翻譯學院
SCHOOL OF TRANSLATION
香港恒生大學
THE HANG SENG UNIVERSITY
OF HONG KONG

深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY
OF HONG KONG

AI VS Human Translation: The Inevitable Replacement?

Dr Sai-cheong SIU

Director, Deep Learning Research and Application Centre

Associate Professor, School of Translation

The Hang Seng University of Hong Kong

Artificial Intelligence

- New Translation Algorithms
- New Translation Applications

AI and People

Impact of AI on Translation and Communication

- What does translation technology mean for your business?
- What are the roles of AI and human translators in multilingual communication?

翻譯學院
SCHOOL OF TRANSLATION
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

Outline

- Overview: Translation Industry and Translation Technology
- Types of Translation Technology: Automatic Translation and More
- AI for Translation: Strengths, Weaknesses and the Way Forward

About Us

Private University

5 Schools
(Business, Communication, Decision Sciences, Humanities and Social Science, and Translation)

The Hang Seng University of Hong Kong

Undergraduate and Postgraduate Programmes

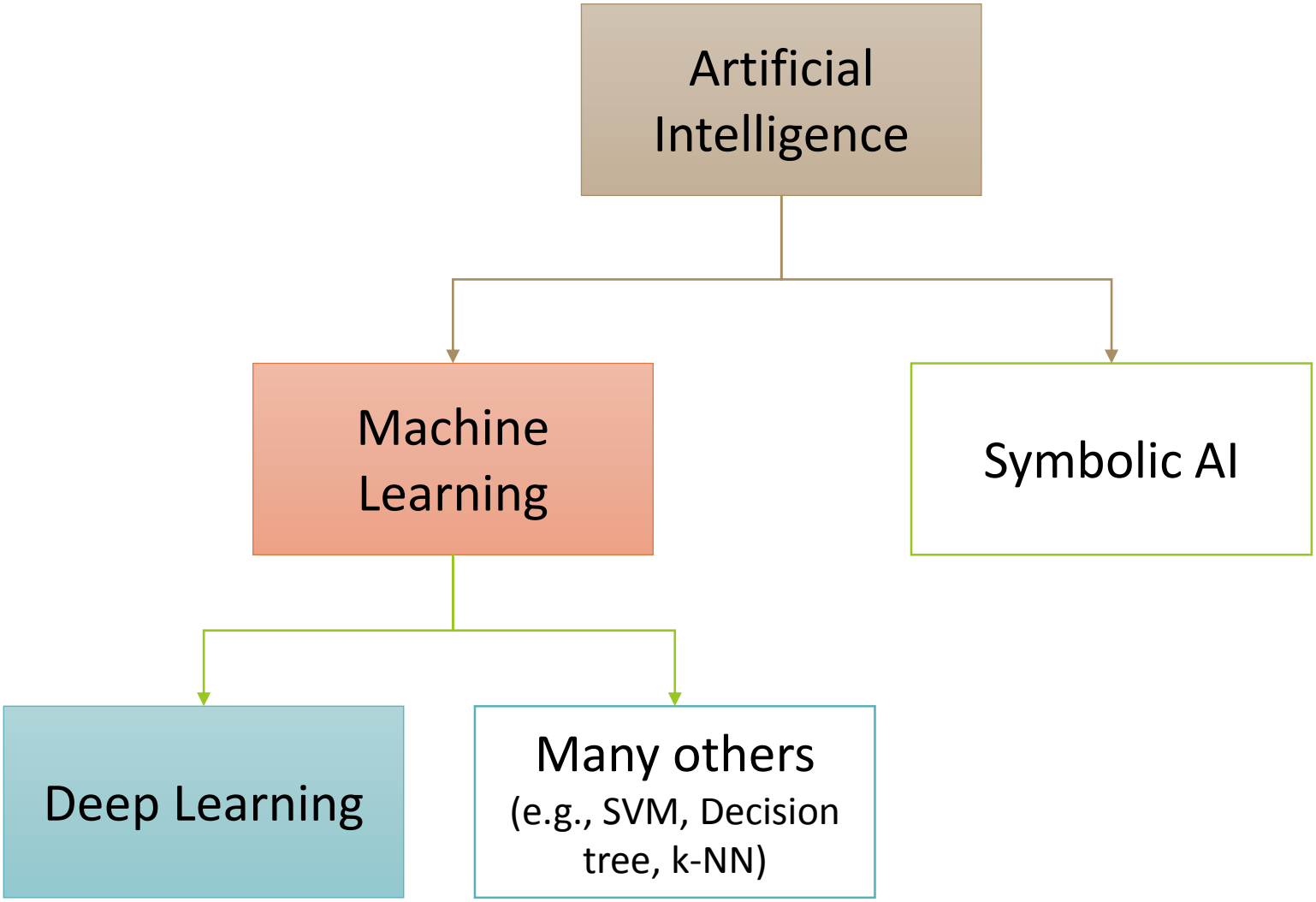
School of Translation

1. Bachelor of Translation with Business
2. Master of Arts in Translation (Business and Legal)
3. Master of Arts in Translation (Computer-aided Translation)
4. Minor in European Studies

Deep Learning Research and Application Centre

翻譯學院
SCHOOL OF TRANSLATION
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG


Artificial Intelligence
Machine Learning
Symbolic AI
Deep Learning
Many others (e.g., SVM, Decision tree, k-NN)

Deep Learning

- Deep Learning technology, which is based on neural networks and large training datasets, has shown great success in many applications.

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

Overview: Translation Industry and Translation Technology

What do translators mean for your business? What does a translator do?

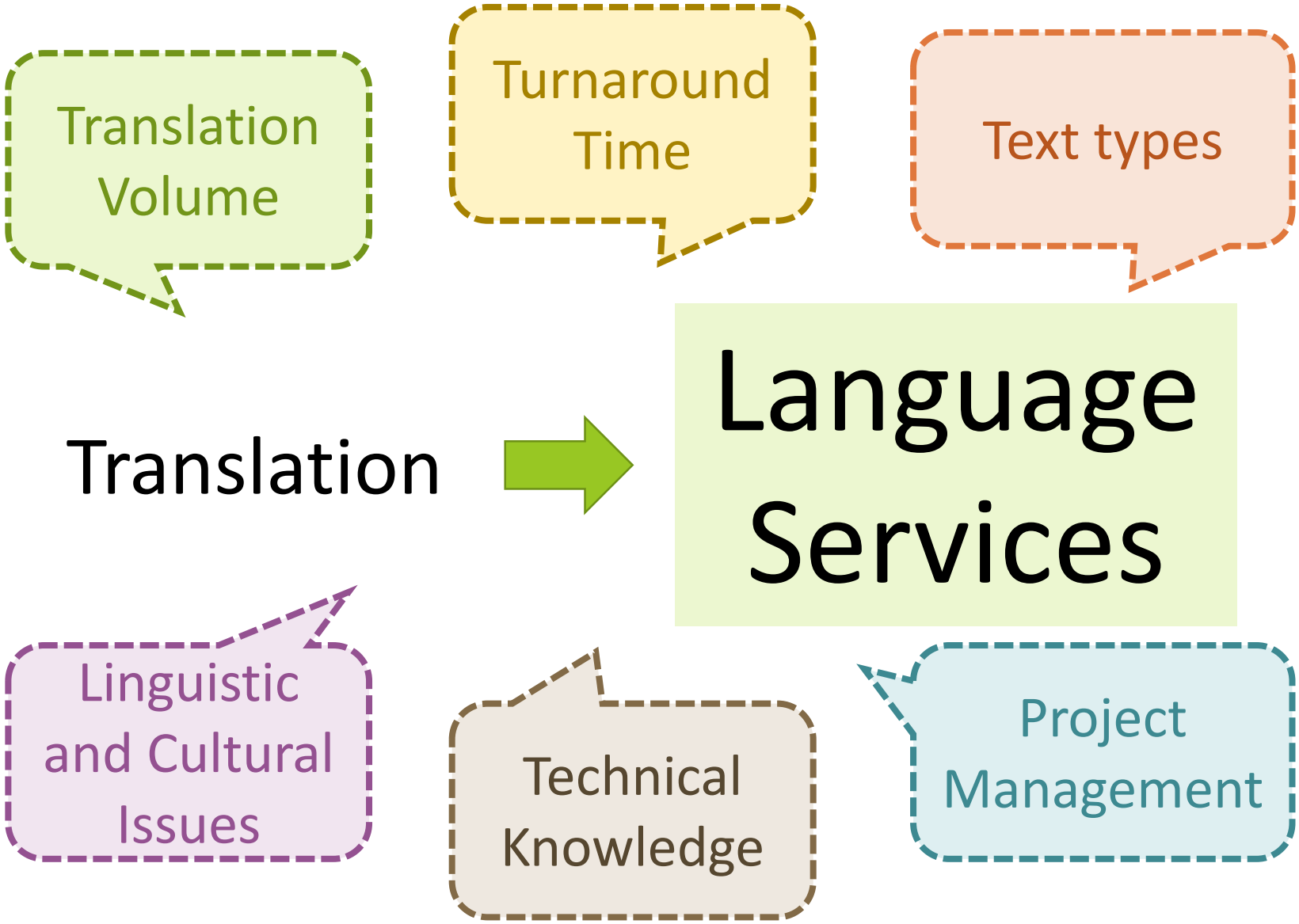
Translation

Interpreting

Localization

Bilingual Editing and Writing

Many more (e.g., transcription, certification, term management…)


Translation Volume
Turnaround Time
Text types
Translation
Language Services
Linguistic and Cultural Issues
Technical Knowledge
Project Management

It’s more than Gxxxle Translate

Translation Technology

Electronic tools for Translation

翻譯學院
SCHOOL OF TRANSLATION
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

Types of Translation Technology

Translation Technology

Electronic Dictionaries

Translation Memories

Terminology Databases

Automatic Translation

Electronic Dictionaries

General Dictionaries

Specialised Dictionaries

Example: HSUHK Business Translation Index

http://www.hsuhkbti.com/

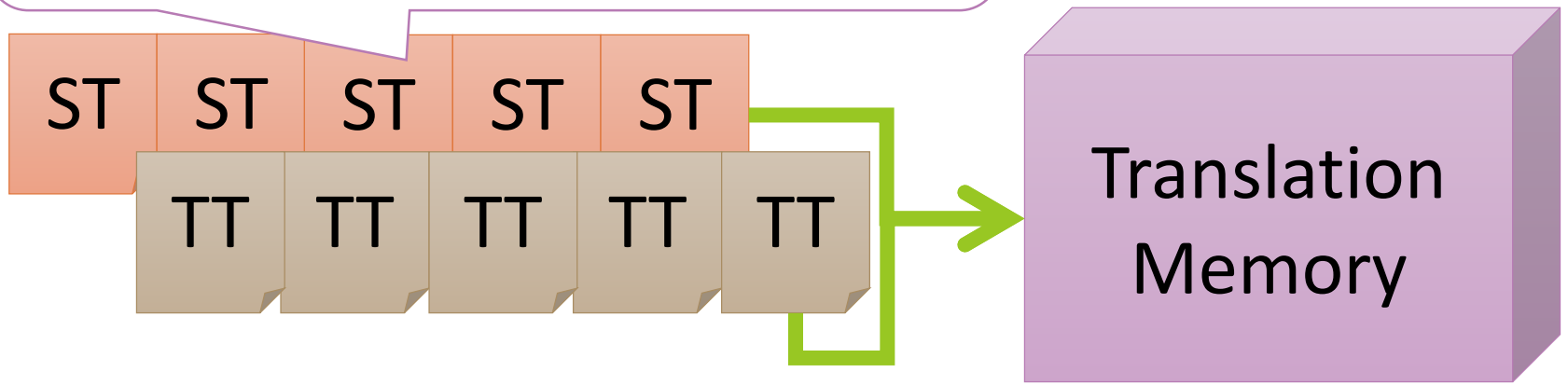
翻譯學院
SCHOOL OF TRANSLATION
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

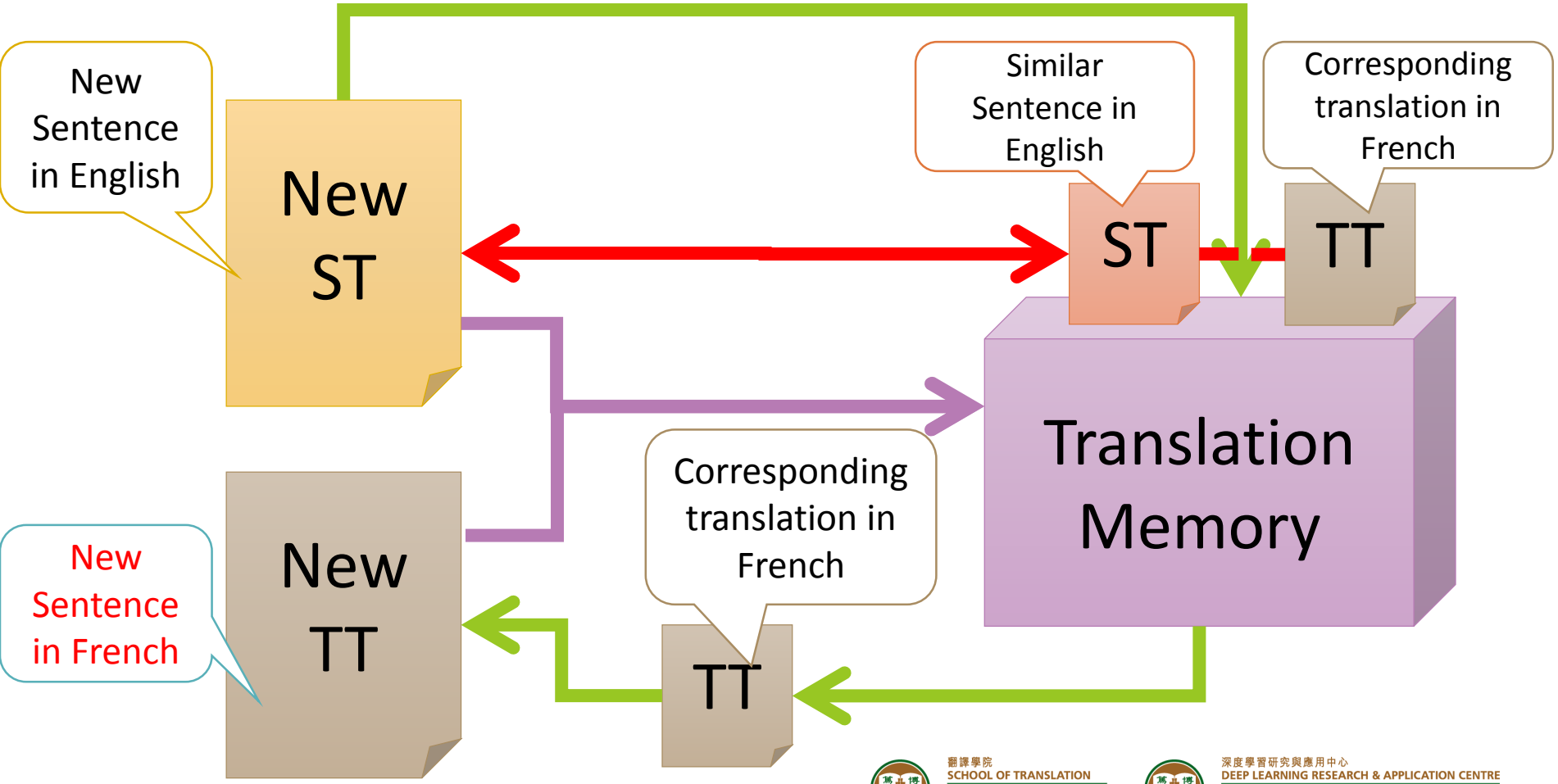
Translation Memory (TM)

ST = Sentence in English

TT = Sentence in French



Translation Memory (TM)

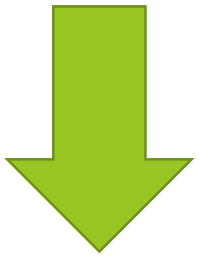


翻譯學院 SCHOOL OF TRANSLATION 香港恒生大學 THE HANG SENG UNIVERSITY OF HONG KONG

深度學習研究與應用中心 DEEP LEARNING RESEARCH & APPLICATION CENTRE 香港恒生大學 THE HANG SENG UNIVERSITY OF HONG KONG

What does TM mean for your business?

