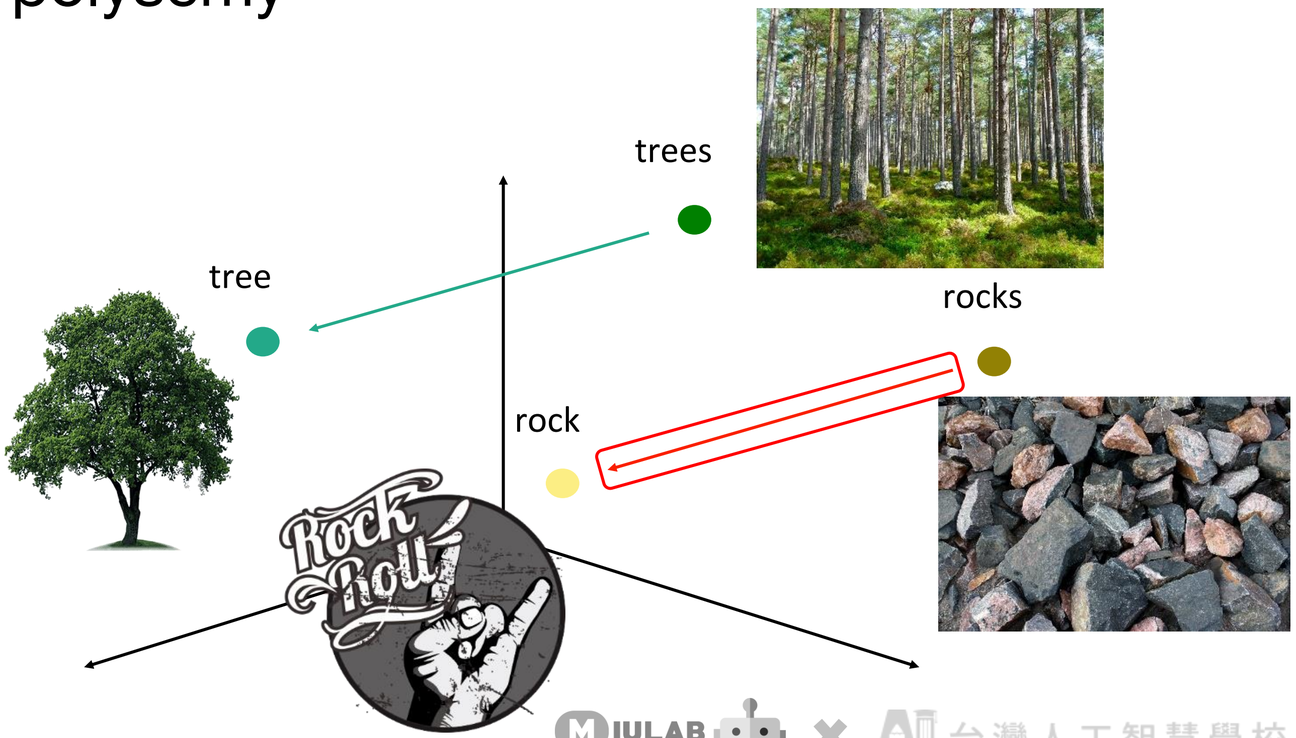
Contextualized Word Embeddings

Background

國立臺灣大學 資訊工程學系
陳縕儂 助理教授
http://vivianchen.idv.tw

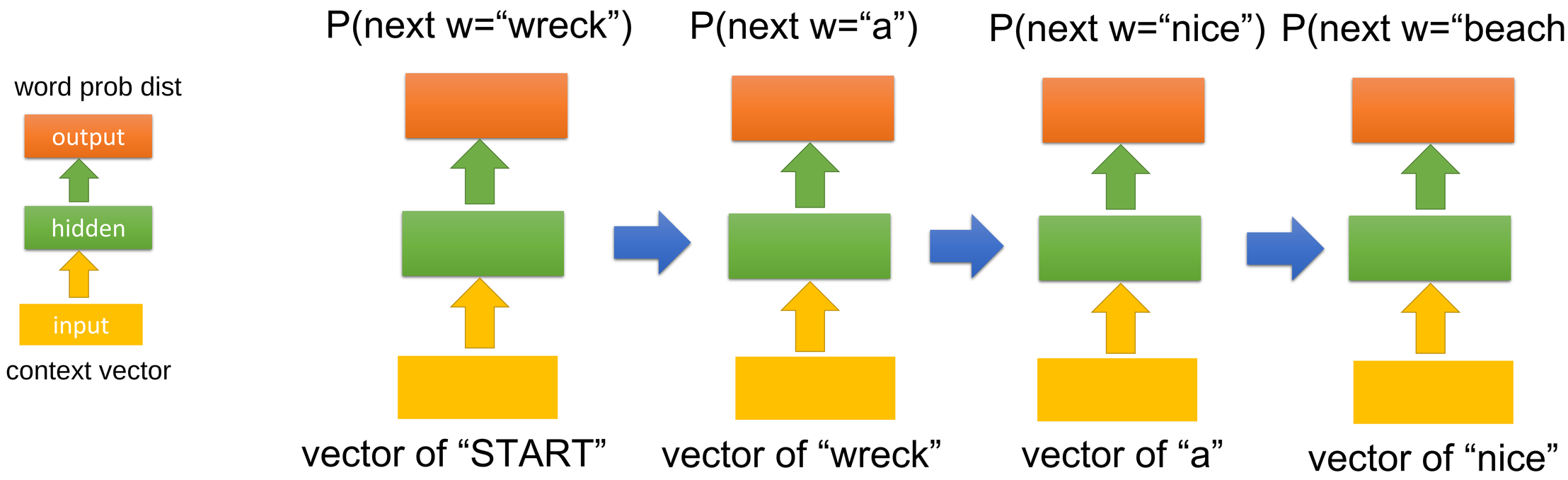
Word Embedding Polysemy Issue

- Words are polysemy
  - ✓ An apple a day, keeps the doctor away.
  - ✓ Smartphone companies including apple, …
- However, their embeddings are NOT polysemy
- Issue
  - ✓ Multi-senses (polysemy)
  - ✓ Multi-aspects (semantics, syntax)



RNNLM

- Idea: condition the neural network on all previous words and tie the weights at each time step

