

Trends in Technology for Translators: Deep Learning and Automatic Translation

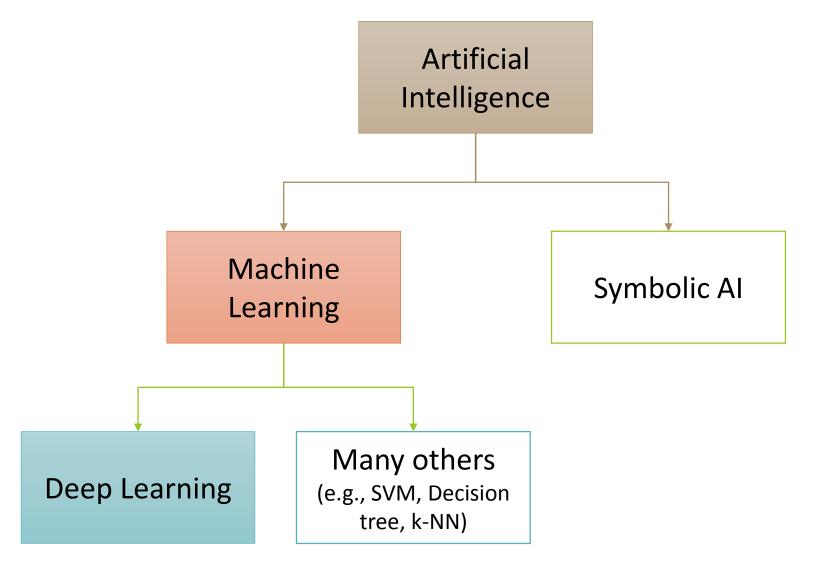
Dr Sai-cheong SIU

Director, Deep Learning Application and Research Centre

The Hang Seng University of Hong Kong

Deep Learning Research and Application Centre

The Hang Seng University of Hong Kong





Deep Learning

AlphaGo

Speech recognition

Sentiment analysis

Summarization

Chatbot

Stock price prediction

Image classification

Object detection

Self-driving cars and many more



Deep Learning Research and Application Centre

- Funding of HK\$6.8 million
- Hardware: Graphics processing unit (GPU)accelerated hardware for much faster training and decoding
- Software: PyTorch, Tensorflow, Keras, Transformer...
- Projects involving deep neural networks



Projects (https://dlc.hsu.edu.hk/)

- Machine translation (e.g., automatic translation of listing documents; recent funding: HK\$1.4 million)
- Smart retail
- English proofreading
- Weather analysis, forecast and communication
- Detection of fake news
- Financial analysis



Trends in Technology for Translators

Outline

- Useful Tools for Professional Translators
- Al and Translation Technology
- Ways to Make Better Use of Automatic Translation Systems

Outline

- Useful Tools for Professional Translators
- Al and Translation Technology
- Ways to Make Better Use of Automatic Translation Systems

Translation Technology

Electronic Dictionaries

Corpora and Concordancers

Translation Databases

Automatic Translation



Electronic Dictionaries

General Dictionaries

Specialised Dictionaries



Example: HSUHK Business Translation Index



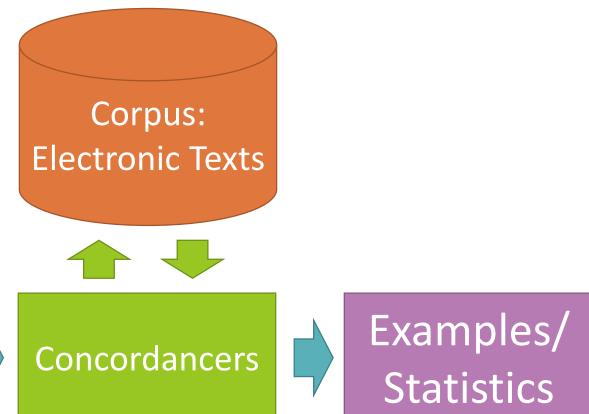
http://www.hsuhkbti.com/SearchB TI?search=technical+analysis





