# *COMP5046*
# *Natural Language Processing*

*Lecture 13: Course Review*


***Dr. Caren Han***

*Semester 1, 2022*
*School of Computer Science,*
*University of Sydney*

# 0 LECTURE PLAN

**Lecture 13: Content Review and Exam Guide**

1. Overview
2. Learning Content Summary
3. Final Exam Guide

# COMP5046 Natural Language Processing



**Survey (5 mins)**

Have your say! ☺

*If you are reading this outside of the lecture, please do this now!*

- *Unit of Study Evaluation survey at:*

*https://student-surveys.sydney.edu.au/students/*

- *Your answers are completely anonymous*
- *Please write specific comments*
- *I personally read every single comment*
- *Thank you very much again!*

# COMP5046 Natural Language Processing



## What we learned in this course!

- Week 1: Introduction to Natural Language Processing (NLP)
- Week 2: Word Embeddings (Word Vector for Meaning)
- Week 3: Word Classification with Machine Learning I
- Week 4: Word Classification with Machine Learning II

NLP and Machine Learning

- Week 5: Language Fundamental
- Week 6: Part of Speech Tagging
- Week 7: Dependency Parsing
- Week 8: Language Model

NLP Techniques

- Week 9: Information Extraction: Named Entity Recognition
- Week 10: Advanced NLP: Attention and Reading Comprehension
- Week 11: Advanced NLP: Transformer and Machine Translation
- Week 12: Advanced NLP: Pretrained Model

Advanced Topic

- Week 13: Exam Review

# COMP5046 Natural Language Processing



## What we learned in this course!

- Week 1: Introduction to Natural Language Processing (NLP)
- Week 2: Word Embeddings (Word Vector for Meaning)
- Week 3: Word Classification with Machine Learning I
- Week 4: Word Classification with Machine Learning II

NLP and Machine Learning

- Week 5: Language Fundamental
- Week 6: Part of Speech Tagging
- Week 7: Dependency Parsing
- Week 8: Language Model

NLP Techniques

- Week 9: Information Extraction: Named Entity Recognition
- Week 10: Advanced NLP: Attention and Reading Comprehension
- Week 11: Advanced NLP: Transformer and Machine Translation
- Week 12: Advanced NLP: Pretrained Model

Advanced Topic

- Week 13: Exam Review

# 2 Lecture Summary

# Lecture Summary



## Lecture 1 Summary

### Word Representation with WordNet

- Use e.g. WordNet, a thesaurus containing lists of synonym sets and hypernyms ("is a" relationships).
- Problems with resources like WordNet

① up-to-date, require human to maintain

② ..

### Representing words as discrete symbols

- One-hot vector ☺

  motel = [0 0 0 0 0 0 0 0 0 0 **1** 0 0 0 0]

  hotel = [0 0 0 0 0 0 0 **1** 0 0 0 0 0 0 0]

In web search, if user searches for "**Seattle motel**", we would like to match documents containing "**Seattle hotel**"

- Problems with words as discrete symbols

# Lecture Summary

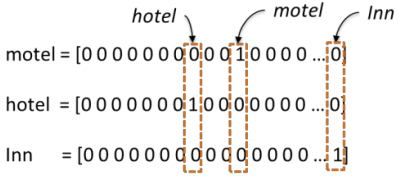


## Lecture 1 Summary

### Count based Word representation

*Important to know: how each techniques works / pros and cons*

- ***One-hot encoding***




sparse

semantic similarity

- ***Bag of Words***

[1, 1, 1, 1, 2, 2, 2, 1, 1, 3]

| a | are | been | day | have | how | nice | see | to | you |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 3 |


no order just frequency

- ***Term Frequency-Inverse Document Frequency***

***TF*** is a count of how many times a word occurs in a given document (= bag of words)

***IDF*** is the number of times a word occurs in a corpus of documents

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

pros cons

# Lecture Summary

## Lecture 2 Summary

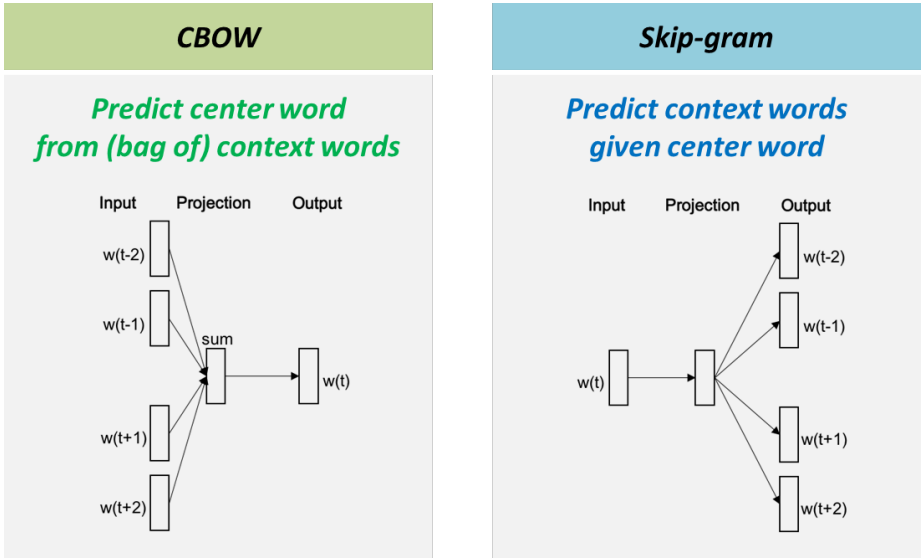
### Prediction-based Word Representation

*Important to know: how each techniques works / pros and cons*

CBOW & SG

context-based word prediction-based word

- ***Word2Vec (CBOW, Skip-Gram)***




***Key parameter for training methods: window size, negative samples***

# 2 Lecture Summary

Fasttext vs Wordvec
↓
OOV problem

## Lecture 2 Summary

### Prediction-based Word Representation

*Important to know: how each techniques works / pros and cons*

- ***FastText with n-gram embedding***


- ***Glove***

  *"Methods like skip-gram may do better on the analogy ask, but they poorly utilize the statistics of the corpus since they train on separate local context windows instead of on global co-occurrence counts."*

# Lecture Summary



## Lecture 3 Summary

### Word Embedding Evaluation

- Intrinsic Evaluation
- Extrinsic Evaluation

| Type | How to work / Benefit |
|---|---|
| Intrinsic | Evaluation on a specific/intermediate subtask |
| | • Fast to compute<br>• Helps to understand that system<br>• Not clear if really helpful unless correlation to real task is established |
| Extrinsic | Evaluation on a real task |
| | • Can take a long time to compute accuracy<br>• Unclear if the subsystem is the problem or its interaction or other subsystems |

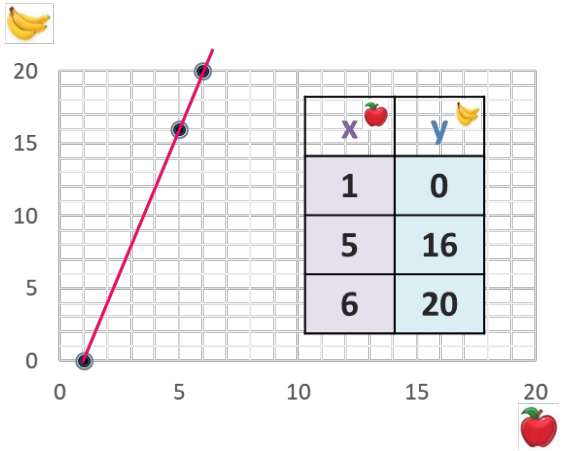
# Lecture Summary

## Lecture 3 Summary

### Deep Learning for Natural Language Processing

- Perceptron and Neural Network (NN)
- Multilayer Perceptron

*Input: x=number of apple given by Lisa*
*Output: y=number of banana received by Lisa*
*Parameters: Need to be estimated*




*Model* $y = Wx + b$ (*weight*, *bias*)

*How can we find the parameters, w and b?*

# Lecture Summary

## Lecture 3 Summary

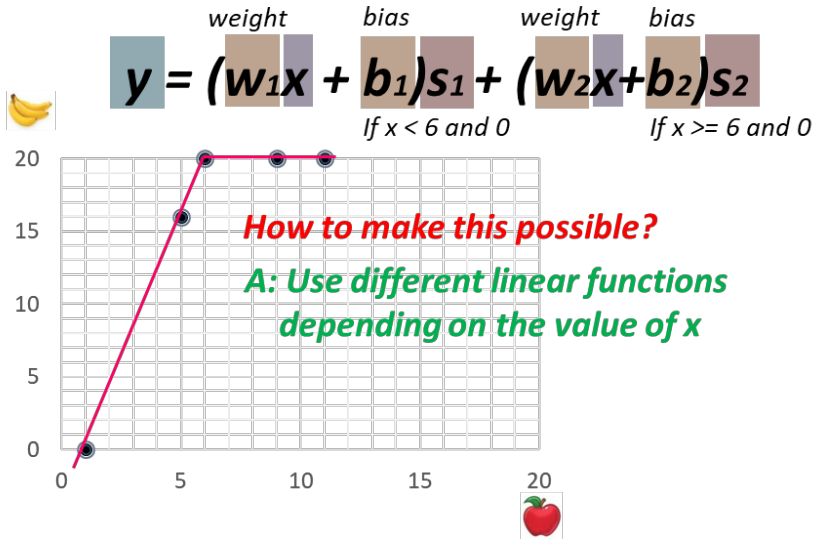
**Deep Learning for Natural Language Processing**

