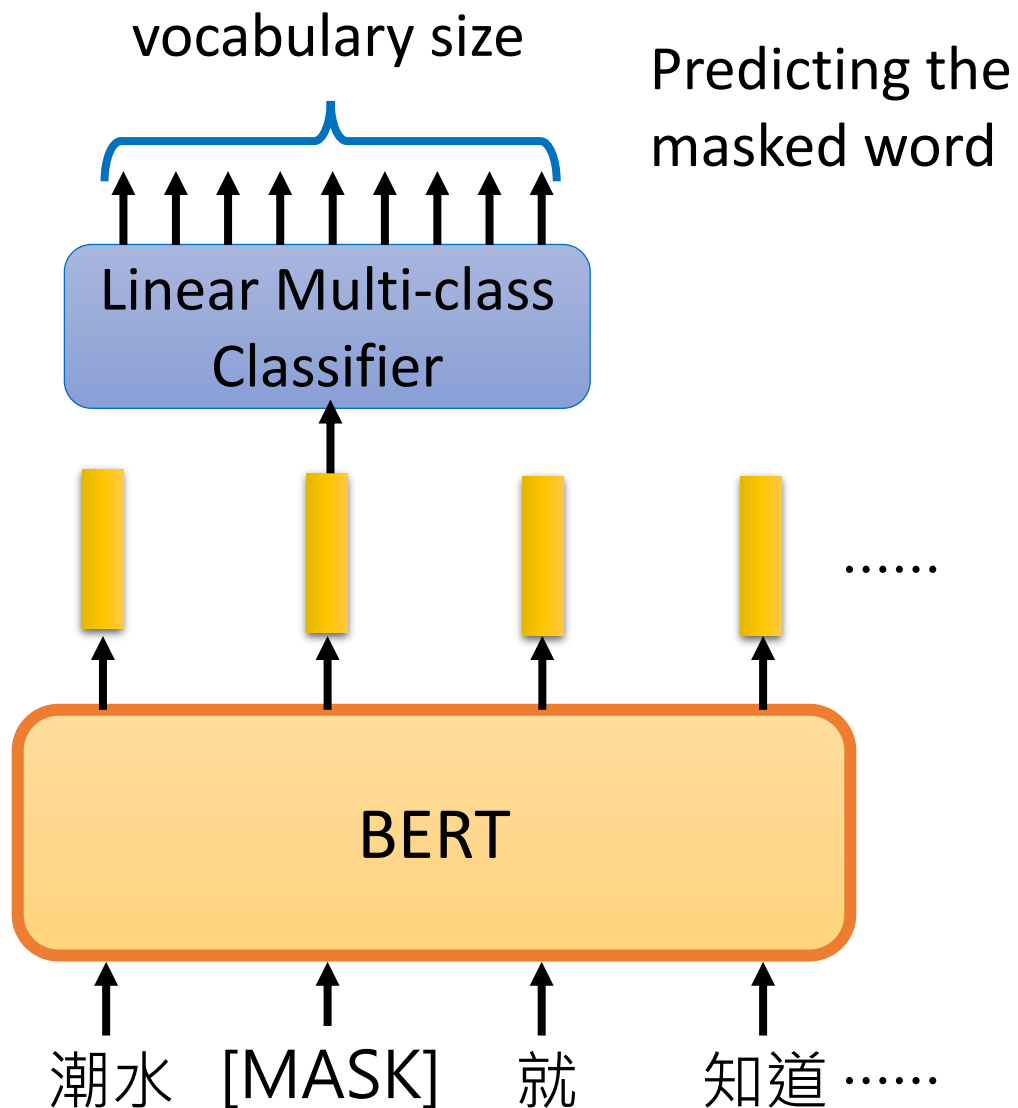


- BERT = Encoder of Transformer  
Learned from a large amount of text  
without annotation