Corpora and Concordancers

- 1. ST Analysis
- 2. TL Expressions
- 3. TT Analysis



Expression





Example 1: Legal Corpus

http://legalcorpus.weebly.com/



Example 2: Ngram Viewer

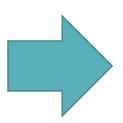
https://books.google.com/ngrams



Translation Databases: Translation Memories and Termbases

Translation Memories: Previous TT

Terminology
Databases:
Multilingual
Terms



Reuse of TL Sentences

Consistent Terminology

> Quality Control

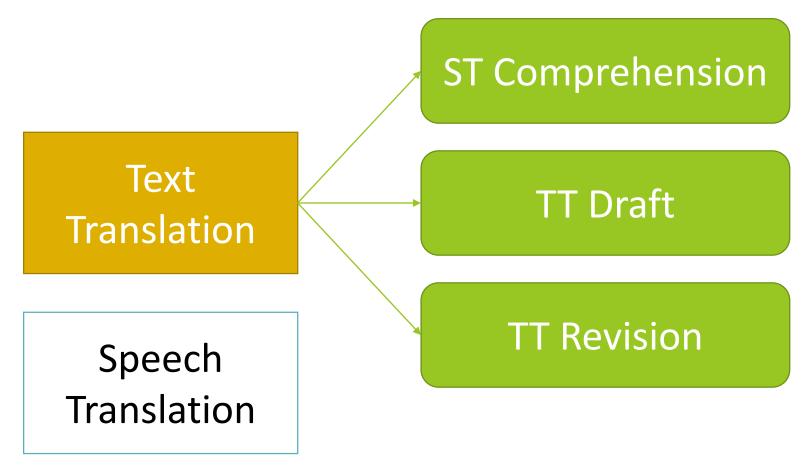


Example: Memsource

https://www.memsource.com/



Automatic Translation





Outline

- Useful Tools for Professional Translators
- Al and Translation Technology
- Ways to Make Better Use of Automatic Translation Systems

Methods

Rule-based Approach

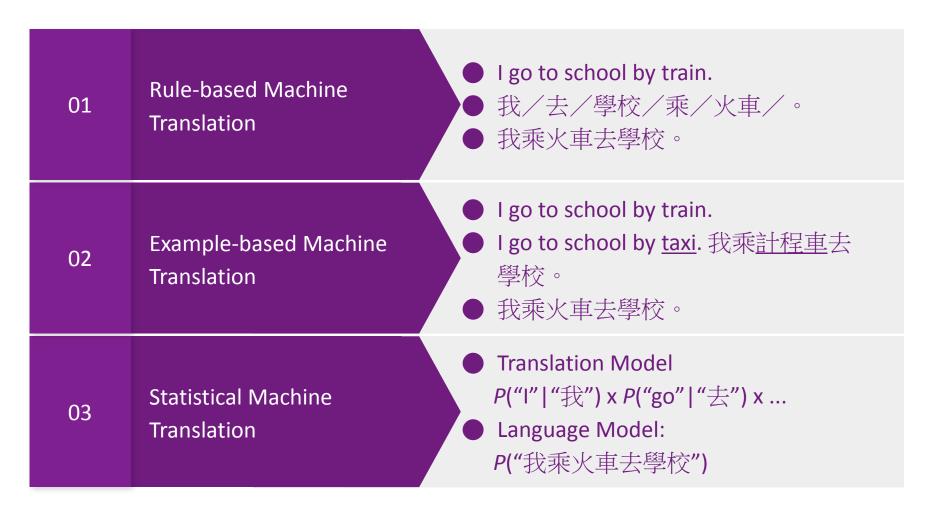
Example-based Approach

Statistical Approach

Deep Learning Approach



Traditional Approaches to Machine Translation



Methods



Rule-based Approach

Example-based Approach

Statistical Approach

Deep Learning Approach



Rule-based Machine Translation (RBMT)

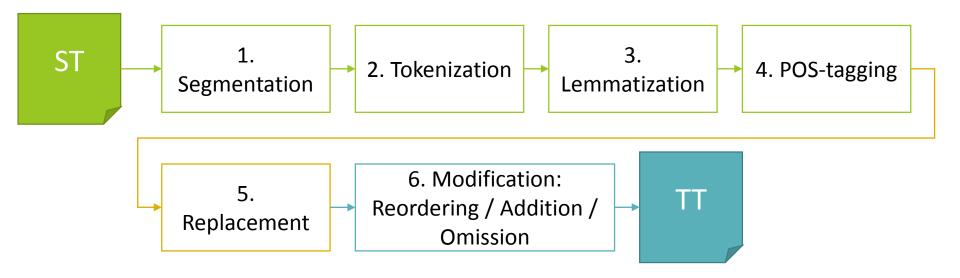
- Direct MT
- Transfer MT



Direct MT

- This is considered to be the first generation of MT systems.
- The ST is treated as a string of words. The SL words are then replaced by TL words, and the TL words are reordered.

Direct MT





Segmentation

the process of splitting a text into sentences

Tom was a student. He went to school by bus. He played football with them.



Tom was a student.

He went to school by bus.