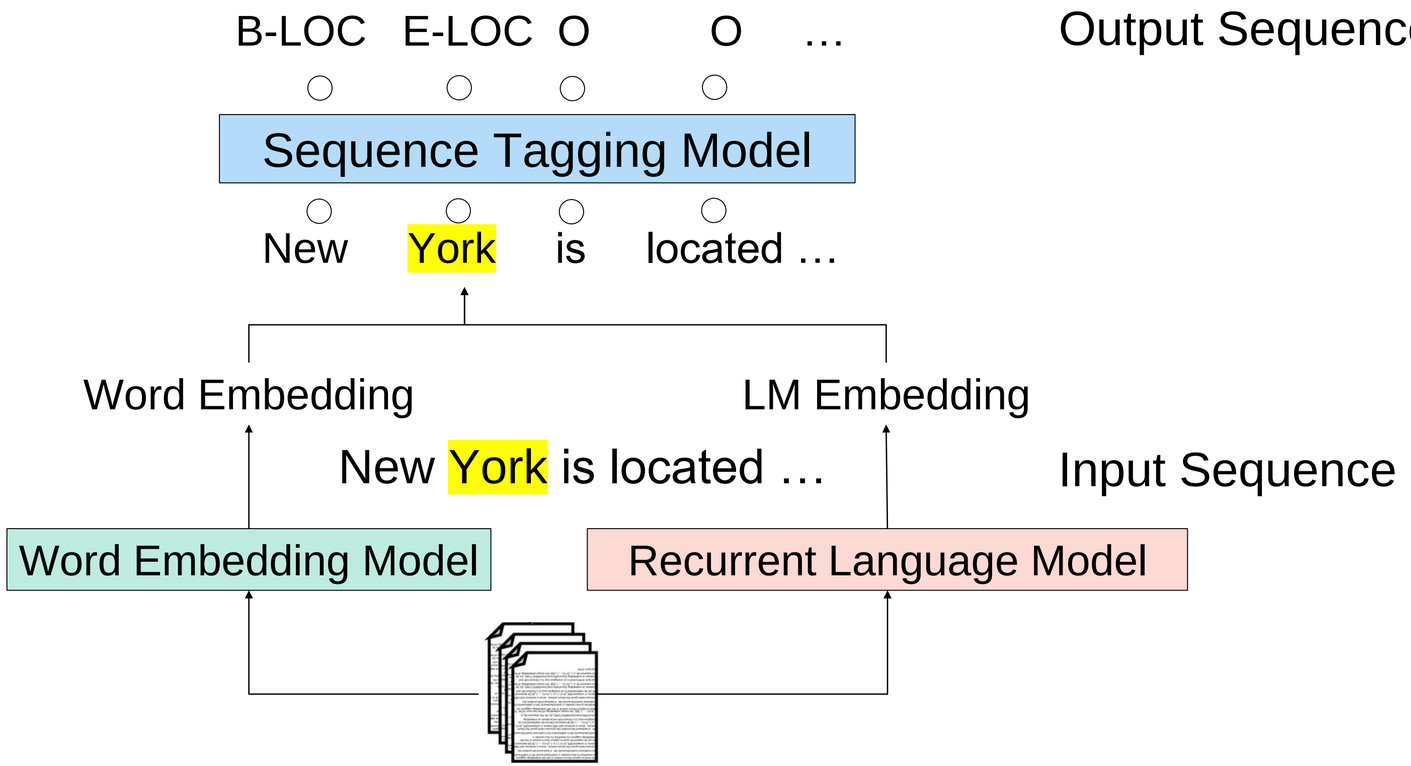


This LM producing **context-specific word representations** at each position

MIULAB × 台灣人工智慧學校

TagLM – “Pre-ELMo”