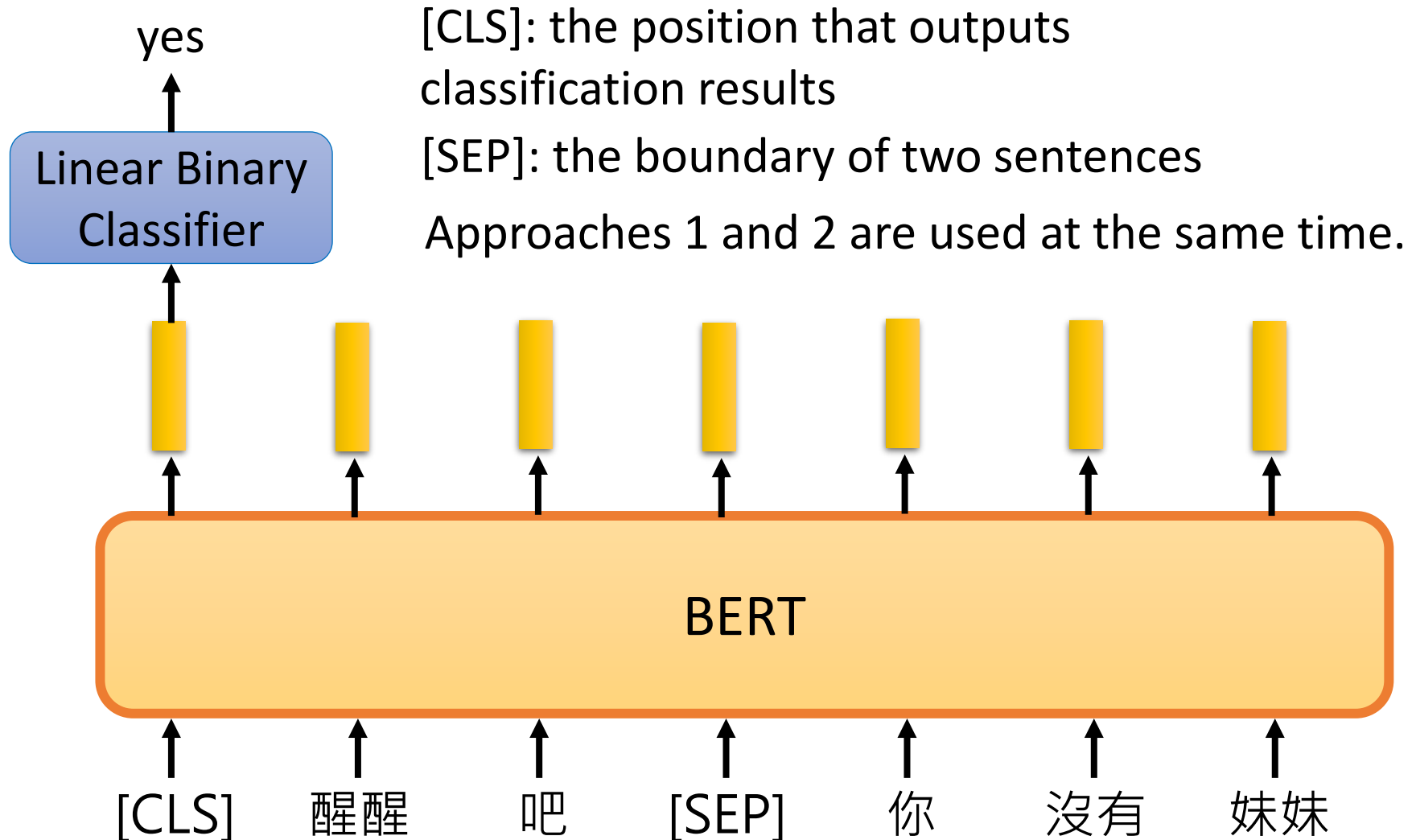
# Training of BERT

- Approach 1:  
Masked LM



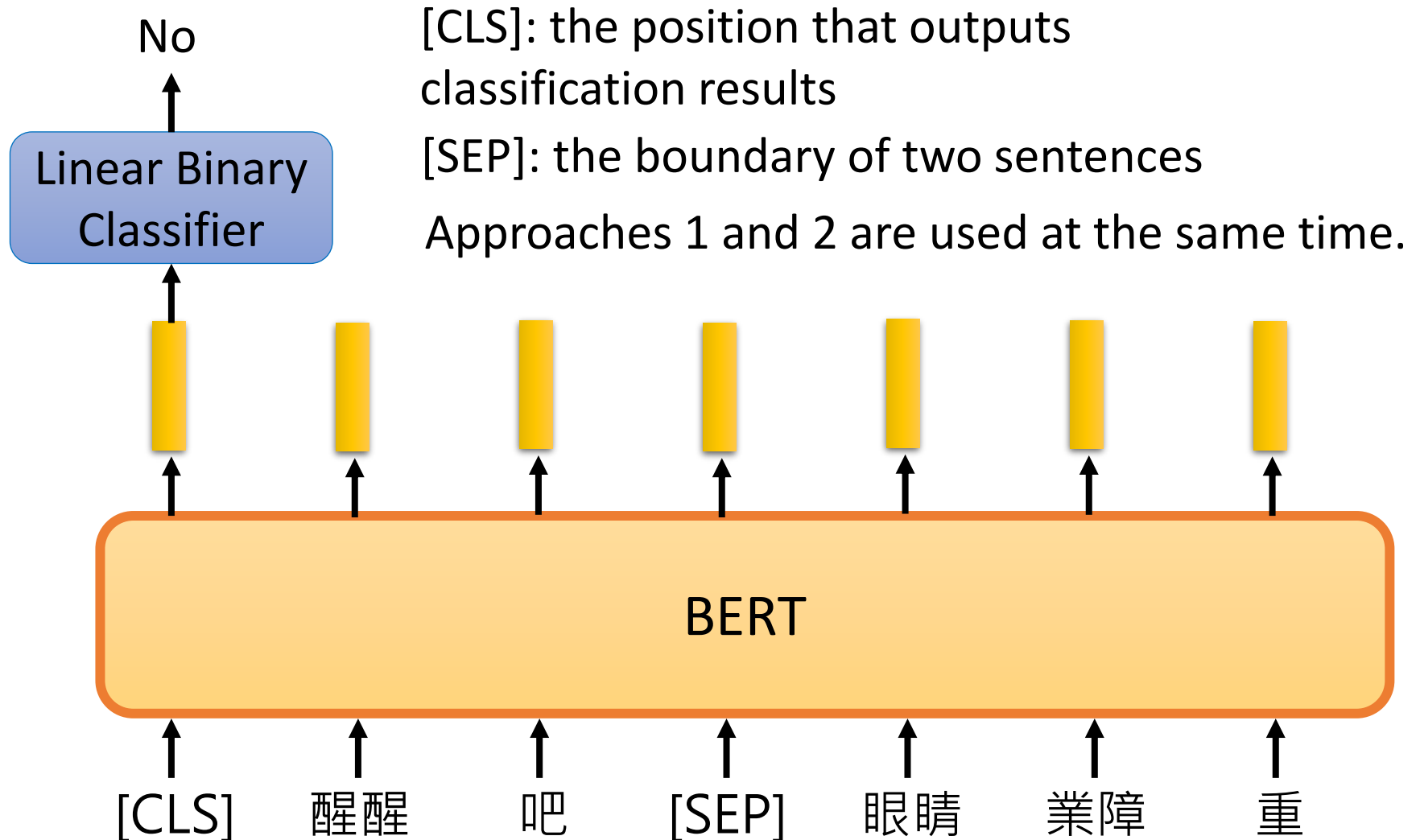
# Training of BERT

## Approach 2: Next Sentence Prediction



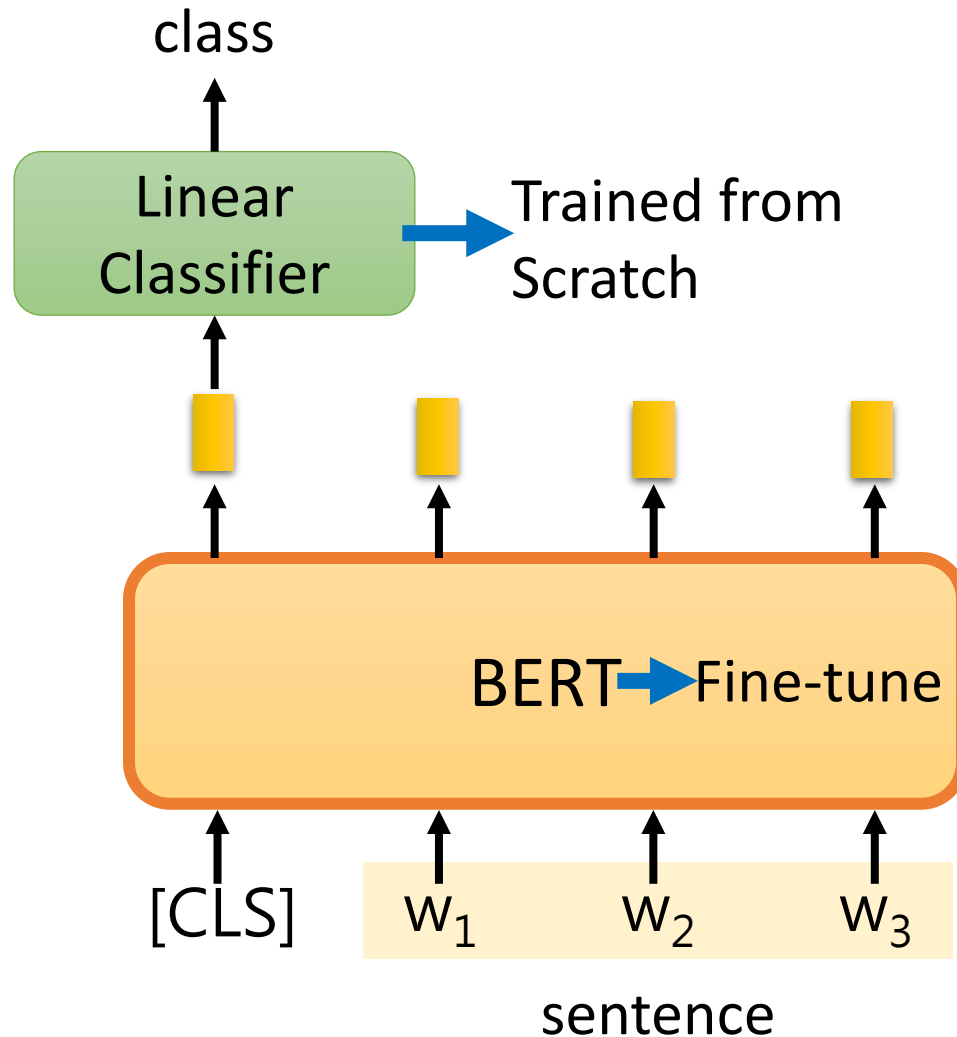
# Training of BERT

## Approach 2: Next Sentence Prediction





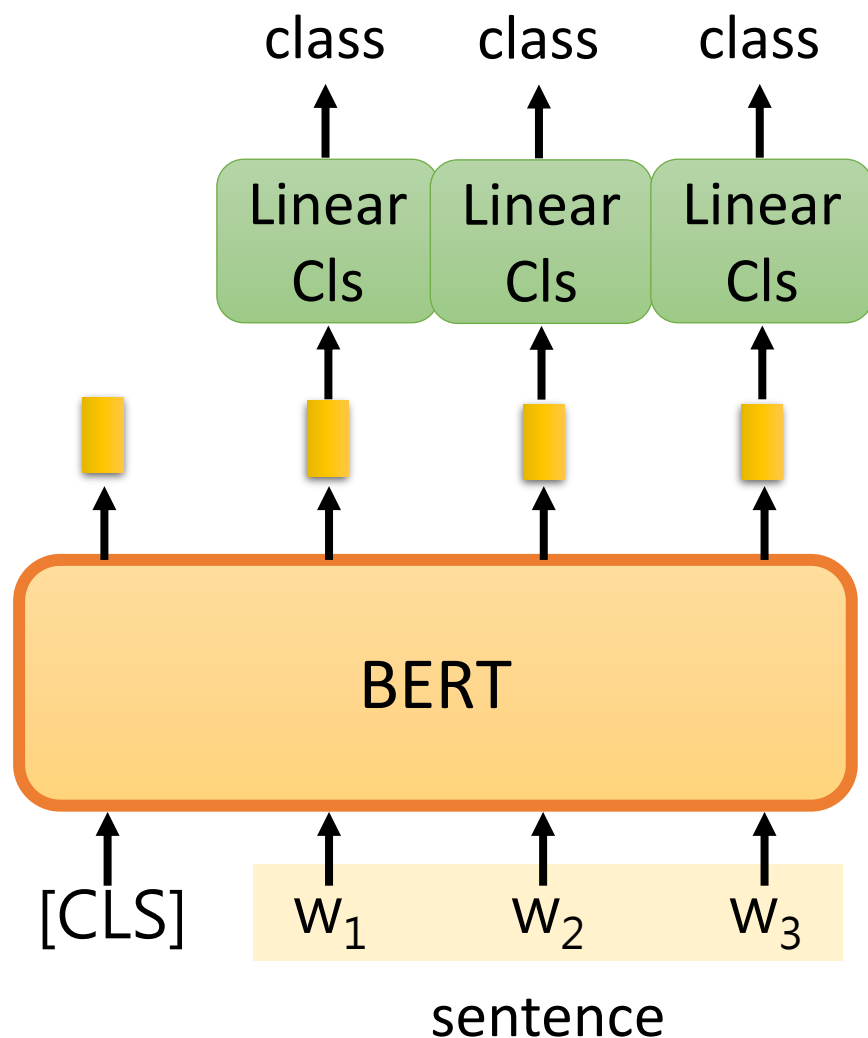
# How to use BERT – Case 1



Input: single sentence,  
output: class

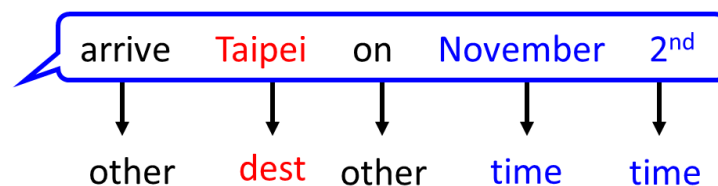
Example:  
Sentiment analysis (our  
HW),  
Document Classification

# How to use BERT – Case 2

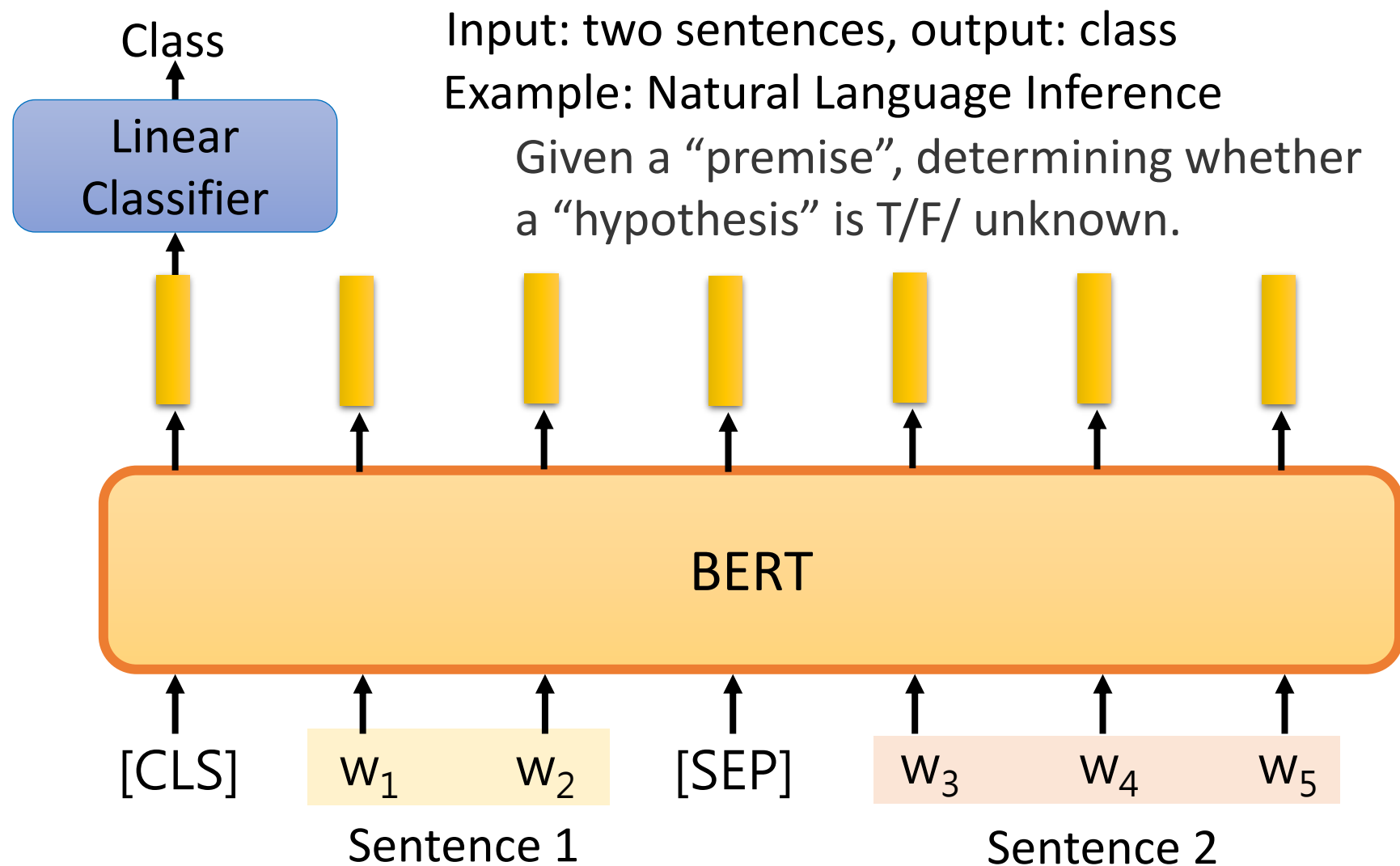


Input: single sentence,  
output: class of each word

Example: Slot filling



# How to use BERT – Case 3

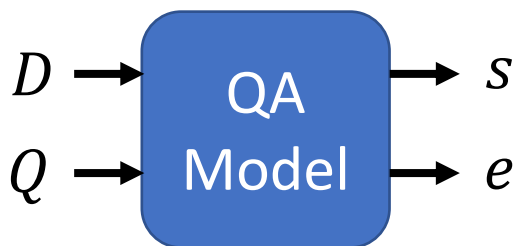


# How to use BERT – Case 4

- Extraction-based Question Answering (QA) (E.g. SQuAD)

**Document:**  $D = \{d_1, d_2, \dots, d_N\}$

**Query:**  $Q = \{q_1, q_2, \dots, q_N\}$



output: two integers ( $s, e$ )

**Answer:**  $A = \{q_s, \dots, q_e\}$

In meteorology, precipitation is any product of the condensation of 17 spheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **grau-pel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain 77 at 79 locations are called "showers".