He played football with them.



Tokenization

the process of dividing a text into words

He played football with them.



He

played

football

with

them

em .



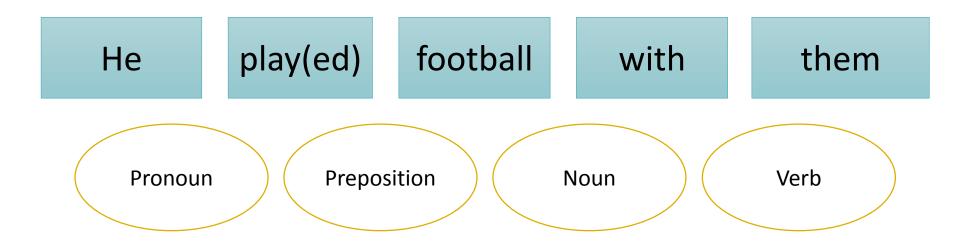
Lemmatization

- the process of grouping the identical, related or inflected forms of a word together as a single item (lemma)
- Example: do, does, doing, did → do



POS-tagging

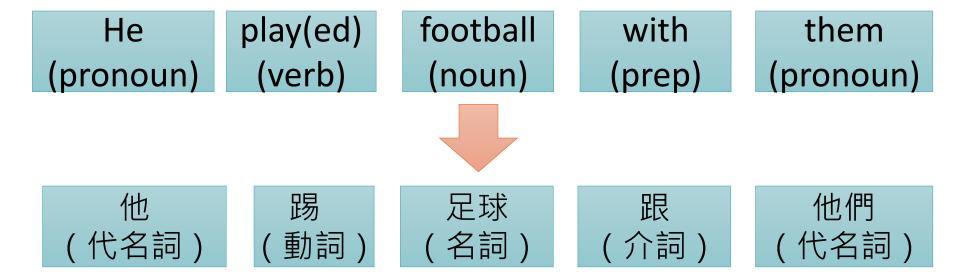
 the process of assigning a part-of-speech (POS) to every lexical item in the input





Replacement

the process of replacing the SL words with TL words





Reordering and Modification

 the process of reordering the TL words and polishing the TT

他 (代名詞) 踢 (動詞) 足球 (名詞) 與 (介詞)

他們 (代名詞)



他 (代名詞)

跟 (介詞) 他們 (代名詞)

踢 (動詞)

足球 (名詞)

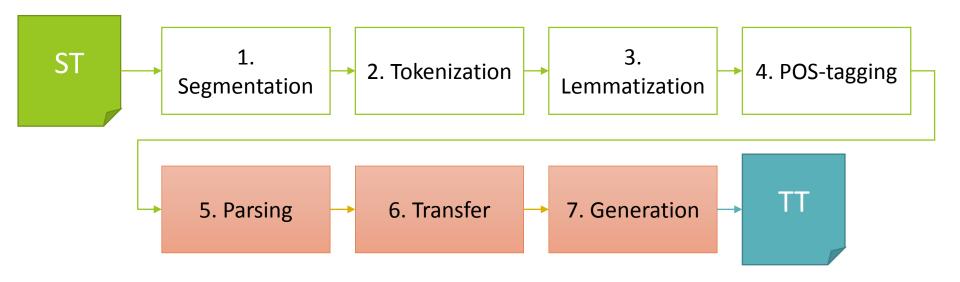


Transfer MT

- The basic idea: Analysis, Transfer and Generation
- Input sentences are analyzed and translated into an abstract internal representation retaining their features in the SL.
- The representation is transferred to a representation with the features in TL and generates the TT.



Transfer MT





Parsing

 the process of analyzing the input and generating the SL representation (e.g., a parse tree) that identifies the function of each word/word group



Transfer

• the conversion of the SL representation into its equivalent TL representation



Generation

the production of the TL text from the TL representation



- ST: The boy went to Didneyland.
- Case conversion: the boy went to didneyland.
- Segmentation
- <1> the boy went to didneyland.



Tokenization

- <1> the
- <2> boy
- <3> went
- <4> to
- <5> didneyland
- <6> .



Lemmatization

- <1> the
- <2> boy
- •<3> go (past tense)
- <4> to
- •<5> didneyland
- <6> .