- Perceptron and Neural Network (NN)
- Multilayer Perceptron

Nonlinear Neural Network

*Data*

| x | y |
|---|---|
| 1 | 0 |
| 5 | 16 |
| 6 | 20 |
| 9 | 20 |
| 11 | 20 |



# Lecture Summary

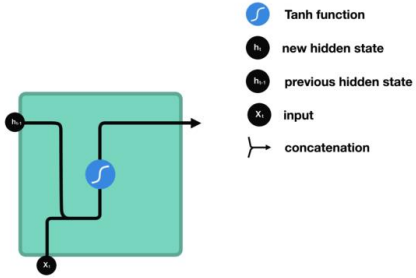


## Lecture 4 Summary

### Seq2Seq Learning

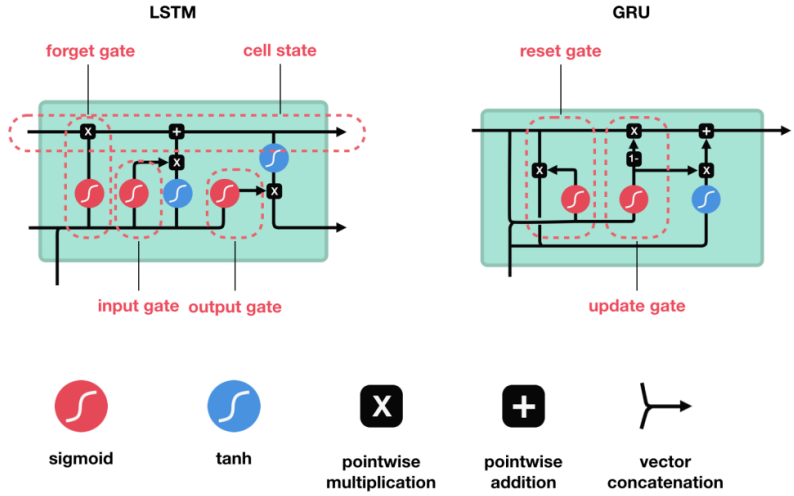
### Seq2Seq Deep Learning

1. RNN (Recurrent Neural Network)
2. LSTM (Long Short-Term Memory)
3. GRU (Gated Recurrent Unit)




$$h_t = f_W(h_{t-1}, x_t)$$

New hidden state — Previous state — input — A function with parameters W



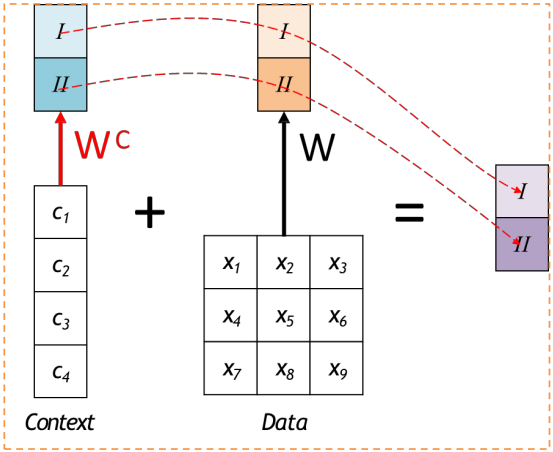
# Lecture Summary

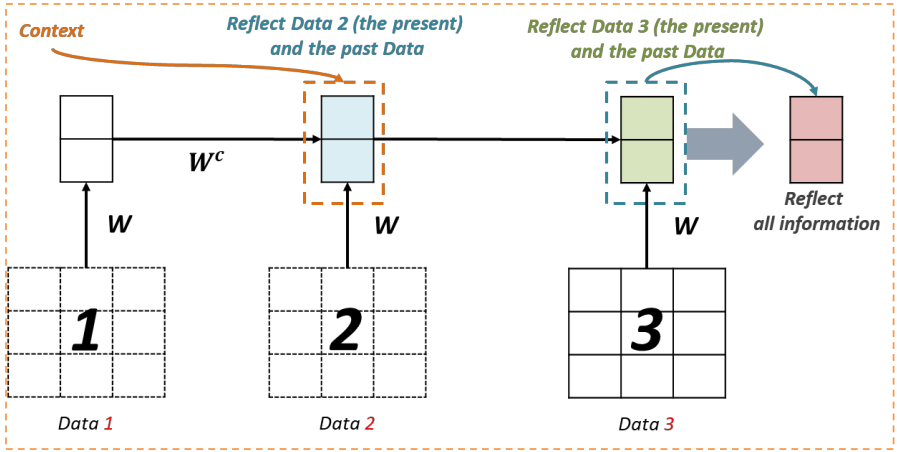
## Lecture 4 Summary

### Seq2Seq Learning

### Seq2Seq Deep Learning

1. RNN (Recurrent Neural Network)
2. LSTM (Long Short-Term Memory)
3. GRU (Gated Recurrent Unit)



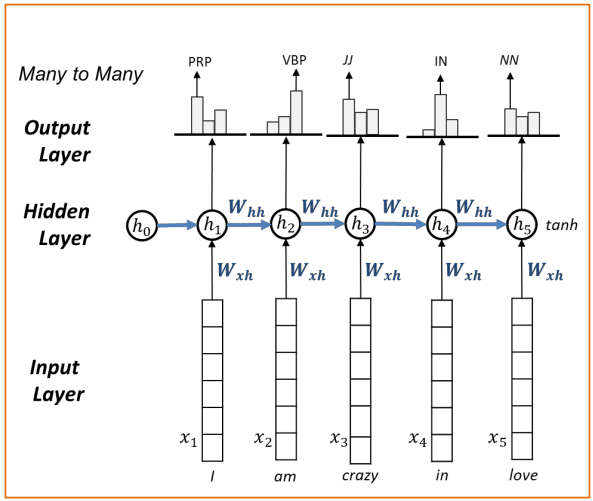
# 2 Lecture Summary

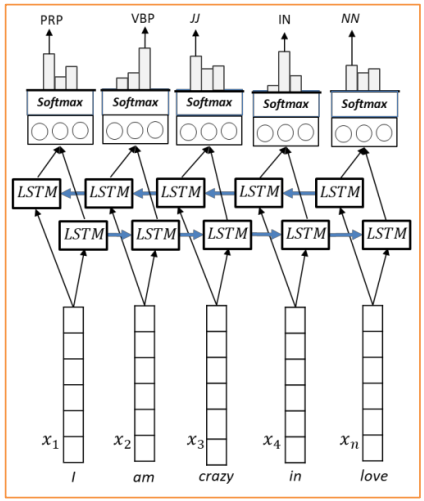
## Lecture 4 Summary

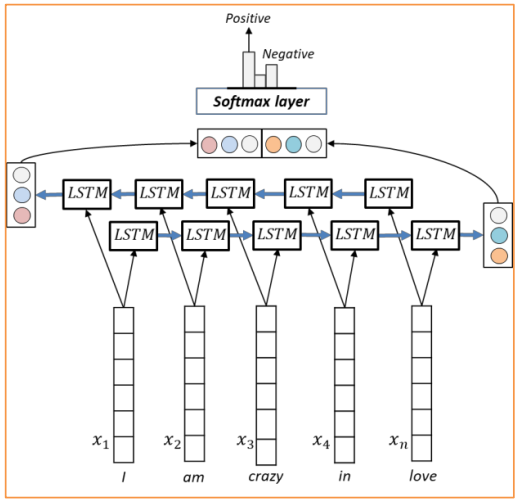
### Seq2Seq Learning

### Seq2Seq Deep Learning

1. RNN (Recurrent Neural Network)
2. LSTM (Long Short-Term Memory)
3. GRU (Gated Recurrent Unit)



# Lecture Summary

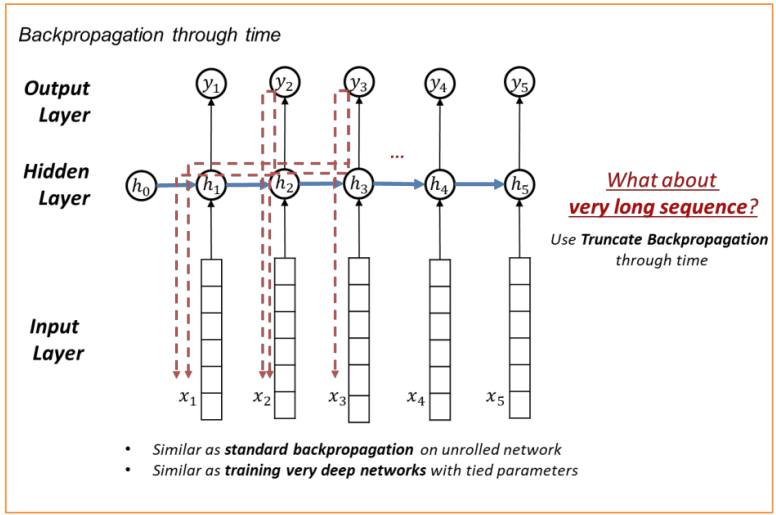


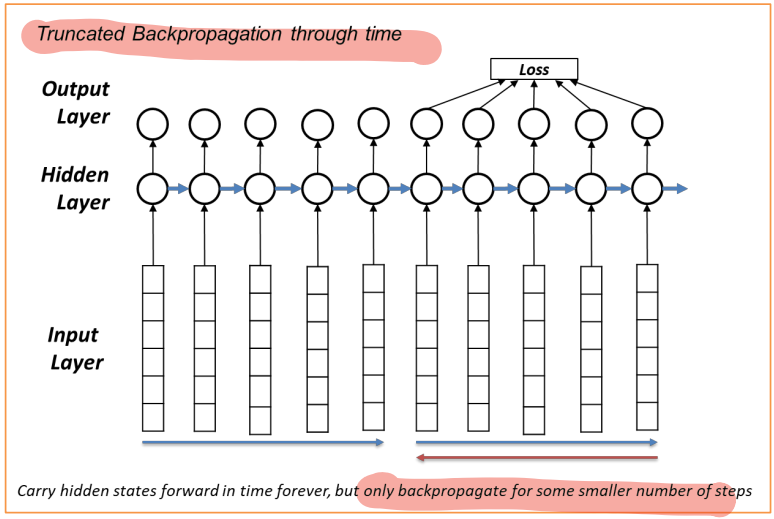
## Lecture 4 Summary

### Seq2Seq Learning

### Seq2Seq Deep Learning

1. RNN (Recurrent Neural Network)
2. LSTM (Long Short-Term Memory)
3. GRU (Gated Recurrent Unit)



# Lecture Summary

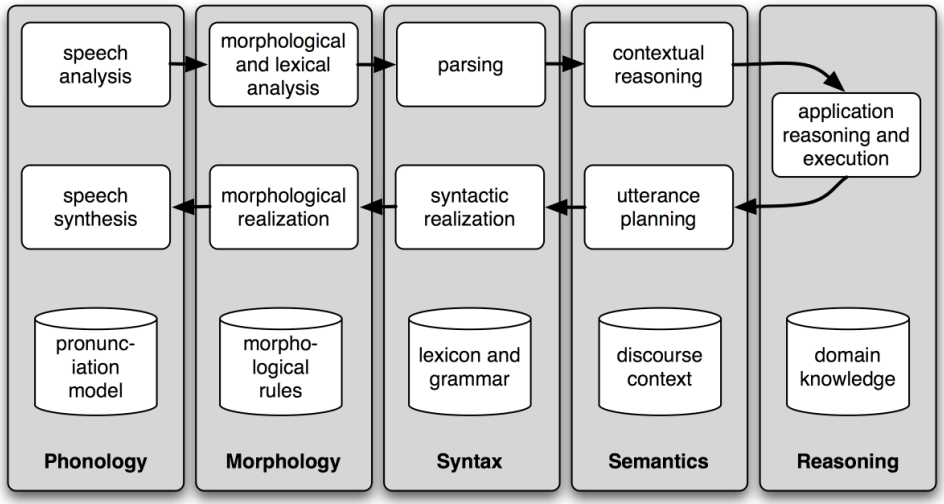
## Lecture 5 Summary

### Language Fundamental

1. Phonology, Morphology, Syntax, Semantics, Pragmatics

### Text Preprocessing

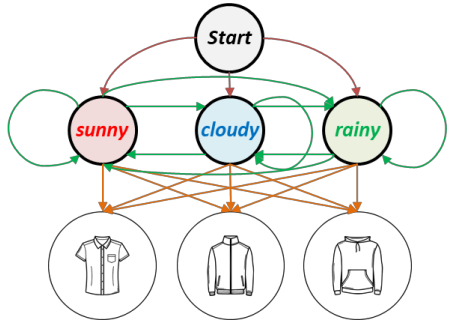
1. Tokenization