What causes precipitation to fall?

**gravity**

$s = 17, e = 17$

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?

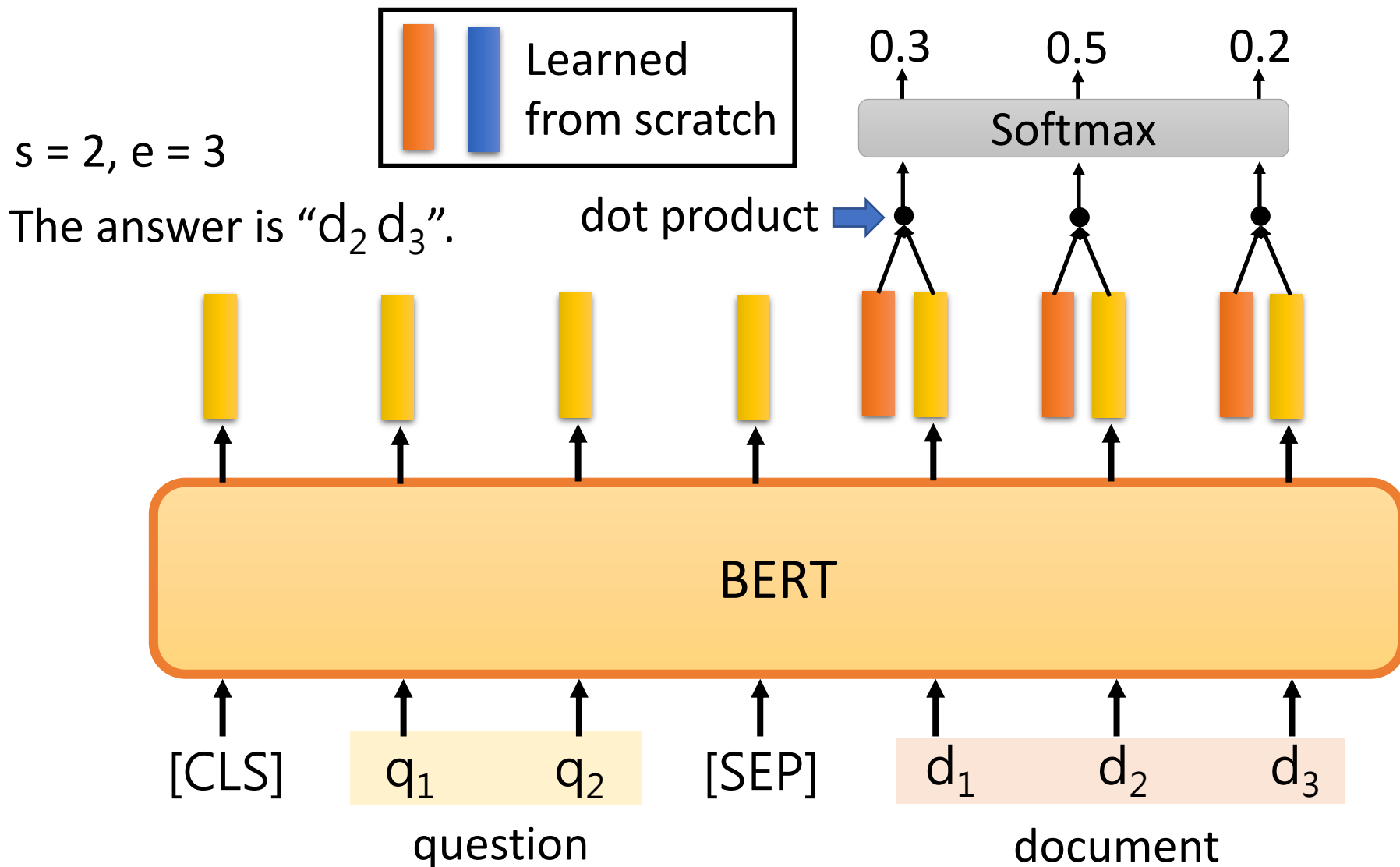
**grau-pel**

Where do water droplets collide with ice crystals to form precipitation?