POS-tagging

```
•<1> the (definite article)
```

```
• <2> boy (noun, singular)
```

```
• <3> go (verb, past tense)
```

```
• <4> to (preposition)
```

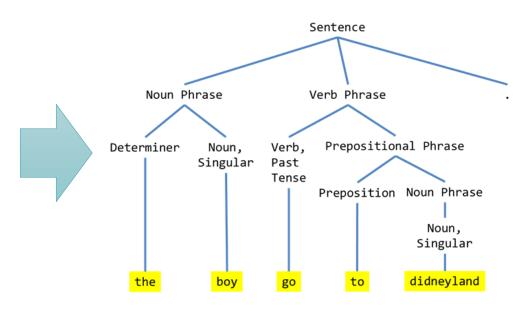
```
•<5> didneyland (noun, singular)
```

```
• <6> .
```

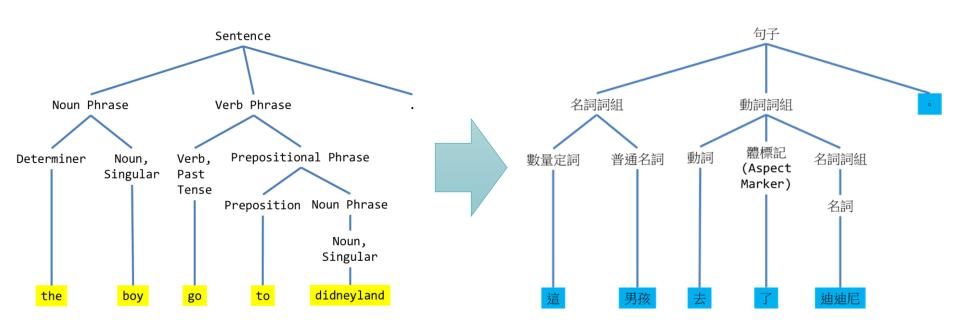


Parsing

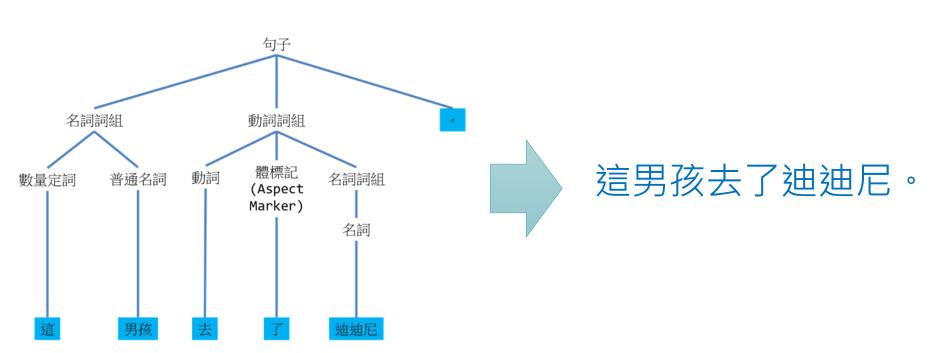
```
<1> the (definite article)
<2> boy (noun, singular)
<3> go (verb, past tense)
<4> to (preposition)
<5> didneyland (noun, singular)
<6> .
```



Transfer



Generation





Methods

Rule-based Approach



Example-based Approach

Statistical Approach

Deep Learning Approach



Finding appropriate translation examples



Reordering the examples retrieved with the assistance of a language model

The best example?

ST: The boy went to school.

Step 1: Retrieving the "closest" example

_		
Example	English (Source	Chinese (Target
	Language)	Language)
1	The boy is naughty.	那男孩很頑皮。
2	They went to the	他們到圖書館去。
	library.	
3	He goes to school by	他乘火車上學去。
	train.	
4	They went to school.	他們到學校去。



Step 2: Using the example

ST: The boy went to school.

Example: They went to school.

Useful words: XXX went to school.

Translation: The boy 到學校去。



Step 3: Modifying the example

The boy → 那男孩 (from dictionary/corpus)

Draft: 那男孩到學校去。

1. Dictionary/Rules:The → 那
Boy → 男孩

2. Other examples in the database:

A. By comparison

The boy is naughty.
→那男孩真頑皮。
He didn't see the boy.

→他看不見那男孩。

B. By Word AlignmentThe/boy/is/tall.→ 那/男孩/長得/高。

3. Translation Model:

P(The|那), P(The|該), P(boy|男孩), P(boy|百厭星)... 4. Hybrid Approach:

1 + 2 + 3

Step 4: Apply the language model and check the readability of the draft.