Translation Speed



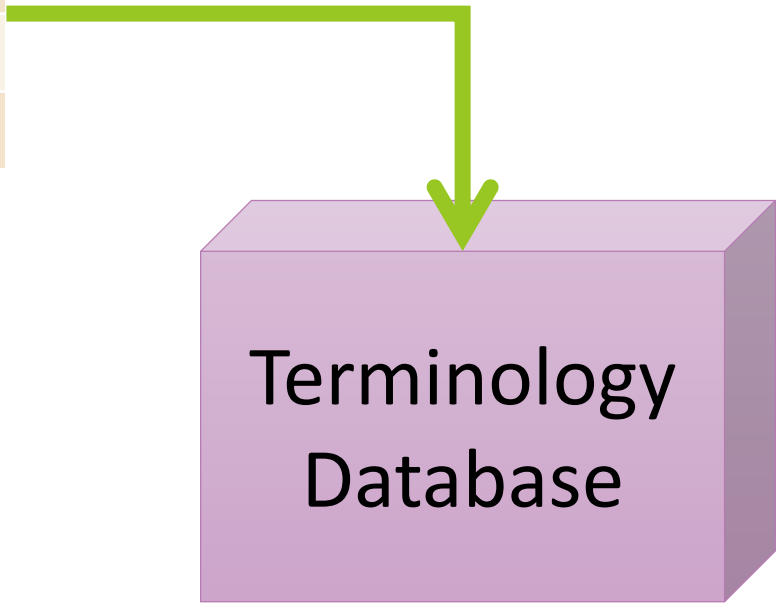
Translation Cost

Translation Style

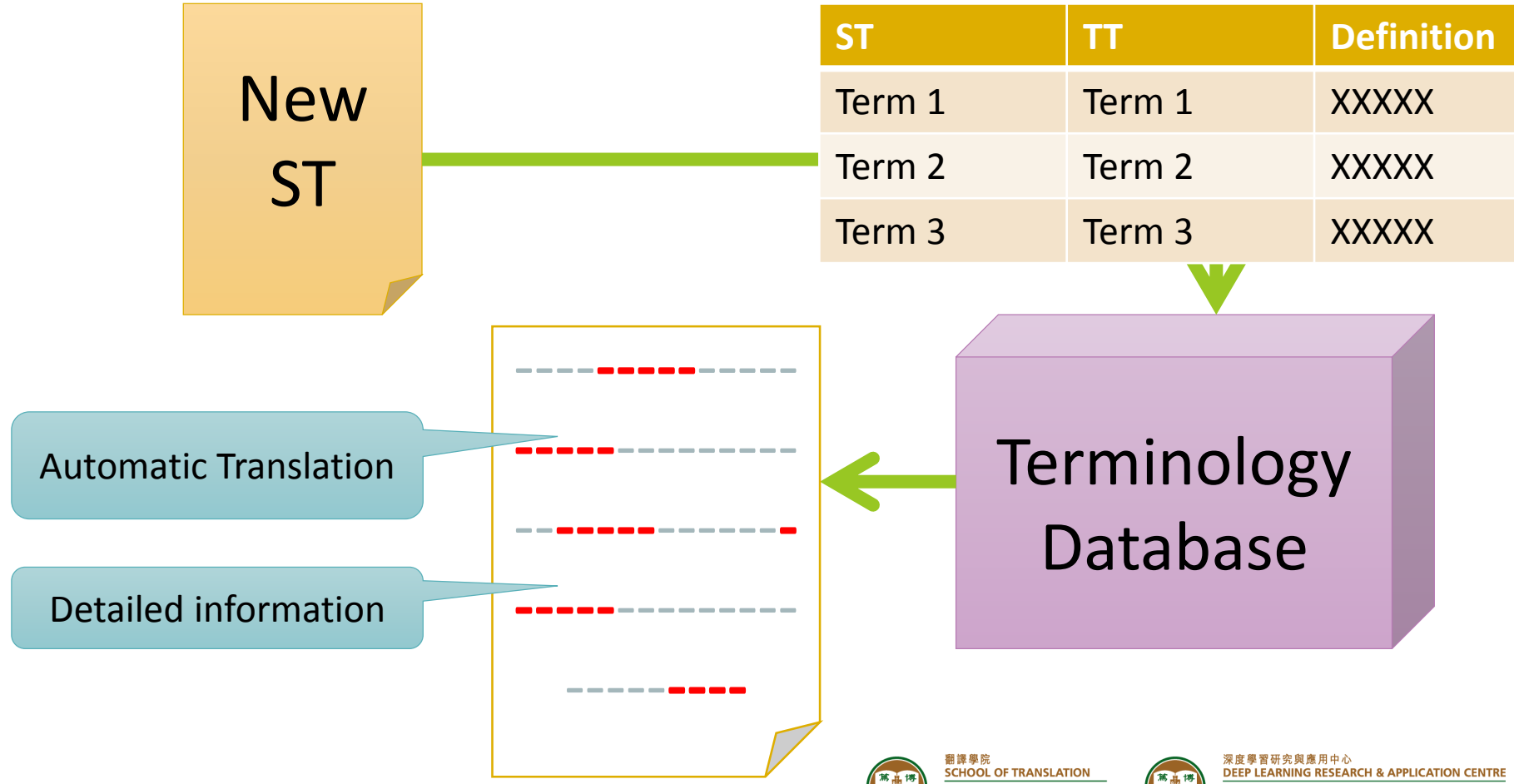
Translation Flexibility

Terminology Database (Termbase)

ST	TT	Definition
Term 1	Term 1	XXXXX
Term 2	Term 2	XXXXX
Term 3	Term 3	XXXXX



Terminology Database (Termbase)



翻譯學院 SCHOOL OF TRANSLATION 香港恒生大學 THE HANG SENG UNIVERSITY OF HONG KONG

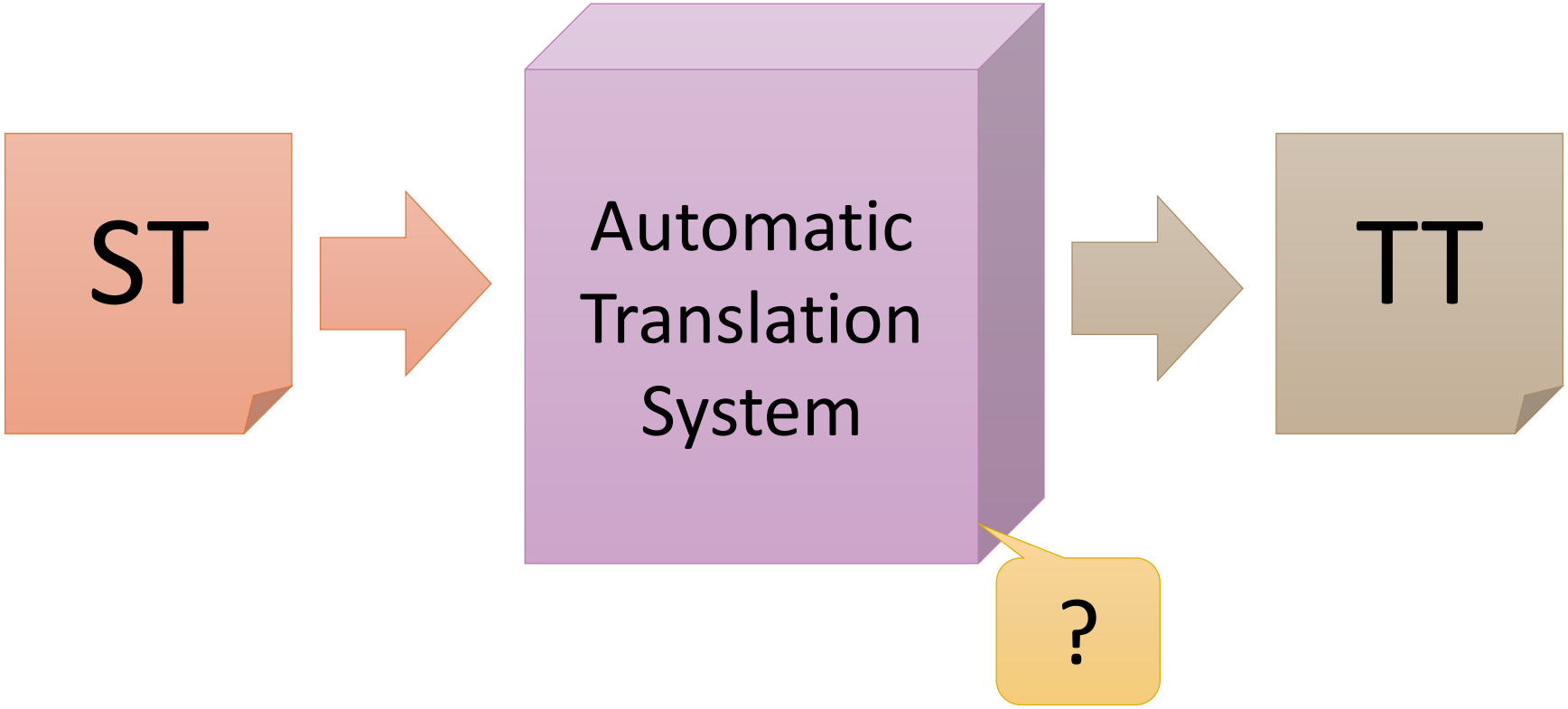
深度學習研究與應用中心 DEEP LEARNING RESEARCH & APPLICATION CENTRE 香港恒生大學 THE HANG SENG UNIVERSITY OF HONG KONG

What do Termbases mean for your business?

Accuracy

Consistency

Machine Translation (MT)