**within a cloud**

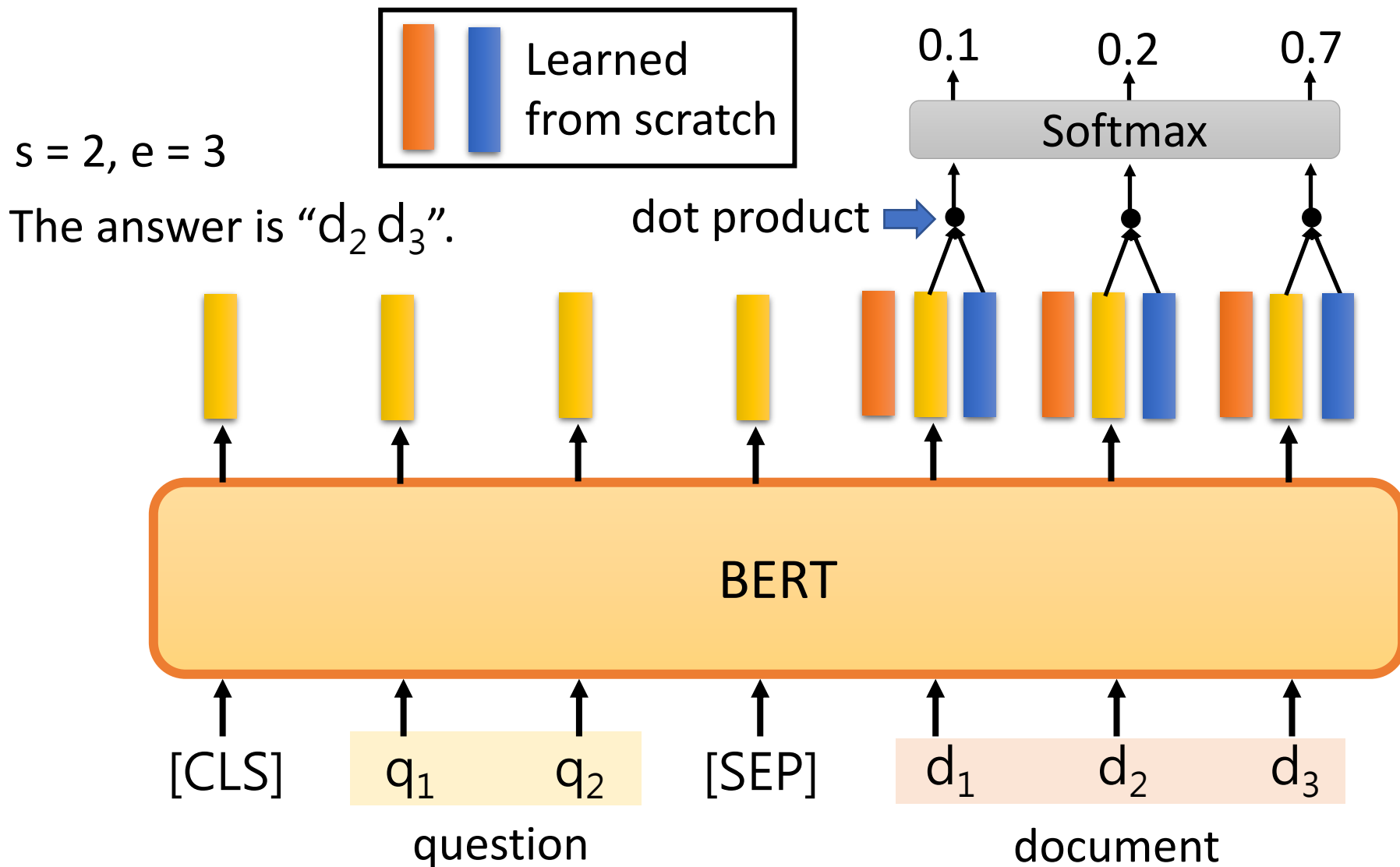
$s = 77, e = 79$

# How to use BERT – Case 4





# How to use BERT – Case 4



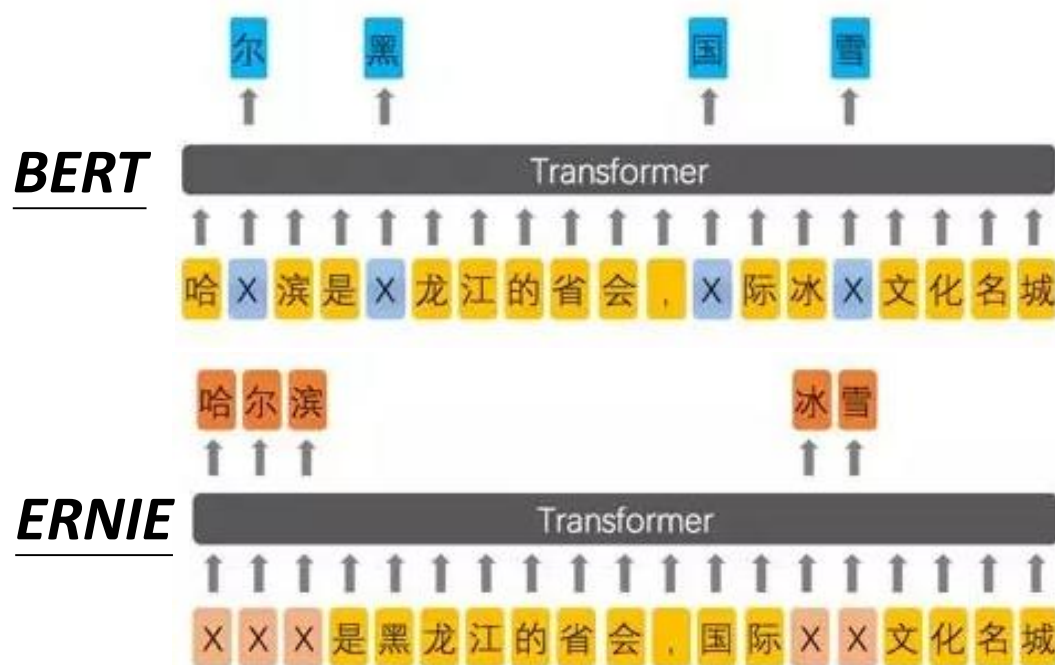
# BERT 屠榜 .....

Rank	Model	EM	F1
	Human Performance <i>Stanford University</i> (Rajpurkar & Jia et al. '18)	86.831	89.452
1 Mar 20, 2019	BERT + DAE + AoA (ensemble) <i>Joint Laboratory of HIT and iFLYTEK Research</i>	87.147	89.474
2 Mar 15, 2019	BERT + ConvLSTM + MTL + Verifier (ensemble) <i>Layer 6 AI</i>	86.730	89.286
3 Mar 05, 2019	BERT + N-Gram Masking + Synthetic Self-Training (ensemble) <i>Google AI Language</i> <a href="https://github.com/google-research/bert">https://github.com/google-research/bert</a>	86.673	89.147
4 May 21, 2019	XLNet (single model) <i>XLNet Team</i>	86.346	89.133
5 Apr 13, 2019	SemBERT(ensemble) <i>Shanghai Jiao Tong University</i>	86.166	88.886

SQuAD 2.0

# Enhanced Representation through Knowledge Integration (ERNIE)

- Designed for Chinese



Source of image:

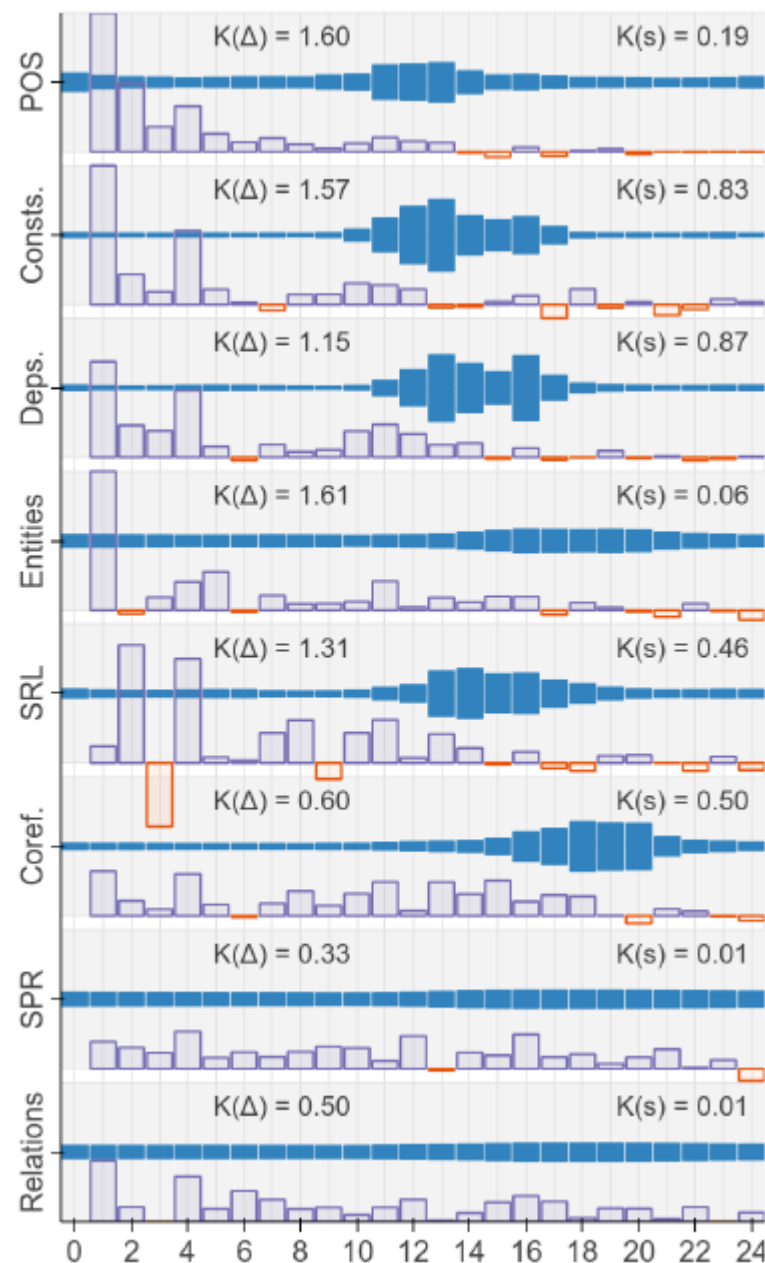
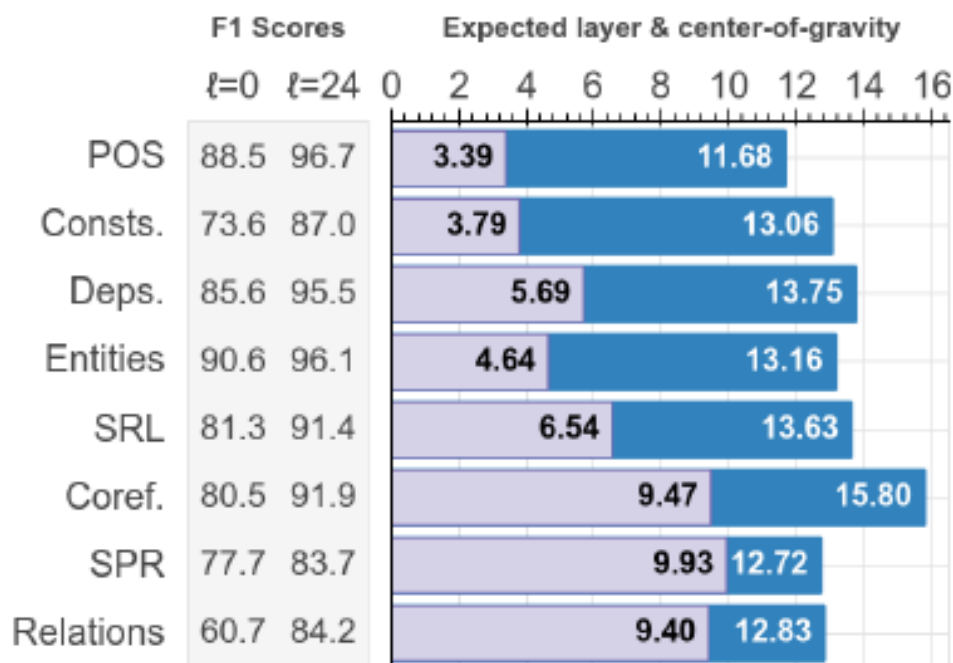
<https://zhuanlan.zhihu.com/p/59436589>

<https://arxiv.org/abs/1904.09223>

# What does BERT learn?

<https://arxiv.org/abs/1905.05950>

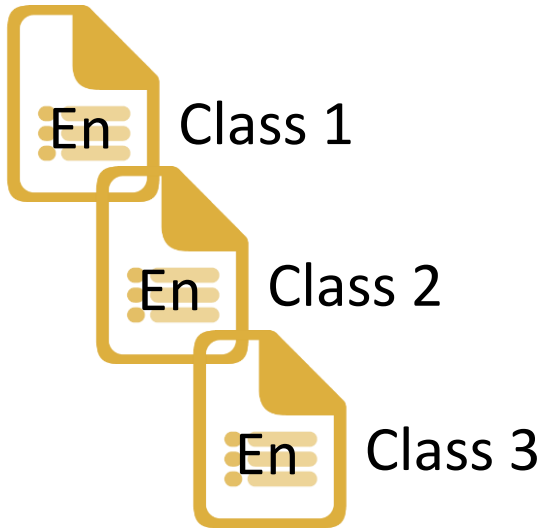
<https://openreview.net/pdf?id=SJzSgnRcKX>



