



# 外文系轉職 AI 工程師之路

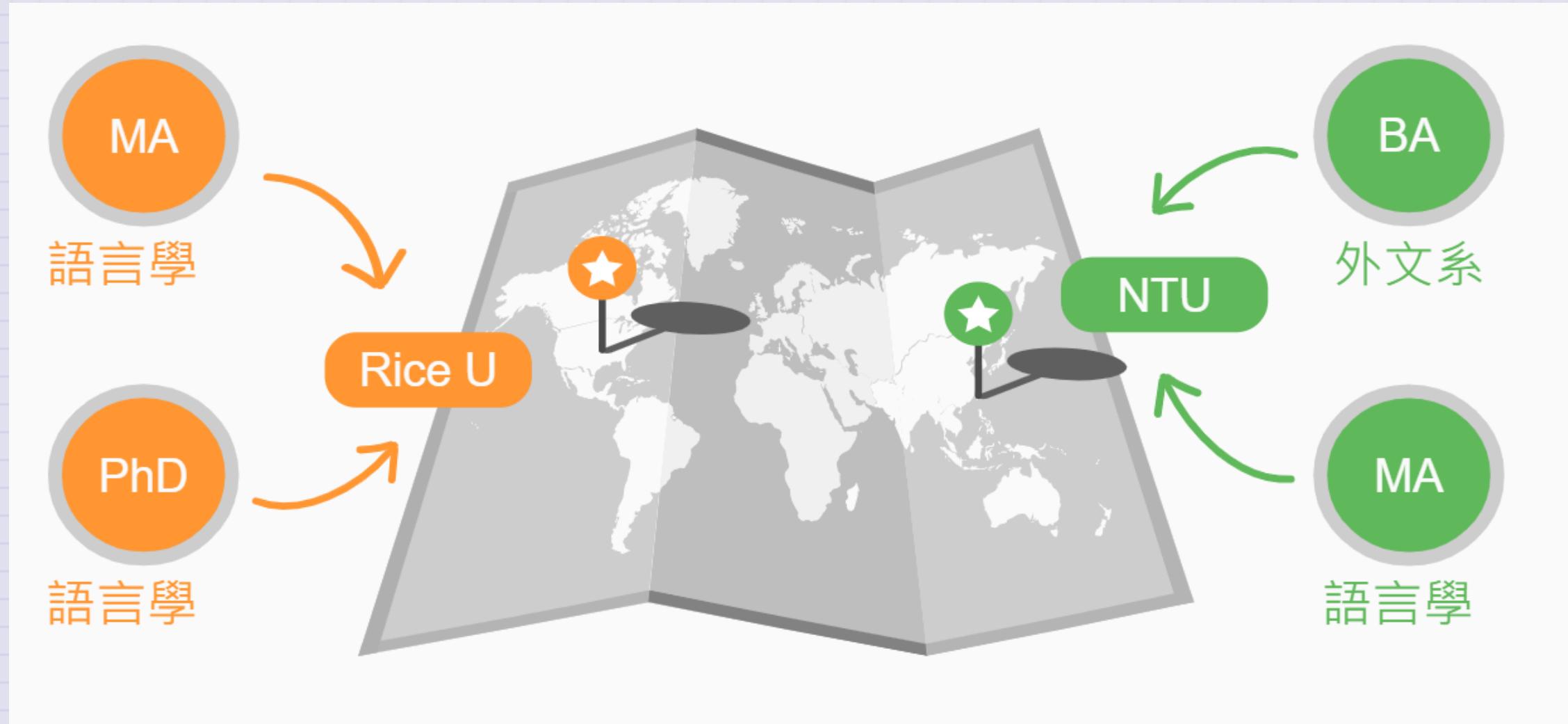
- 講者：[江豪文博士](#)
- 日期：2024-09-11

# 大綱

- 教育背景
- 工作經歷
- 自學歷程
- 問答時間

# 教育背景





# English Major

外文系  
做什麼



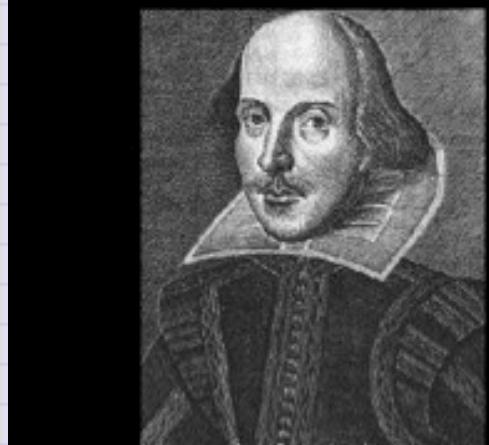
What my friends think I do



What my parents think I do



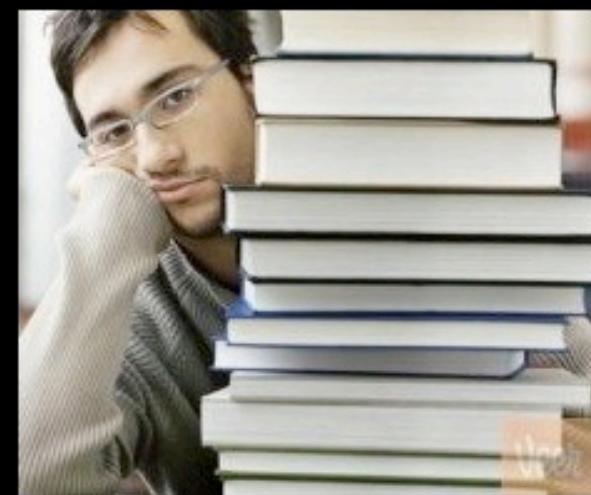
