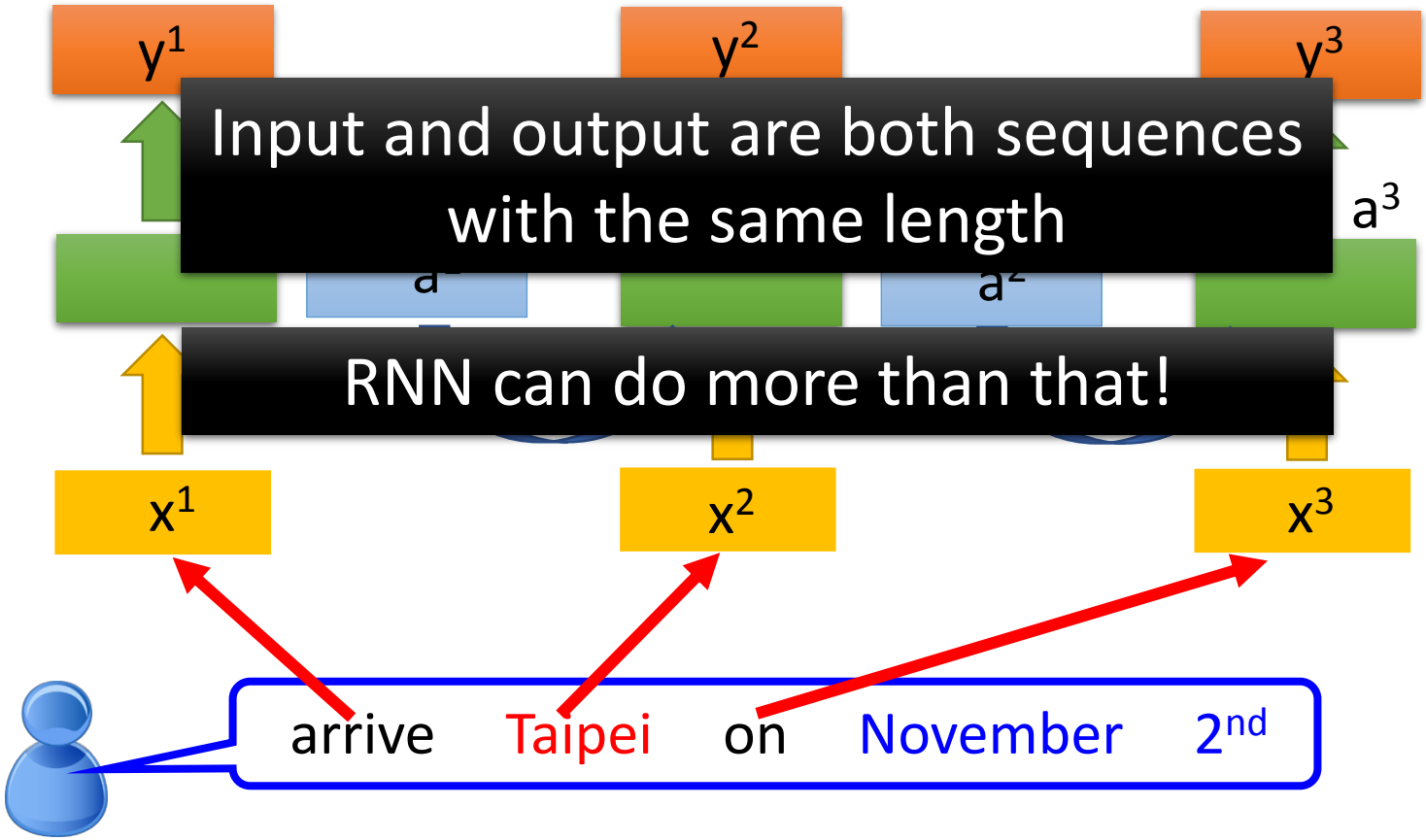
# More Applications ……

Probability of "arrive" in each slot

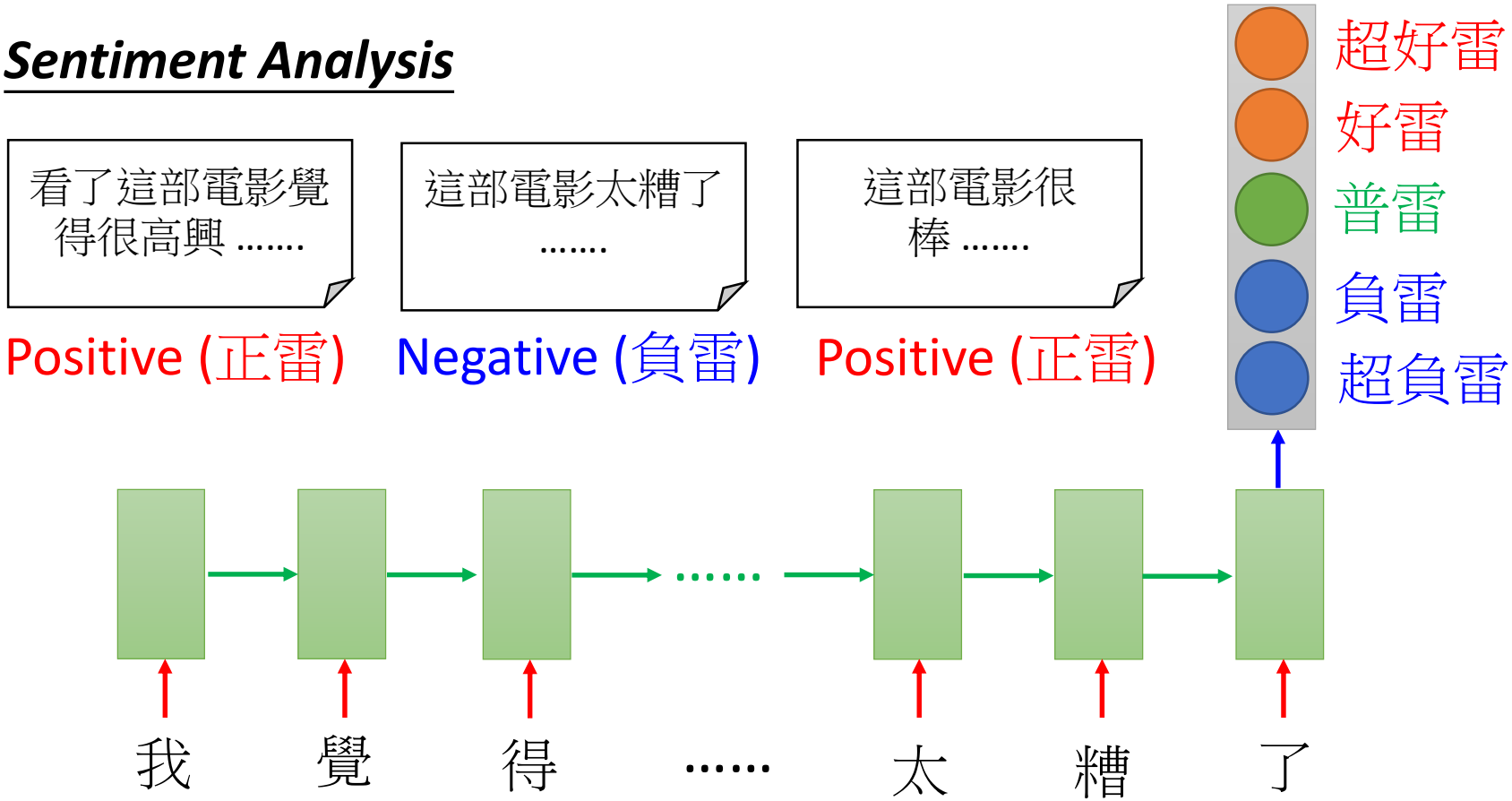
Probability of "Taipei" in each slot

Probability of "on" in each slot



# Many to one

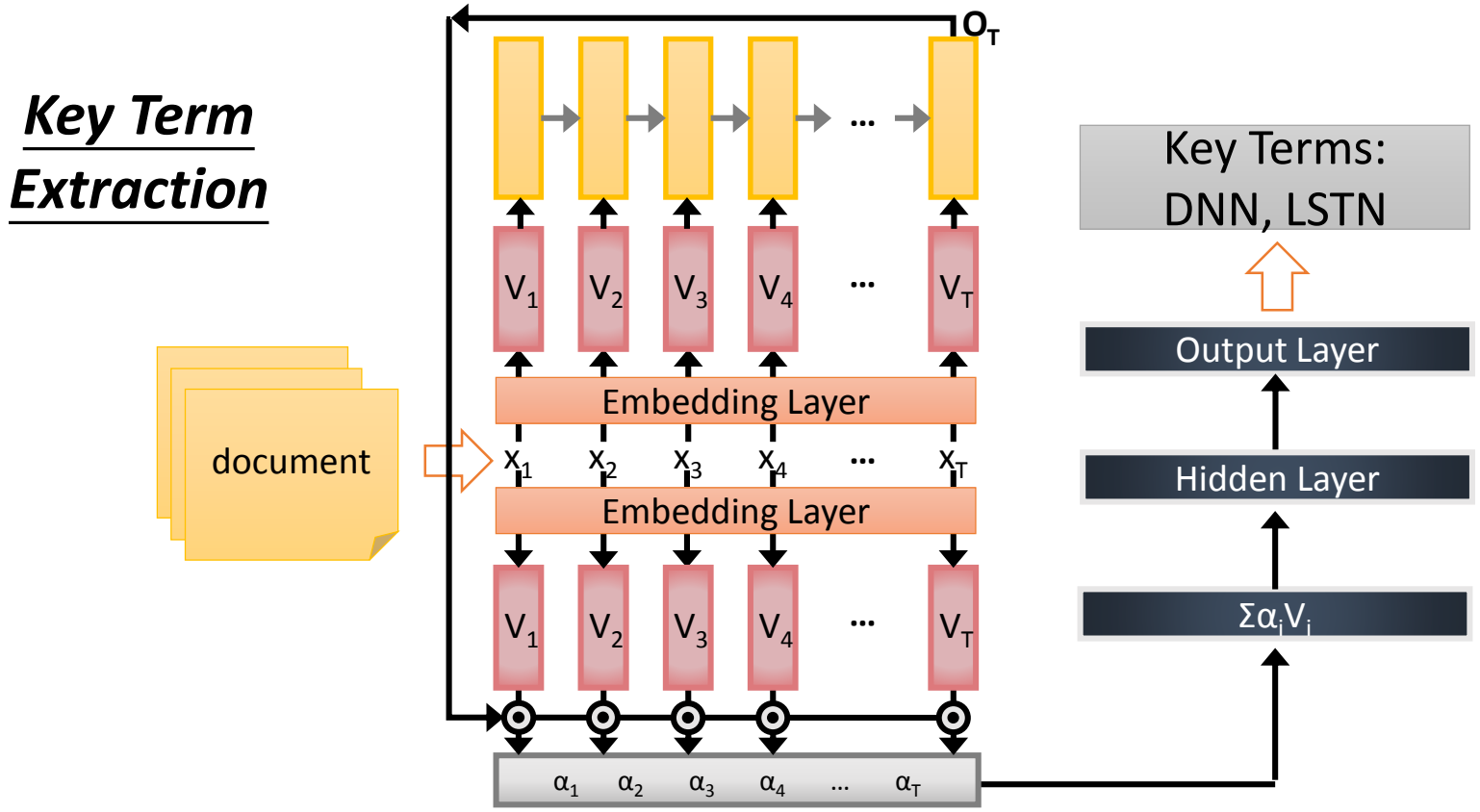
- Input is a vector sequence, but output is only one vector



# Many to one

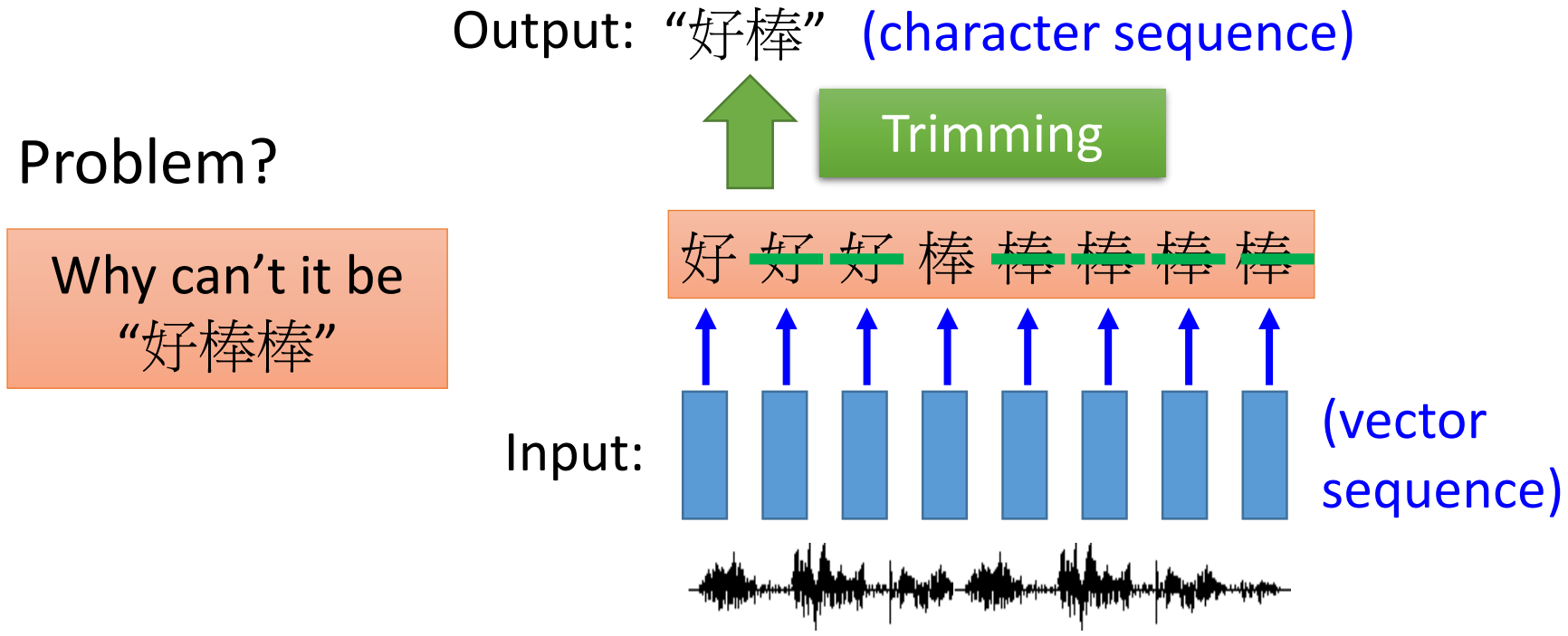
[Shen & Lee, Interspeech 16]