2. Cleaning and Normalisation
3. Stemming and Lemmatisation
4. Stopword
5. Regular Expression



# Lecture Summary

## Lecture 6 Summary

1. Part-of-Speech Tagging N-N
2. Baseline Approaches
   1. Rule-based Model
   2. Look-up Table Model
   3. N-Gram Model
3. Probabilistic Approaches
   1. Hidden Markov Model
   2. Conditional Random Field
4. Deep Learning Approaches




*Predicting the **weather (hidden variable)** based on the type of **clothes that someone wears (observed)***

**Initial**

| | |
|---|---|
| Rainy | 0.6 |
| Cloudy | 0.3 |
| Sunny | 0.1 |

**Transitions**

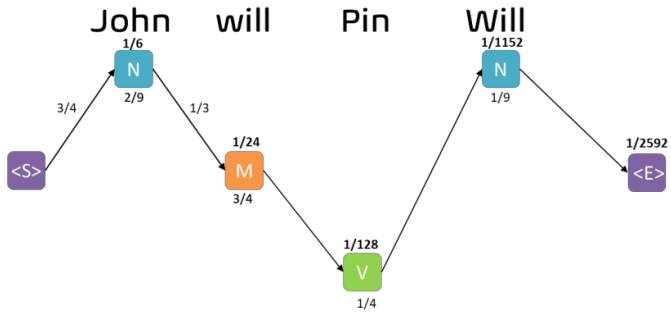
| | | Tomorrow | | |
|---|---|---|---|---|
| | | Rainy | Cloudy | Sunny |
| Today | Rainy | 0.6 | 0.3 | 0.1 |
| | Cloudy | 0.4 | 0.3 | 0.3 |
| | Sunny | 0.1 | 0.4 | 0.5 |

**Emissions**

| | Shirts | Jacket | Hoodies |
|---|---|---|---|
| Rainy | 0.8 | 0.19 | 0.01 |
| Cloudy | 0.5 | 0.4 | 0.1 |
| Sunny | 0.01 | 0.2 | 0.79 |

| | N | V | M |
|---|---|---|---|
| Emma | 4/9 | 0 | 0 |
| John | 2/9 | 0 | 0 |
| Will | 1/9 | 0 | 3/4 |
| Pin | 2/9 | 1/4 | 0 |
| Can | 0 | 0 | 1/4 |
| Meet | 0 | 2/4 | 0 |
| Pat | 0 | 1/4 | 0 |

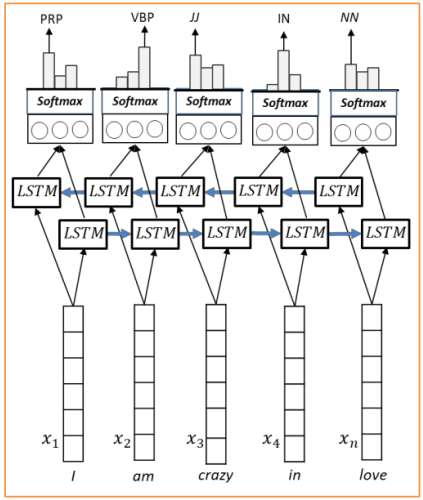
| | N | V | M | <E> |
|---|---|---|---|---|
| <S> | 3/4 | 0 | 1/4 | 0 |
| N | 1/9 | 1/9 | 3/9 | 4/9 |
| V | 4/4 | 0 | 0 | 0 |
| M | 1/4 | 3/4 | 0 | 0 |

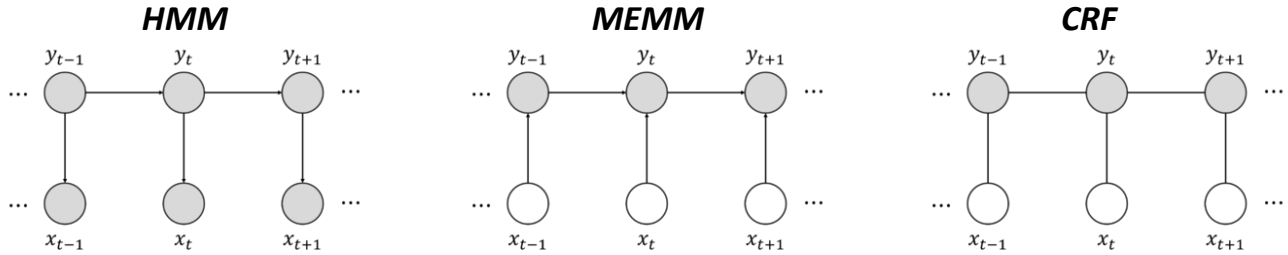


# Lecture Summary

## Lecture 6 Summary

1. Part-of-Speech Tagging
2. Baseline Approaches
   1. Rule-based Model
   2. Look-up Table Model
   3. N-Gram Model
3. Probabilistic Approaches
   1. Hidden Markov Model
   2. Conditional Random Field
4. Deep Learning Approaches