Methods

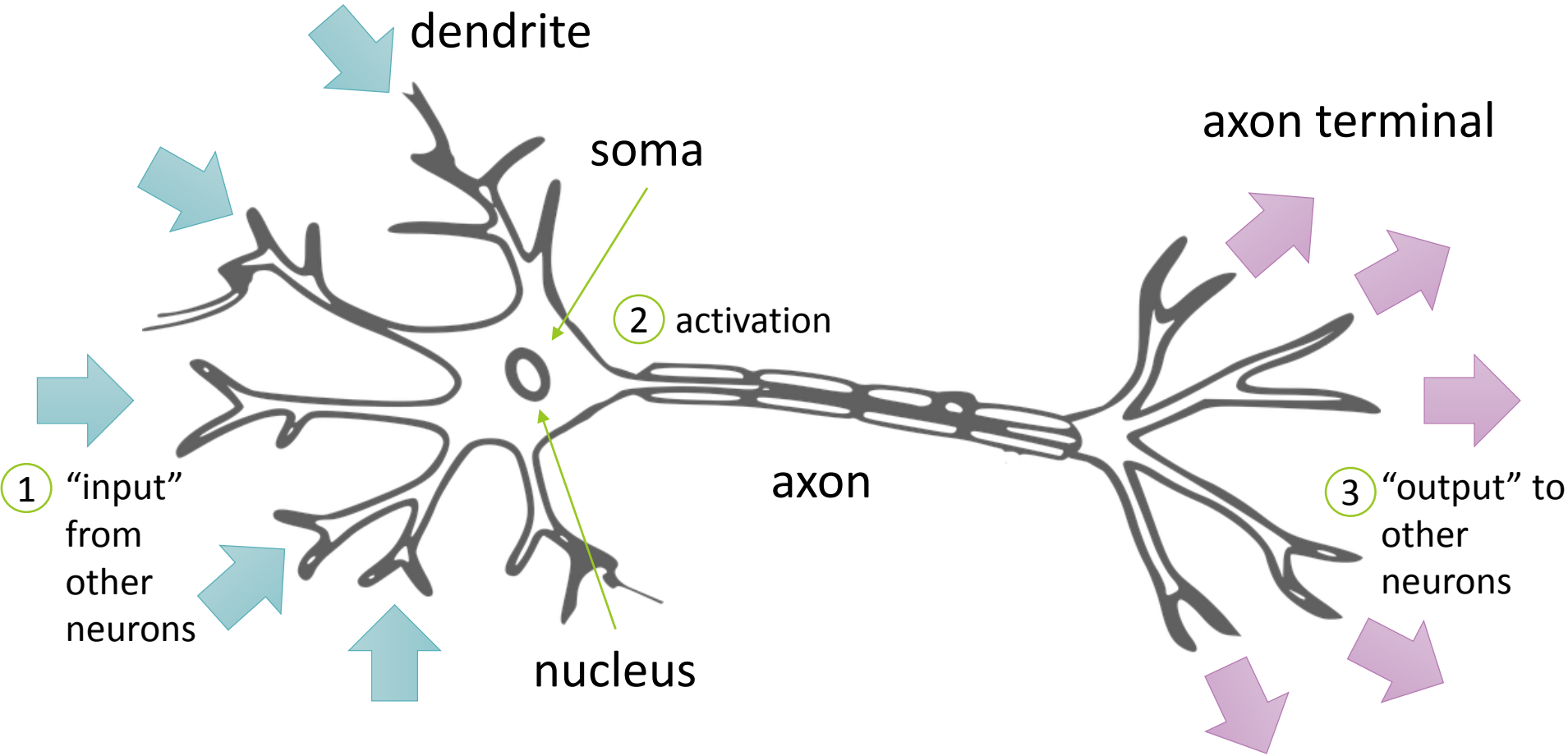
Rule-based Approach

Example-based Approach

Statistical Approach

Deep Learning Approach

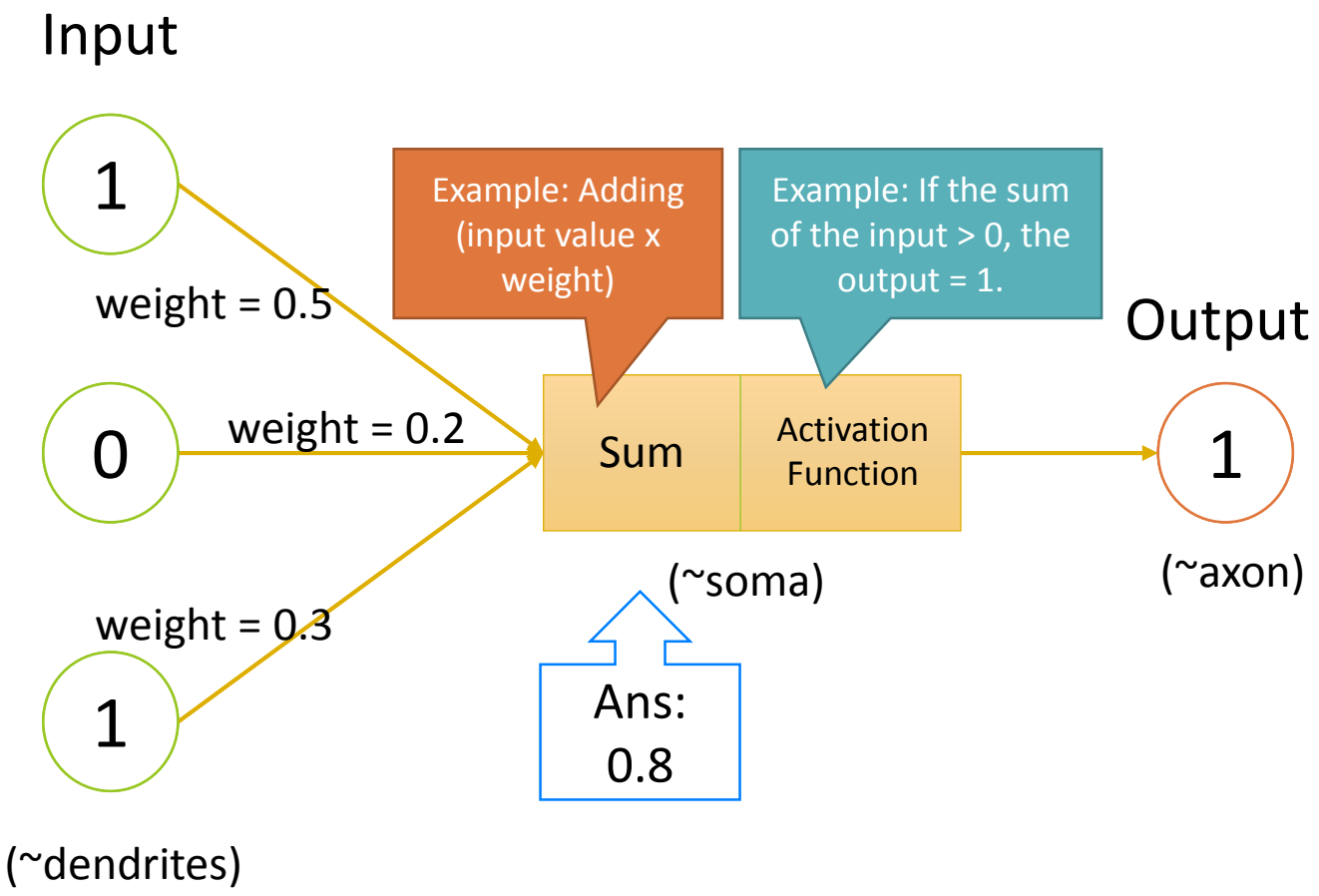
Biological Neuron



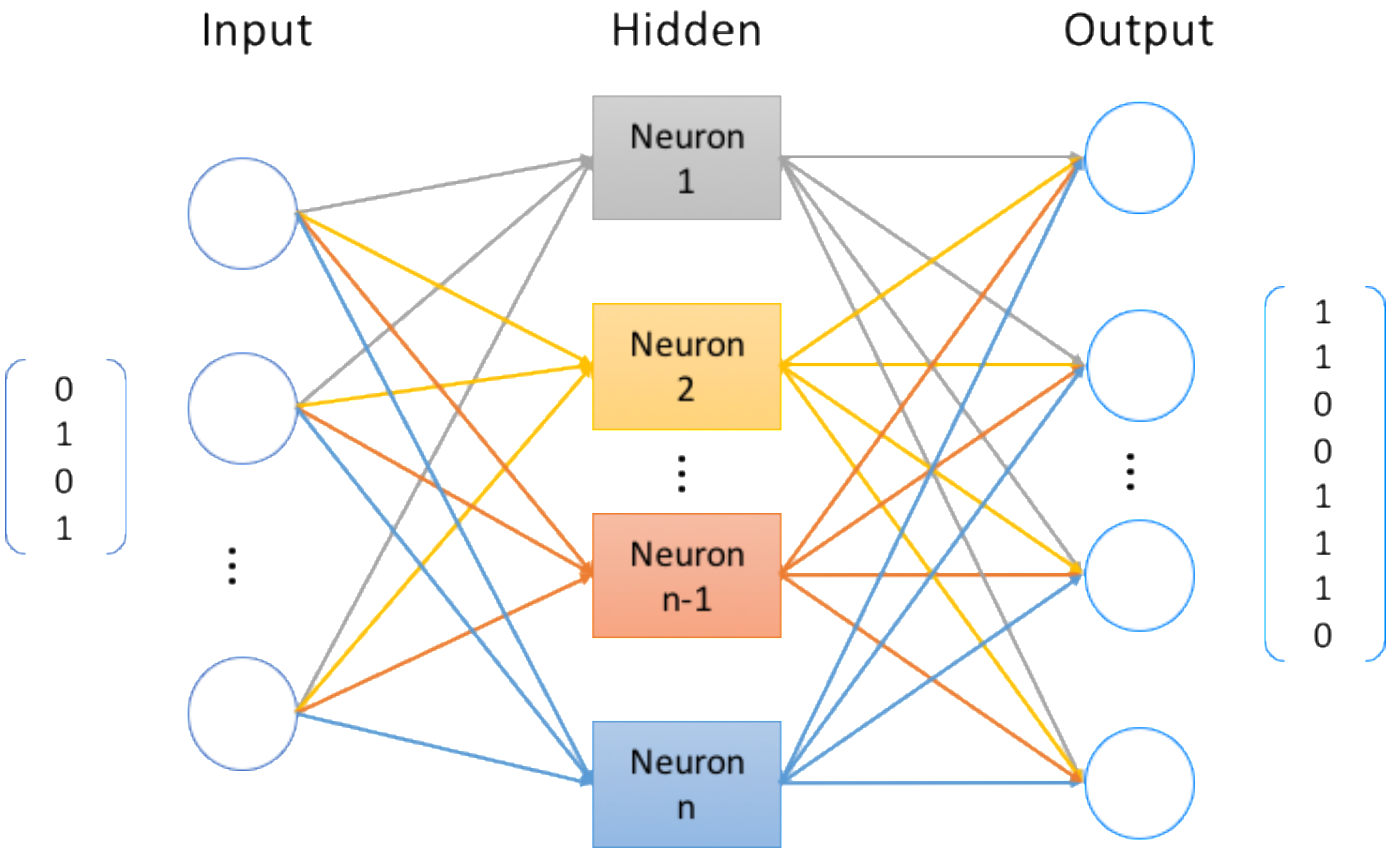
翻譯學院
SCHOOL OF TRANSLATION
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

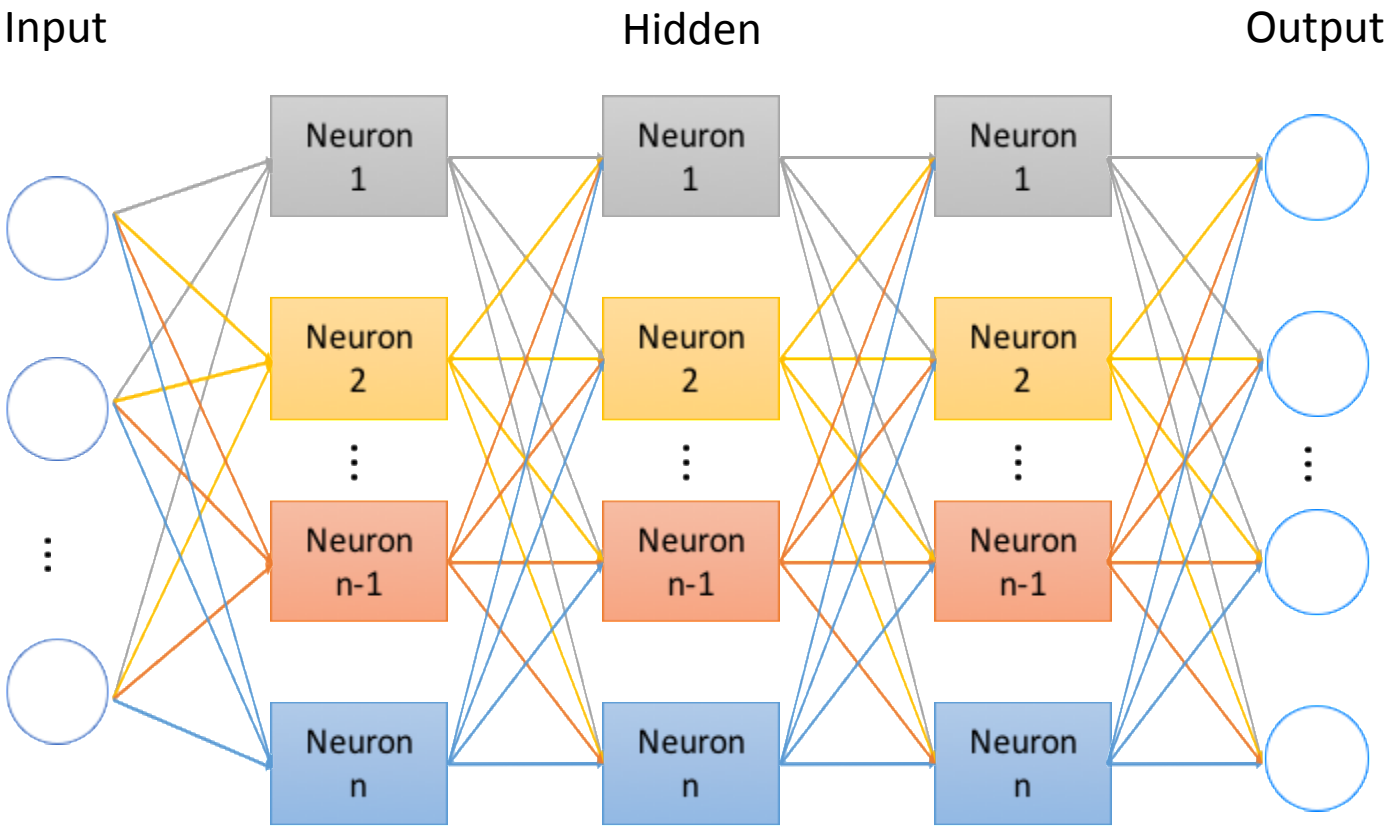
Artificial Neuron



Artificial Neural Network: Many Neurons

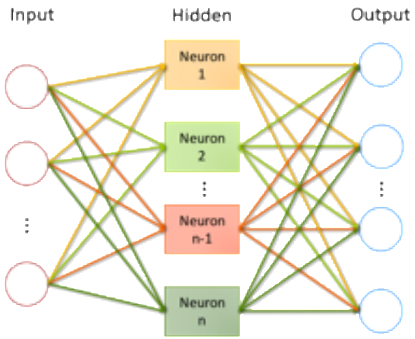


Deep Learning: Multiple Hidden Layers

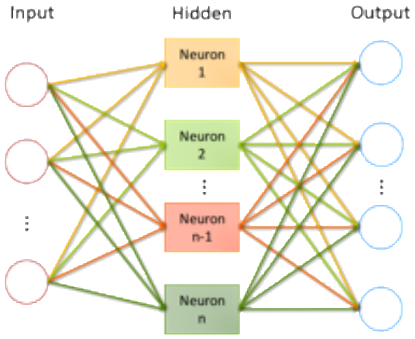


Neural Machine Translation (NMT)

ST

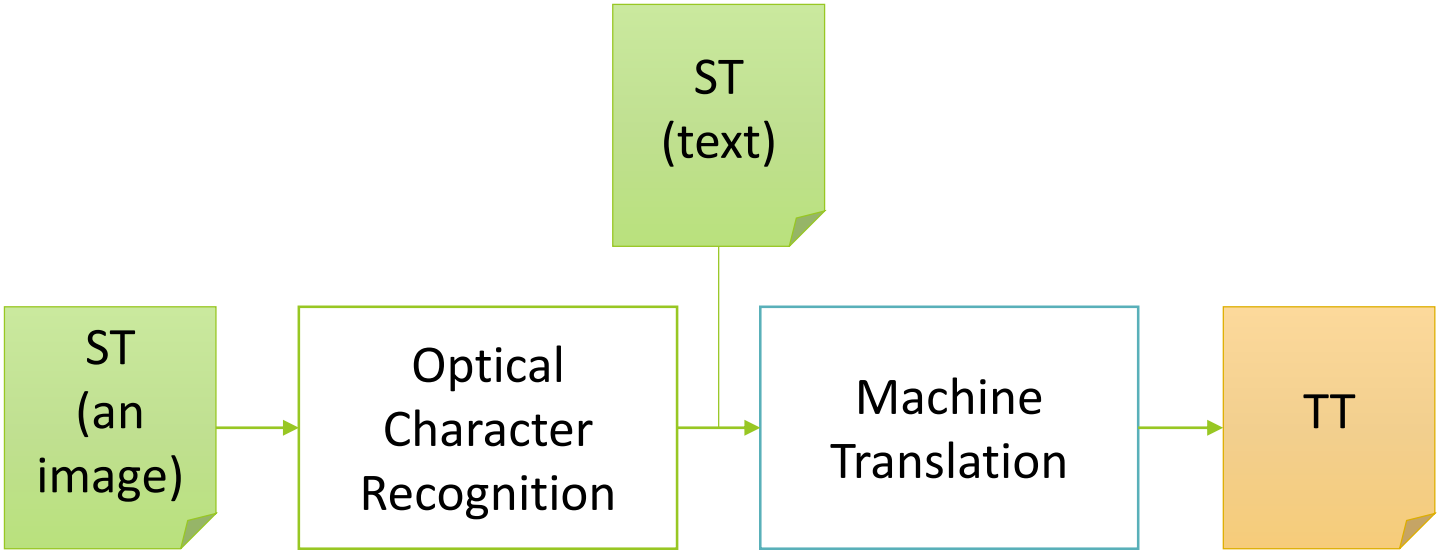


3
0
6
2
4
7
0
0
3
0
6
2
4
7
7
0

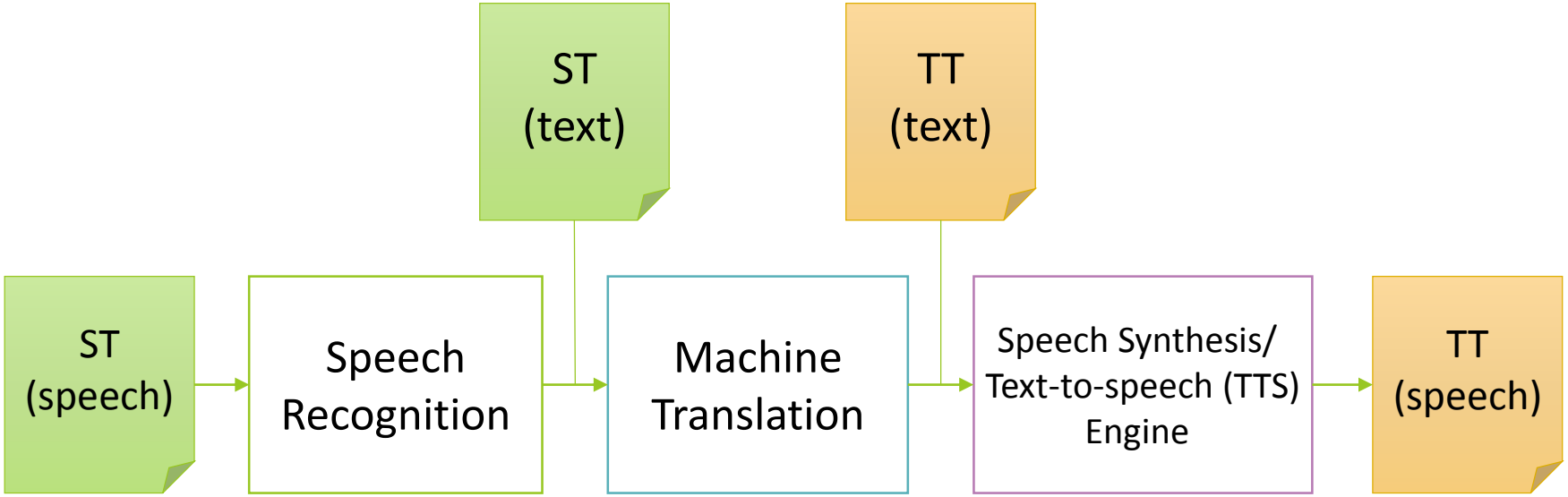


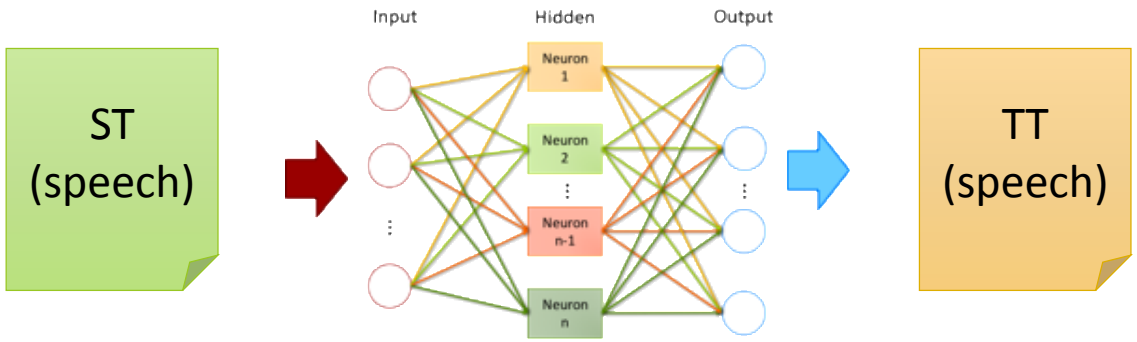
TT

Image Translation



Speech Translation





Other Tools

- Corpora and Concordancers
- Localization Tools
- Translation Servers
- Checkers
- Tool for summarization
- Computer-aided Editing Tools