- Input is a vector sequence, but output is only one vector



# Many to Many (Output is shorter)

- Both input and output are both sequences, ***but the output is shorter.***
  - E.g. ***Speech Recognition***

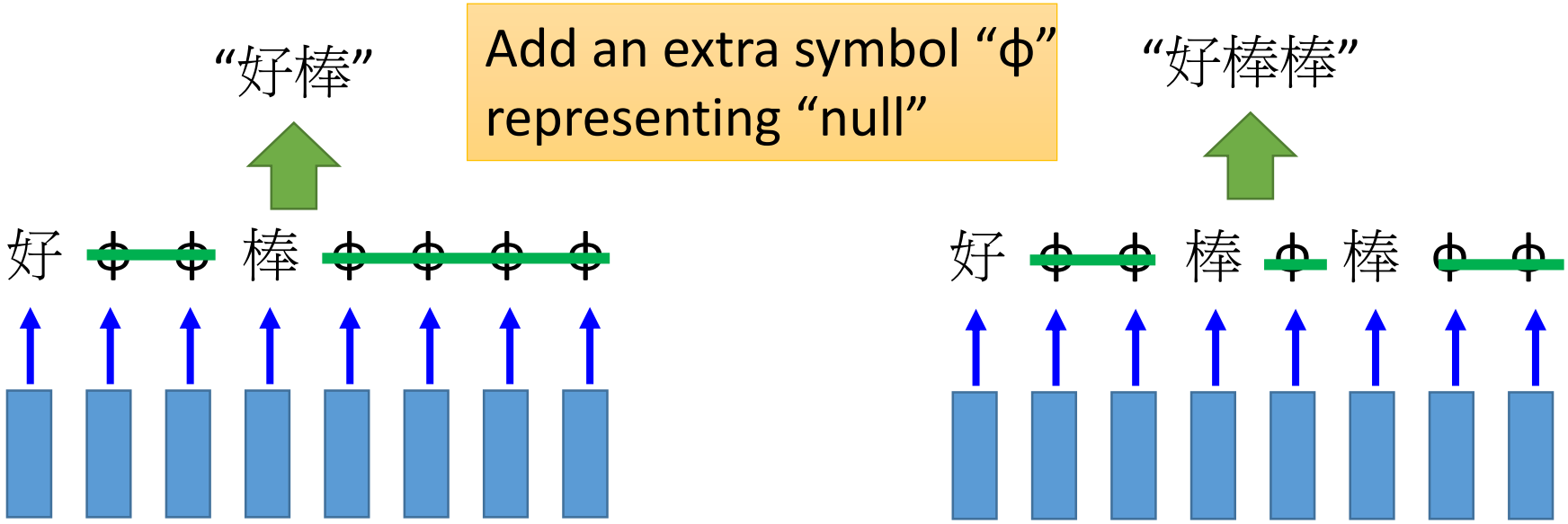



Problem?

Why can’t it be “好棒棒”

# Many to Many (Output is shorter)

- Both input and output are both sequences, ***but the output is shorter.***
- Connectionist Temporal Classification (CTC) [Alex Graves, ICML'06][Alex Graves, ICML'14][Haşim Sak, Interspeech'15][Jie Li, Interspeech'15][Andrew Senior, ASRU'15]



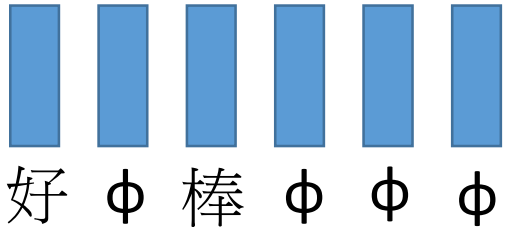
# Many to Many (Output is shorter)

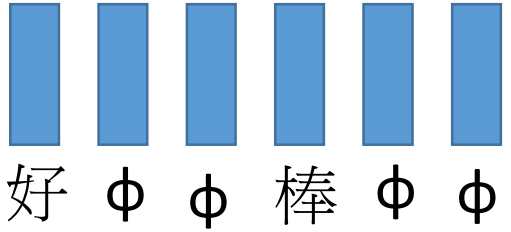
- CTC: Training

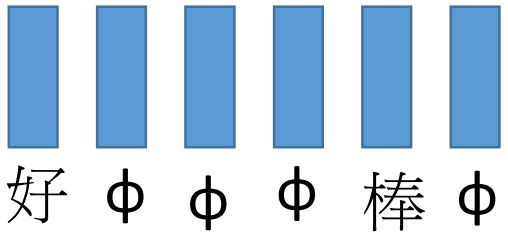
Acoustic Features:

Label: 好 棒

All possible alignments are considered as correct.

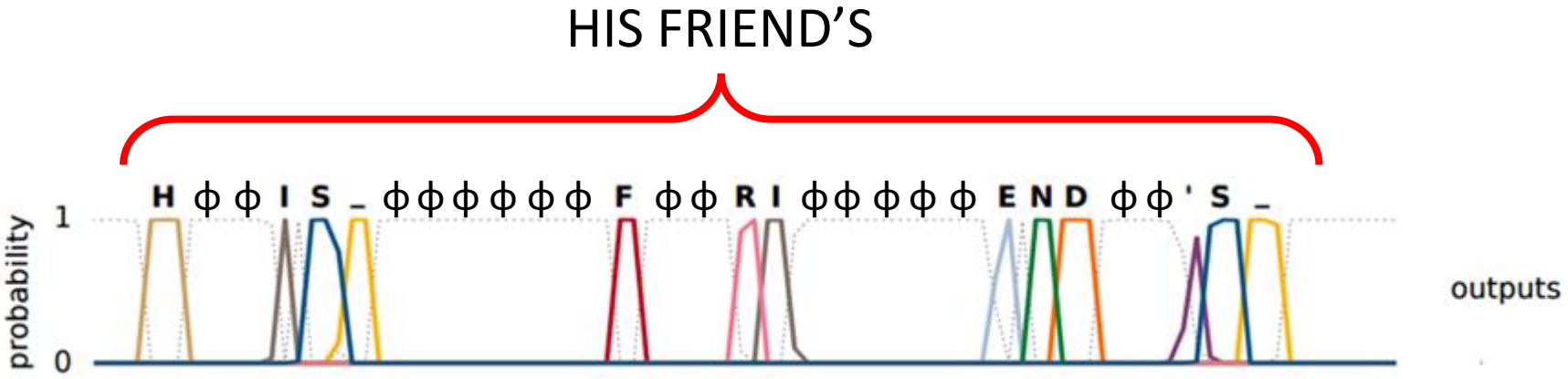


# Many to Many (Output is shorter)

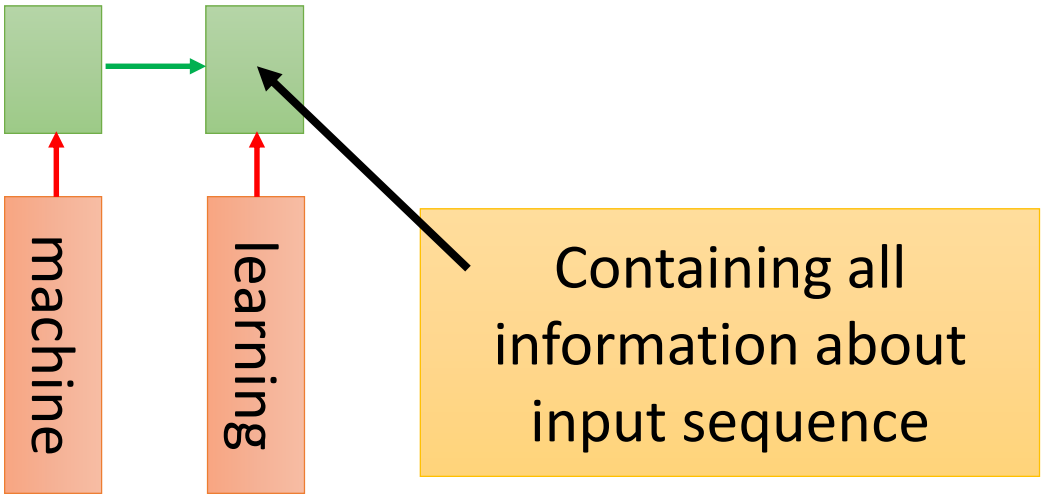
- CTC: example




Graves, Alex, and Navdeep Jaitly. "Towards end-to-end speech recognition with recurrent neural networks." *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*. 2014.

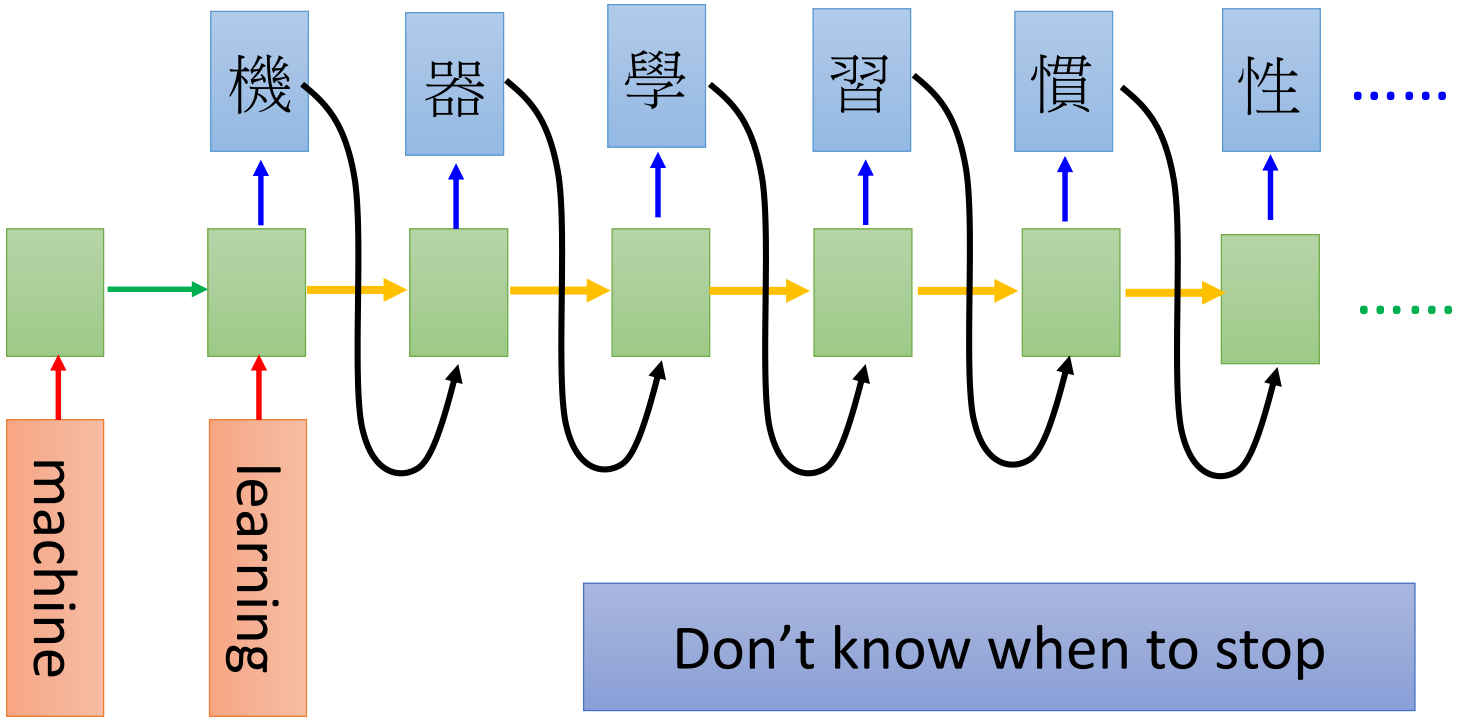
# Many to Many (No Limitation)

- Both input and output are both sequences ***with different lengths***. → ***Sequence to sequence learning***
  - E.g. ***Machine Translation*** (machine learning→機器學習)