Output: 那男孩到學校去。



Methods

Rule-based Approach

Example-based Approach

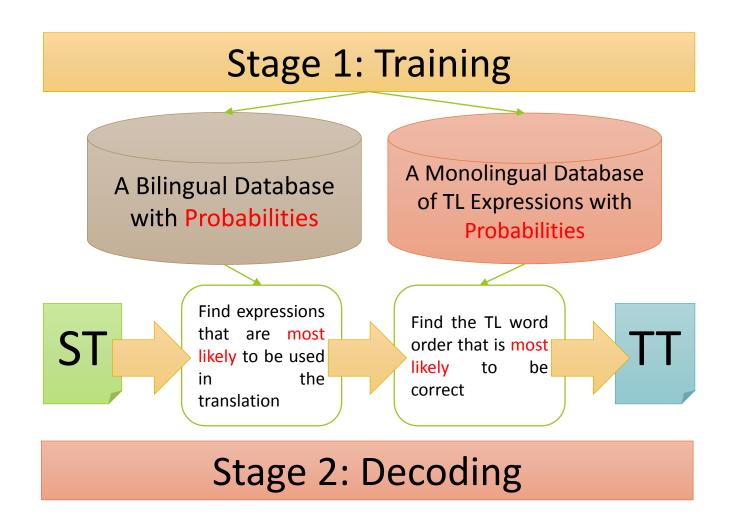


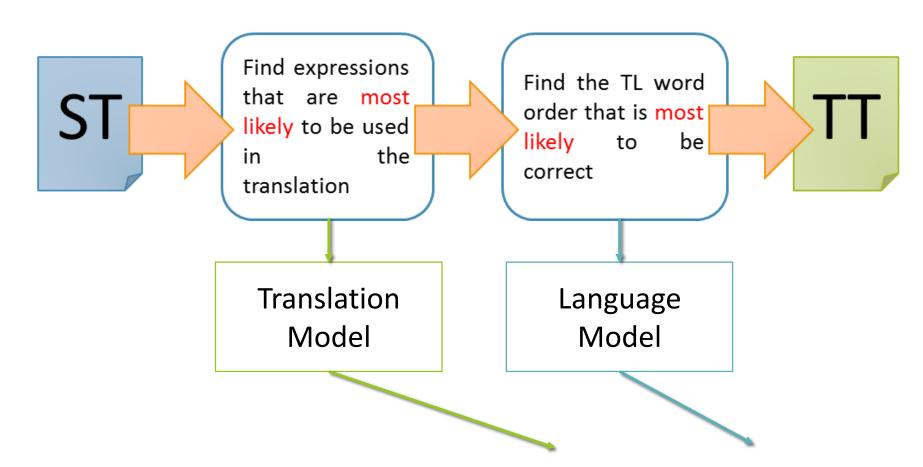
Statistical Approach

Deep Learning Approach



Statistical Machine Translation





$T'= \arg \max_{T} P(S|T) \times P(T)$



Language Model

Do you prefer fish ? Elephants love her . Give me fish . Dogs love swimming . We love dogs .



Monolingual Corpus in the Target Language



Translation Model

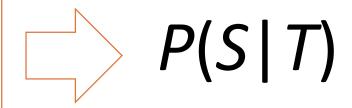
Source Language

我喜歡貓。大象鍾意我。人們喜歡狗。 他們嗜魚。我們喜歡猴子。你喜歡豬。

Target Language

I love cats. Elephants love me. People love dogs. They prefer fish. We love monkeys. You prefer pigs.

Bilingual Corpus





Monolingual (in TL)

Bilingual

Data

Do you love dogs ?
Elephants loves her .
Give me fish .
They love elephants .
I go swimming .
We love dogs .



Tables

I	0.04
you	0.04
dogs	0.08
fish	0.04
elephants	0.08
love	0.12
loves	0.04
we	0.04
they	0.04

我喜歡貓。→ I love cats.

大象喜歡我。→ Elephants love me.

他喜歡狗。→ He loves dogs.

他們喜歡魚。→ They love fish.

我們喜歡猴子。→ We love monkeys.

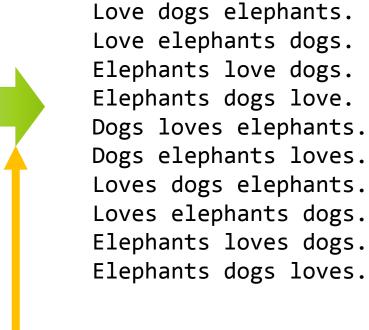
你喜歡豬。→ You love pigs.



我	I	0.50
我	me	0.50
你	you	1.00
他	he	1.00
貓	cats	1.00
狗	dogs	1.00
魚	fish	1.00
豬	pigs	1.00
猴子	monkeys	1.00
大象	elephants	1.00
喜歡	love	0.83
喜歡	loves	0.17
我們	we	1.00
他們	they	1.00
0	•	1.00

User's Input

狗喜歡大象。



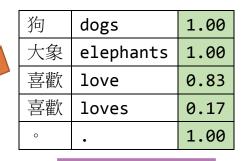
Dogs love elephants.