# Lecture Summary

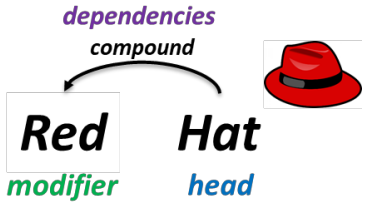
## Lecture 7 Summary

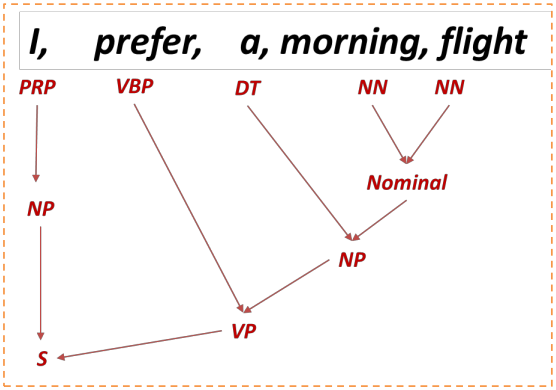
**Linguistic Structure**

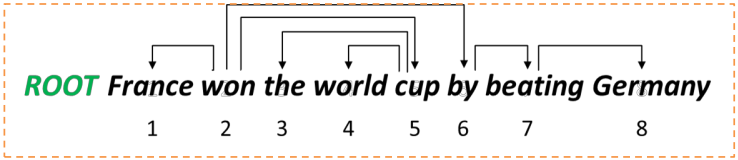
**Dependency Structure**

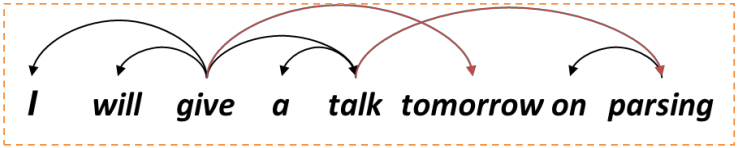
**Dependency Parsing Algorithms**

1. Transition-based Dependency Parsing
2. Deep Learning-based Dependency Parsing



# Lecture Summary



## Lecture 7 Summary

**Linguistic Structure**

**Dependency Structure**

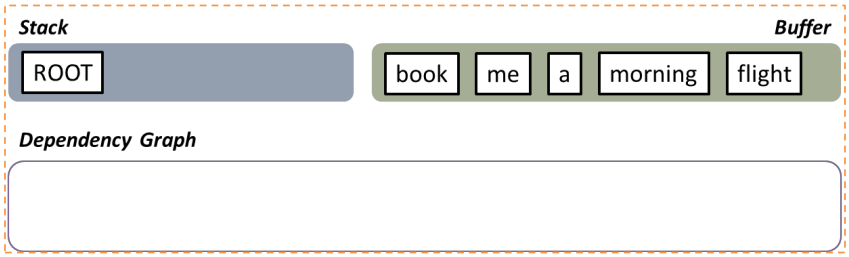
**Dependency Parsing Algorithms**

1. Transition-based Dependency Parsing
2. Deep Learning-based Dependency Parsing

***Graph-based Dependency Parsing***

Transition

~~*Transaction*~~***-based Dependency Parsing***




*How to choose next action?*

***Unlabeled Attachment Score**
***Labeled Attachment Score**

# Lecture Summary



## Lecture 8 Summary

Language Model

**Traditional Language Model**

Neural Language Model

Natural Language Generation

Other NLG Approaches

Language Model and NLG Evaluation

***P(An, adorable, little, boy, is, spreading, smiles)***

*= P(An) × P(adorable|An) × P(little|An adorable) × P(boy|An adorable little) × P(is|An adorable little boy) × P(spreading|An adorable little boy is) × P(smiles|An adorable little boy is spreading)*

***Trained Corpus***

*An adorable little boy is .....*

*..... An adorable little boy is*

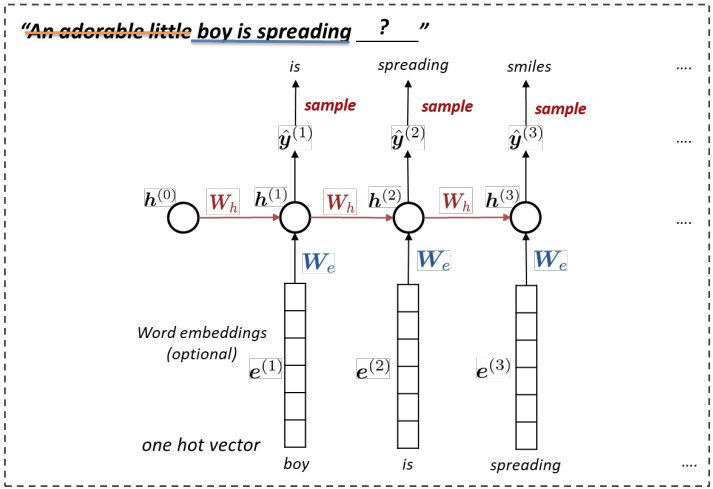
*An adorable little boy laughed*

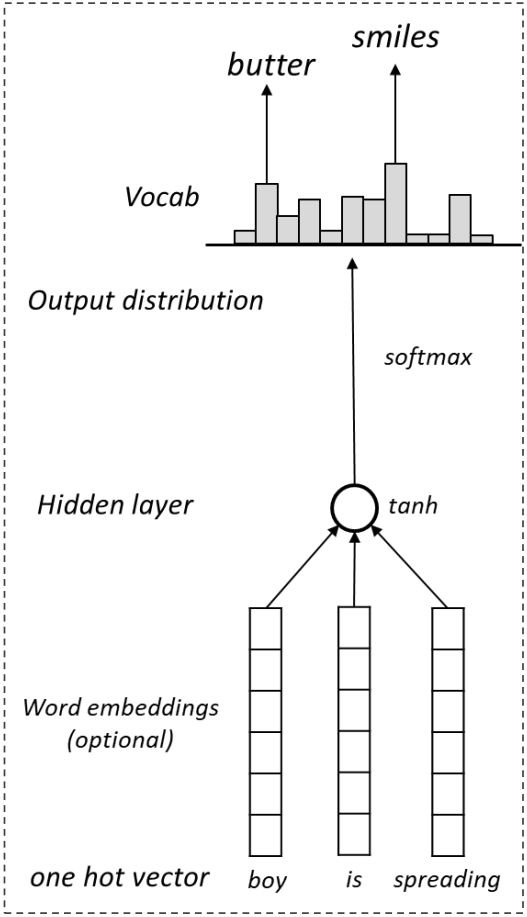
- ***Count-based***
- ***N-gram Language Models***

# Lecture Summary

## Lecture 8 Summary

Language Model
**Traditional Language Model**
Neural Language Model
Natural Language Generation
Other NLG Approaches
Language Model and NLG Evaluation



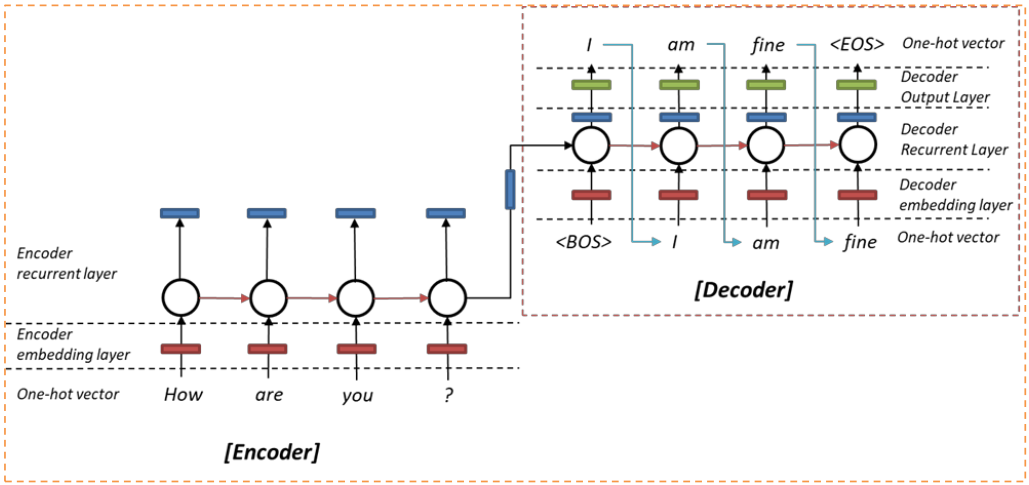



# Lecture Summary



## Lecture 8 Summary

Language Model
Traditional Language Model
Neural Language Model
**Natural Language Generation**
Other NLG Approaches
Language Model and NLG Evaluation



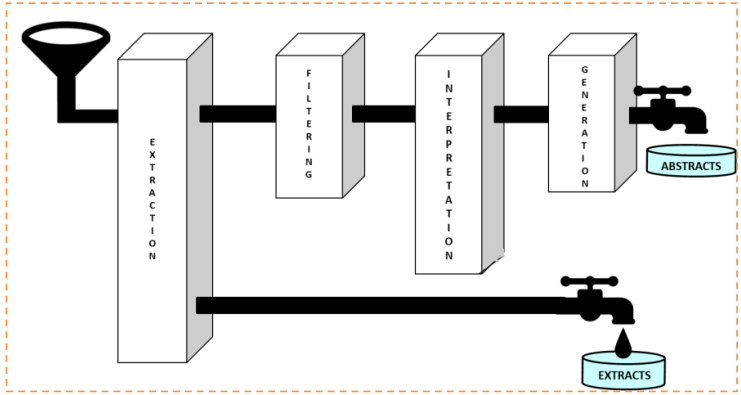

***Decoding Algorithm***
*1: Greedy Decoding*
*2: Beam Search*

# Lecture Summary

## Lecture 8 Summary

Language Model
Traditional Language Model
Neural Language Model
Natural Language Generation
**Other NLG Approaches**
Language Model and NLG Evaluation

"Flights from *ORIGIN* to *DEST* on *DEPT_DATE* *DEPT_TIME*. Just one moment please"

# Lecture Summary



## Lecture 8 Summary

Language Model
Traditional Language Model
Neural Language Model
Natural Language Generation
Other NLG Approaches
**Language Model and NLG Evaluation**