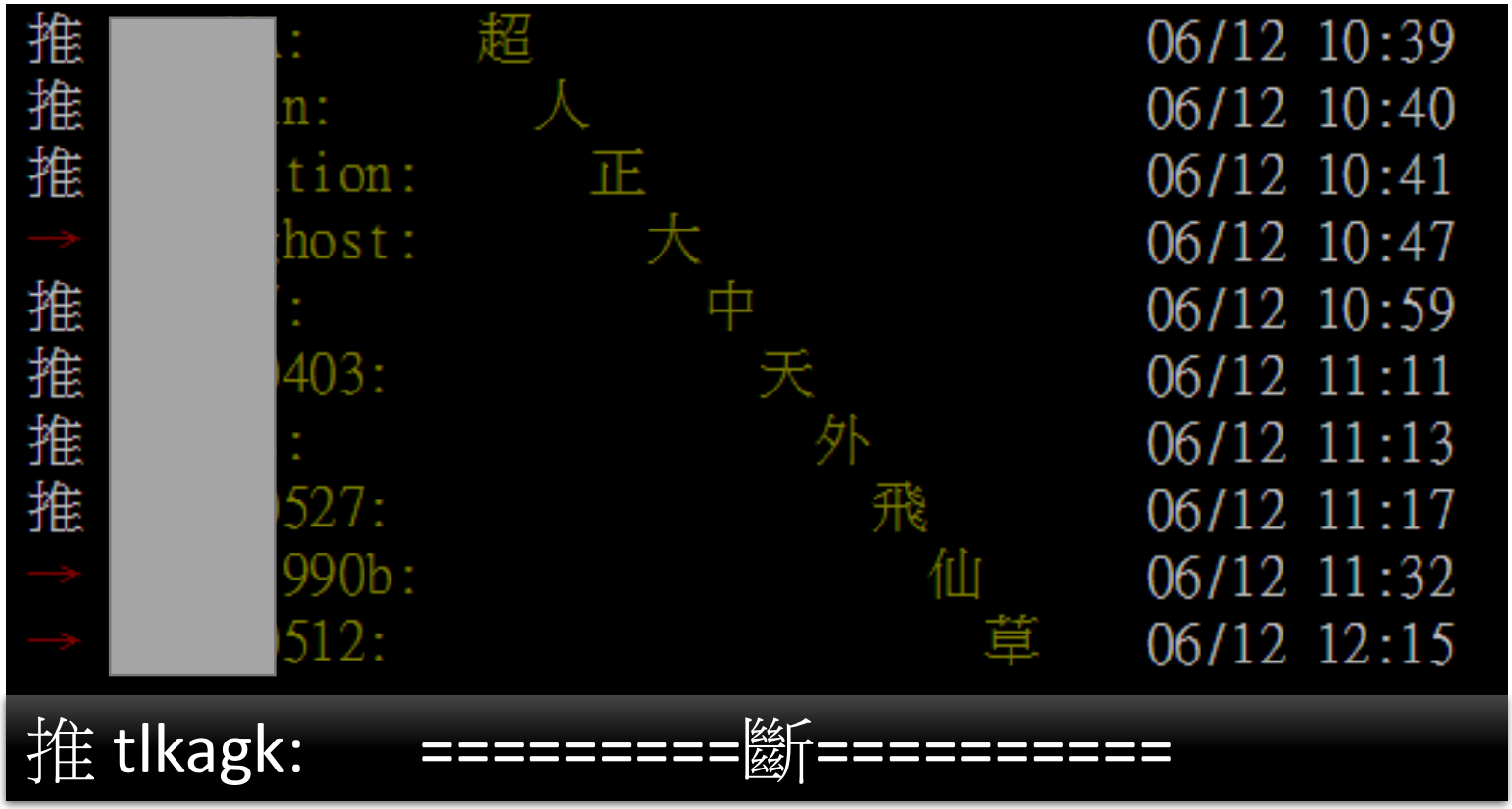


# Many to Many (No Limitation)

- Both input and output are both sequences ***with different lengths***. → ***Sequence to sequence learning***
  - E.g. ***Machine Translation*** (machine learning→機器學習)



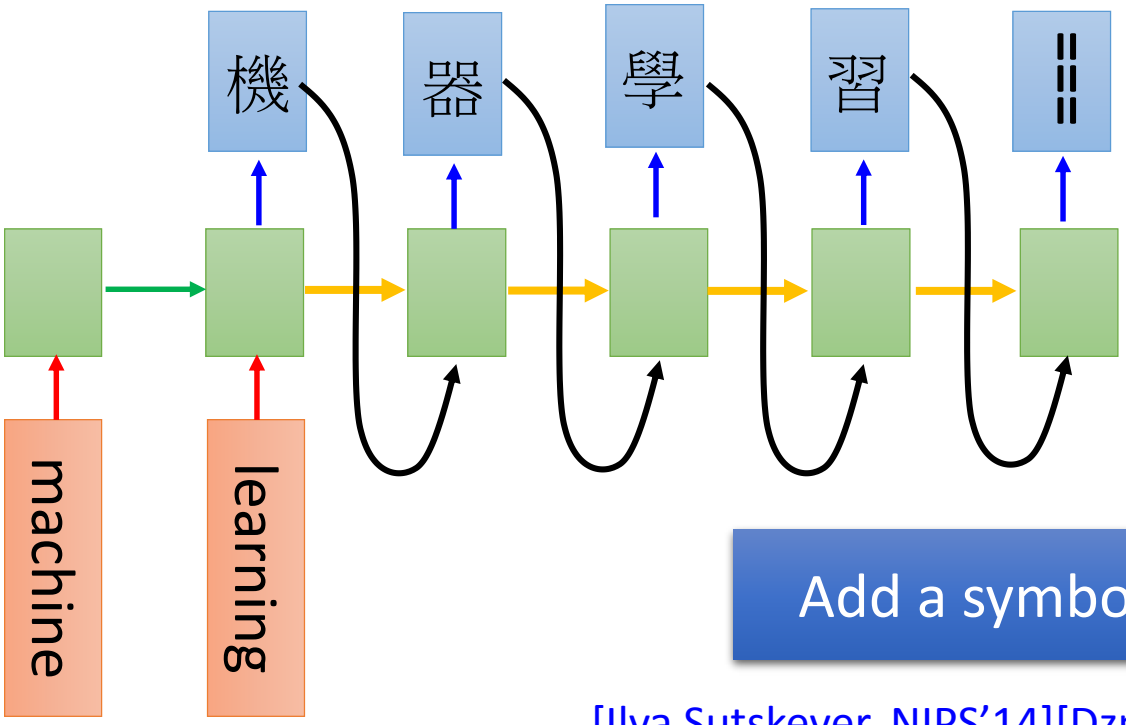
# Many to Many (No Limitation)




接龍推文是ptt在推文中的一種趣味玩法，與推齊有些類似但又有所不同，是指在推文中接續上一樓的字句，而推出連續的意思。該類玩法確切起源已不可知(鄉民百科)

# Many to Many (No Limitation)

- Both input and output are both sequences ***with different lengths***. → ***Sequence to sequence learning***
  - E.g. ***Machine Translation*** (machine learning→機器學習)




Add a symbol "===" (斷)

[Ilya Sutskever, NIPS'14][Dzmitry Bahdanau, arXiv'15]

https://arxiv.org/pdf/1612.01744v1.pdf

# Many to Many (No Limitation)

- Both input and output are both sequences ***with different lengths***. → ***Sequence to sequence learning***
  - E.g. ***Machine Translation*** (machine learning→機器學習)


(a) Machine translation alignment

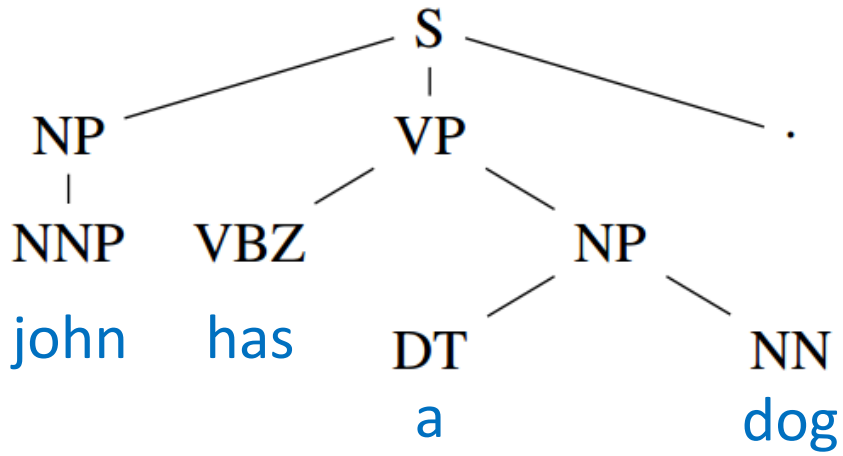
(b) Speech translation alignment

Figure 1: Alignments performed by the attention model during training

# Beyond Sequence

- Syntactic parsing

John has a dog . →

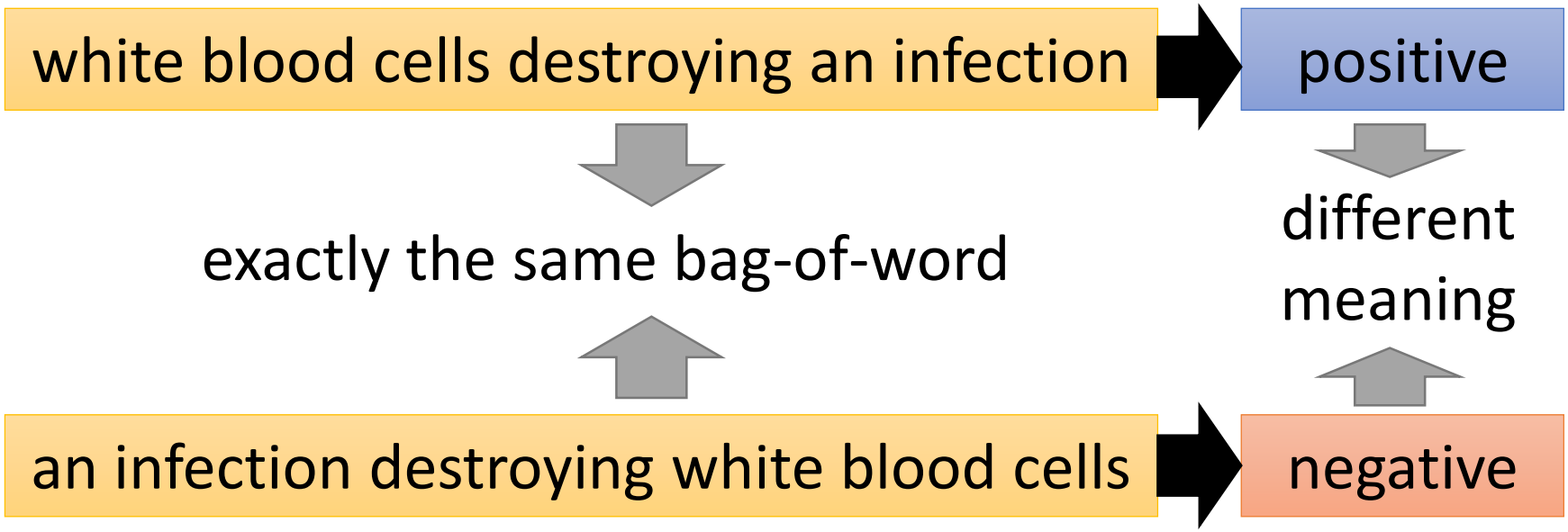



John has a dog . → (S (NP NNP )$_{\text{NP}}$ (VP VBZ (NP DT NN )$_{\text{NP}}$ )$_{\text{VP}}$ . )$_{\text{S}}$

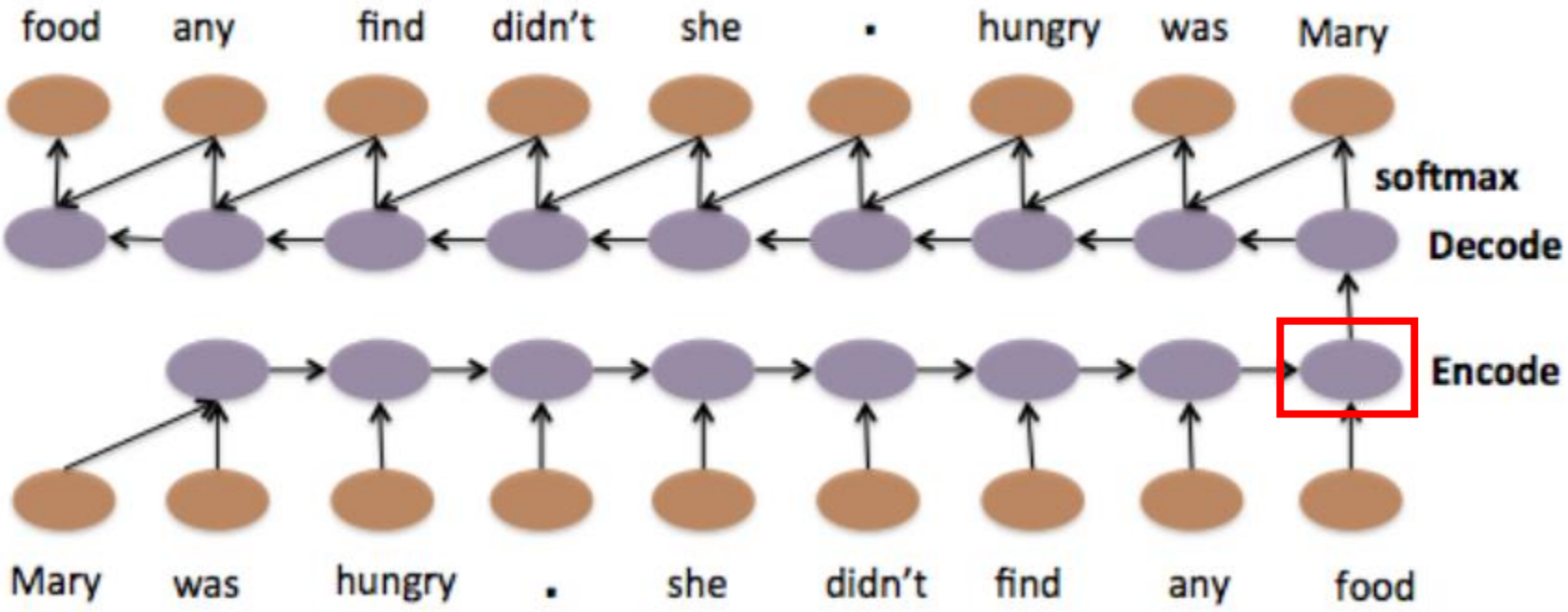
Oriol Vinyals, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, Geoffrey Hinton, Grammar as a Foreign Language, NIPS 2015

# Sequence-to-sequence Auto-encoder - Text

- To understand the meaning of a word sequence, the order of the words can not be ignored.