Dogs elephants love.

MT Output

Dogs love elephants.

我	I	0.50
我	me	0.50
你	you	1.00
他	he	1.00
貓	cats	1.00
狗	dogs	1.00
魚	fish	1.00
豬	pigs	1.00
猴子	monkeys	1.00
大象	elephants	1.00
喜歡	love	0.83
喜歡	loves	0.17
我們	we	1.00
他們	they	1.00
٥		1.00



aogs	0.08
elephants	0.08
love	0.12
loves	0.04

Bilingual Table

Monolingual Table

Methods

Rule-based Approach

Example-based Approach

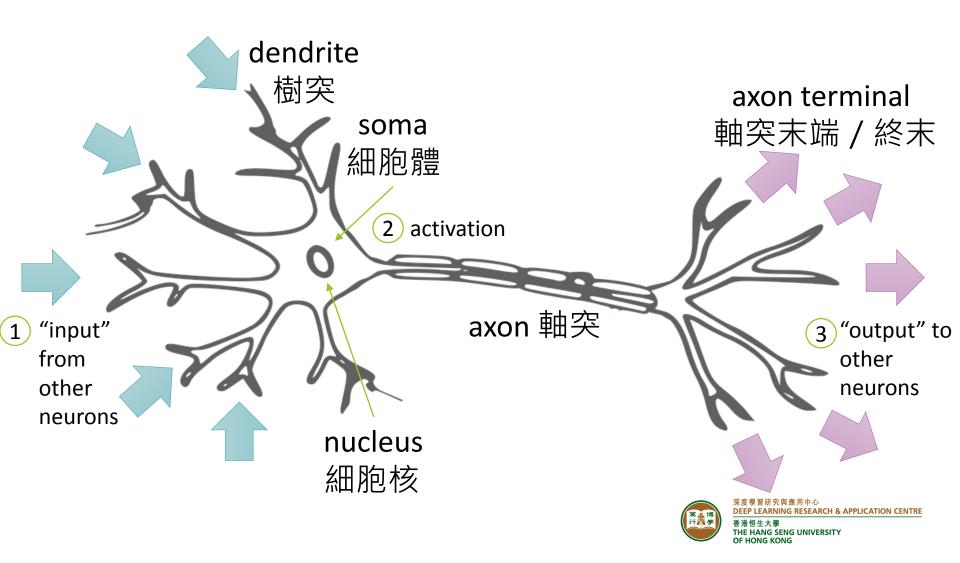
Statistical Approach



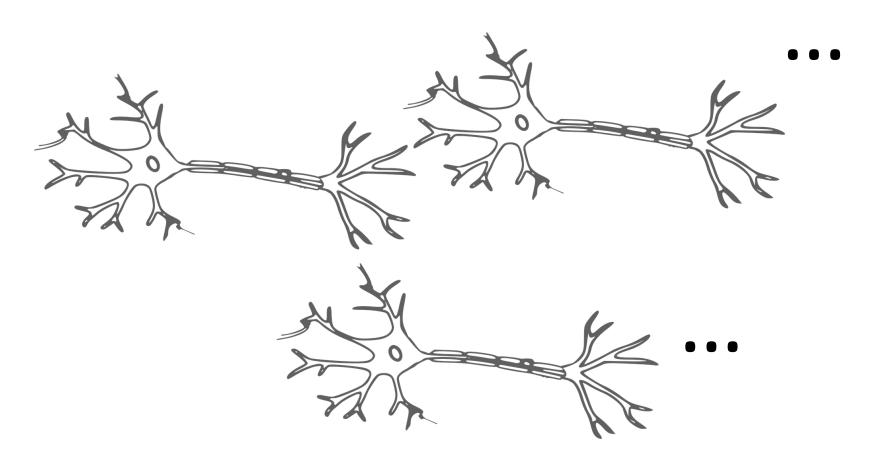
Deep Learning Approach



Biological Neuron



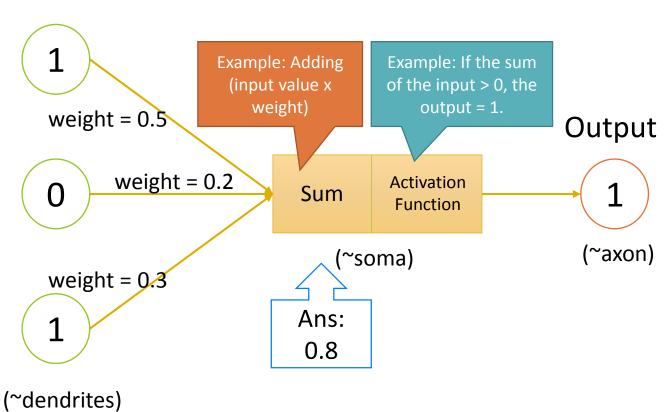
Biological Neural Network





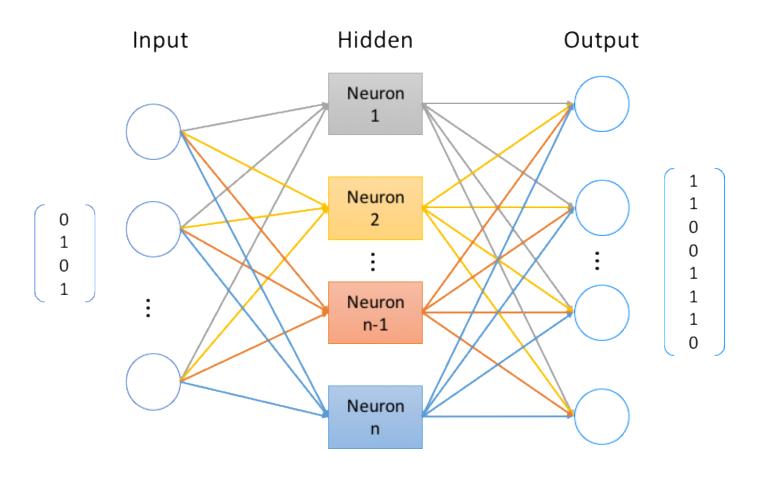
Artificial Neuron

Input



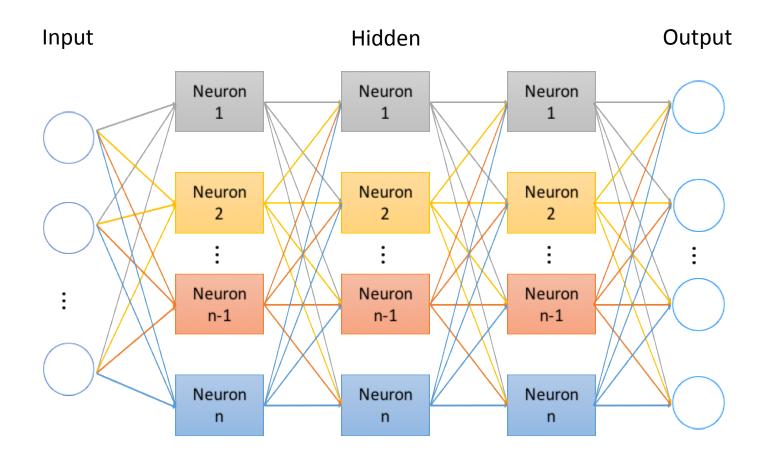


Artificial Neural Network: Many Neurons





Deep Learning: Multiple Hidden Layers





Automatic Translation as a Prediction Problem: Given X, what is the next word?

```
Given the following:
```

```
ST: I / go / to / school / by / bus / . /
TT: 我 / 乘
```

Which of the following is most likely to be the next word?

公車

火車

學校

咖啡

電腦

翻譯



Deep Learning and Automatic Translation

```
Given [I] / [go] / [to] / [school] / [by] / [train] / [.], what is the next word?
我
[I] / [go] / [to] / [school] / [by] / [train] / [.] / [我] \Rightarrow ?
乘
[l] / [go] / [to] / [school] / [by] / [train] / [.] / [我] / [乘] ⇒?
火重
[I] / [go] / [to] / [school] / [by] / [train] / [.] / [我] / [乘] / [公車] ⇒?
去
[I] / [go] / [to] / [school] / [by] / [train] / [.] / [我] / [乘] / [公車] / [去] ⇒?
學校
[l] / [go] / [to] / [school] / [by] / [train] / [.] / [我] / [乘] / [公車] / [去] / [
學校]/⇒?
```

Translation: 我乘火車去學校。



Outline

- Useful Tools for Professional Translators
- Al and Translation Technology
- Ways to Make Better Use of Automatic Translation Systems

Be a Smart User!