AI for Translation: Strengths, Weaknesses and the Way Forward

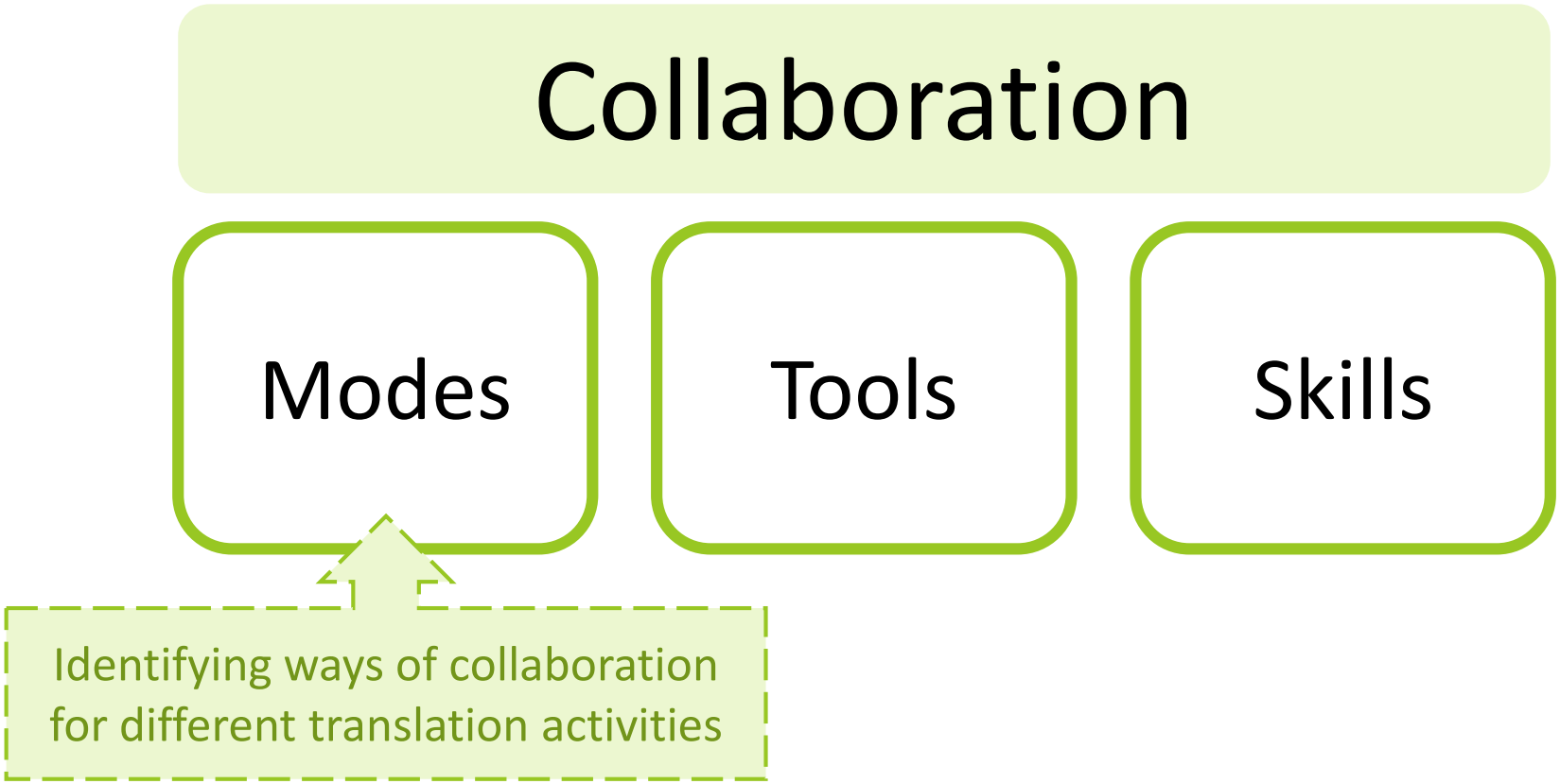
Strengths

Speed	Gist Translation	Translation Draft
Translation of Repetitive Elements	Term Consistency	Language Learning

Weaknesses

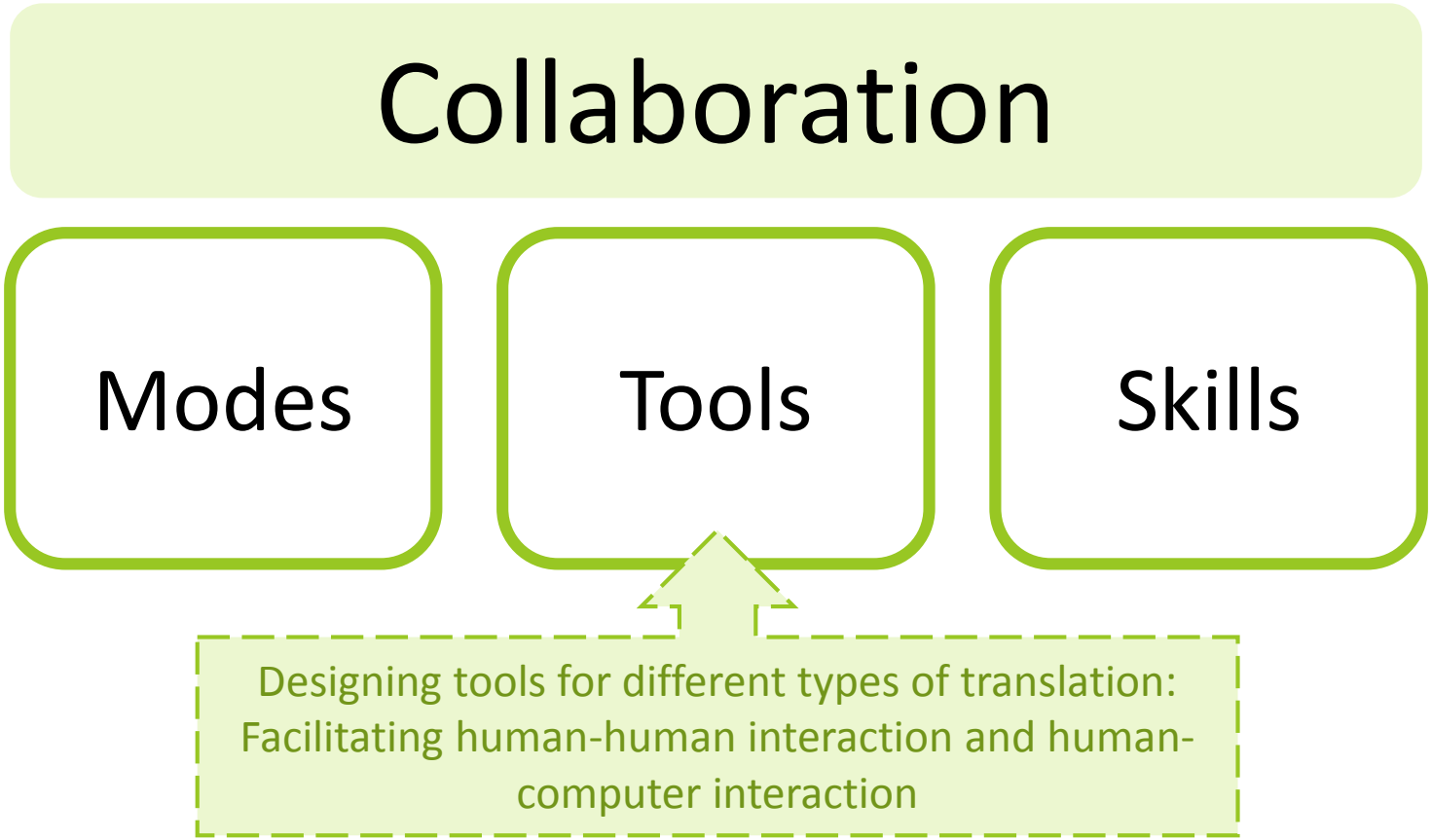
Accuracy	Ambiguity	Limited Context
Cultural Issues	Limited Text Types	Lack of Reasoning

The Way Forward



	Levels of User-Computer Interaction		
Human Translation	Machine-aided/assisted Human Translation (MAHT)	Human-aided/assisted Machine Translation (HAMT)	Machine Translation (MT)

The Way Forward



Tools

Language Professionals

Translation Researchers

General Public

General Translation

Domain-specific Translation

Other Language Services

Integrated Tools for Translation of Financial Documents

Significance of financial translation: IPO documents as an example

Largest IPO Centre: US$36.5 billion in 2018

Integral Part of Initial Public Offerings

Bilingual Prospectuses

IPO documents are usually prepared by lawyers and accountants in English and need to be translated to Chinese as required by HKEX. The translation task, usually handled by printing companies, is costly and labor-intensive.

Tool 1: IPO Translate

Tool 2: FinTranX

Feature 1
Deep Learning Technology

Feature 2
Domain-specific Machine Translation

Feature 3
Tool Integration for Interaction

1. Automatic Translation: Generation of translation draft
2. Translation Memory: Reuse of previous translations
3. Terminology Databases: Translation of terms
4. External Databases: Online research

The Way Forward

Collaboration

- Modes
- Tools
- Skills

Developing skills for better collaboration in the era of AI-aided Translation

翻譯學院
SCHOOL OF TRANSLATION
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

Language Skills
Reading, Writing, Listening and Speaking

Translation Skills
Translation Methods and Translation Theories

Evaluation

Editing

Project Management

Tool Operation and Customization

Database Management

AI and Machine Learning

翻譯學院
SCHOOL OF TRANSLATION
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

AI VS Human Translation: Inevitable Collaboration

Language professionals familiar with AI
Language professionals NOT familiar with AI

Thank you very much!

- Email: scsiu@hsu.edu.hk

Additional Information

Research on Financial Translation

2014: Hang Seng Management College Business Translation Index

2015: “Automated Pre-editing and post-editing: A Hybrid Approach to the Computerized Translation of IPO Prospectuses”, Journal of Translation Technology 1(1)

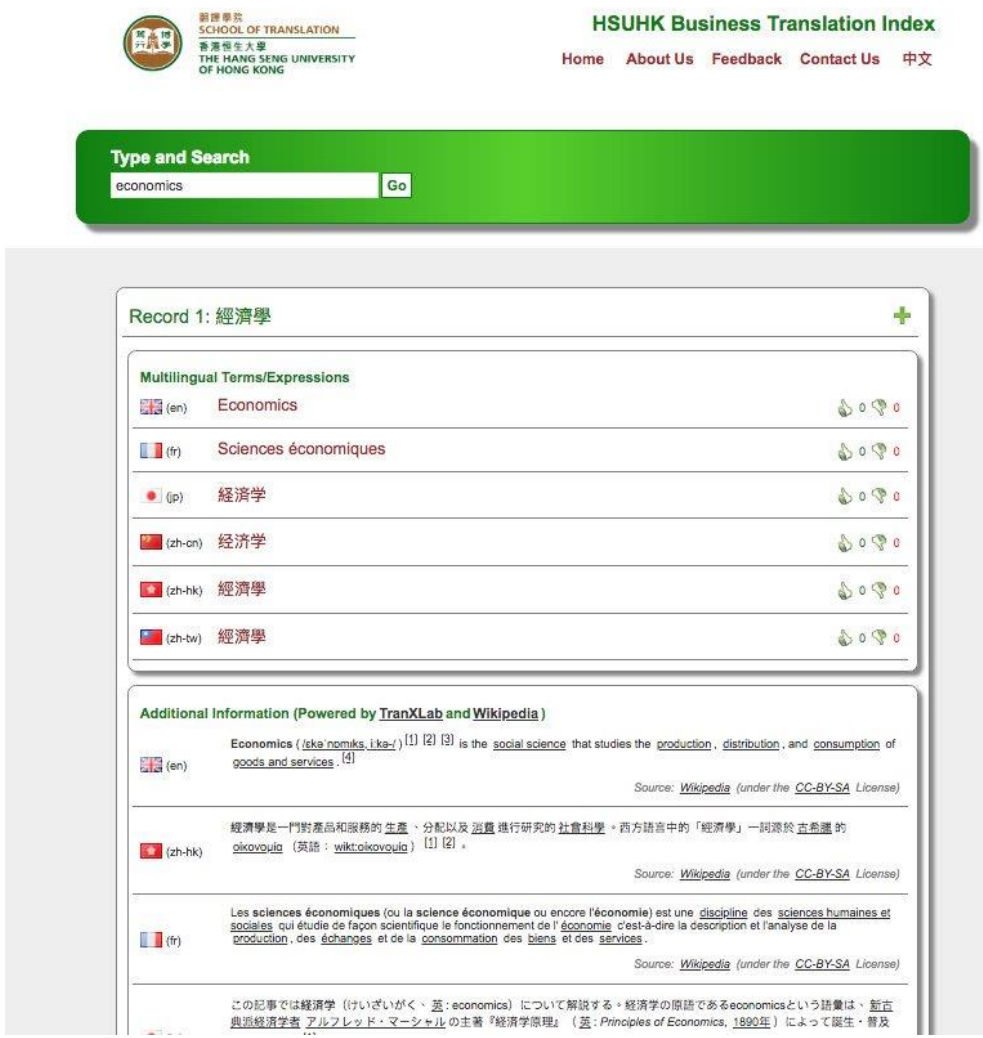
2016: Neural Machine Translation

2017: Establishment of Deep Learning Centre

2018: Project Supported by the Innovation and Technology Fund

2019: Development of an Integrated Platform for Translation of IPO Documents

Example: HSUHK Business Translation Index



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

深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

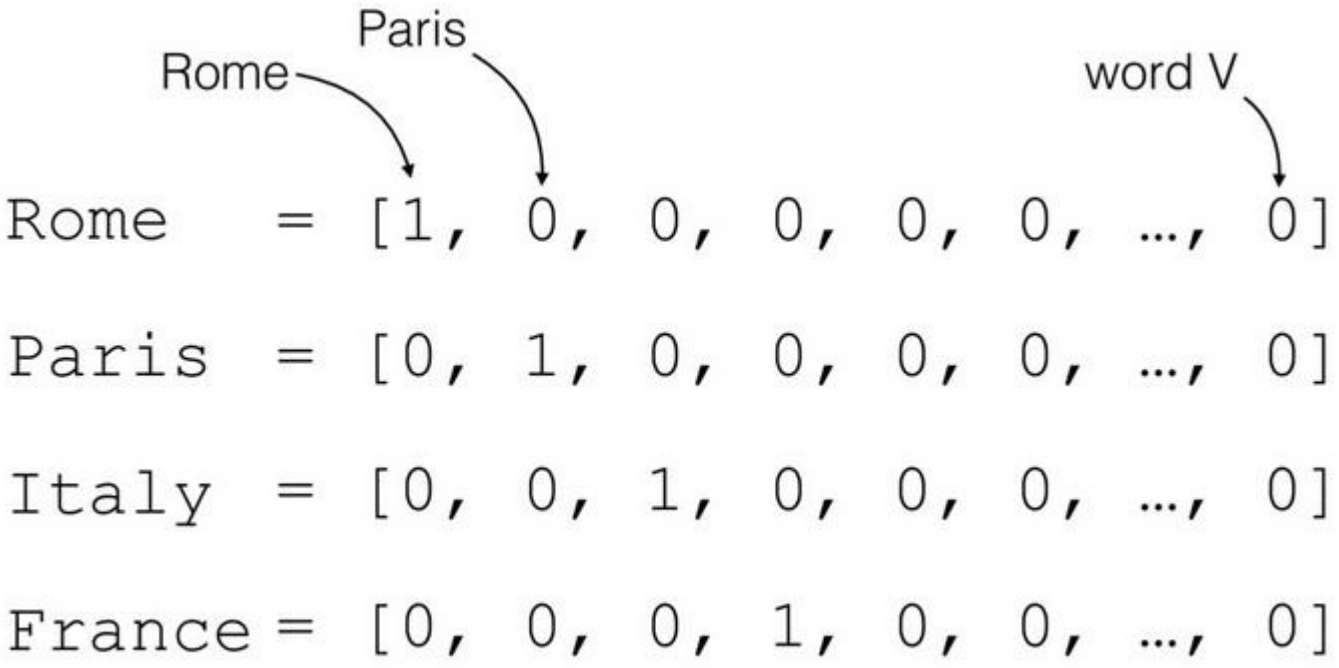
MT Methods

Traditional Approaches to Machine Translation

01	Rule-based Machine Translation	● I go to school by train. ● 我／去／學校／乘／火車／。 ● 我乘火車去學校。
02	Example-based Machine Translation	● I go to school by train. ● I go to school by taxi. 我乘計程車去學校。 ● 我乘火車去學校。
03	Statistical Machine Translation	● Translation Model *P*(“I”|“我”) x *P*(“go”|“去”) x ... ● Language Model: *P*(“我乘火車去學校”)