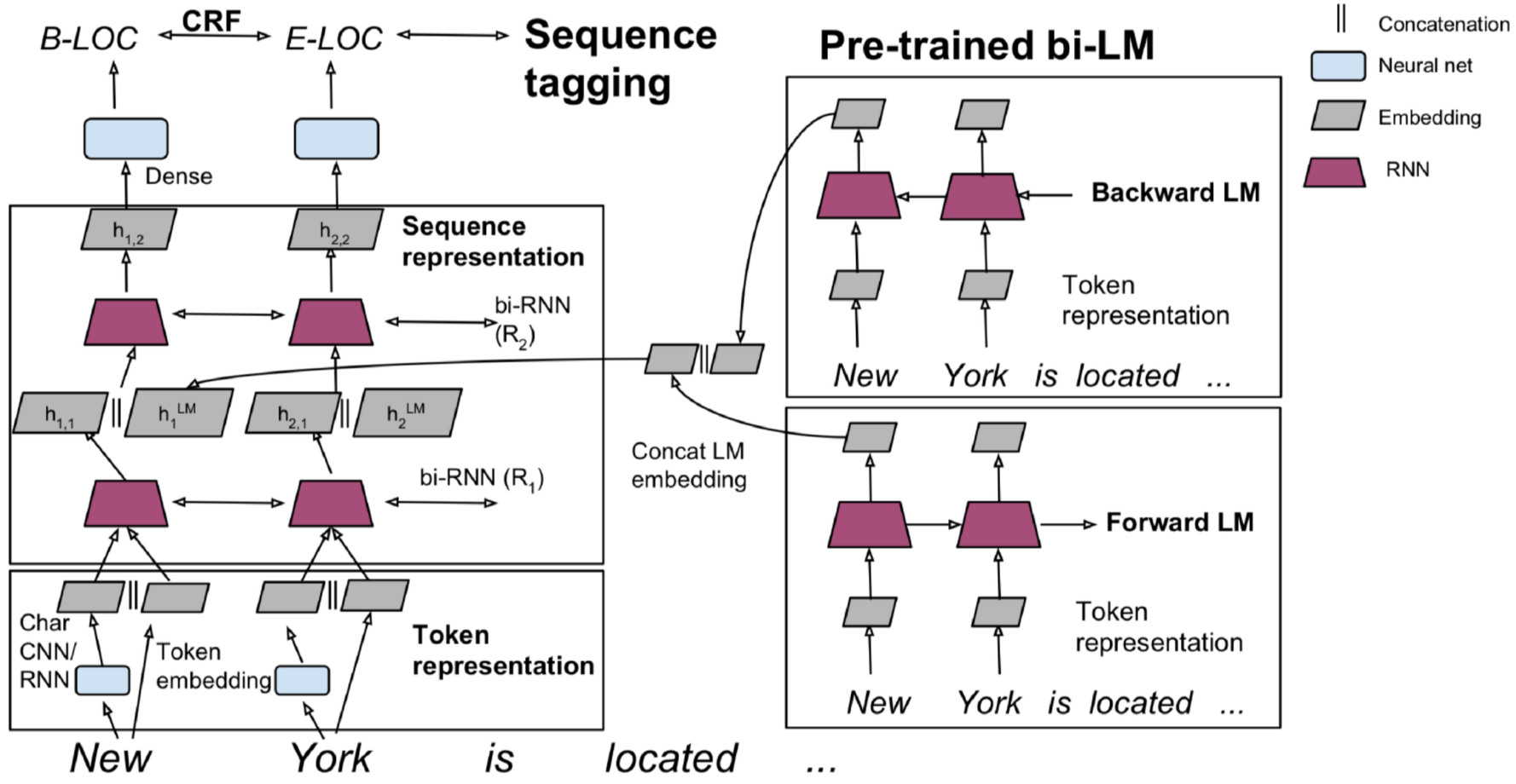
- Idea: train NLM on big unannotated data and provide the context-specific embeddings for the target task → **semi-supervised learning**