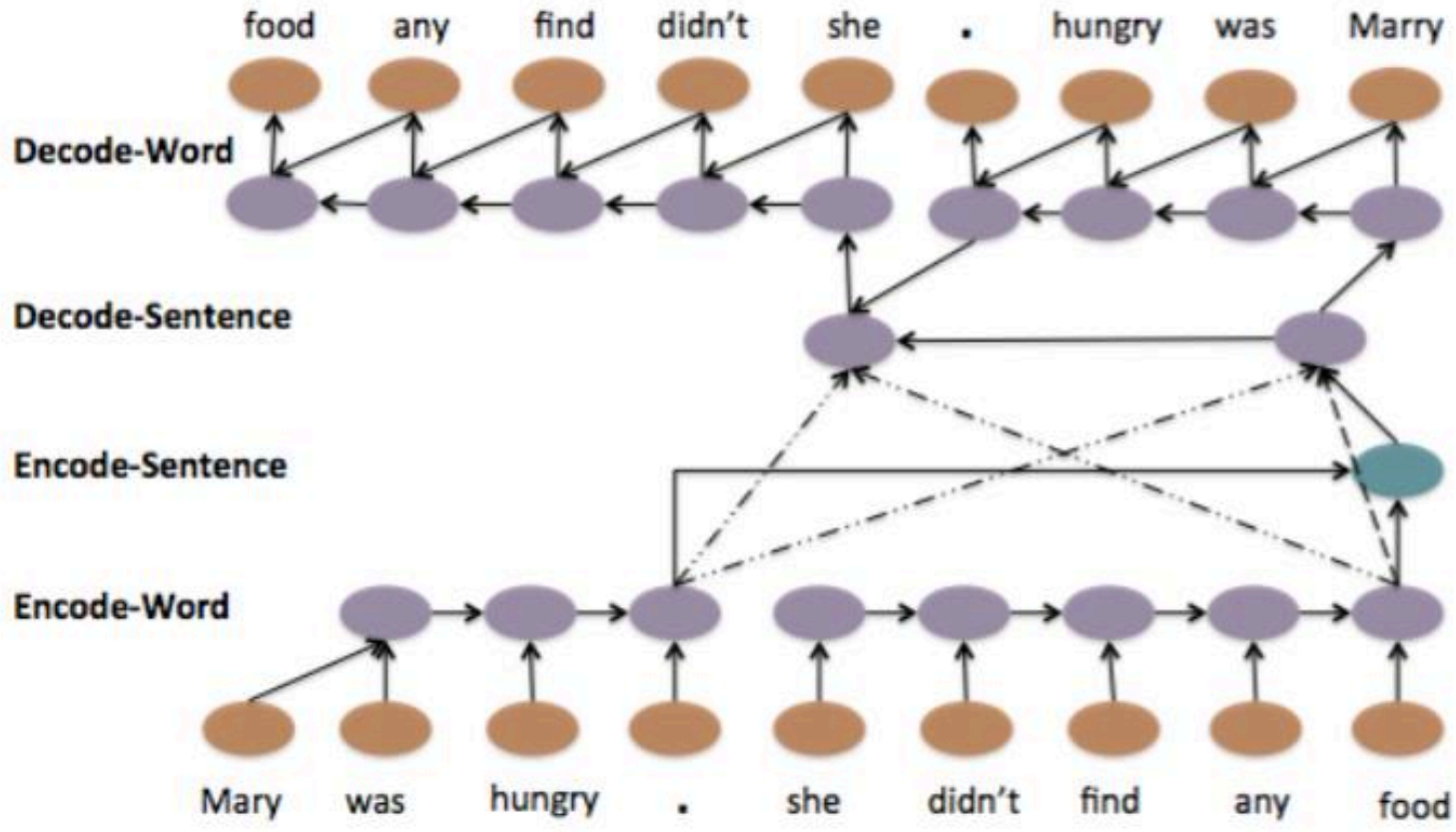


# Sequence-to-sequence Auto-encoder - Text




Li, Jiwei, Minh-Thang Luong, and Dan Jurafsky. "A hierarchical neural autoencoder for paragraphs and documents." *arXiv preprint arXiv:1506.01057*(2015).

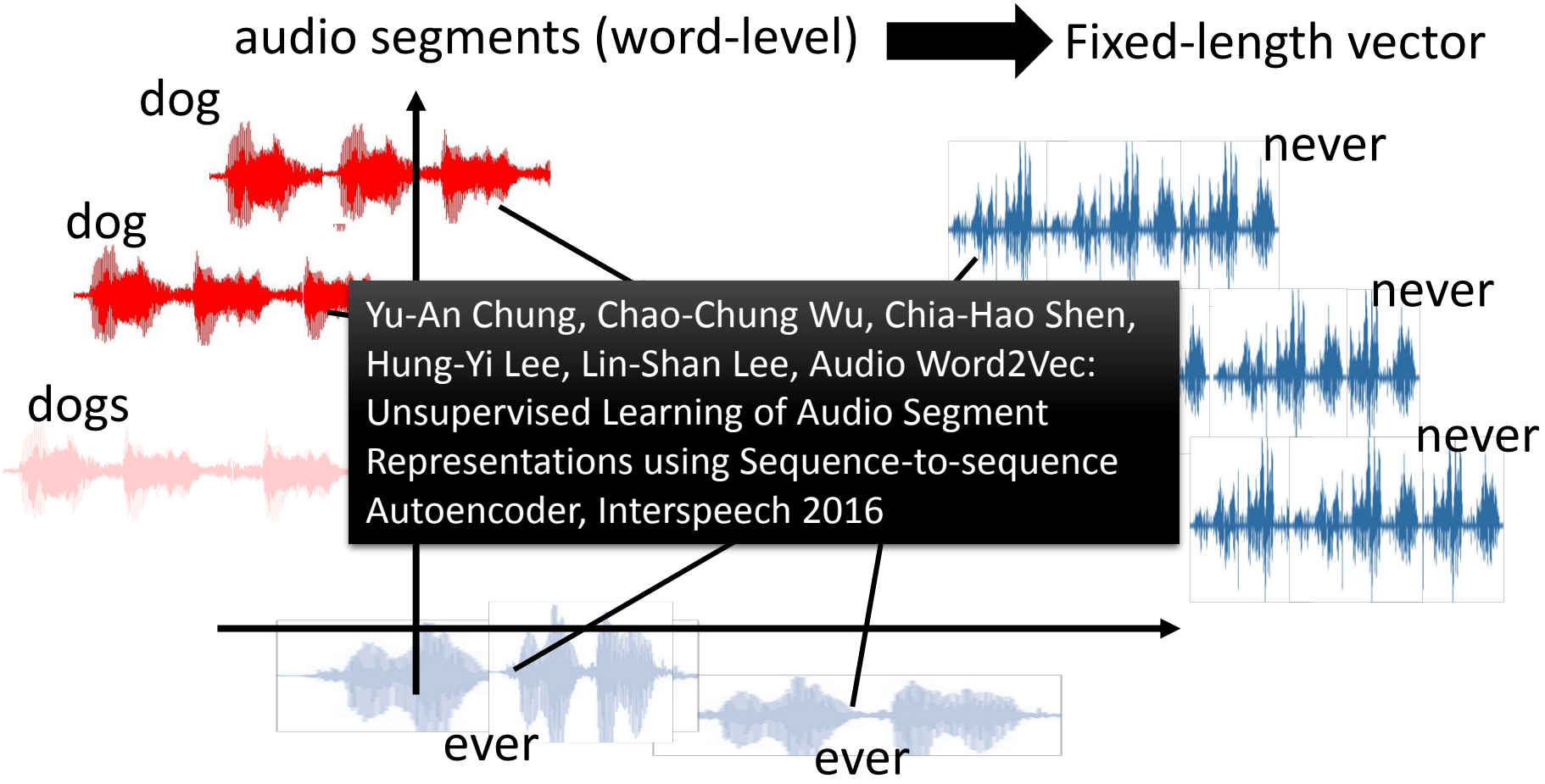
# Sequence-to-sequence Auto-encoder - Text




Li, Jiwei, Minh-Thang Luong, and Dan Jurafsky. "A hierarchical neural autoencoder for paragraphs and documents." *arXiv preprint arXiv:1506.01057*(2015).

# Sequence-to-sequence Auto-encoder - Speech

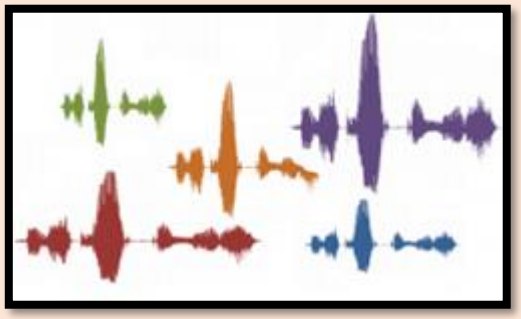
- Dimension reduction for a sequence with variable length



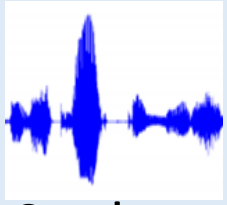
# Sequence-to-sequence Auto-encoder - Speech