Word Embedding

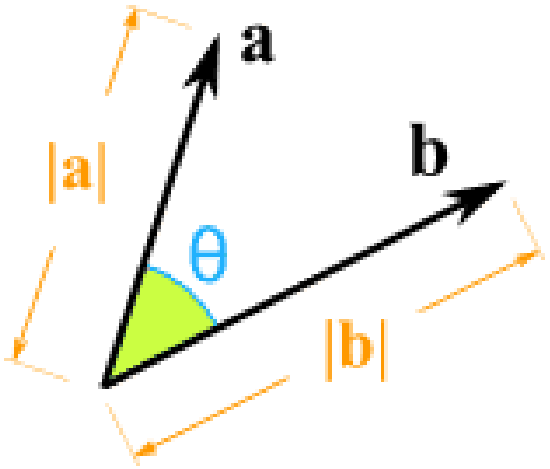
Word Embedding (1): One-hot Vector

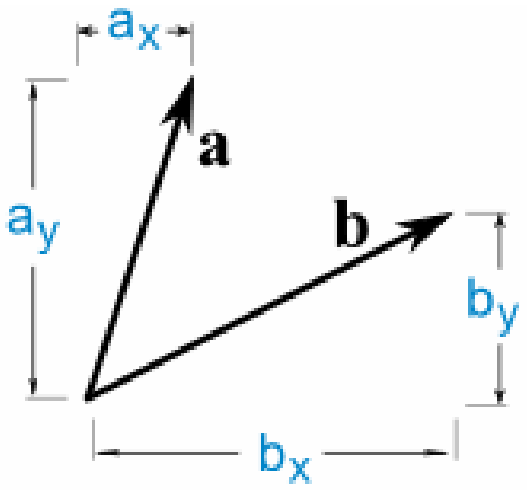


For more information, visit
https://medium.com/@athif.shaffy/one-hot-encoding-of-text-b69124bef0a7

Problem

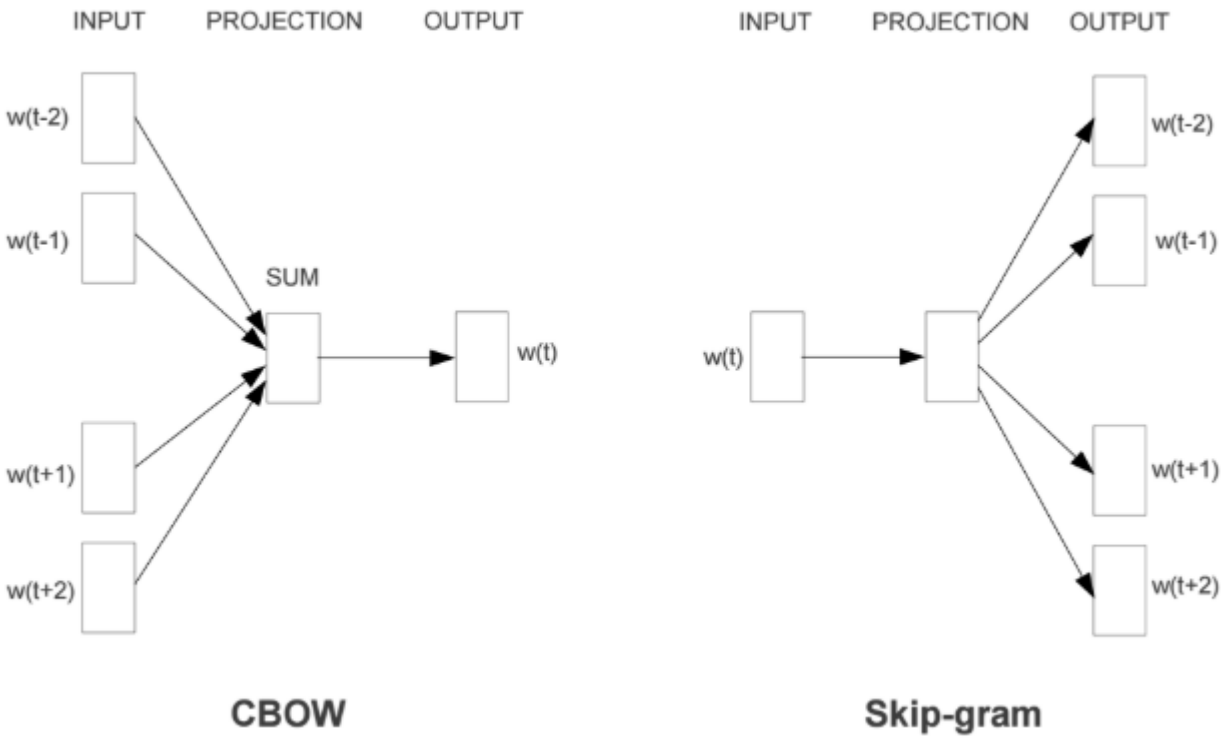
$$a \cdot b = a_x \times b_x + a_y \times b_y = |a||b|\cos(\theta)$$





For more information, visit
https://www.mathsisfun.com/algebra/vectors-dot-product.html