NLG cannot use perplexity directly
natural & answer correctly

***Perplexity***

$$\text{perplexity} = \prod_{t=1}^{T} \left( \frac{1}{P_{\text{LM}}(\boldsymbol{x}^{(t+1)} | \ \boldsymbol{x}^{(t)}, \ldots, \boldsymbol{x}^{(1)})} \right)^{1/T 1/T}$$

*So, Lower Perplexity is better!*

# Lecture Summary



## Lecture 9 Summary

### Information Extraction

### Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

### Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation

***Rule-based NER***

1. *Determining **which person holds what office in what organization***
2. *Determining **where an organization is located***

| Unigram | Mr. | Scott | Morrison | flew | to | Beijing |
|---|---|---|---|---|---|---|
| Lowercase unigram | mr. | scott | morrison | flew | to | beijing |
| POS tag | nnp | nnp | nnp | vbd | to | nnp |
| length | 3 | 5 | 4 | 4 | 2 | 7 |
| In first-name gazetteer | no | yes | no | no | no | no |
| In location gazetteer | no | no | no | no | no | yes |
| 3-letter suffix | Mr. | ott | son | lew | - | ing |
| 2-letter suffix | r. | tt | on | ew | to | ng |
| 1-letter suffix | . | t | n | w | o | g |
| Tag predictions | O | B-per | I-per | O | O | B-loc |

# Lecture Summary



## Lecture 9 Summary

### Information Extraction

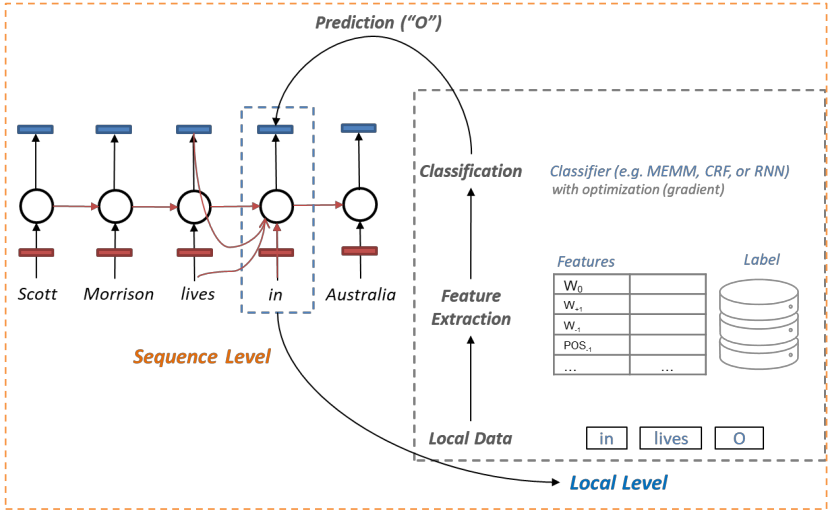
### Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

### Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation

| | Josiah | tells | Caren | John | Smith | is | a | student |
|---|---|---|---|---|---|---|---|---|
| IO encoding | PER | O | PER | PER | PER | O | O | O |
| IOB encoding | B-PER | O | B-PER | B-PER | I-PER | O | O | O |



# Lecture Summary

## Lecture 9 Summary

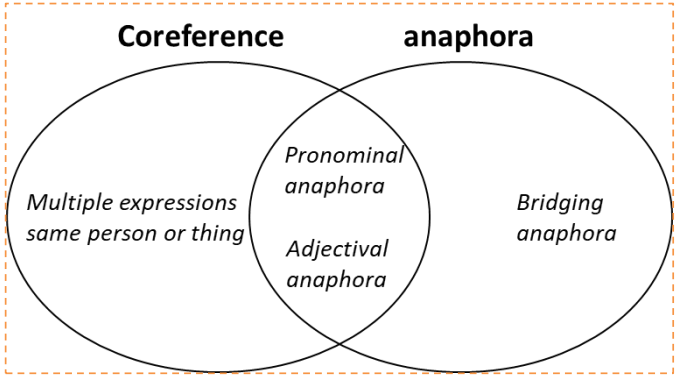
### Information Extraction

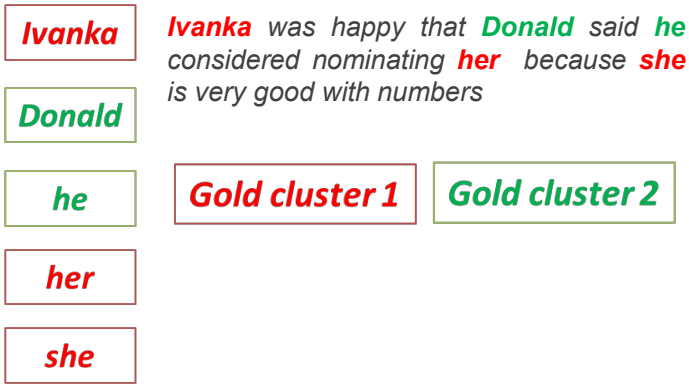
### Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

### Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation



# Lecture Summary

## Lecture 9 Summary

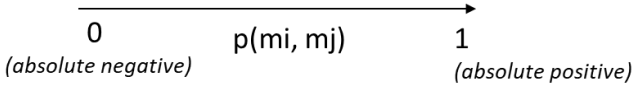
### Information Extraction

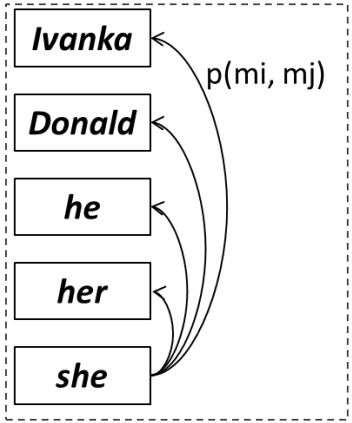
### Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

### Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation



# Lecture Summary



## Lecture 9 Summary

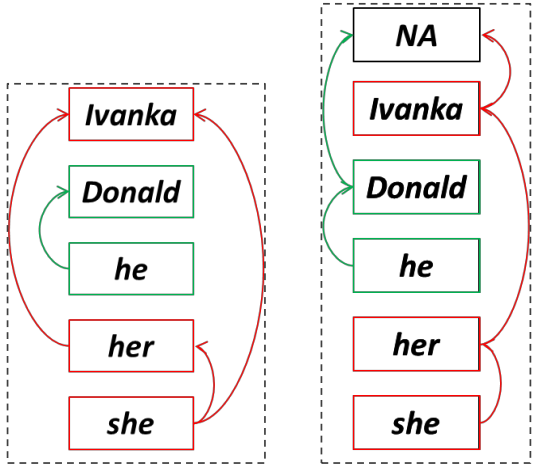
### Information Extraction

### Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

### Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation




$$J = -\sum_{i=2}^{N}\sum_{j=1}^{i} y_{ij} \log p(m_j, m_i)$$

Coreferent mentions pairs should get high probability, others should get low probability

Iterate through mentions

Iterate through candidate antecedents (previously occurring mentions)

# Lecture Summary

## Lecture 9 Summary

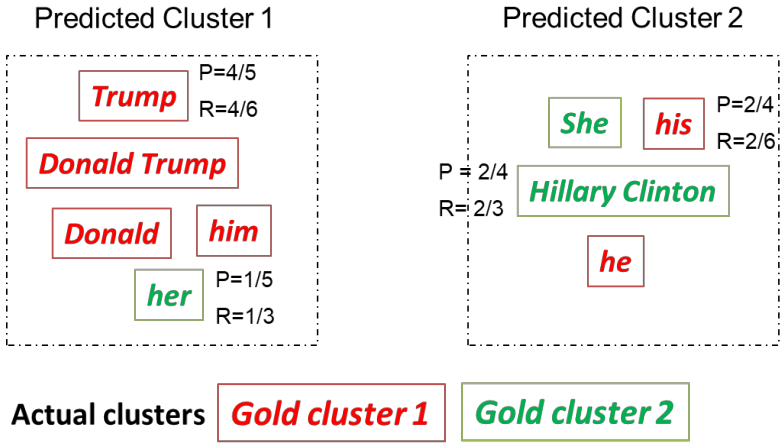
### Information Extraction

### Named Entity Recognition (NER)

1. Traditional NER
2. Sequence Model for NER
3. NER Evaluation

### Coreference Resolution

1. Mention-pair
2. Mention Ranking model
3. Coreference Evaluation



# Lecture Summary

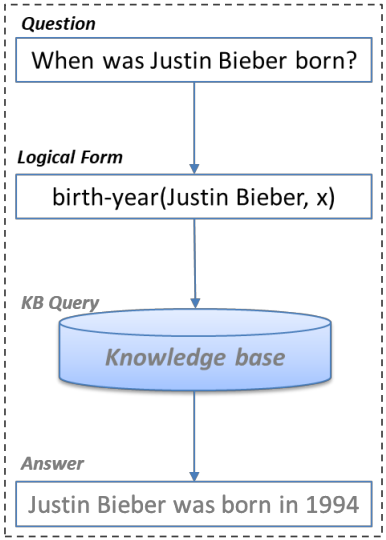
## Lecture 10 Summary

Question Answering

**Knowledge-based Question Answering**

IR-based Question Answering and Reading Comprehension

Additional: Visual Question Answering




based on each knowledge base

| Question | Logical Form |
|---|---|
| When was Justin Bieber born? | birth-year(Justin Bieber, x) |
| What is the largest state? | argmax(λx.state(x),λx.size(x)) |

| Subject | Predicate (relation) | Object |
|---|---|---|
| Justin Bieber | birth-year | 1994 |
| Frédéric Chopin | birth-year | 1810 |
| ... | ... | ... |