What my professors think I do



What society thinks I do



What I think I do



What I actually do

# Linguistics Student

語言學  
學什麼

外文系轉職AI工程師之路



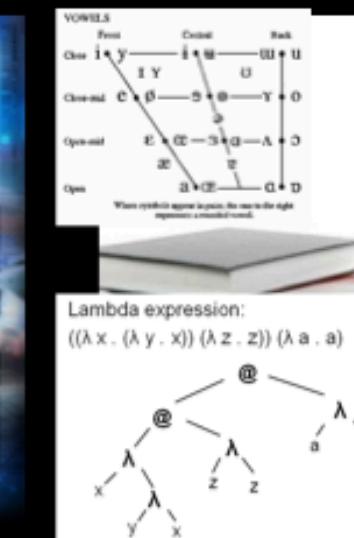
What society thinks I do.



What I think I do.



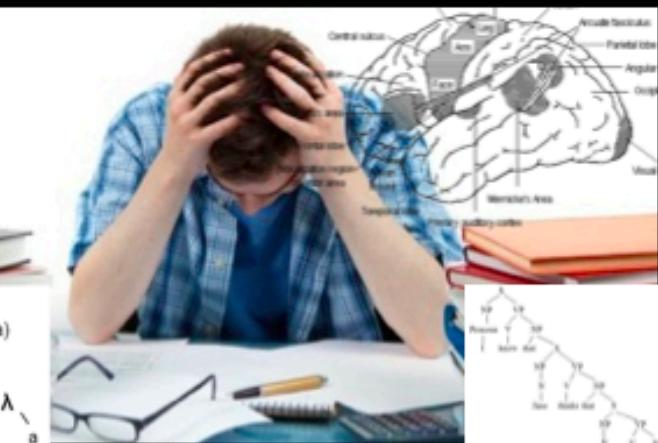
What my friends think I do.