Word Embedding (2): Continuous Bag of Words and Skip-gram



For more information, visit
https://mubaris.com/2017/12/14/word2vec/

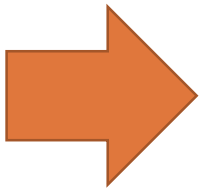
NMT

Methods

Rule-based Approach

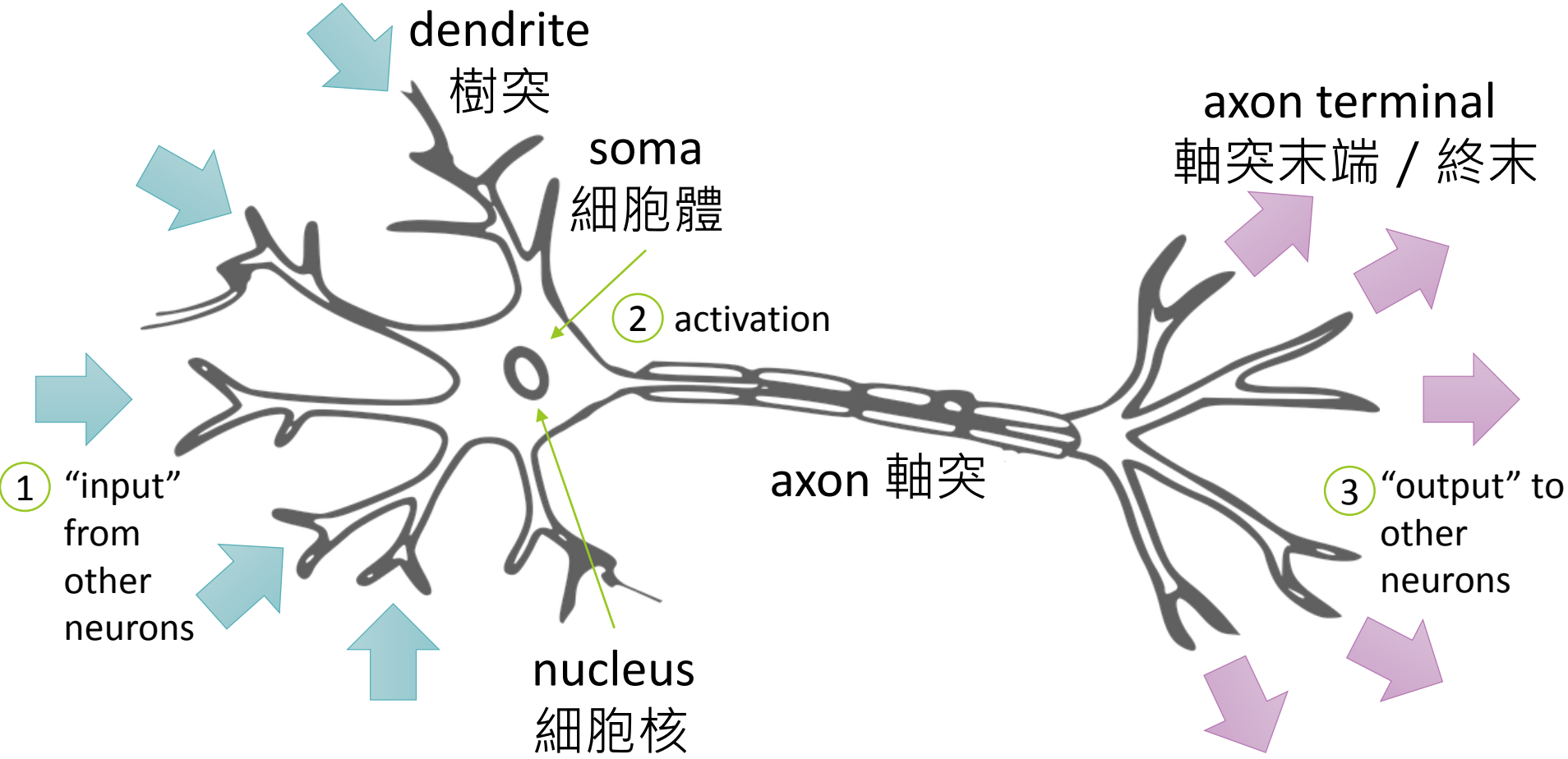
Example-based Approach

Statistical Approach



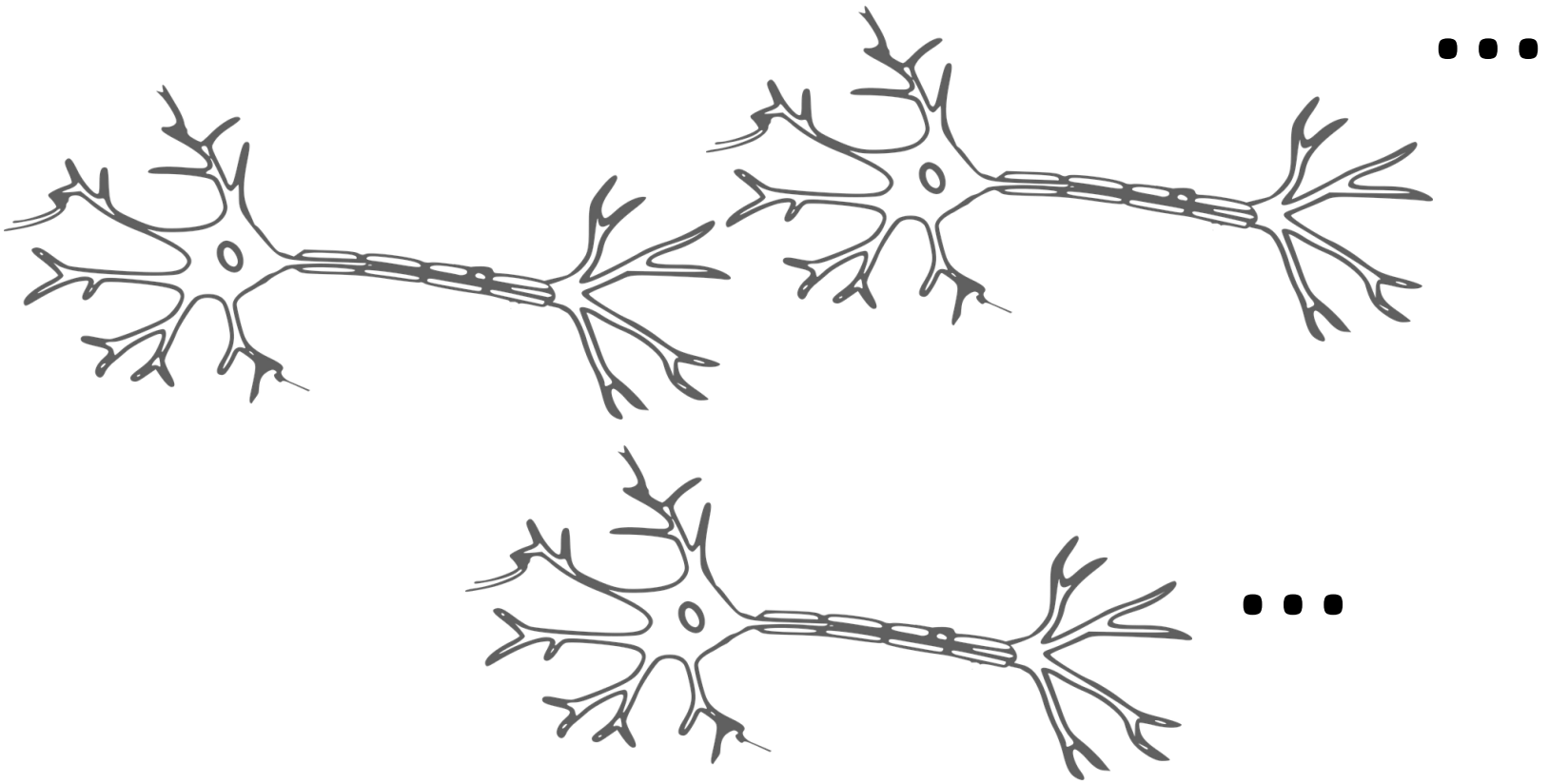
Deep Learning Approach

Biological Neuron

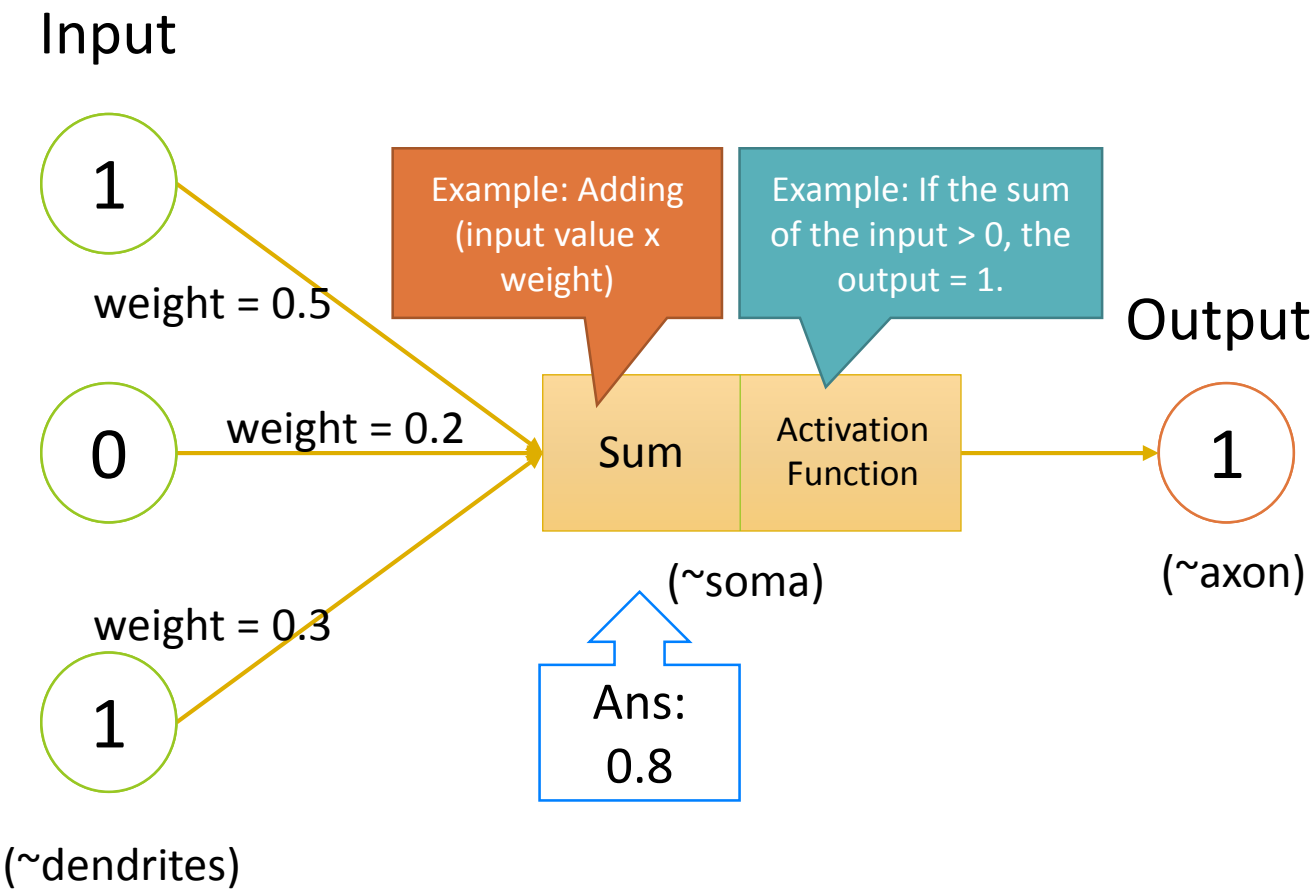


深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

Biological Neural Network

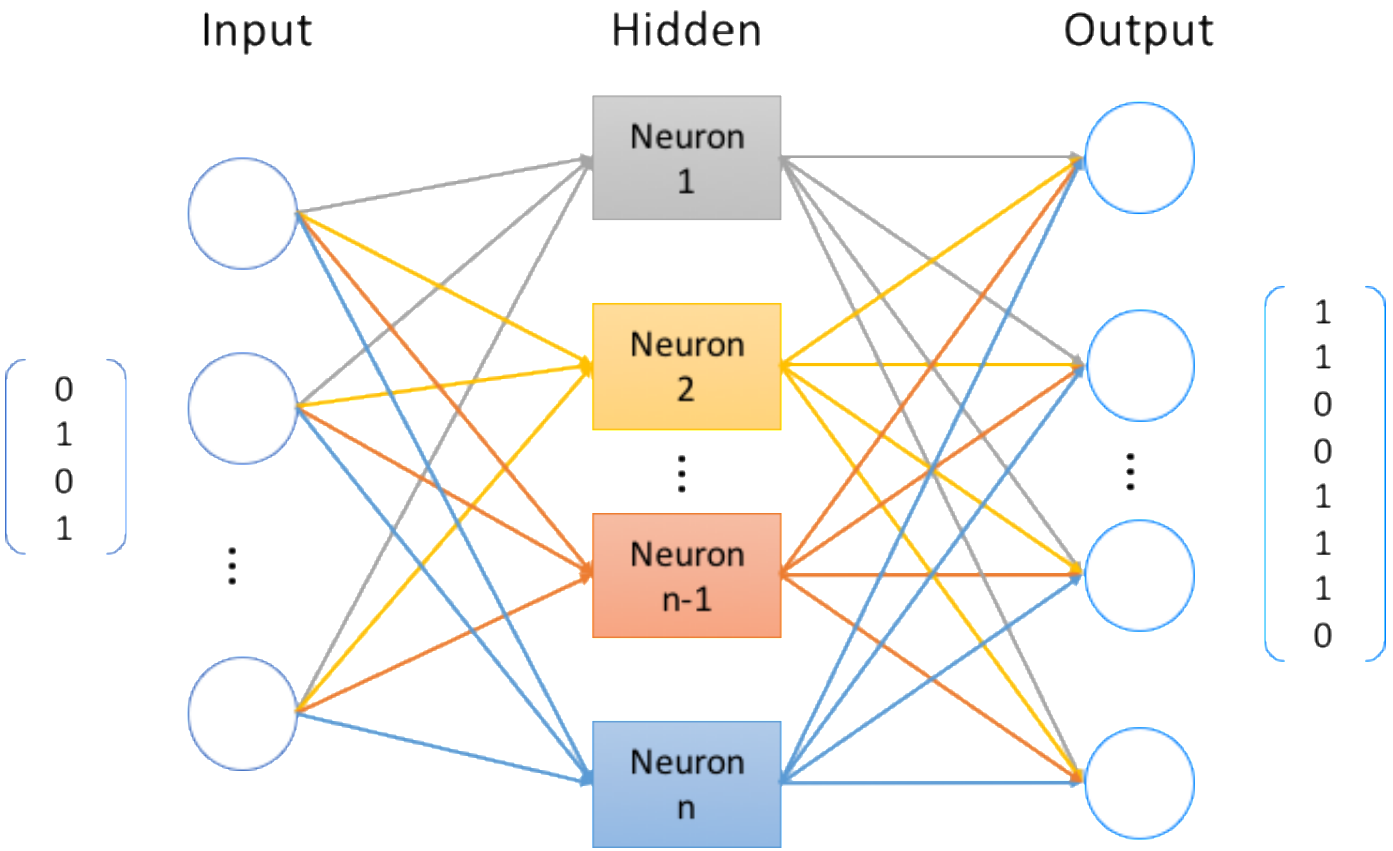


Artificial Neuron



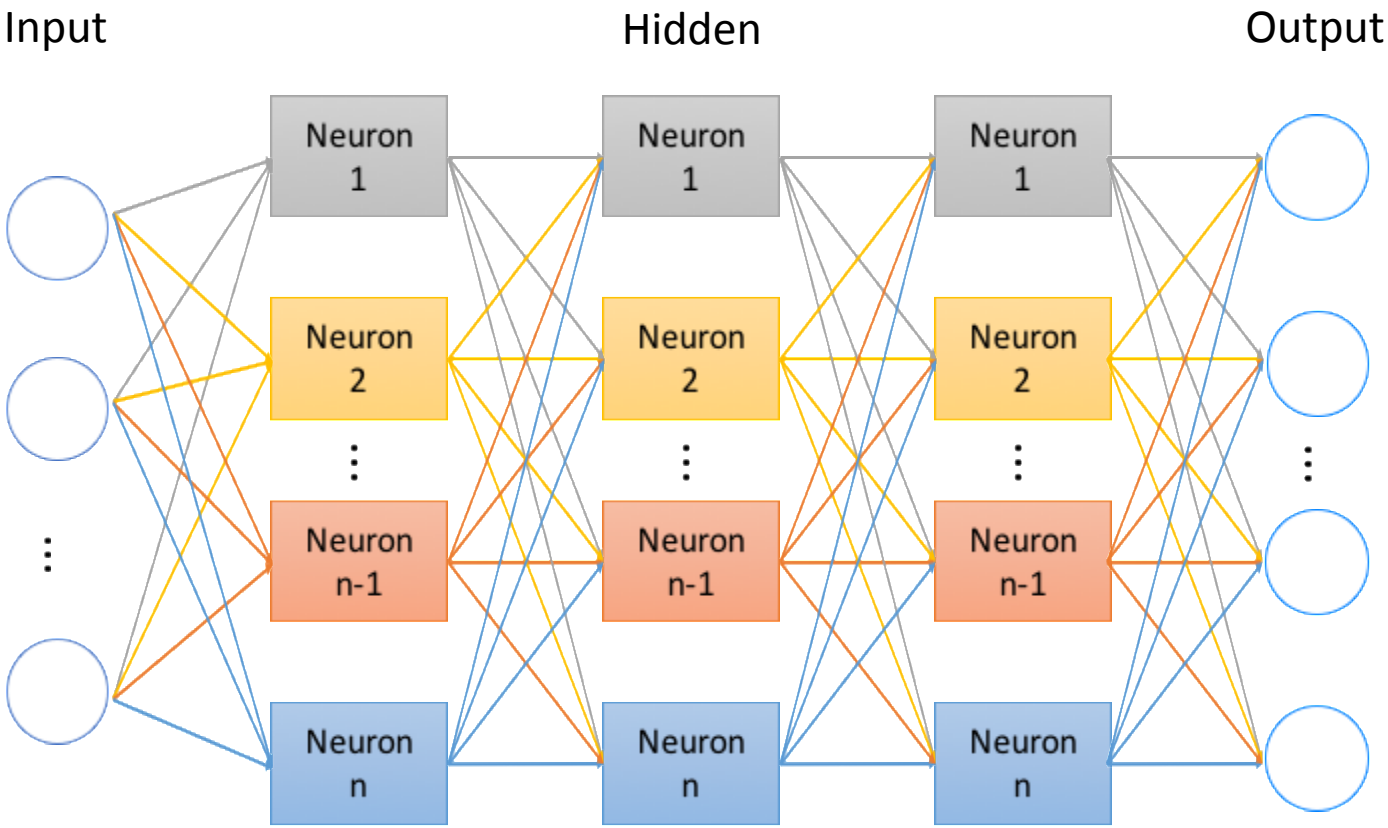
深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

Artificial Neural Network: Many Neurons



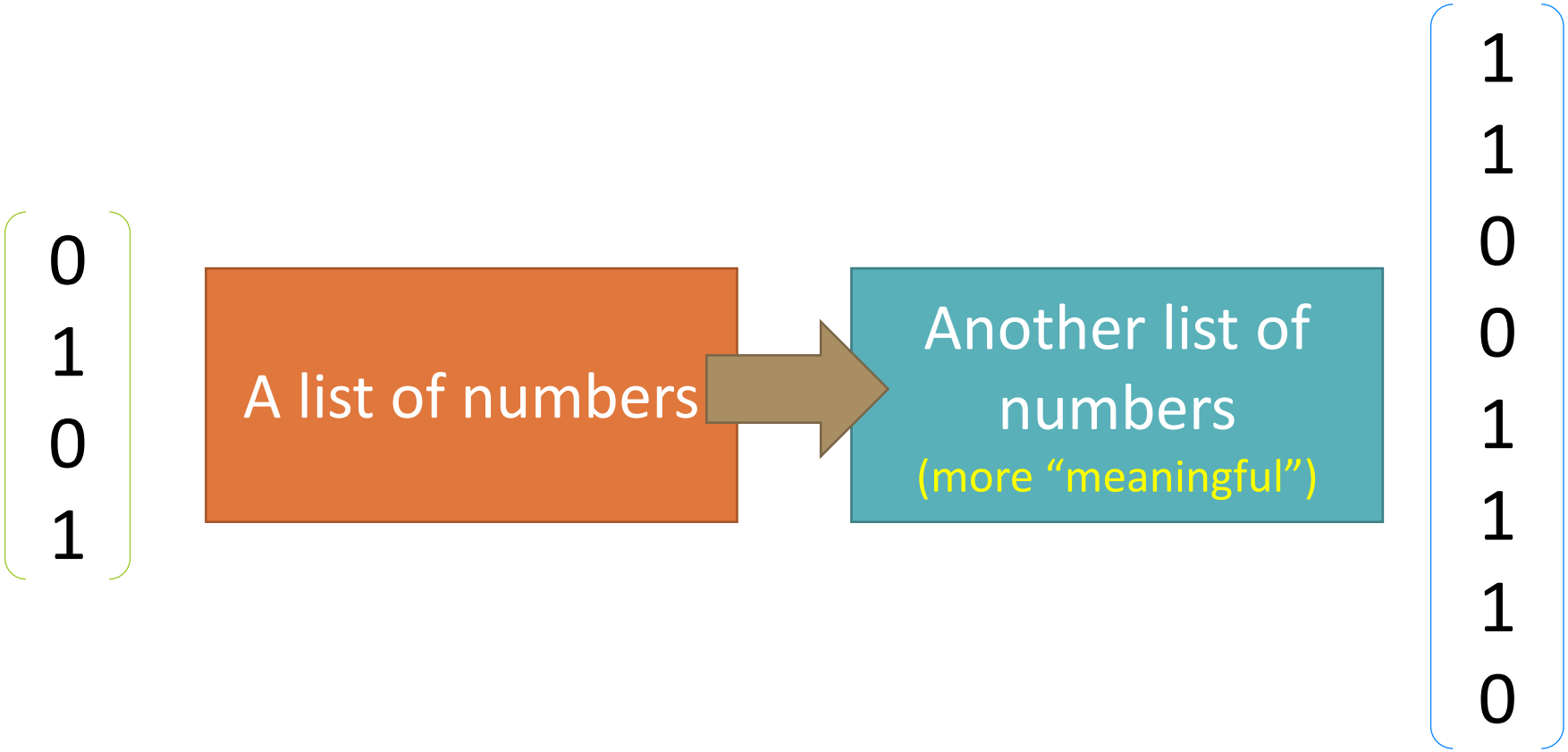
深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY OF HONG KONG

Deep Learning: Multiple Hidden Layers



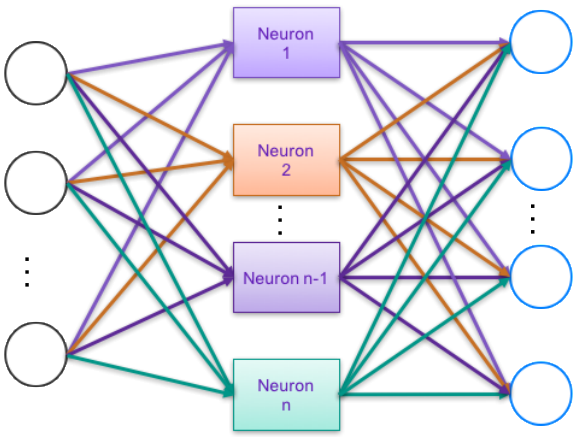
深度學習研究與應用中心
DEEP LEARNING RESEARCH & APPLICATION CENTRE
香港恒生大學
THE HANG SENG UNIVERSITY
OF HONG KONG

Artificial Neural Network



Artificial Neural Network and Image Processing

0
1
0
1
0
1
1
0



Probability	that the picture is a …
0.55	cat
0.300.	tiger
10	dog
0.05	man

Multi-class Classification

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] ⇒ ?

乘

[I] / [go] / [to] / [school] / [by] / **[train]** / [.] / [我] / [乘] ⇒ ?