# Lecture Summary

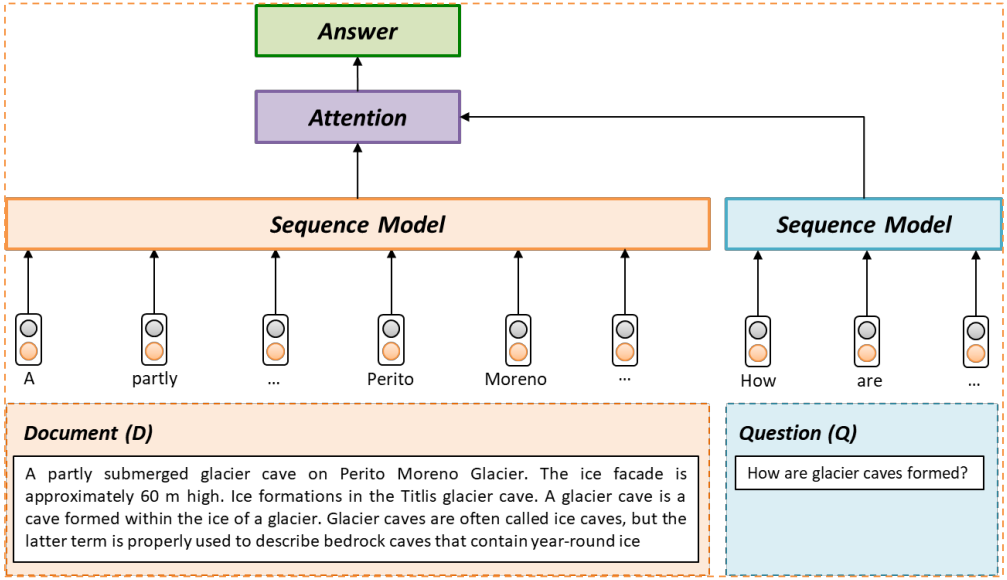
## Lecture 10 Summary

Question Answering

Knowledge-based Question Answering

**IR-based Question Answering and Reading Comprehension**

Additional: Visual Question Answering



# Lecture Summary

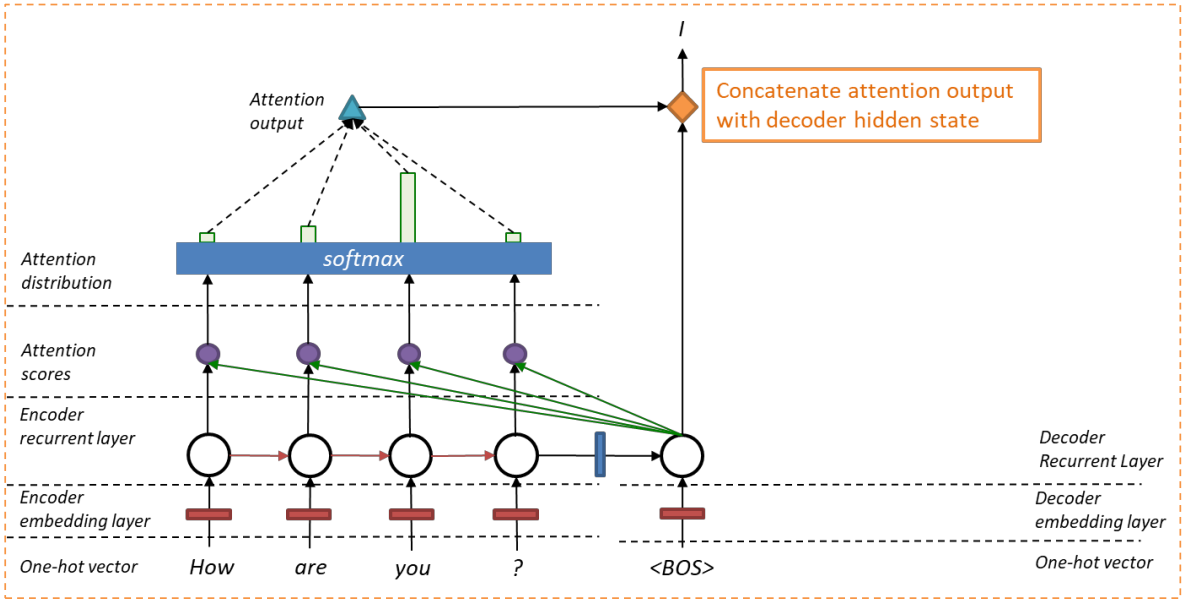
## Lecture 10 Summary

Question Answering

Knowledge-based Question Answering

**IR-based Question Answering and Reading Comprehension**

Additional: Visual Question Answering



# Lecture Summary

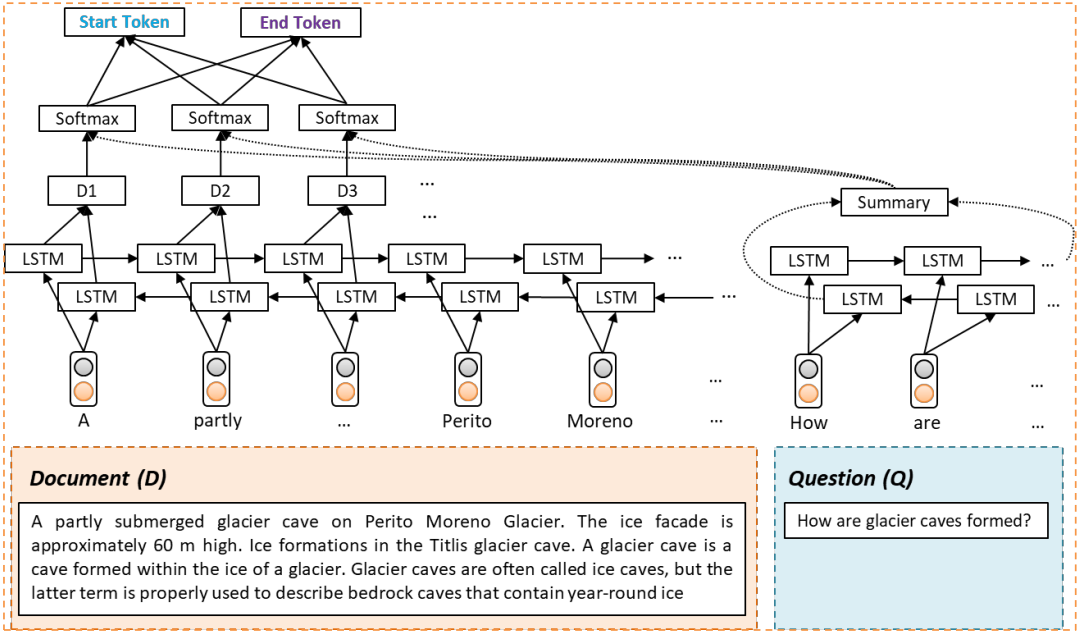
## Lecture 10 Summary

Question Answering

Knowledge-based Question Answering

**IR-based Question Answering and Reading Comprehension**

Additional: Visual Question Answering



# Lecture Summary

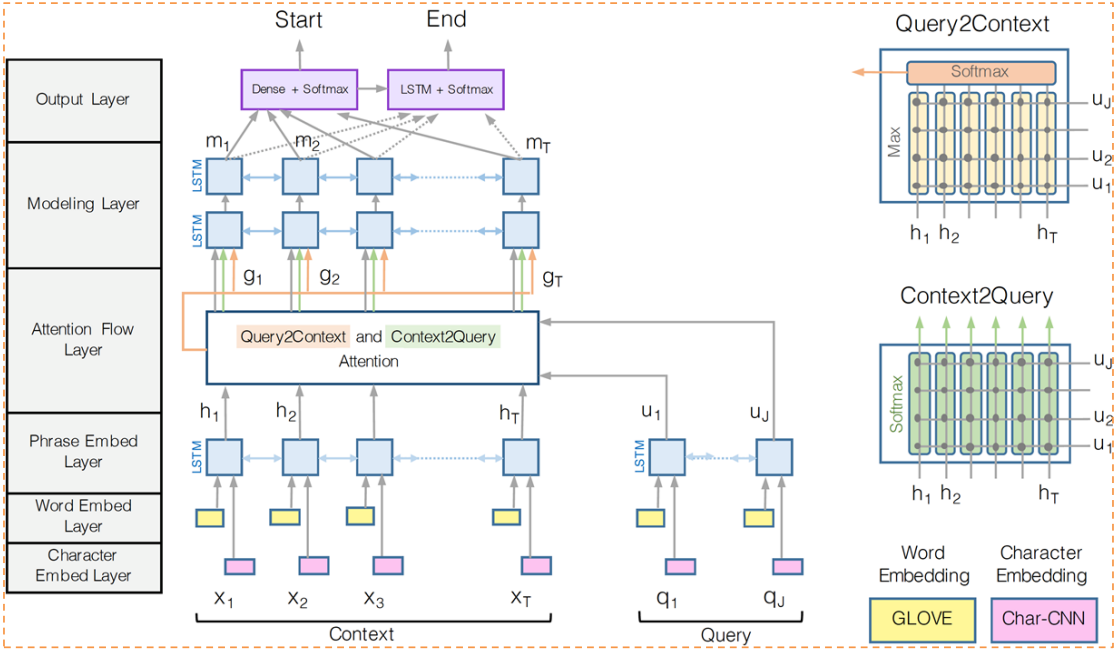
## Lecture 10 Summary

Question Answering