Audio archive divided into variable-length audio segments

***Off-line***



Audio Segment to Vector



Spoken Query

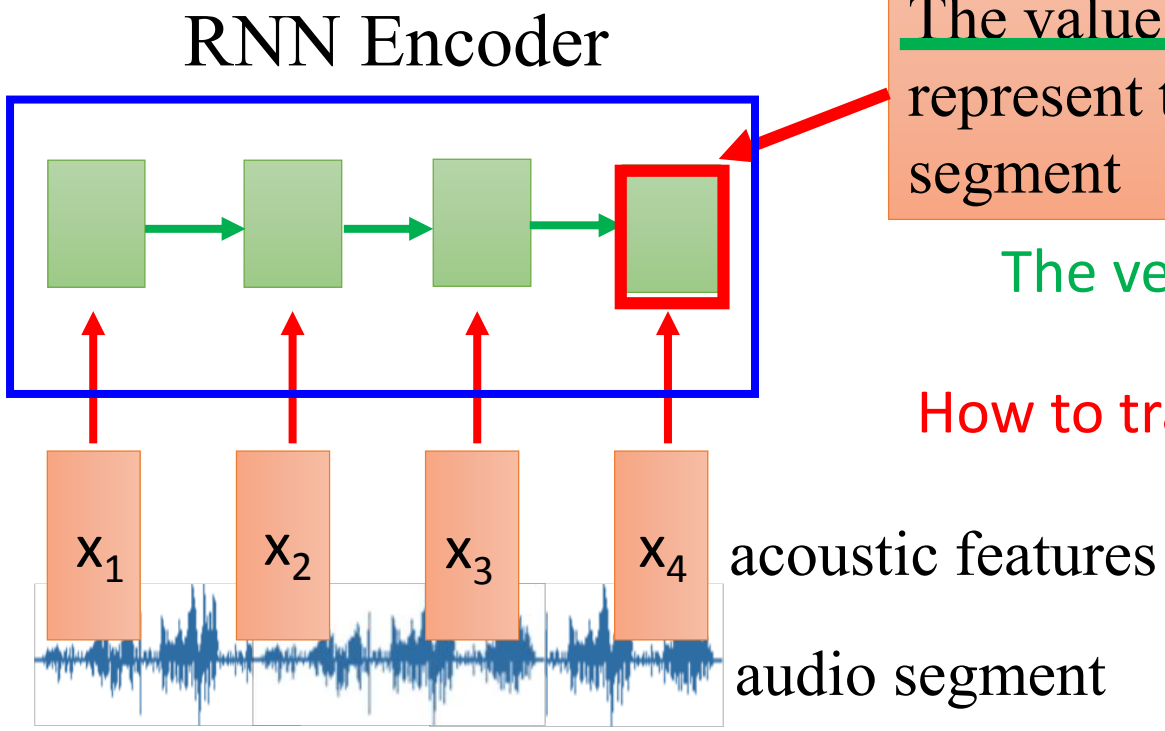
Audio Segment to Vector

Similarity

Search Result

***On-line***

# Sequence-to-sequence Auto-encoder - Speech

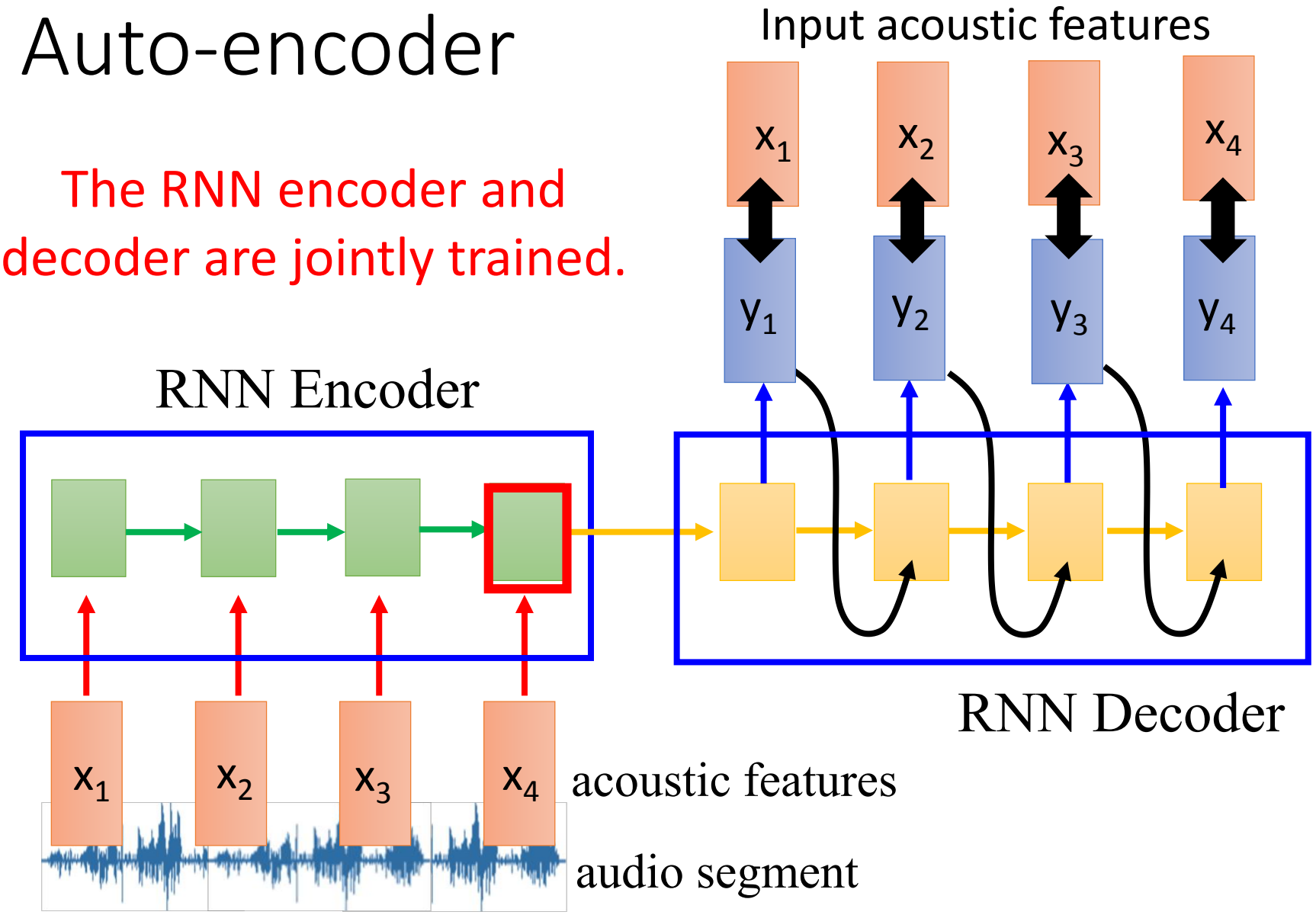

The values in the memory represent the whole audio segment

The vector we want

How to train RNN Encoder?

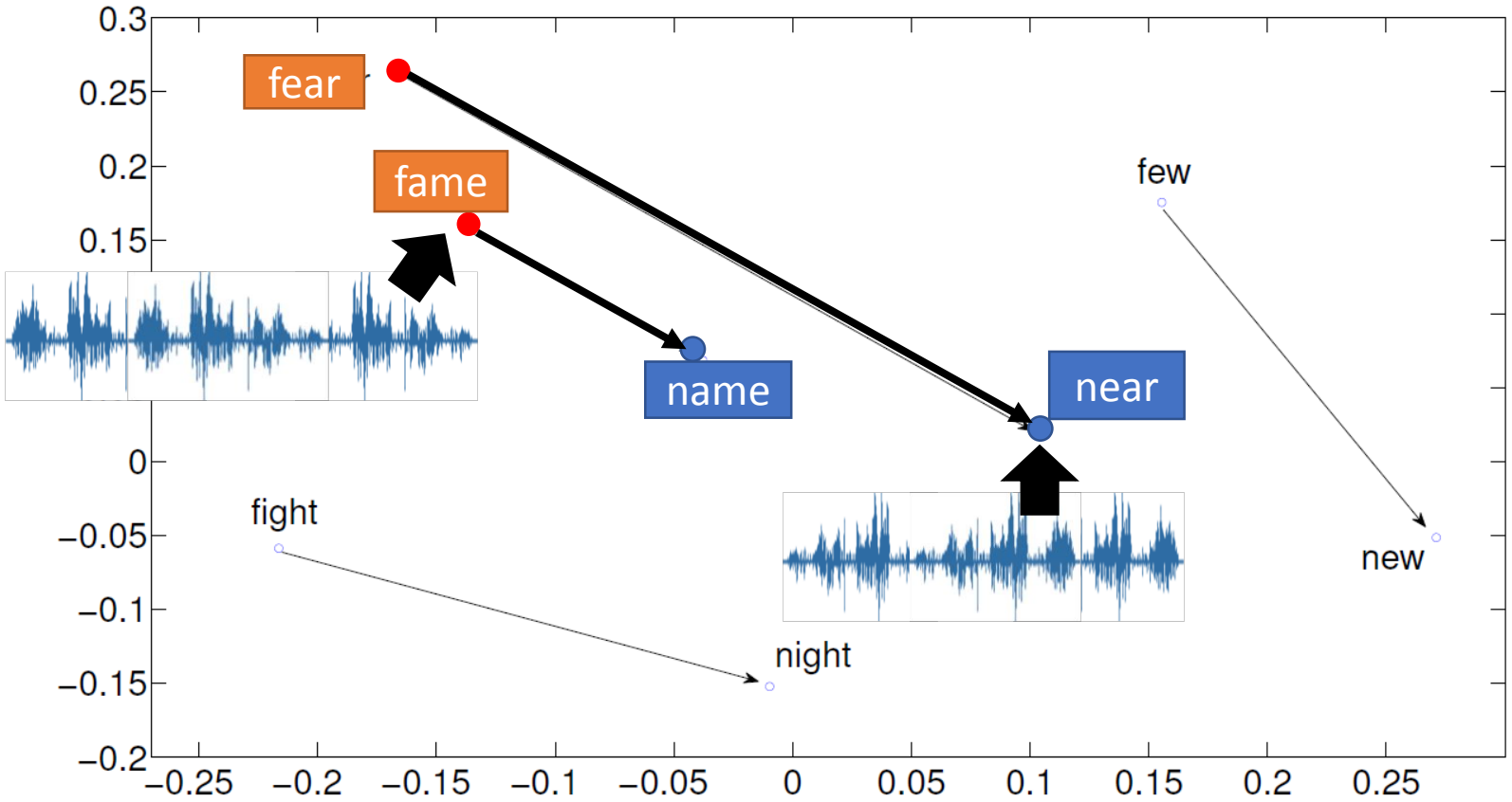
# Sequence-to-sequence Auto-encoder

The RNN encoder and decoder are jointly trained.

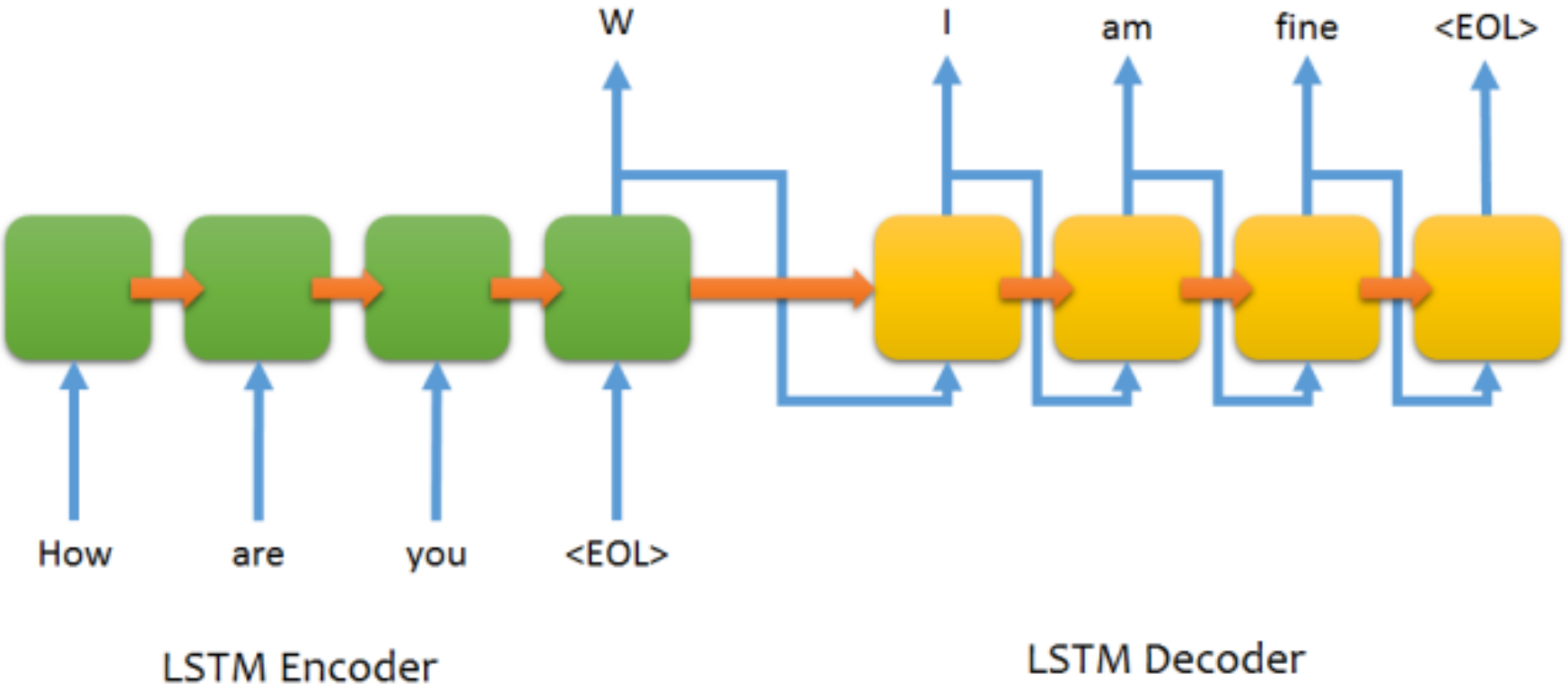


# Sequence-to-sequence Auto-encoder - Speech

- Visualizing embedding vectors of the words



# Demo: Chat-bot




電視影集 (~40,000 sentences)、美國總統大選辯論

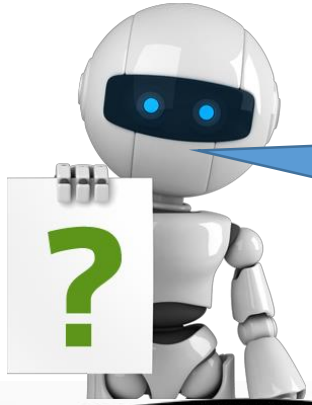
# Demo: Chat-bot

- Develop Team
  - Interface design: Prof. Lin-Lin Chen & Arron Lu
  - Web programming: Shi-Yun Huang
  - Data collection: Chao-Chuang Shih
  - System implementation: Kevin Wu, Derek Chuang, & Zhi-Wei Lee (李致緯), Roy Lu (盧柏儒)
  - System design: Richard Tsai & Hung-Yi Lee

NTU IOX center

# Demo: Video Caption Generation

Video




A girl is running.


A group of people is knocked by a tree.


A group of people is walking in the forest.

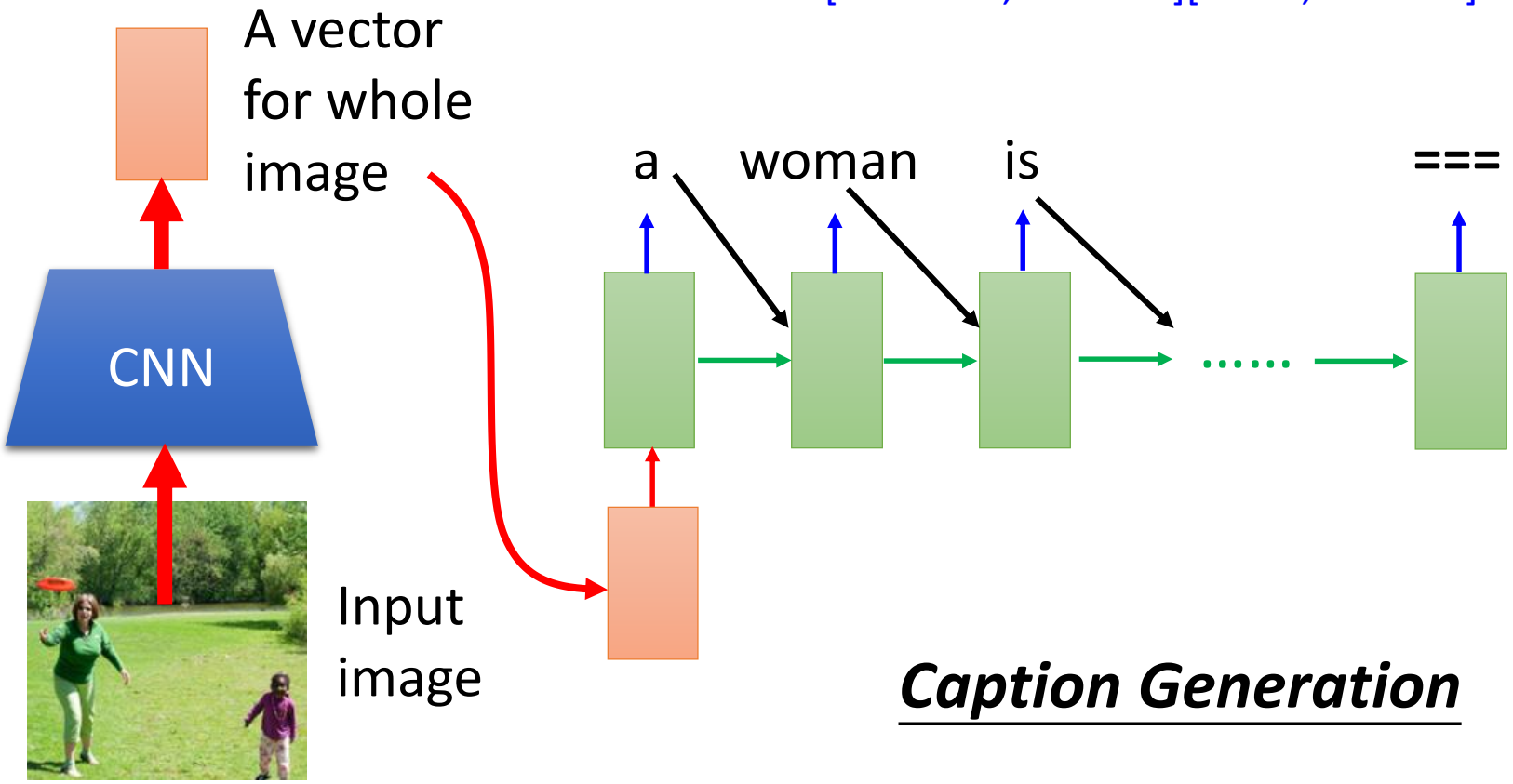
# Demo: Video Caption Generation

- Can machine describe what it see from video?
- Demo: 台大語音處理實驗室 曾柏翔、吳柏瑜、盧宏宗
- Video: 莊舜博、楊棋宇、黃邦齊、萬家宏