火車

[I] / **[go]** / **[to]** / [school] / [by] / [train] / [.] / [我] / [乘] / [火車] ⇒ ?

去

[I] / [go] / [to] / **[school]** / [by] / [train] / [.] / [我] / [乘] / [火車] / [去] ⇒ ?

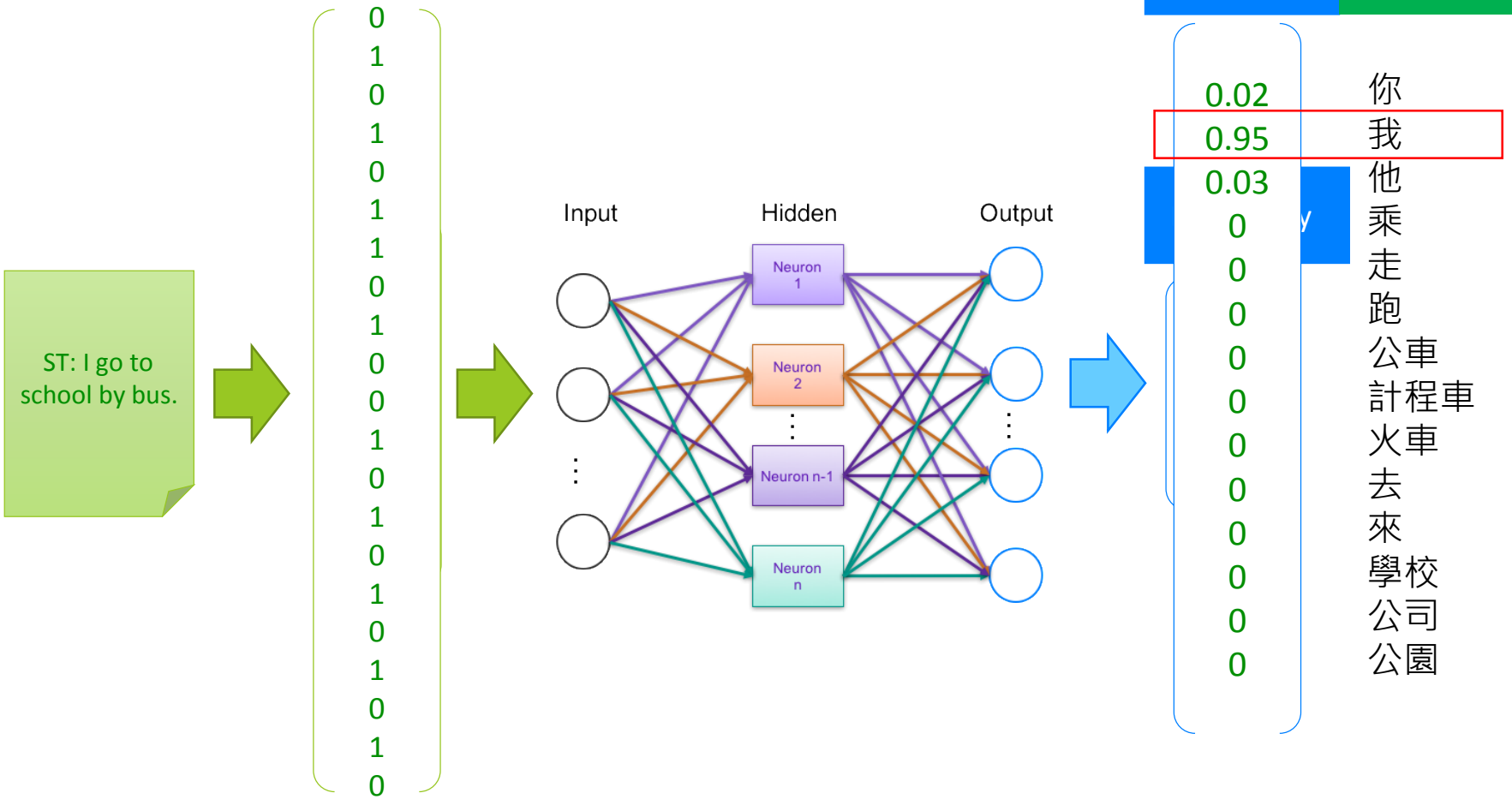
學校

[I] / [go] / [to] / [school] / [by] / [train] / **[.]** / [我] / [乘] / [火車] / [去] / [學校] / ⇒ ?

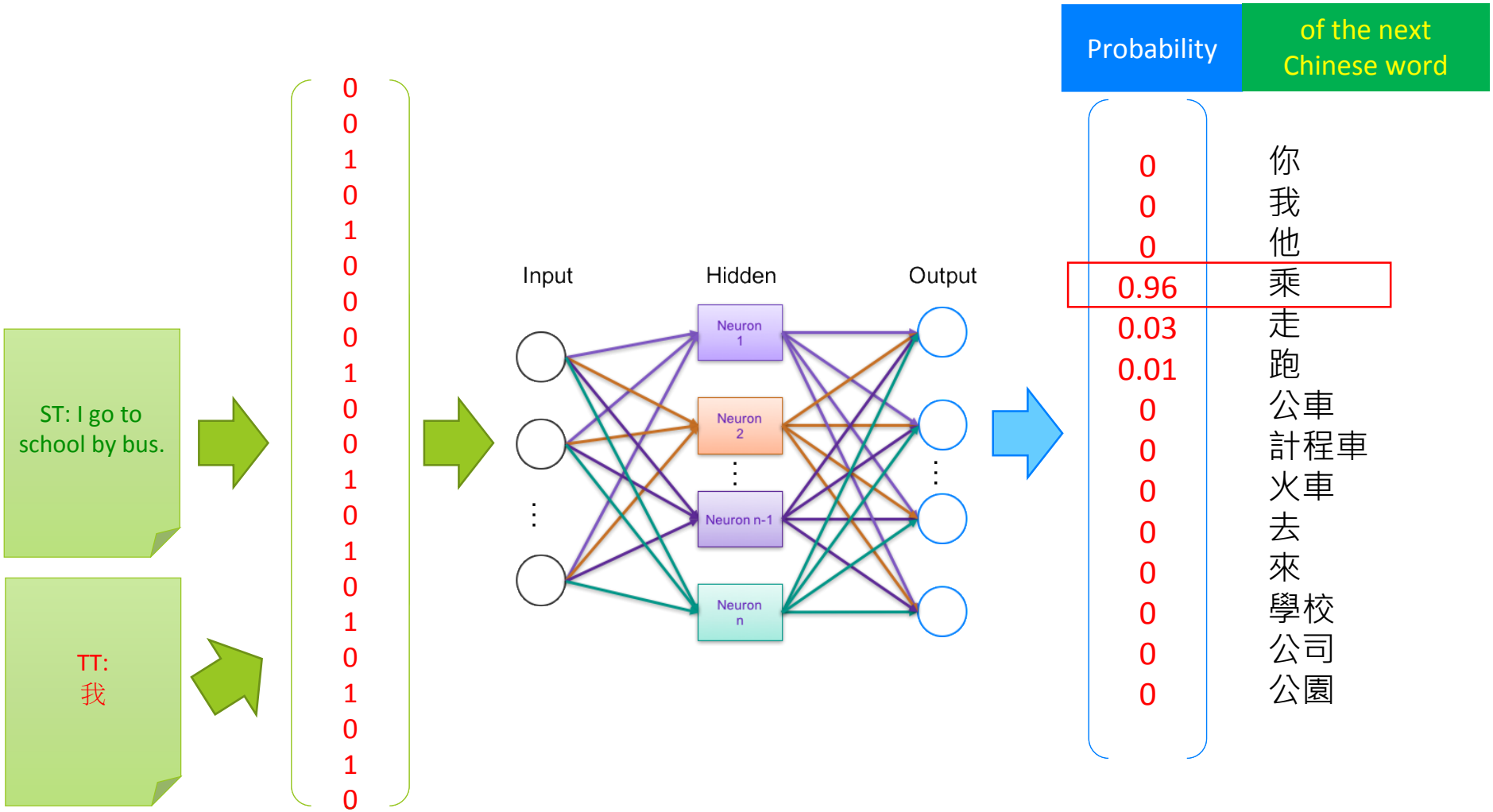
。

Translation: 我乘火車去學校。

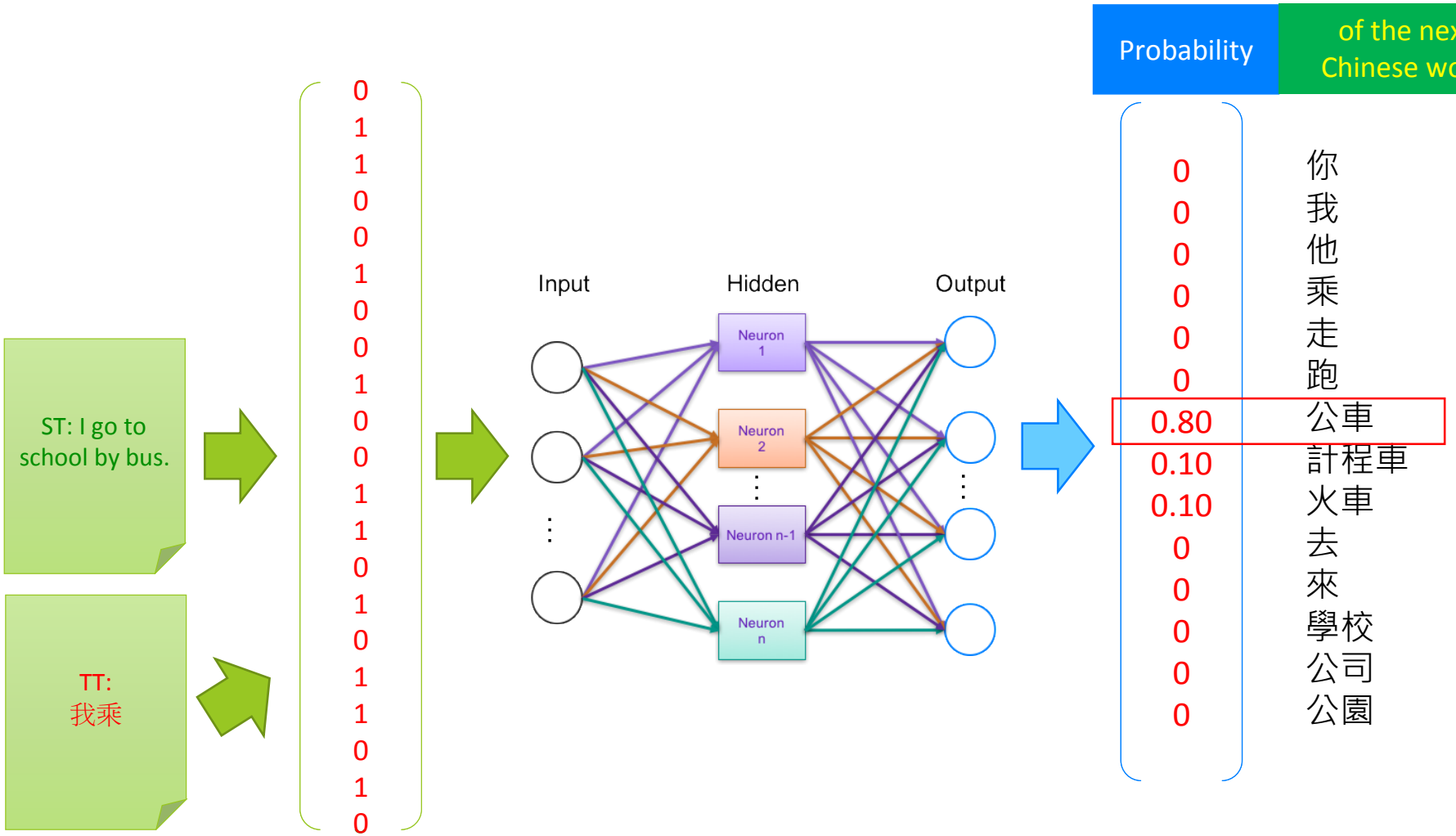
Neural Machine Translation as Multi-class Classification