Peters et al., “Semi-supervised sequence tagging with bidirectional language models,” in *ACL*, 2017.

TagLM Model Detail

- Leveraging pre-trained LM information



Peters et al., "Semi-supervised sequence tagging with bidirectional language models," in *ACL*, 2017.

TagLM on Name Entity Recognition

The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.

Model	Description	CONLL 2003 F1
Klein+, 2003	MEMM softmax markov model	86.07
Florian+, 2003	Linear/softmax/TBL/HMM	88.76
Finkel+, 2005	Categorical feature CRF	86.86
Ratinov and Roth, 2009	CRF+Wiki+Word cls	90.80
Peters+, 2017	BLSTM + char CNN + CRF	90.87
Ma and Hovy, 2016	BLSTM + char CNN + CRF	91.21
TagLM (Peters+, 2017)	LSTM BiLM in BLSTM Tagger	**91.93**

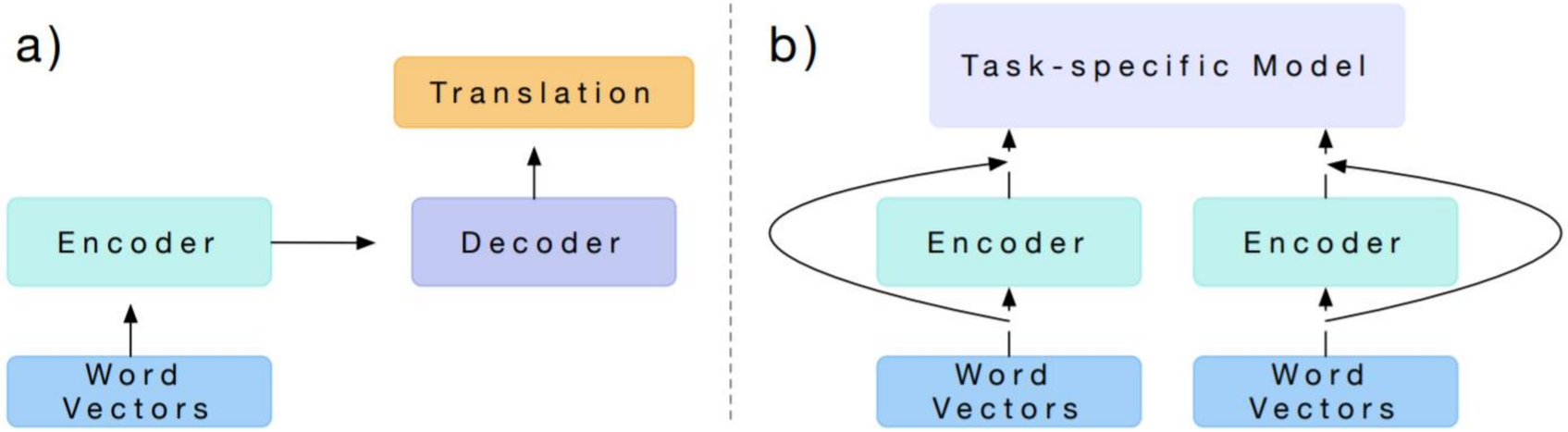
Peters et al., “Semi-supervised sequence tagging with bidirectional language models,” in *ACL*, 2017.

CoVe

- Idea: use trained sequence model to provide contexts to other NLP tasks

a) MT is to capture the meaning of a sequence

b) NMT provides the context for target tasks



CoVe vectors outperform GloVe vectors on various tasks

The results are not as strong as the simpler NLM training

McCann et al., “Learned in Translation: Contextualized Word Vectors”, in *NIPS*, 2017.