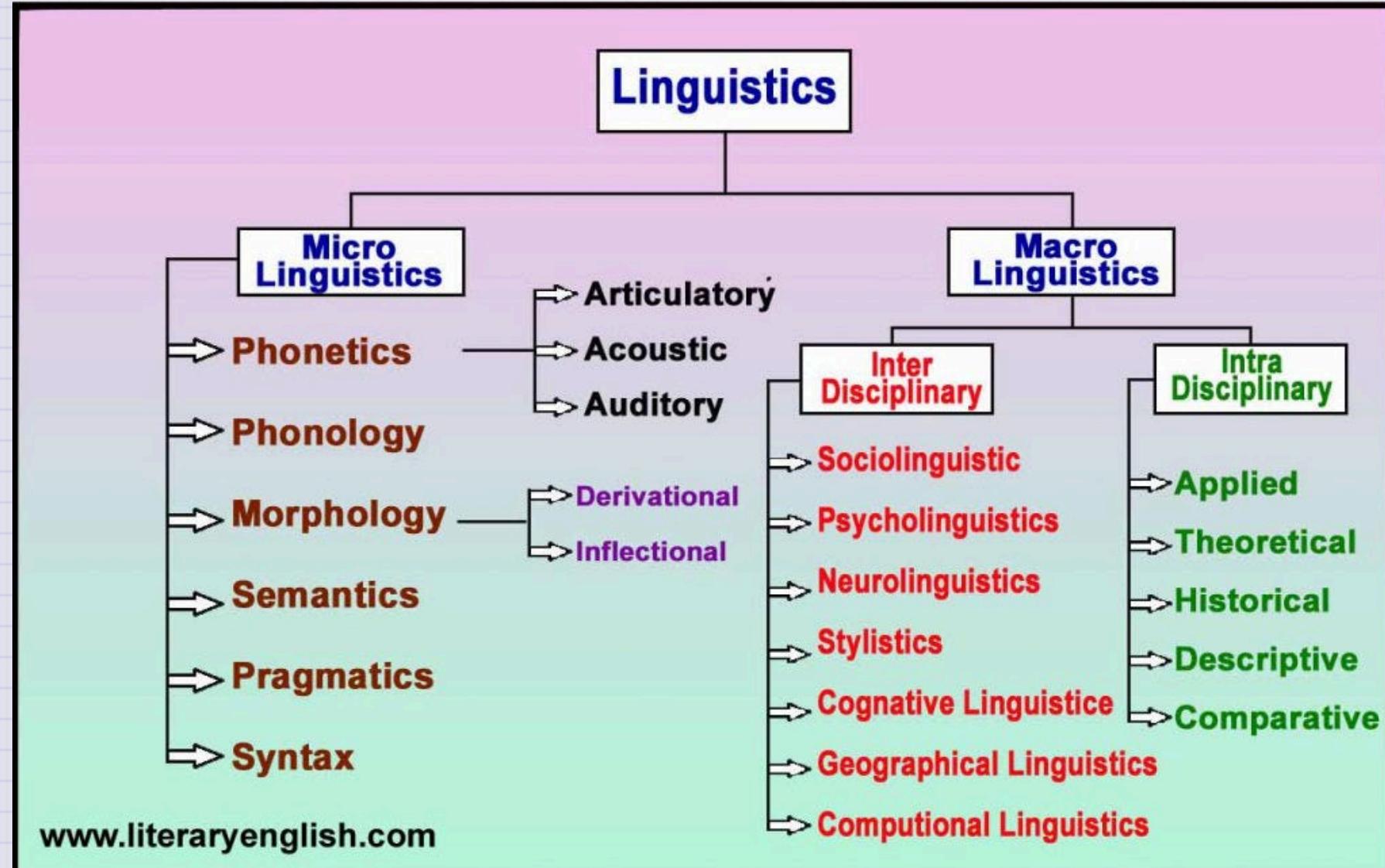
What I really do!



What my parents think I do.