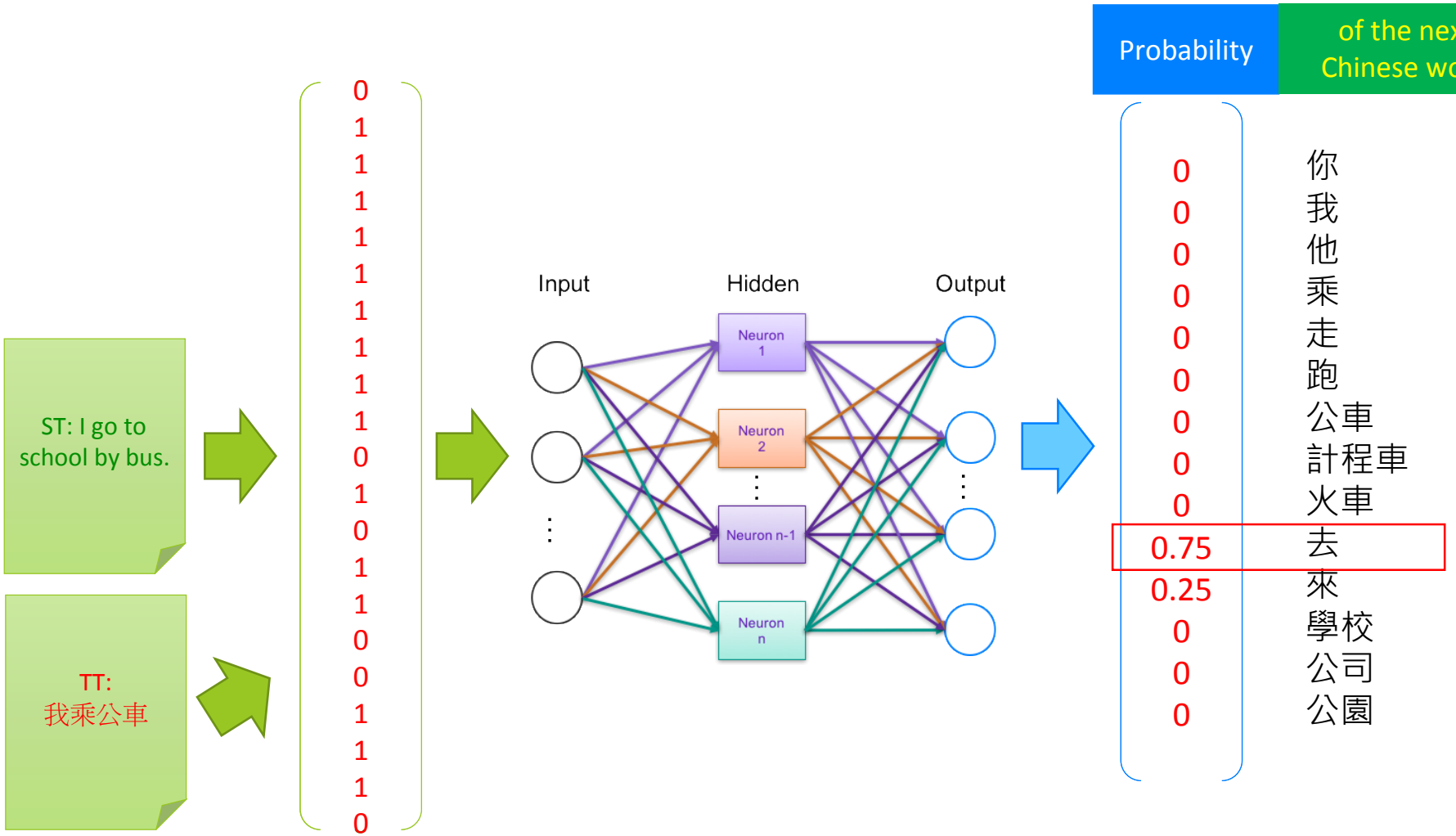
Neural Machine Translation



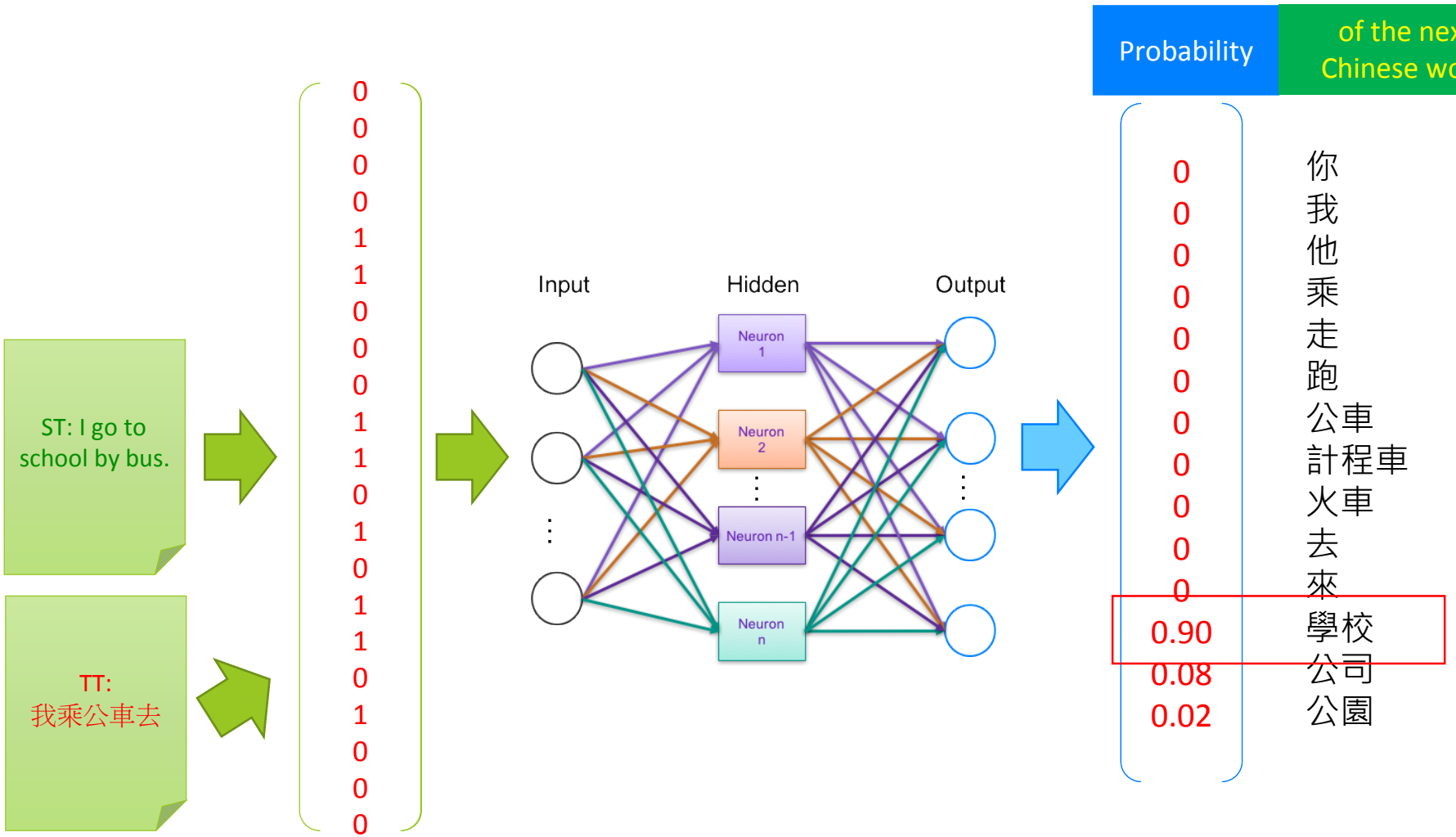
Neural Machine Translation



Neural Machine Translation

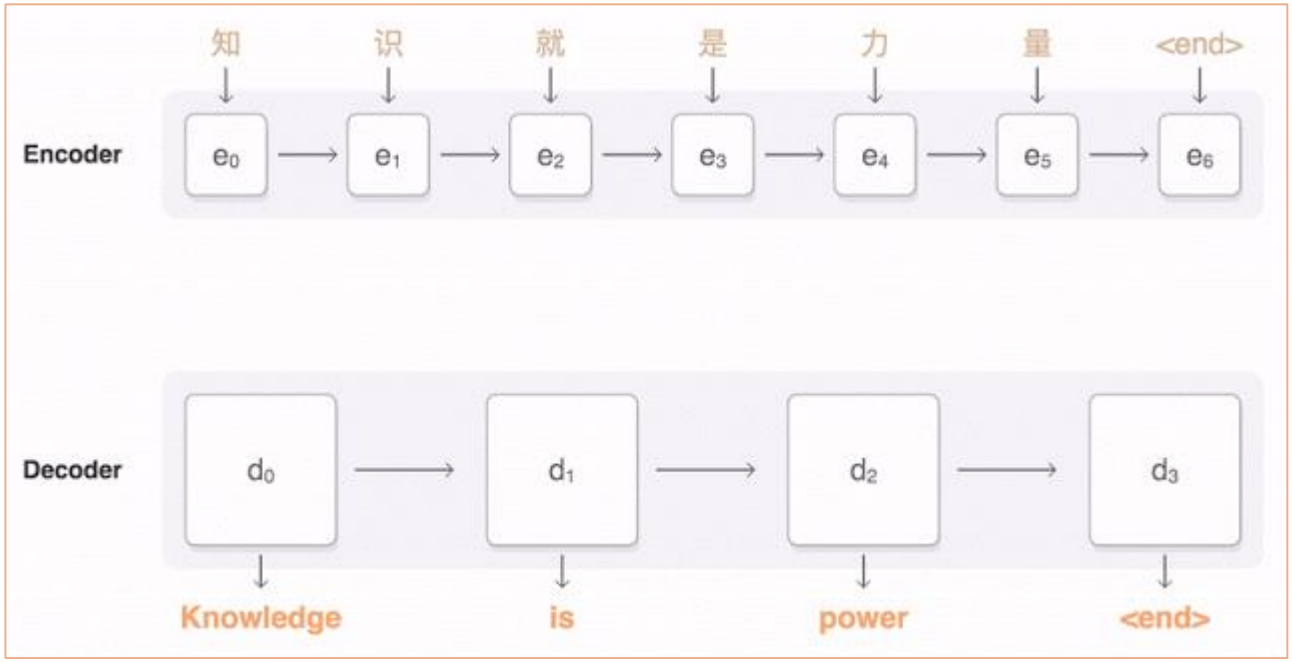


Neural Machine Translation



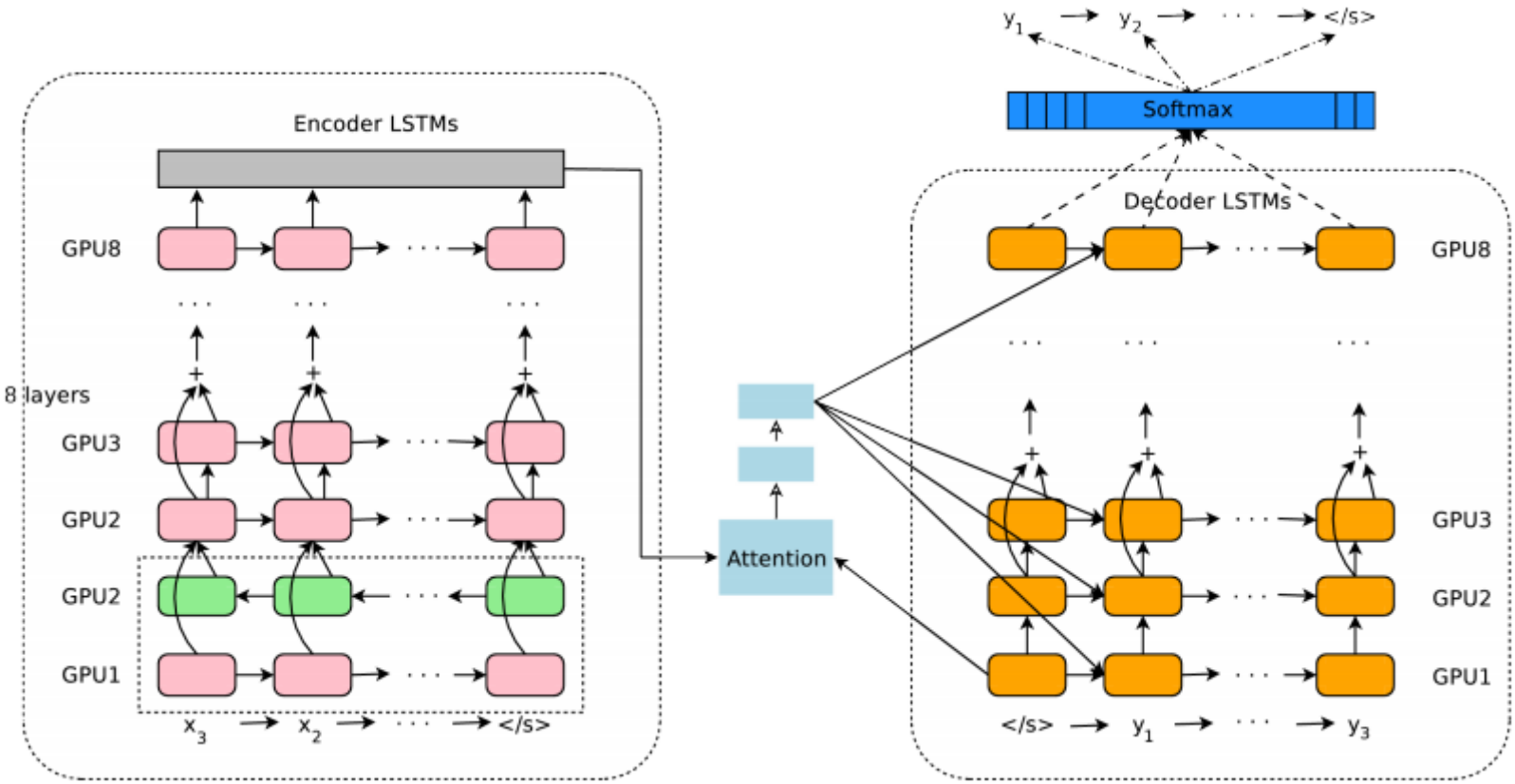
Bidirectional LSTM

RNN Encoder-decoder with Attention



For more information, visit https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html

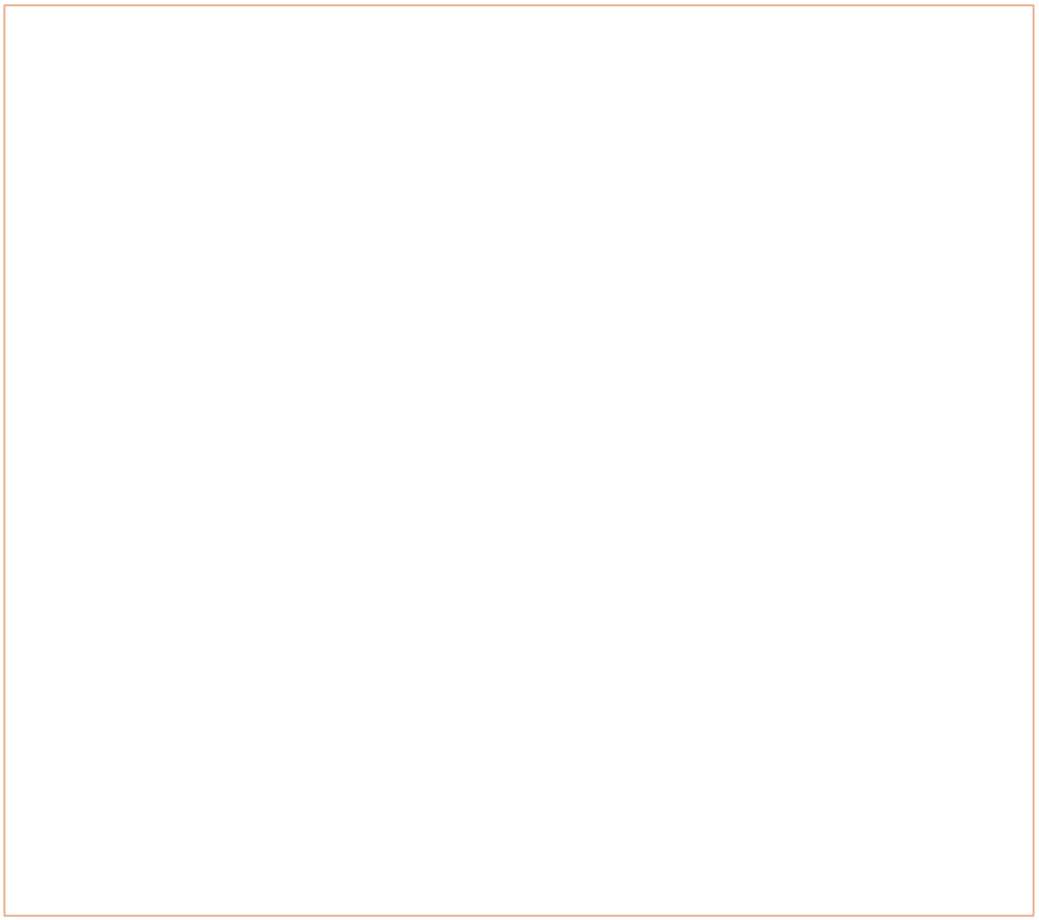
RNN Encoder-decoder with Attention



For more information, visit
https://arxiv.org/pdf/1609.08144.pdf

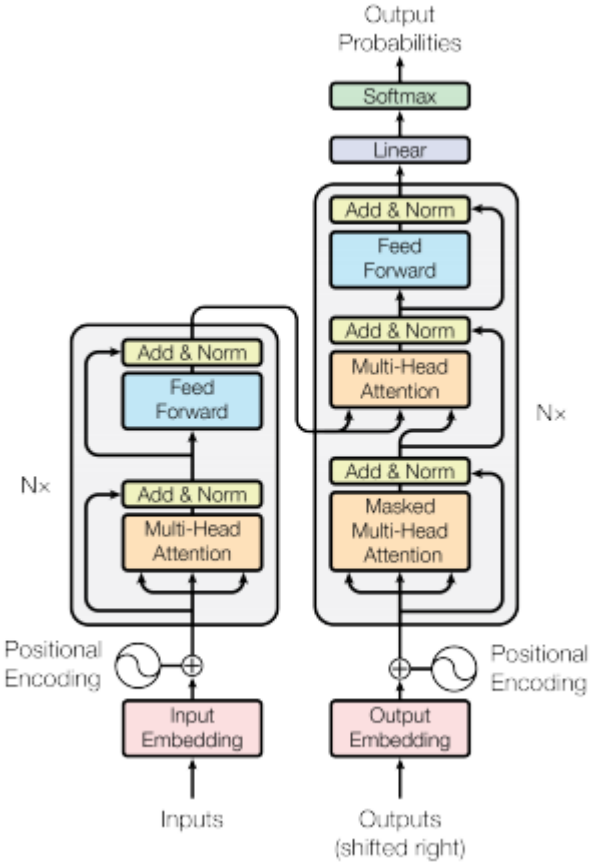
Self-attentional Transformer

Self-attentional Transformer



For more information, visit
https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html

Self-attentional Transformer



For more information, visit
https://arxiv.org/pdf/1706.03762.pdf

Evaluation

Automatic Evaluation

The use of Bilingual Evaluation Understudy (BLEU) metric (Papineni, Roukos, Ward, & Zhu, 2002), a measure of n-gram precision with a value between 0 and 1, comparing a machine translation output with a reference:

$$BLEU = \min(1, e^{1-\frac{r}{c}}) \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$