Knowledge-based Question Answering

**IR-based Question Answering and Reading Comprehension**

Additional: Visual Question Answering



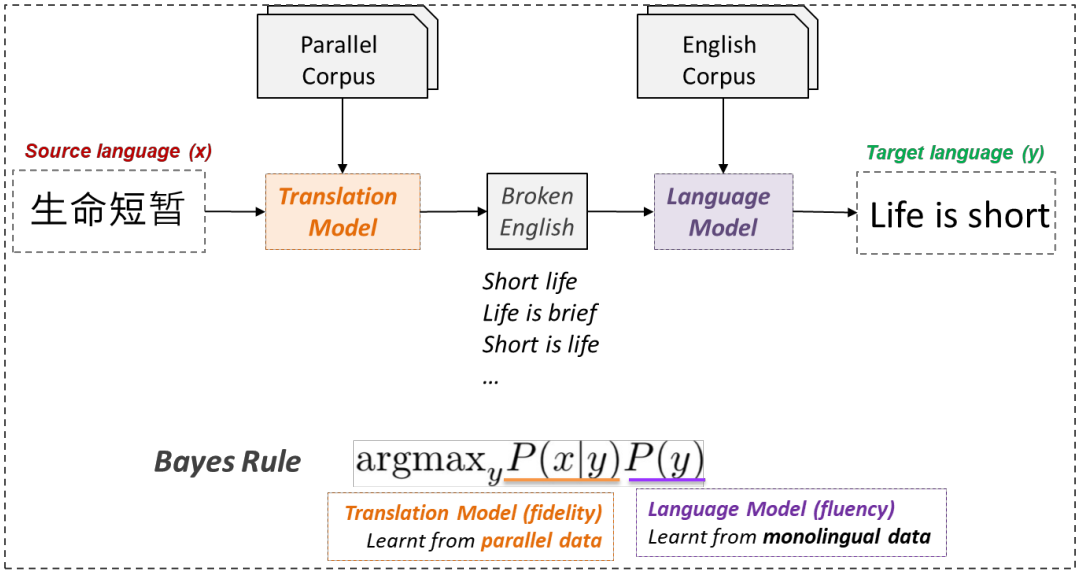
# Lecture Summary



## Lecture 11 Summary

### Machine Translation

- Statistical Machine Translation
- Neural Machine Translation
- Attention and Transformer for MT

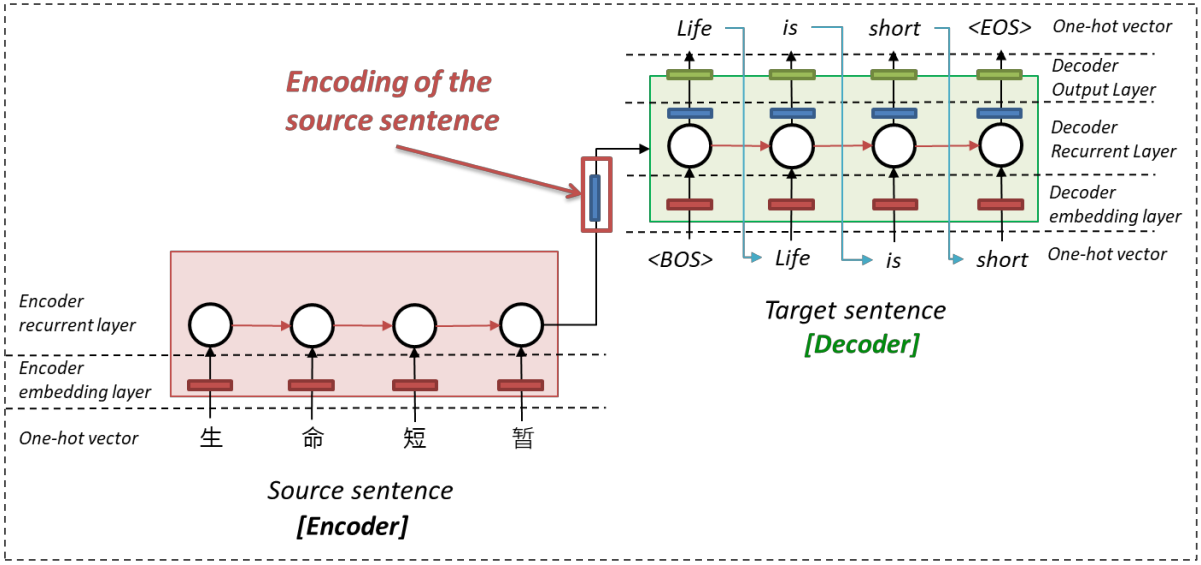


# Lecture Summary

## Lecture 11 Summary

### Machine Translation

- Statistical Machine Translation
- Neural Machine Translation
- Attention and Transformer for MT

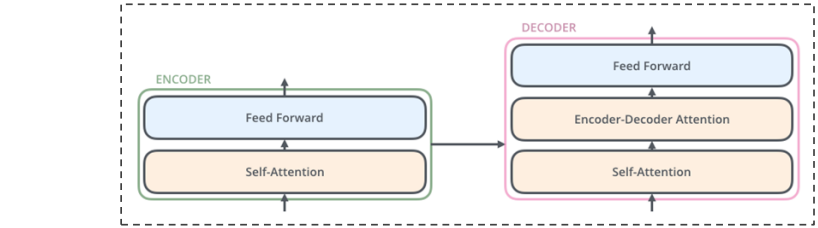


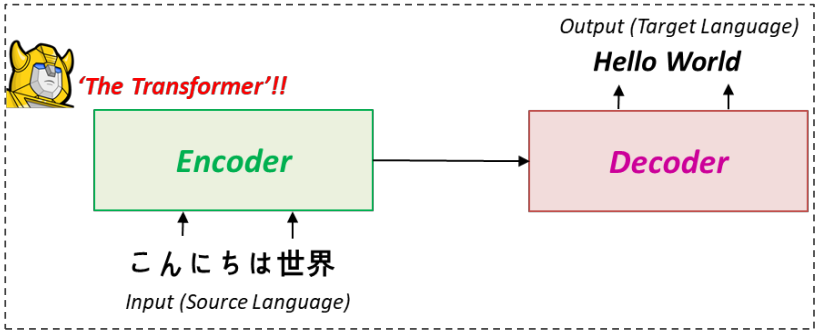
# Lecture Summary

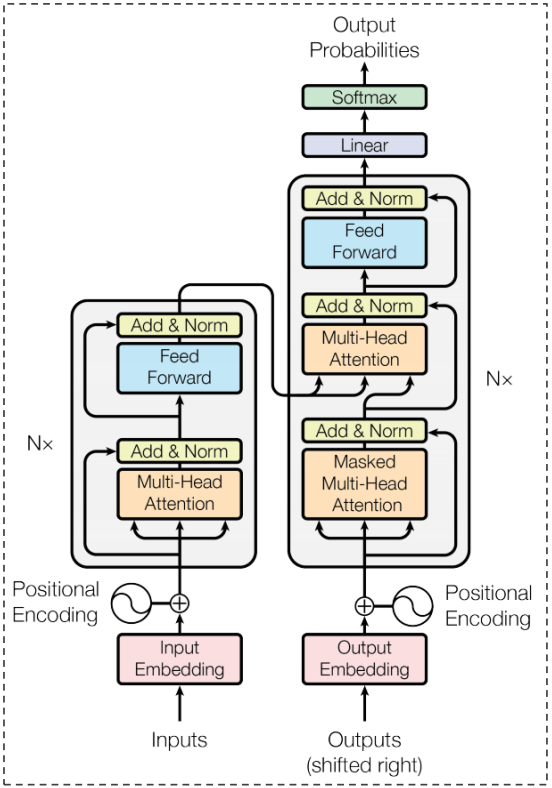
## Lecture 11 Summary

### Machine Translation

- Statistical Machine Translation
- Neural Machine Translation
- Attention and Transformer for MT