# 原來語言學 不是學語言



# 我以為我會 研究的語言

## 外文系轉職AI工程師之路



但我實際上  
研究的語言

外文系轉職AI工程師之路



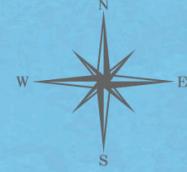


# 台灣南島語

## 臺灣原住民族 16族分布圖

臺灣原住民族文化園區

Namasan Kakivaualjan tua Kakudakudan nua Kacalisian i Taivan

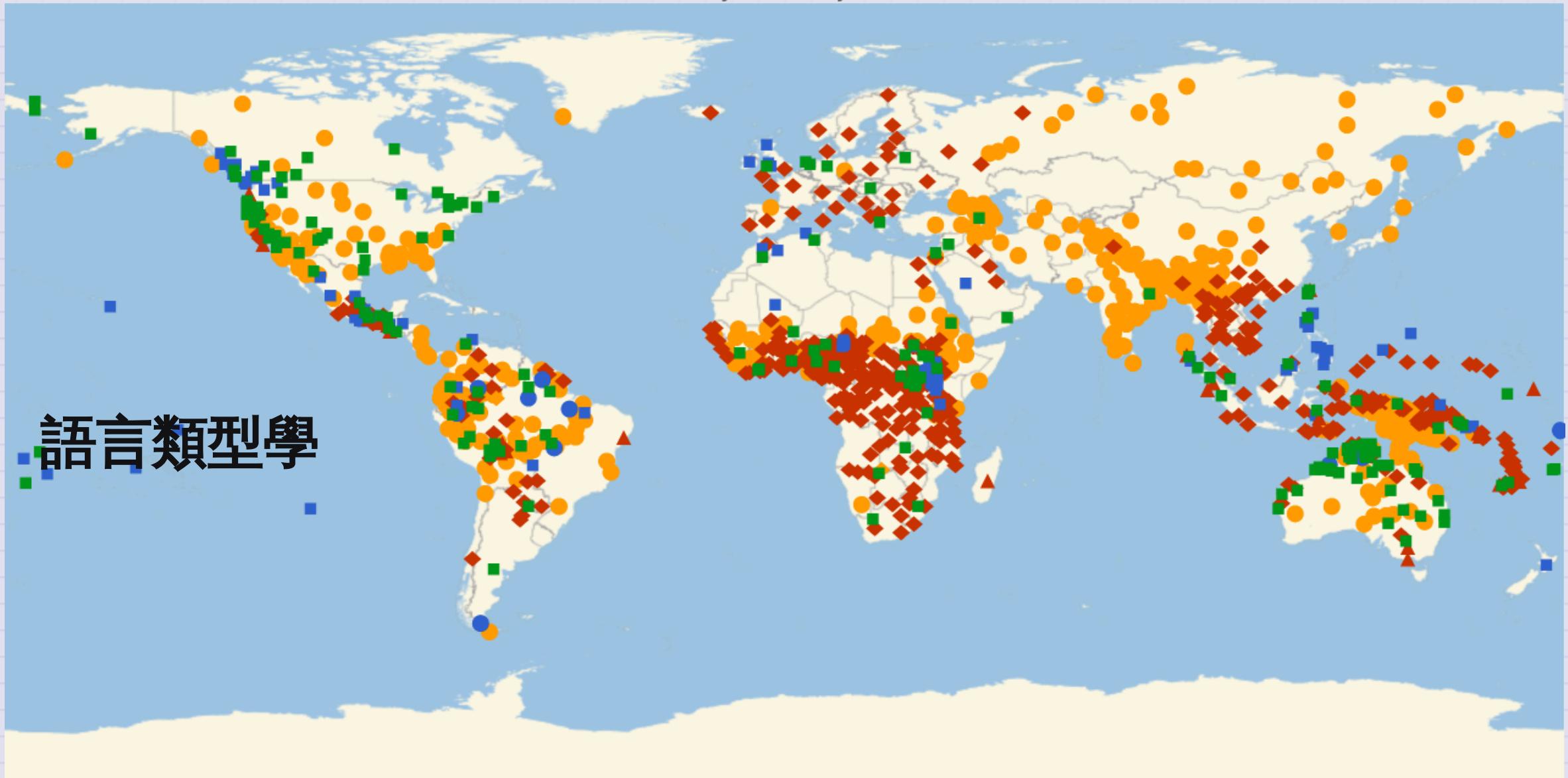


## 學位論文

- 碩士: 噶瑪  
蘭語空間認  
知之研究
- 博士: 台灣  
南島語的名  
物化與領屬  
結構



# Order of Subject, Object and Verb



● 1 SOV   ◆ 2 SVO   ■ 3 VSO   ▲ 4 VOS   ● 5 OVS  
◆ 6 OSV   ■ 7 No dominant order

- 名物化論文 1/2

Jiang, Haowen. 2021a.