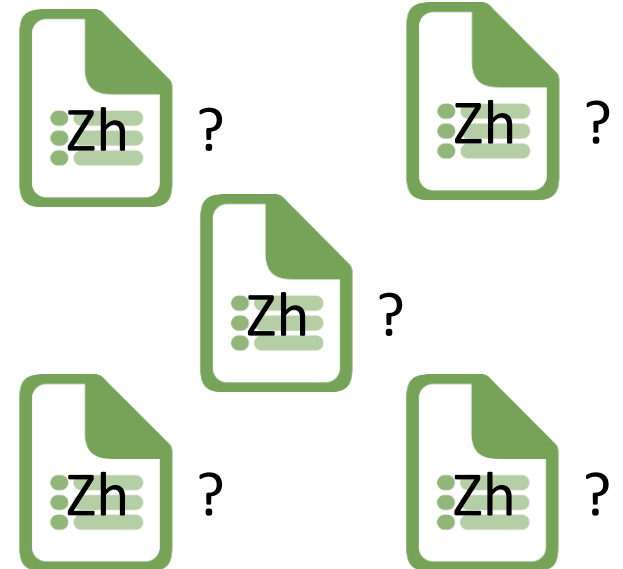
# Multilingual BERT

Trained on 104 languages

Task specific training  
data for English

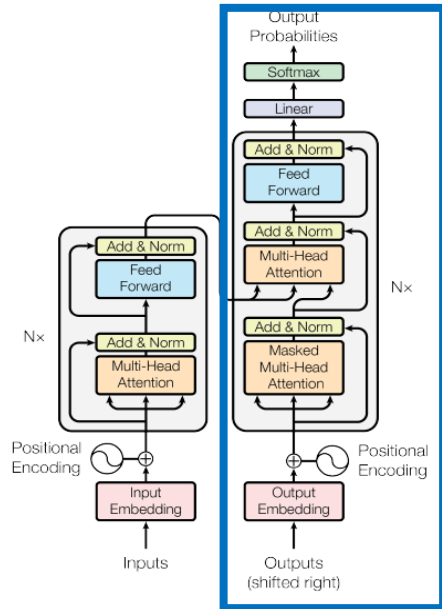


Task specific testing  
data for Chinese





# Generative Pre-Training (GPT)



Transformer  
Decoder

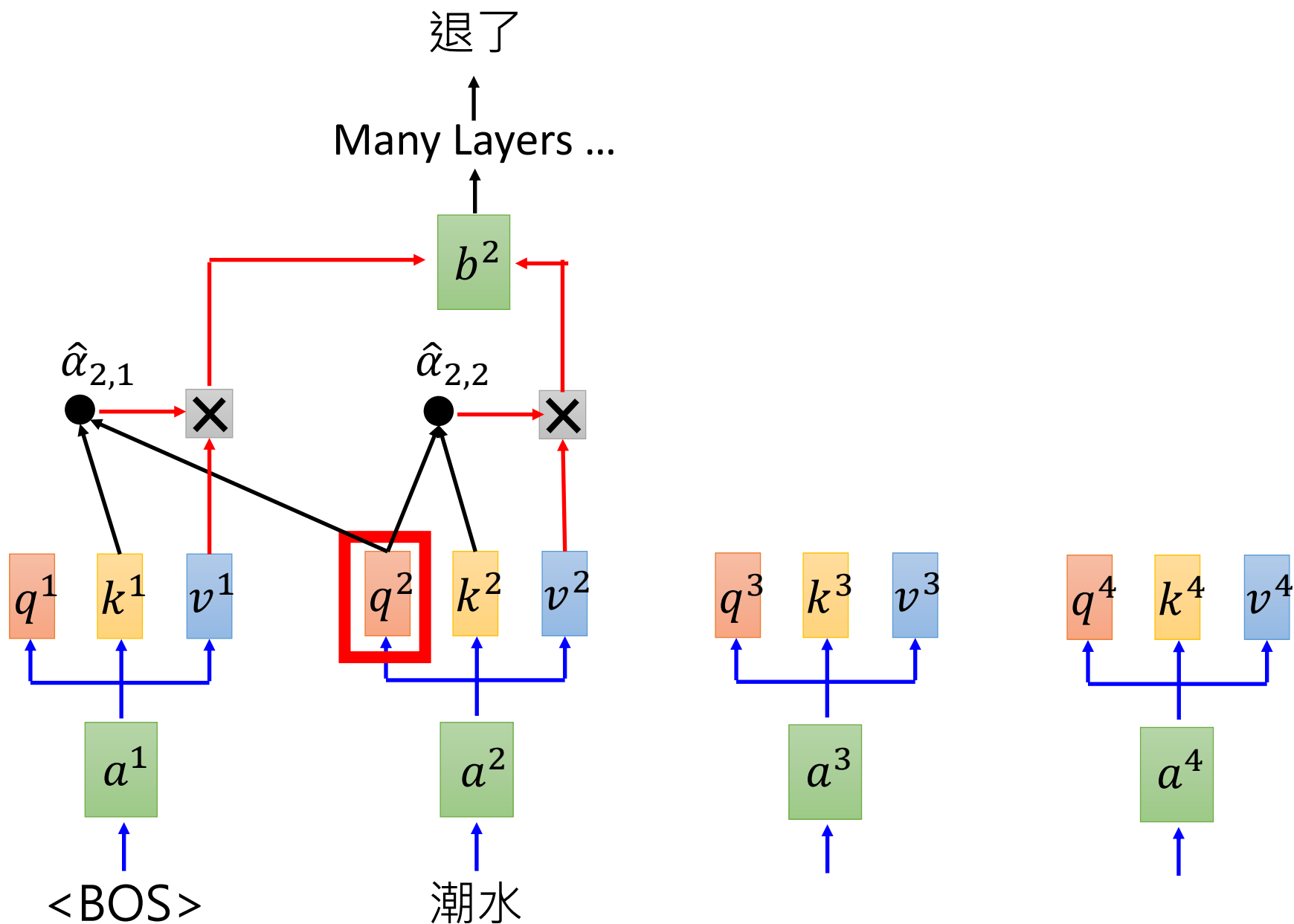
BERT  
(340M)

ELMO  
(94M)

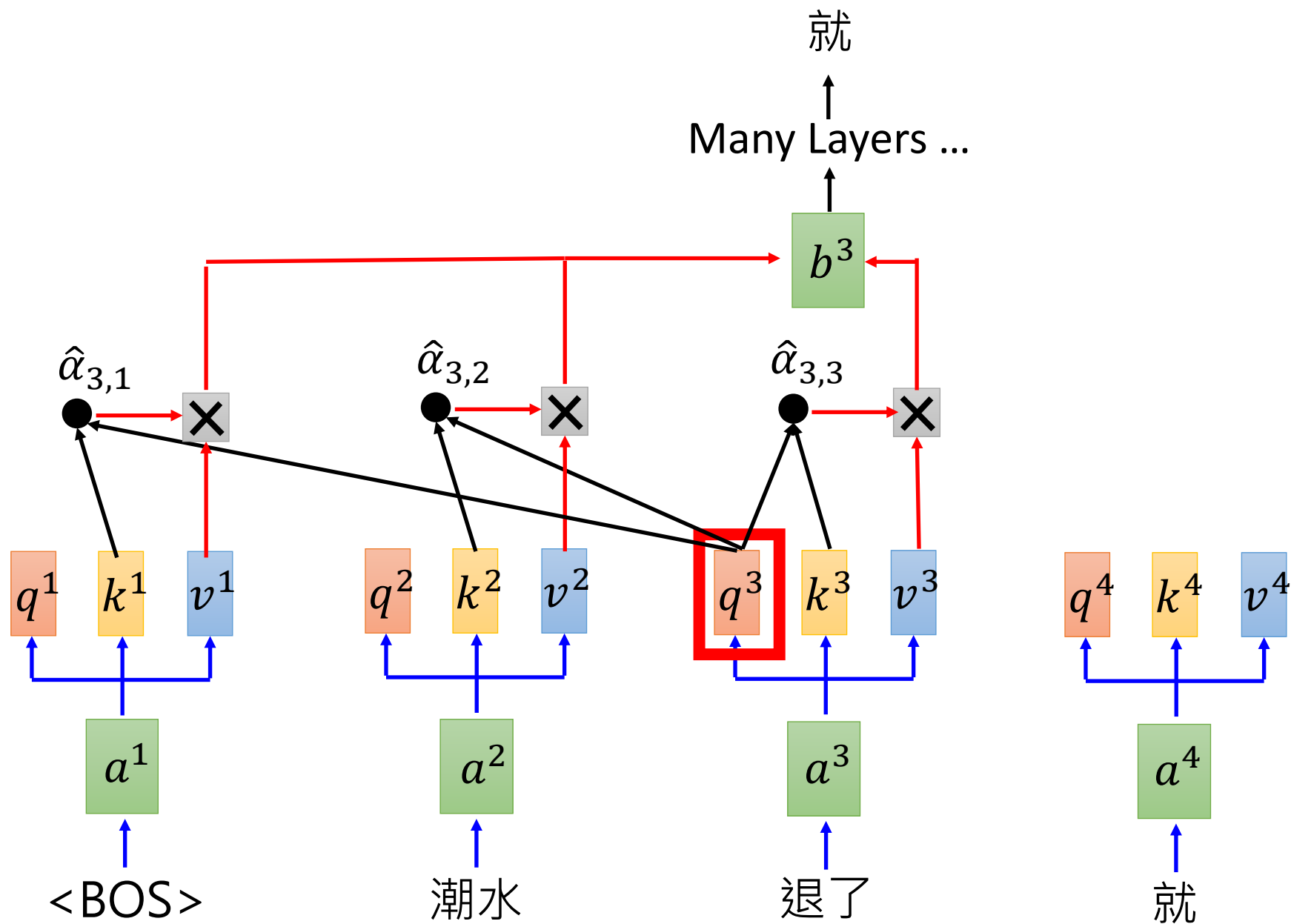


GPT-2  
(1542M)