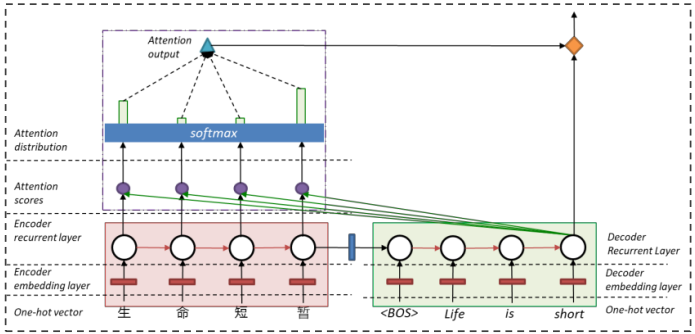



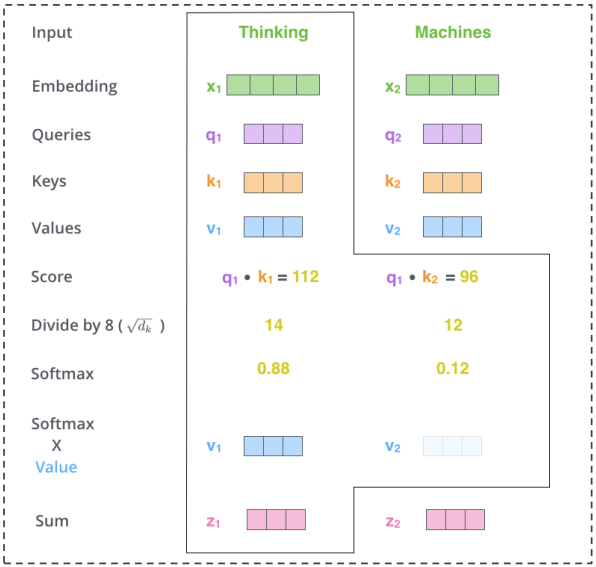
# Lecture Summary

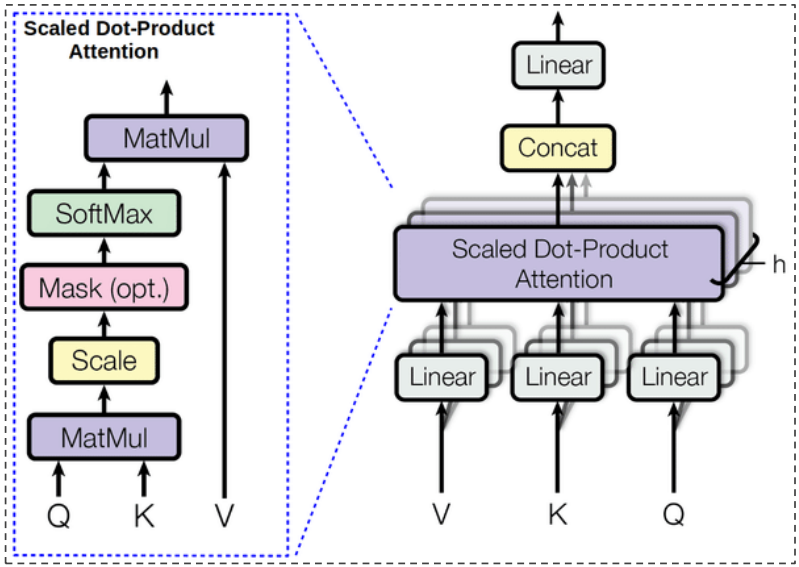
## Lecture 11 Summary

### Machine Translation

- Statistical Machine Translation
- Neural Machine Translation
- Attention and Transformer for MT



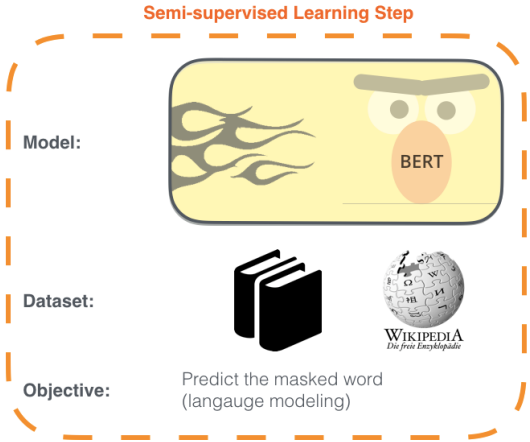



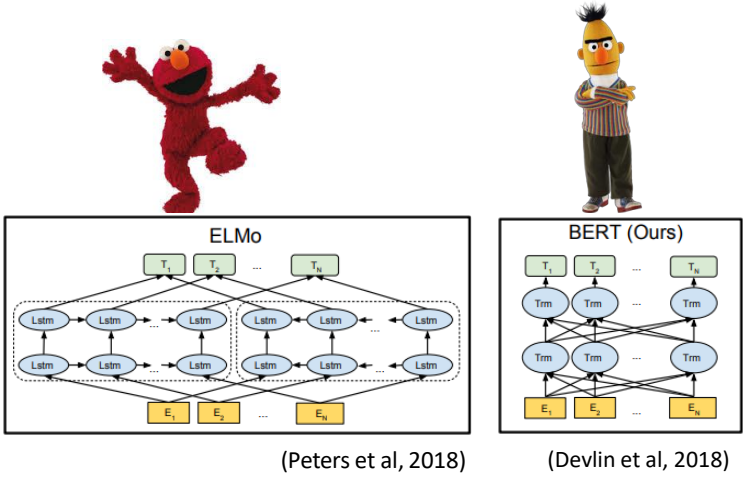
# 2 Lecture Summary

## Lecture 13 Summary

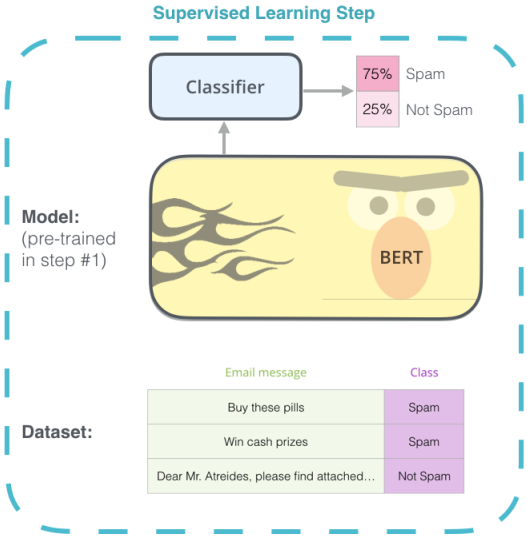
### The Rise of the Pre-trained Model

- Early Stage Pretrained Model
- Masked Language Model
- Next Sentence Prediction
- Post BERT




(Peters et al, 2018) (Devlin et al, 2018)



# Exam Guide



## Assessment Update

| Assessment | Weight | Due | Length |
|---|---|---|---|
| Lab exercises | 10% | Multiple weeks | n/a |
| Assignment 1 | 20% | Week 8 | n/a |
| Assignment 2 | 20% | Week 14 | n/a |
| Final Exam<br>Take-home Exam<br>(Open Book) | 50% | 4 Short Answer Question (32%) | 80 mins |
| | | 1 Essay Question (18%) | 40 mins |

**To pass UoS,** achieve at least 40% (20 out of 50)

**To pass UoS,** achieve at least 40% (20 out of 50)

***Canvas Exam Site***

The canvas exam site is different that the Canvas site we use during the semester for teaching
It is called Final exam for COMP5046
The Exams office will give you access to the exam site
The recommended browsers for Canvas: Chrome and Firefox, not Internet Explorer or others

***Exam Conditions***

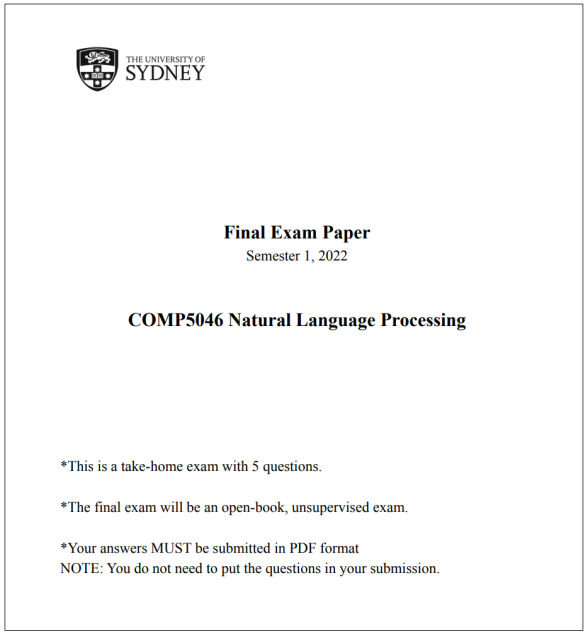
***2 hours***, Online exam (+10 minutes reading time +25 mins upload time)
***Open book*** examination, Unsupervised (Un-proctored)
***Complete ALL questions*** in the space provided on the exam paper
**Five (5) questions (total 50 marks) = four 8 marks questions (two 4 marks sub questions) + one (1) 18 marks question**
**= 32 marks + 18 marks**

# 3 Exam Guide

## Final Exam Preview! (All available in Canvas)

***Let's Have a Look at this Together!***

***Sample Online Exam – Question and Answer***



THE UNIVERSITY OF SYDNEY

**Final Exam Paper**

Semester 1, 2022

**COMP5046 Natural Language Processing**

*This is a take-home exam with 5 questions.

*The final exam will be an open-book, unsupervised exam.

*Your answers MUST be submitted in PDF format
NOTE: You do not need to put the questions in your submission.

# 3 Exam Guide

**Important! Suggestion!**

*The questions are not easily googlalable.*
*Your answer will be checked by the plagiarism checking software.*

*DO NOT Copy from other resources. We expect your own answer!*

*Well organized and be quick*
*Allocate time according to the marks for each question*

# Exam Guide

## Students in different time zones

- Important for international students who could not come to Australia and exchange students who had to return to their country.
- If you are in a different time zone and the exam is too early/too late for you, you can apply for Special Consideration to have special arrangements for the exam:

https://www.sydney.edu.au/students/special-consideration.html

*All Good! Thanks for following and surviving*

*GOOD LUCK in YOUR FINAL EXAM*