Domain-specific Systems

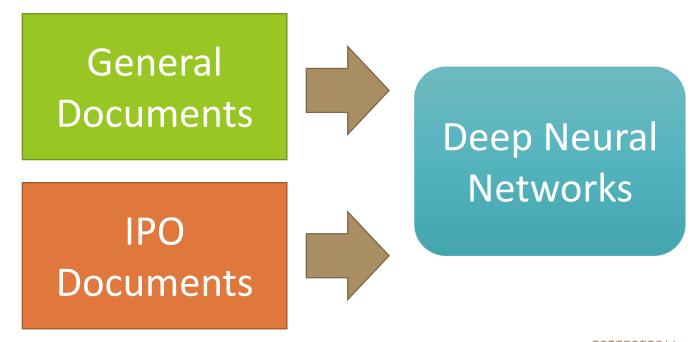
Multiple Systems

Customizable Systems



Domain-specific Systems

- Example: Our Project
- Machine Translation of IPO Documents





Multiple Systems

- Qtranslate: http://quest-app.appspot.com/home
- ImTranslator: http://imtranslator.net/compare/

Customizable Systems

Our Prototype

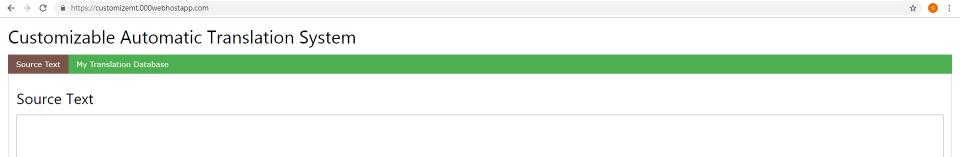
https://customizemt.000webhostapp.com/



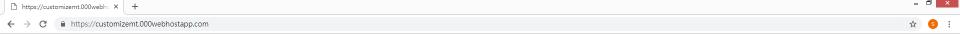
Making MT More Useful

- Automatic Segmentation
- Customizable Translation Database
- Key Expression Identification
- Instant Search



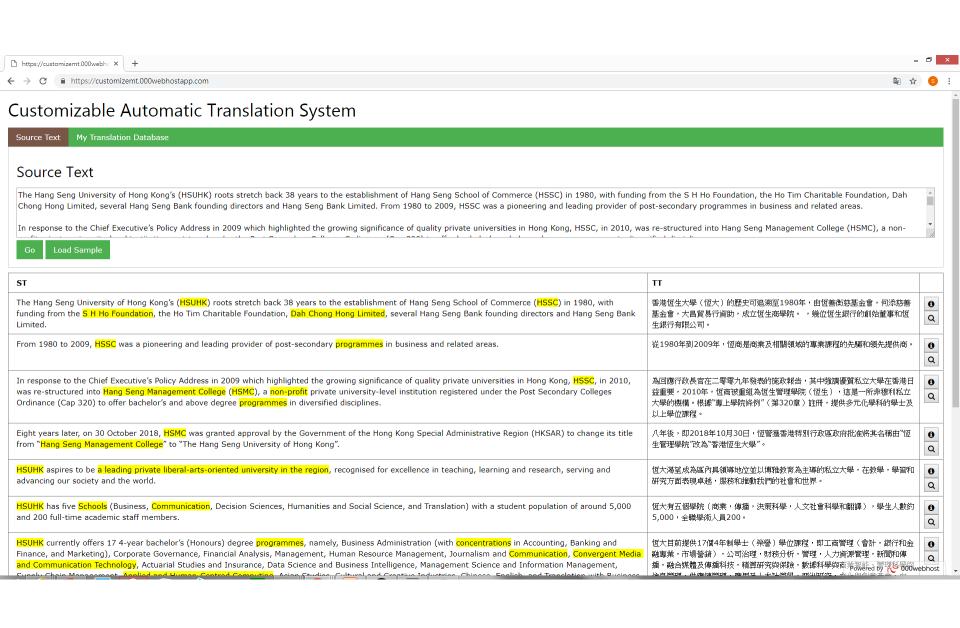


https://customizemt.000webhc × +



Customizable Automatic Translation System







Many other tools

Tools for Medical Translation

Tools for Translation of Government

Documents

Tools for Scientific
Translation

Tools for Translation of Philosophical Texts

Tools for Interpreting

Tools for Subtitling

Tools for Web
Translation

Tools for Bilingual Digital Marketing

Tools for Team
Translation / Project
Management



Short Courses

MA in Translation

Other Research Projects

- Prototype: IPO prospectus
- MT-CAT Integration (with multiple translation suggestions?)
- Computer-aided Interpreting



Thank you very much!

- Email: scsiu@hsu.edu.hk
- Website: https://www.hsuhk-translation-technology.com
- More Workshops:



Our Facebook https://goo.gl/M9rHs5