# Demo: Image Caption Generation

- Input an image, but output a sequence of words

[Kelvin Xu, arXiv'15][Li Yao, ICCV'15]

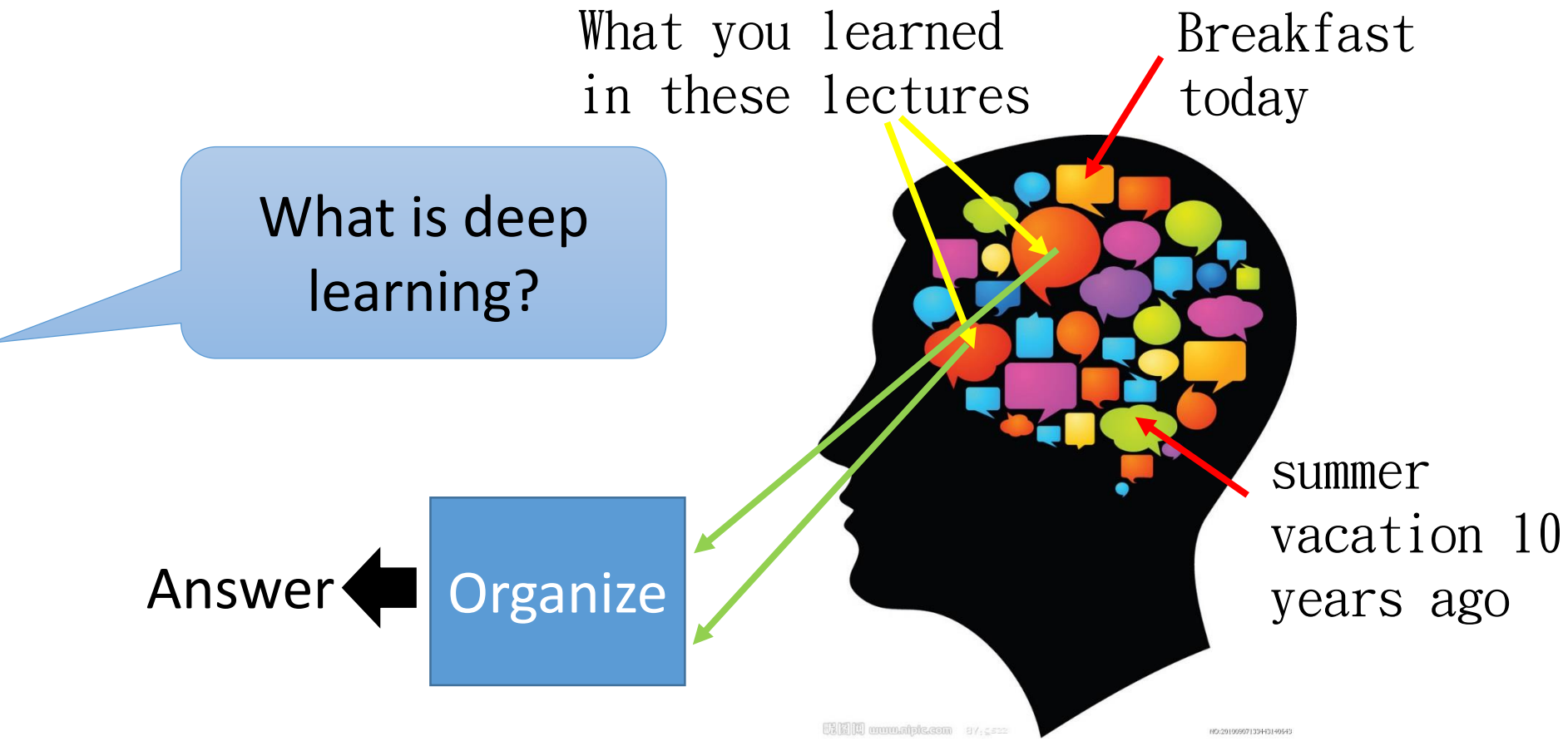


# Demo: Image Caption Generation

- Can machine describe what it see from image?
- Demo:台大電機系 大四 蘇子睿、林奕辰、徐翊祥、陳奕安

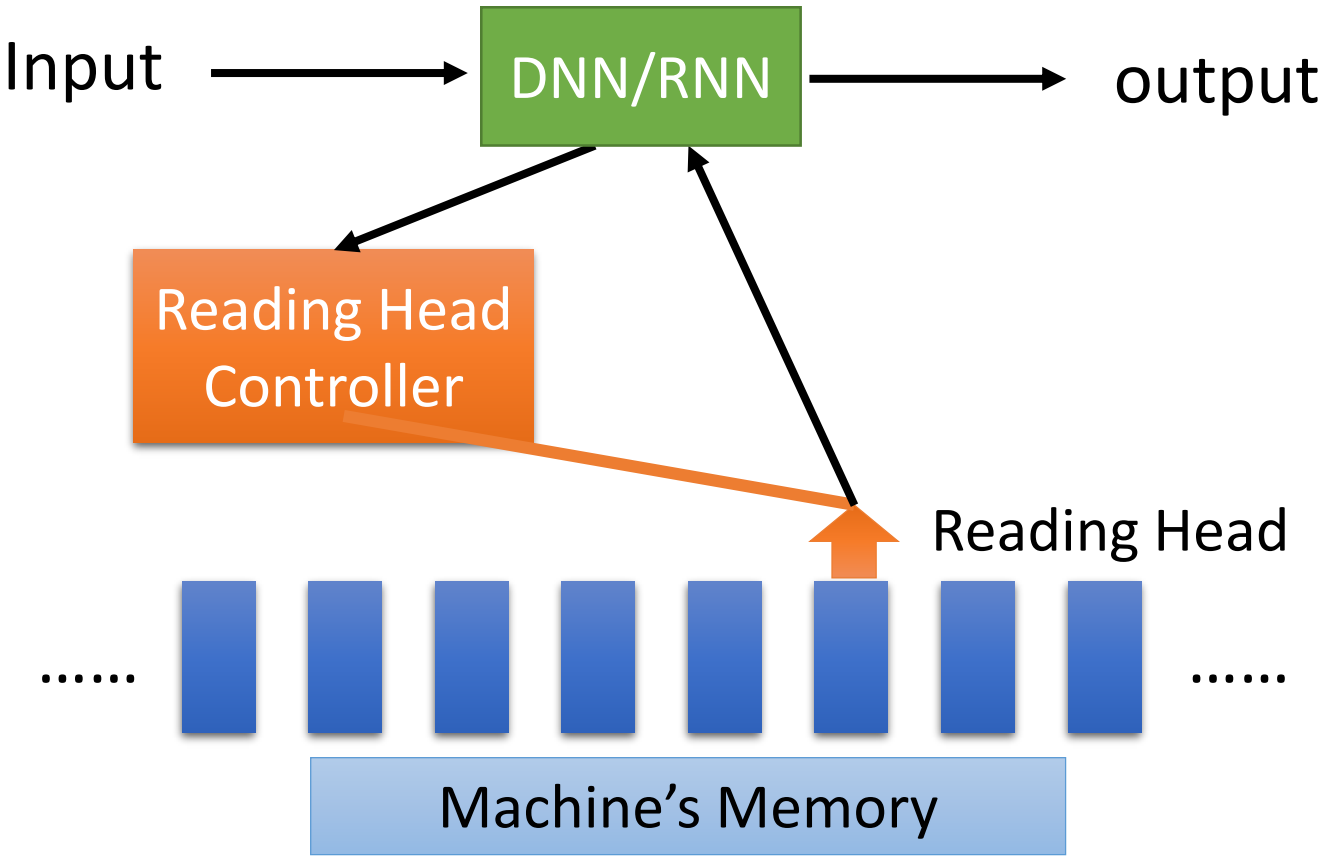
http://news.ltn.com.tw/photo/politics/breakingnews/975542_1