Argument nominalization in  
Formosan languages: A  
functional-typological  
approach. Osaka: Osaka  
University Press.

[>> Amazon](#)

Nominalization Theory  
and  
Linguistic Analysis

Sung-Yeo Chung  
Masayoshi Shibatani

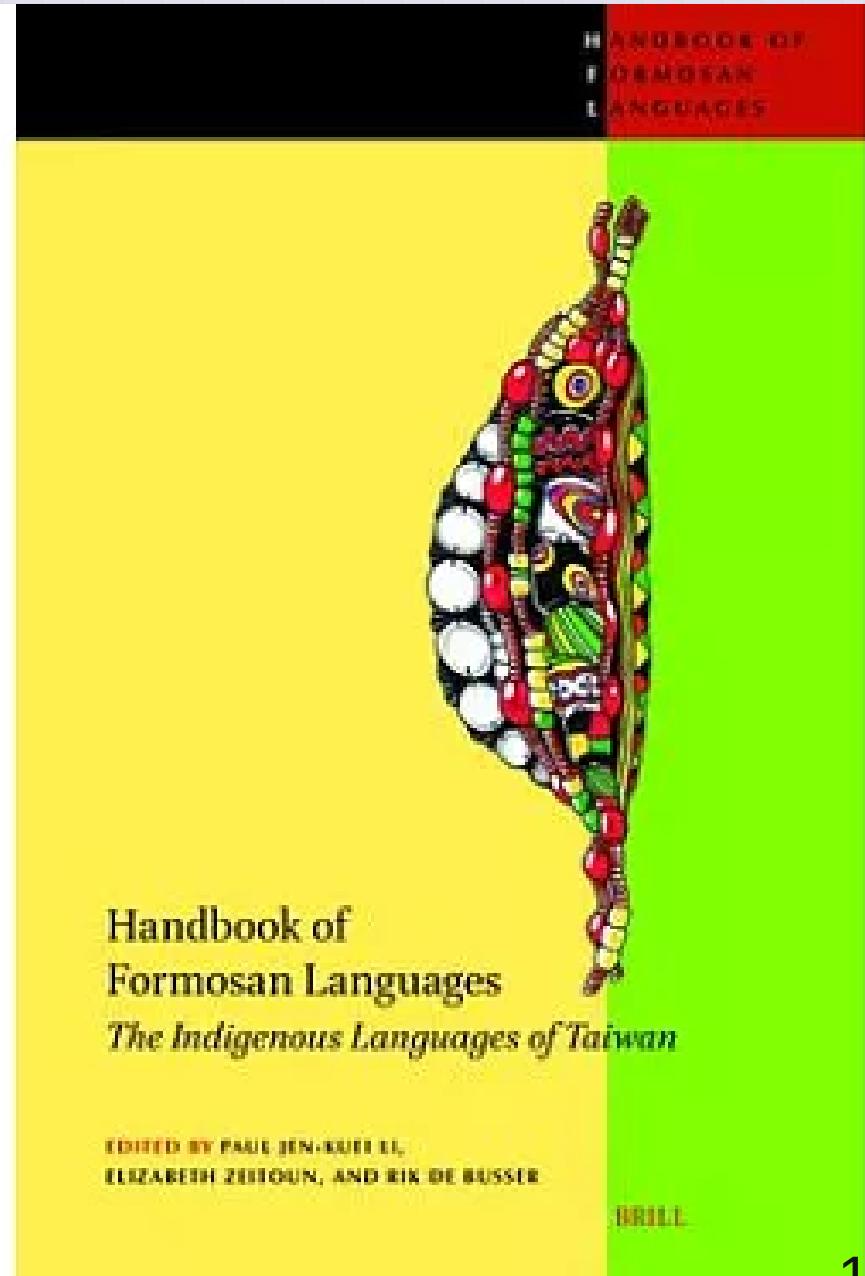
体言化理論と  
言語分析

鄭聖汝  
柴谷方良  
編

- 名物化論文 2/2

Jiang, Haowen. 2023b. 16.  
Nominalization in  
Formosan Languages.  
Leiden: Brill