# Generative Pre-Training (GPT)



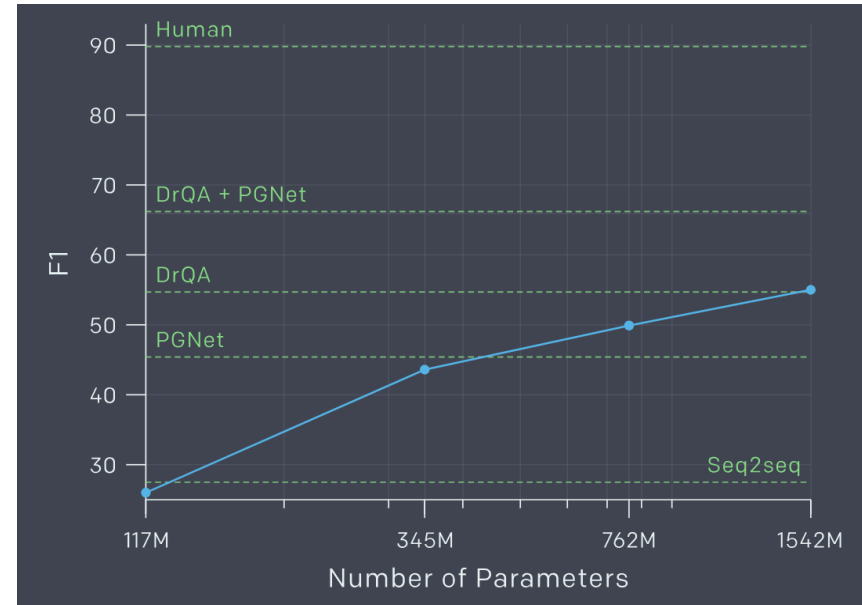
# Generative Pre-Training (GPT)



# Zero-shot Learning?

- **Reading Comprehension**

$d_1, d_2, \dots, d_N,$   
"Q:",  $q_1, q_2, \dots, q_N,$   
"A:"



- **Summarization**  $d_1, d_2, \dots, d_N, \text{"TL;DR:"}$

- **Translation**

English sentence 1

=

French sentence 1

English sentence 2

=

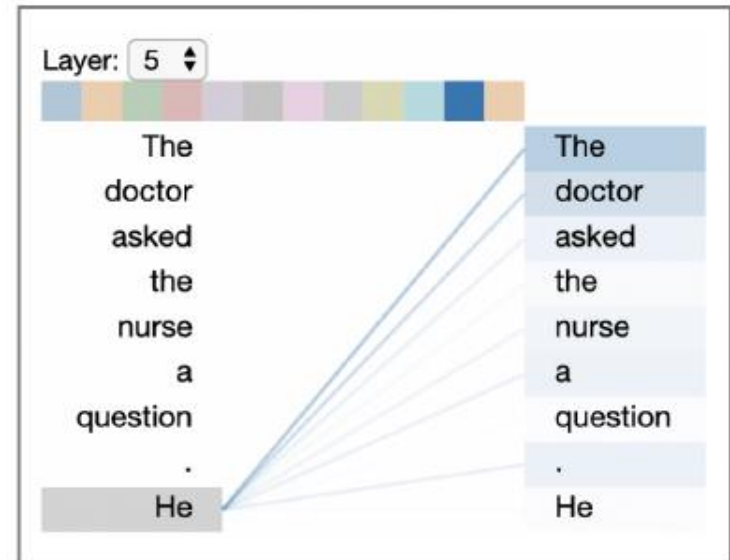
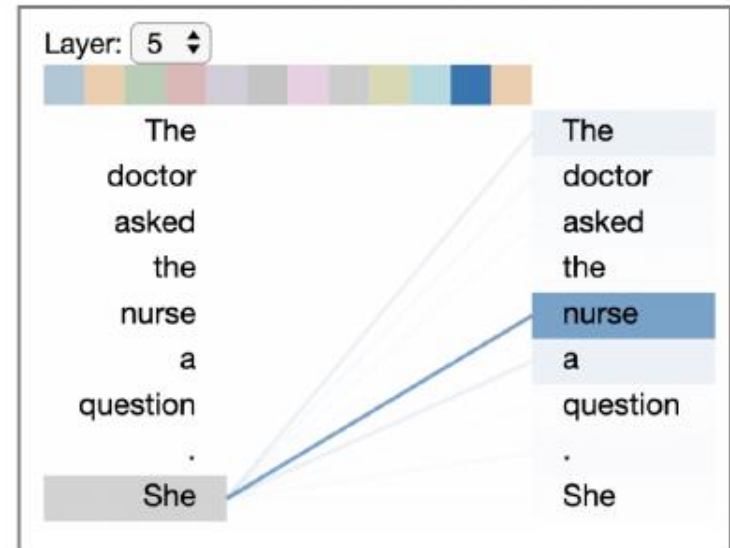
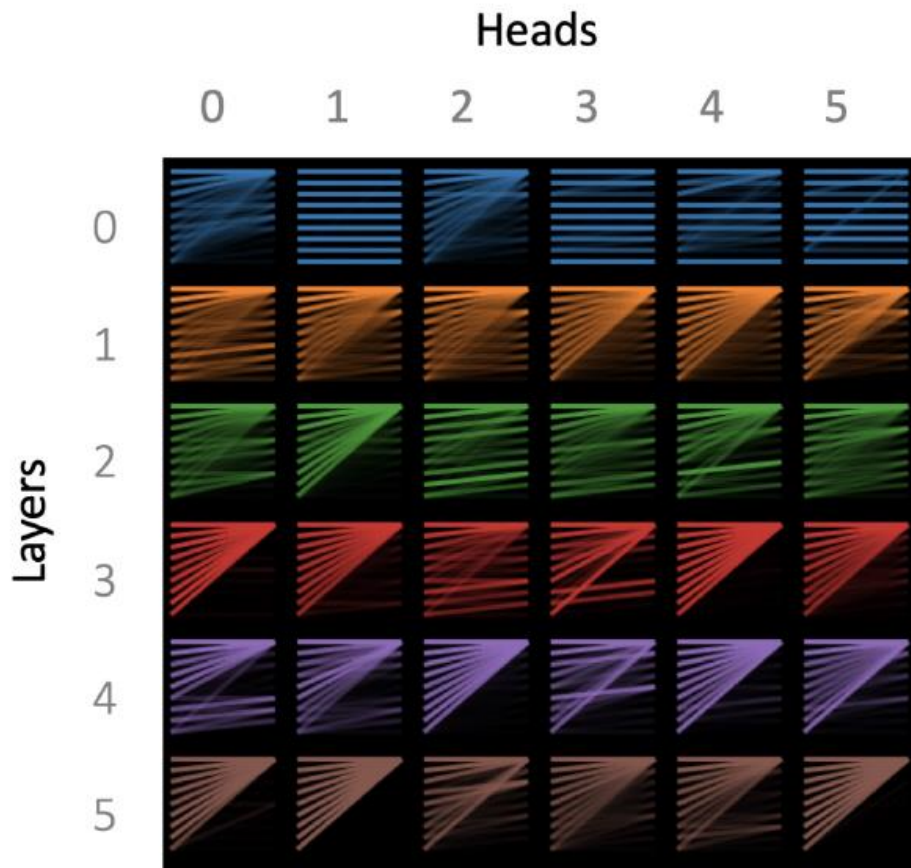
French sentence 2

English sentence 3

=

<https://arxiv.org/abs/1904.02679>  
(The results below are from GPT-2)

# Visualization





EM PROMPT  
-WRITTEN)

*In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*

MODEL  
MPLETION  
(MACHINE-  
10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

<https://talktotransformer.com/>



Credit: Greg Durrett

# Can BERT speak?

- Unified Language Model Pre-training for Natural Language Understanding and Generation
  - <https://arxiv.org/abs/1905.03197>
- BERT has a Mouth, and It Must Speak: BERT as a Markov Random Field Language Model
  - <https://arxiv.org/abs/1902.04094>
- Insertion Transformer: Flexible Sequence Generation via Insertion Operations
  - <https://arxiv.org/abs/1902.03249>
- Insertion-based Decoding with automatically Inferred Generation Order
  - <https://arxiv.org/abs/1902.01370>