# Attention-based Model



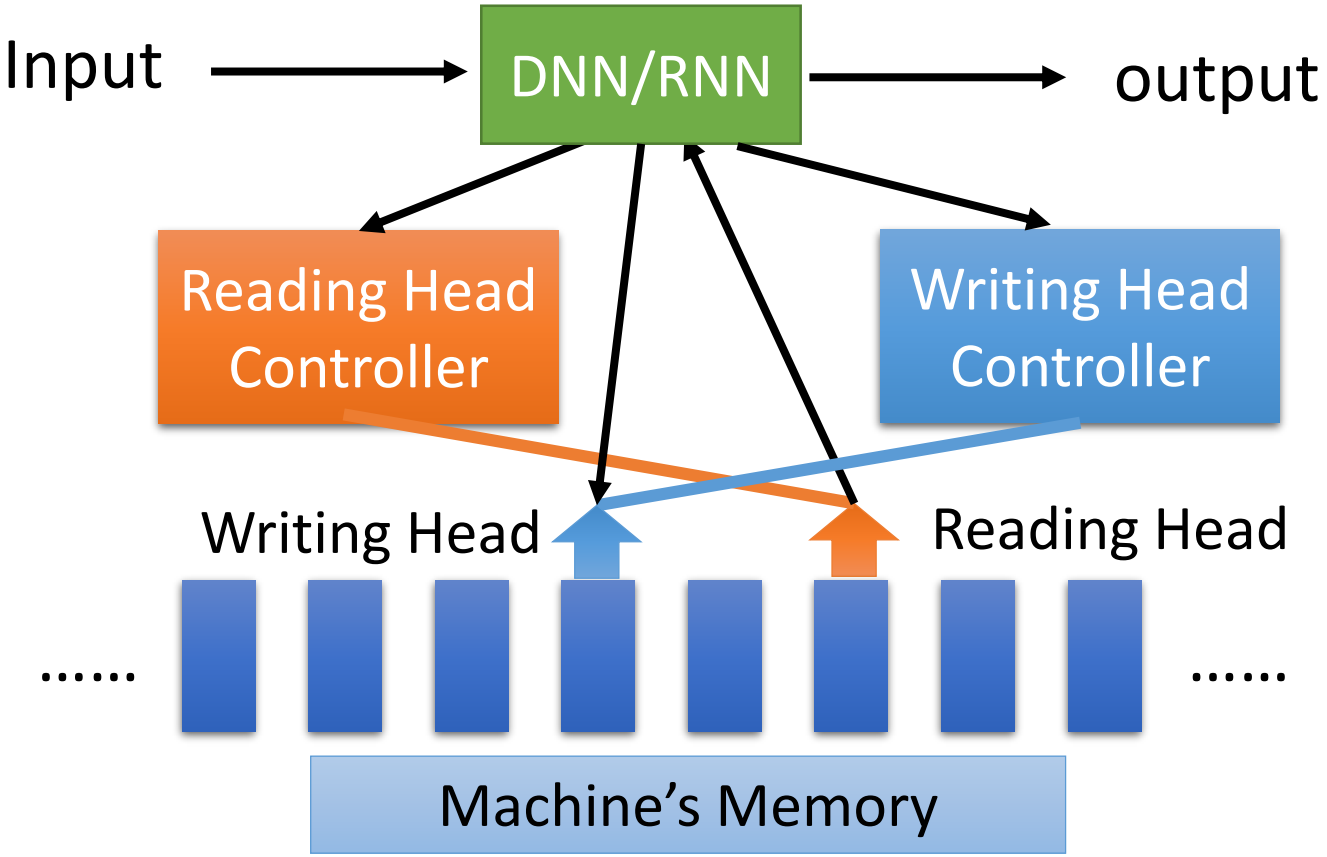

http://henrylo1605.blogspot.tw/2015/05/blog-post_56.html

# Attention-based Model



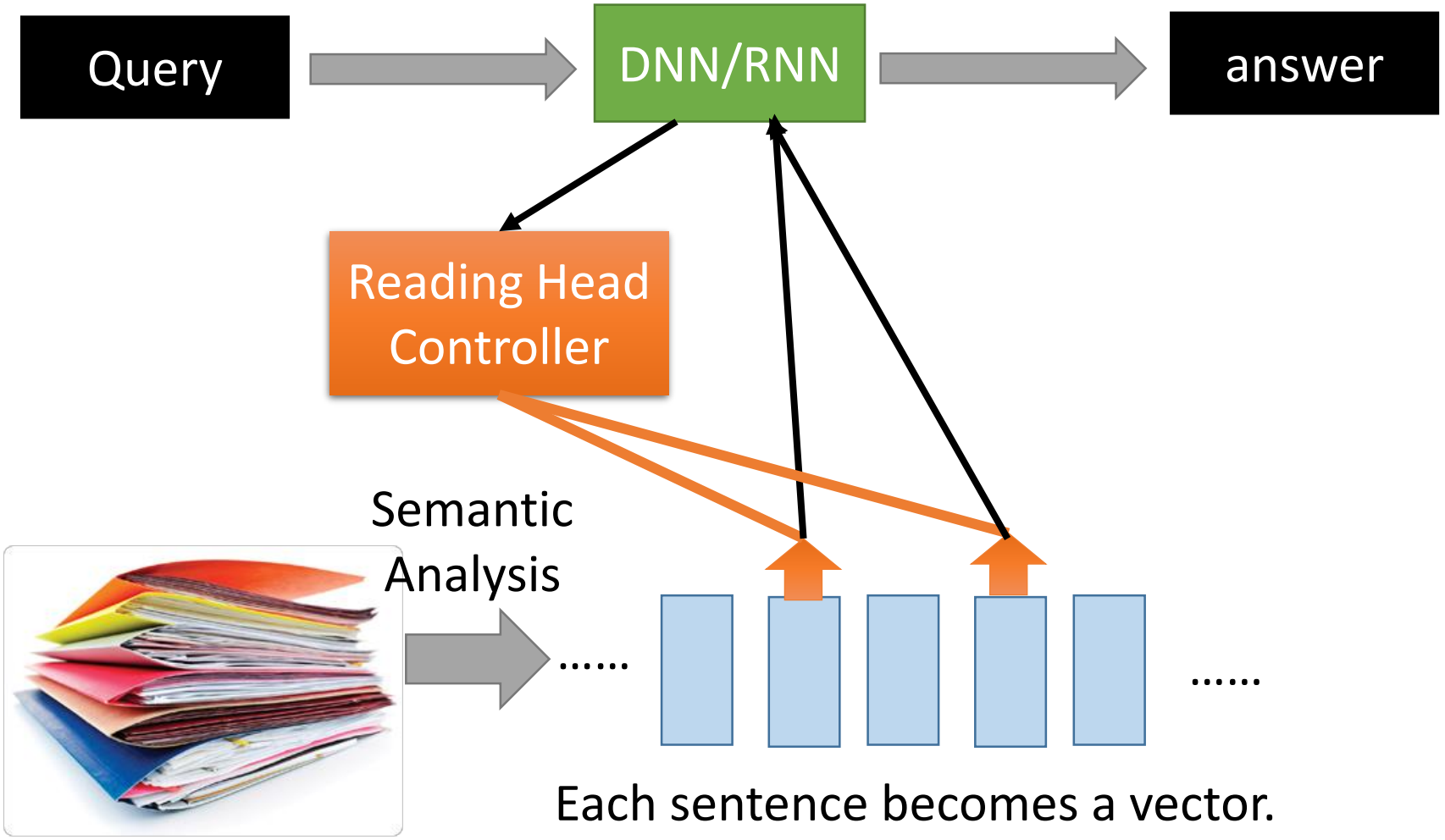

Ref:
http://speech.ee.ntu.edu.tw/~tlkagk/courses/MLDS_2015_2/Lecture/Attain%20(v3).ecm.mp4/index.html

# Attention-based Model v2




Neural Turing Machine

# Reading Comprehension



# Reading Comprehension

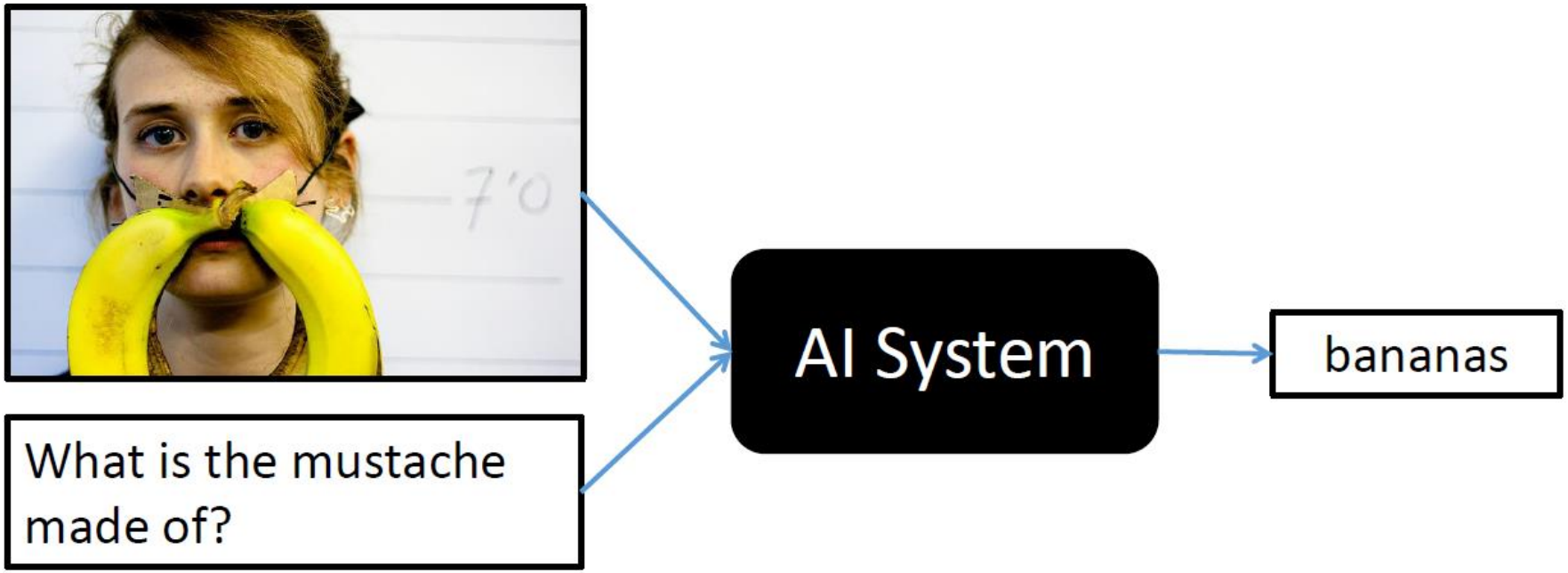
- End-To-End Memory Networks. S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus. NIPS, 2015.

The position of reading head:

| Story (16: basic induction) | Support | Hop 1 | Hop 2 | Hop 3 |
|---|---|---|---|---|
| Brian is a frog. | yes | 0.00 | 0.98 | 0.00 |
| Lily is gray. | | 0.07 | 0.00 | 0.00 |
| Brian is yellow. | yes | 0.07 | 0.00 | 1.00 |
| Julius is green. | | 0.06 | 0.00 | 0.00 |
| Greg is a frog. | yes | 0.76 | 0.02 | 0.00 |
| **What color is Greg? Answer: yellow** | **Prediction: yellow** | | | |

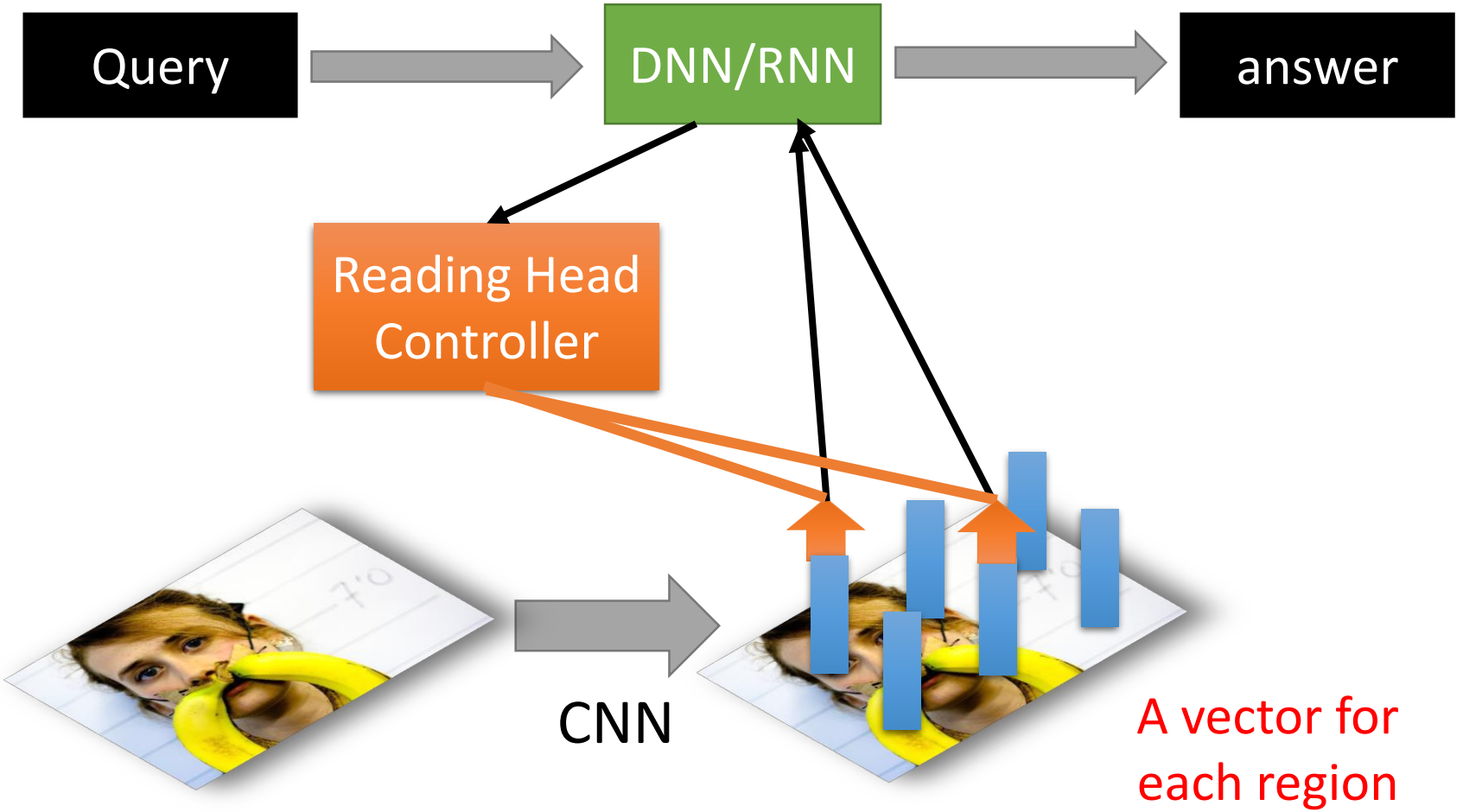
Keras has example:
https://github.com/fchollet/keras/blob/master/examples/babi_memnn.py

# Visual Question Answering




source: http://visualqa.org/

# Visual Question Answering



# Speech Question Answering

- **TOEFL Listening Comprehension Test by Machine**
- Example:

  Audio Story:  (The original story is 5 min long.)

  Question: “ What is a possible origin of Venus’ clouds? ”

  Choices:

  (A) gases released as a result of volcanic activity

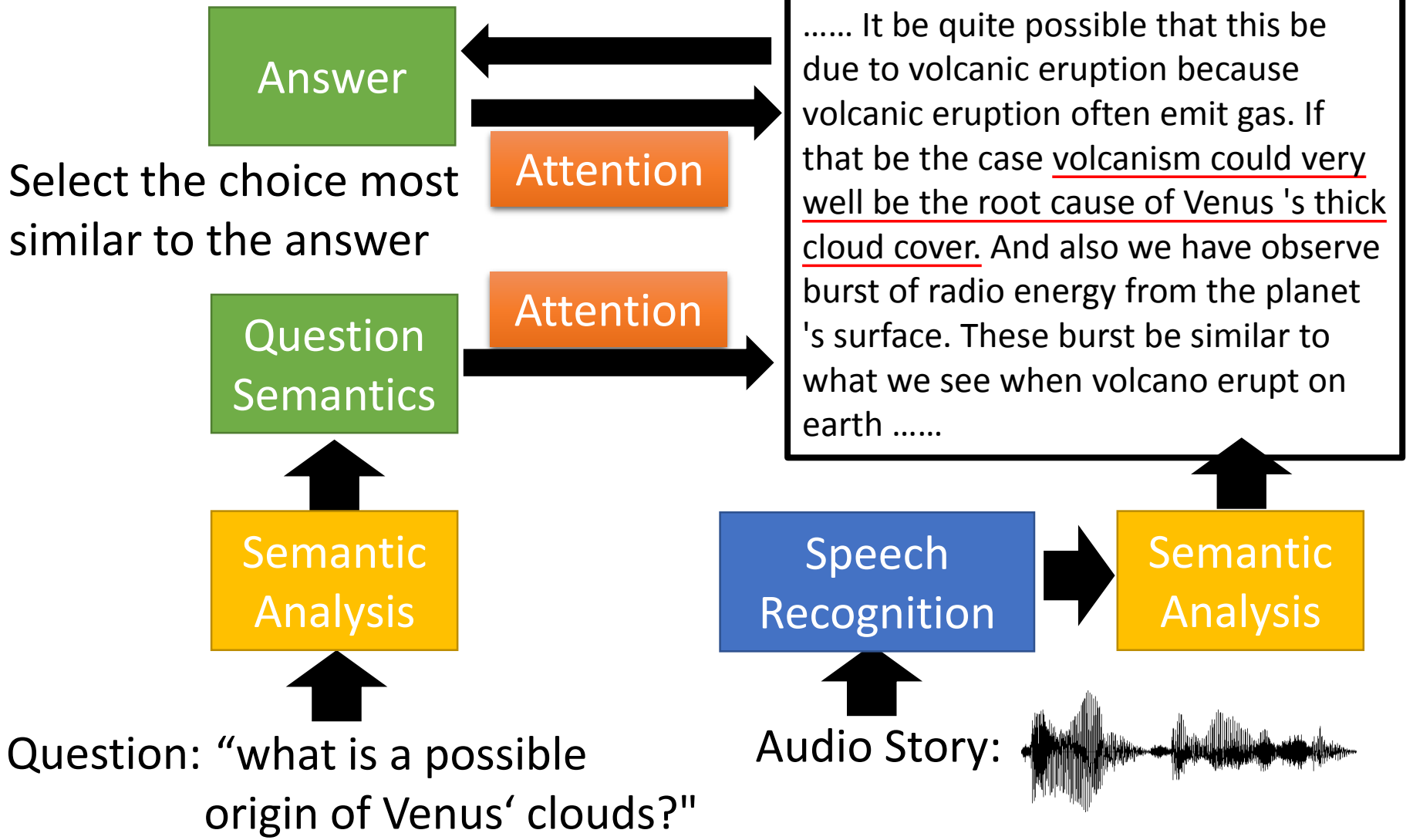
  (B) chemical reactions caused by high surface temperatures

  (C) bursts of radio energy from the plane's surface

  (D) strong winds that blow dust into the atmosphere

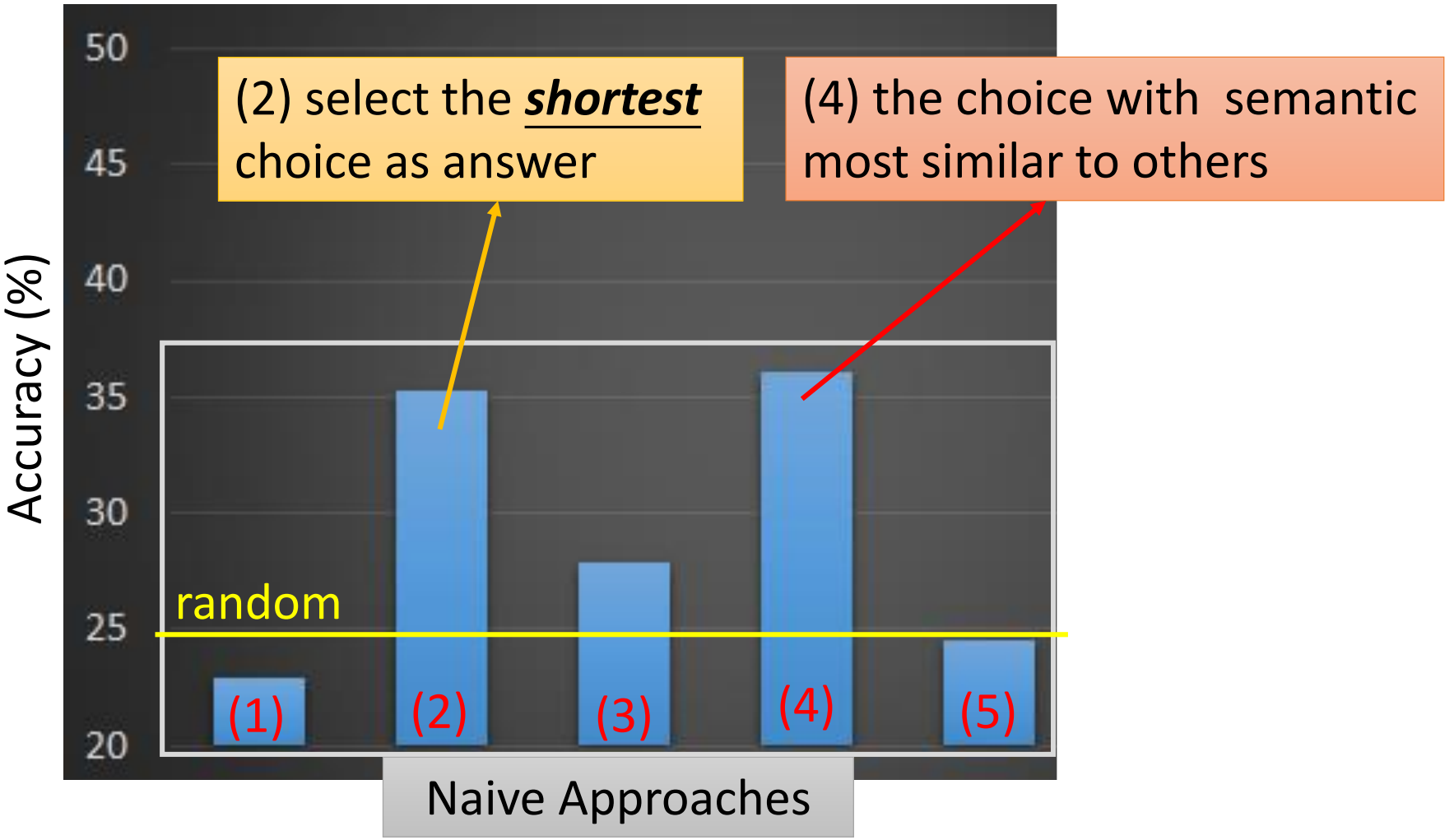
# Model Architecture

Everything is learned from training examples

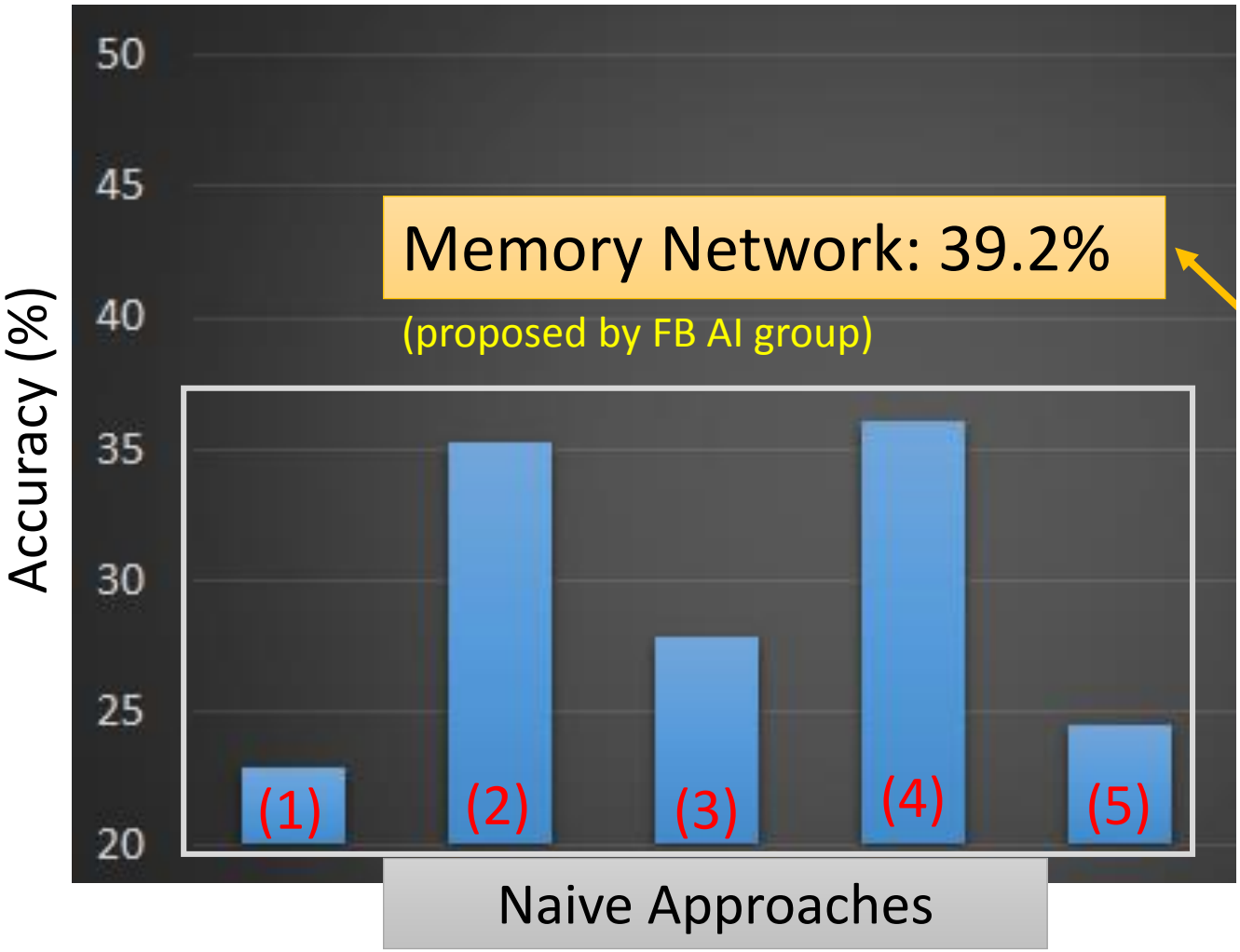


# Simple Baselines

Experimental setup:
717 for training,
124 for validation, 122 for testing



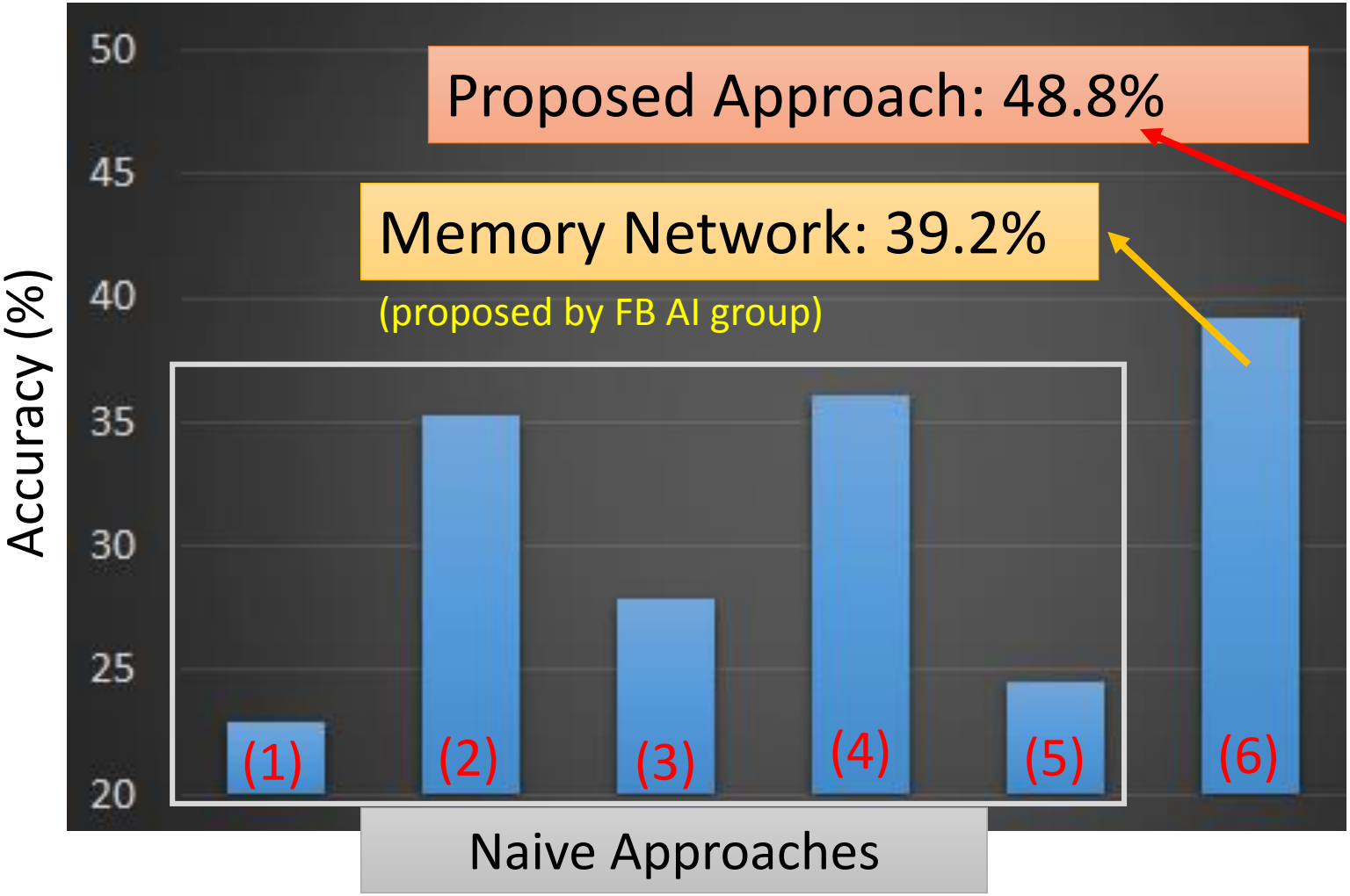
# Memory Network



# Proposed Approach

[Tseng & Lee, Interspeech 16]

[Fang & Hsu & Lee, SLT 16]



# To Learn More ……

- The Unreasonable Effectiveness of Recurrent Neural Networks
  - http://karpathy.github.io/2015/05/21/rnn-effectiveness/
- Understanding LSTM Networks
  - http://colah.github.io/posts/2015-08-Understanding-LSTMs/