[>> Brill](#)



## • 布農語線上詞典

The screenshot displays the homepage of the Indigenous Languages Online Dictionary. On the left, a vertical green sidebar features the title "原住民族語言線上辭典" and a yellow logo. The main content area shows circular icons for various languages: A美語, 泰雅語, 排灣語, 布農語, 卑南語, 魁凱語, 鄒語, 賽夏語, 雅美語, 邵語, 噶瑪蘭語, 太魯閣語, 撒奇萊雅語, 賽德克語, 拉阿魯哇語, and 卡那卡那富語. To the right, there's a navigation bar with links to "網站導覽", "關於本站", and "使用手冊下載". Below it is a "最新消息" section with a message about a recruitment drive for the online dictionary. Further down is a "每日一句" section with a daily quote in噶瑪蘭語. At the bottom right is a "跨族語檢索" search bar with fields for language selection and input, along with social media links for Facebook and eLearning.

## • 標記式語言 XML

```
<?xml version="1.0" encoding="UTF-8"?>
<interlinearGloss>
  <sentence>Я тебя люблю.</sentence>
  <gloss>
    <word>Я</word>
    <gloss>I</gloss>
    <translation>I</translation>
  </gloss>
  <gloss>
    <word>тебя</word>
    <gloss>you</gloss>
    <translation>you</translation>
  </gloss>
  <gloss>
    <word>люблю</word>
    <gloss>love</gloss>
    <translation>love</translation>
  </gloss>
  <freeTranslation>I love you.</freeTranslation>
</interlinearGloss>
```

 檢索

詞項列表

範疇分類

逐詞註解

首頁 / 布農語 /  檢索

# Mazima

詞頻： (1118)

收錄來源：九階教材

## 解釋1 喜歡

- Mazima saikin saicia. 

我喜歡牠。

- 
- Mazima saikin maun mas haisingdulaz. 

我喜歡吃糯米蒸筒飯。

• 通過布農語認證考試 >> 台灣立報

## 10月拚認證 找回母語記憶

記者：呂淑姮台北報導

地點：全臺 全部

「達悟」在達悟族語裡面，代表「人」的意思，指的是在蘭嶼的達悟族人。手上掛著象徵「船之眼」串珠的達悟族人廖清福說，他從15歲離開蘭嶼漁人部落，到台東念書，之後就很少再講族語。

今年53歲的廖清福，高中開始到台東求學，到外地之後沒有機會和環境使用自己的達悟母語，雖然小時候會講，但是當時碰到禁說方言政策，讓族語的回憶漸漸有些模糊。

近幾年來原住民族語復振，廖清福每到飛魚季，一定回到蘭嶼老家，向80多歲的母親認真學習母語，要把達悟族語再找回來，後來成為97年達悟族語認證考試榜首。現在，他也將族語傳承給自己的孩子：「希望他們知道自己的來處。」

行政院原民會18日舉辦「由我們發聲！母語是我們的身分證」記者會。副主委林江義宣布，99年度原住民族語言能力認證考試將於10月30日舉辦，報名日期自即日起到9月17日。

林江義說，今年是第8屆母語認證考試，和前7次不同的地方在於獎勵辦法。只要年滿70歲或者未滿15歲通過認證者、非原住民籍通過認證者、成績滿分、各族別榜首、6人以上團體報名，都可以獲得1千元獎金與表揚。

族語考試不限定年齡與族別，漢人也可以參加。一位曾在台大語言所擔任研究助理的學生江豪文，**通過布農族語認證**，他在台大時參與的研究和布農族語有關，覺得布農語很好聽，所以特別去學習，也通過了考試。

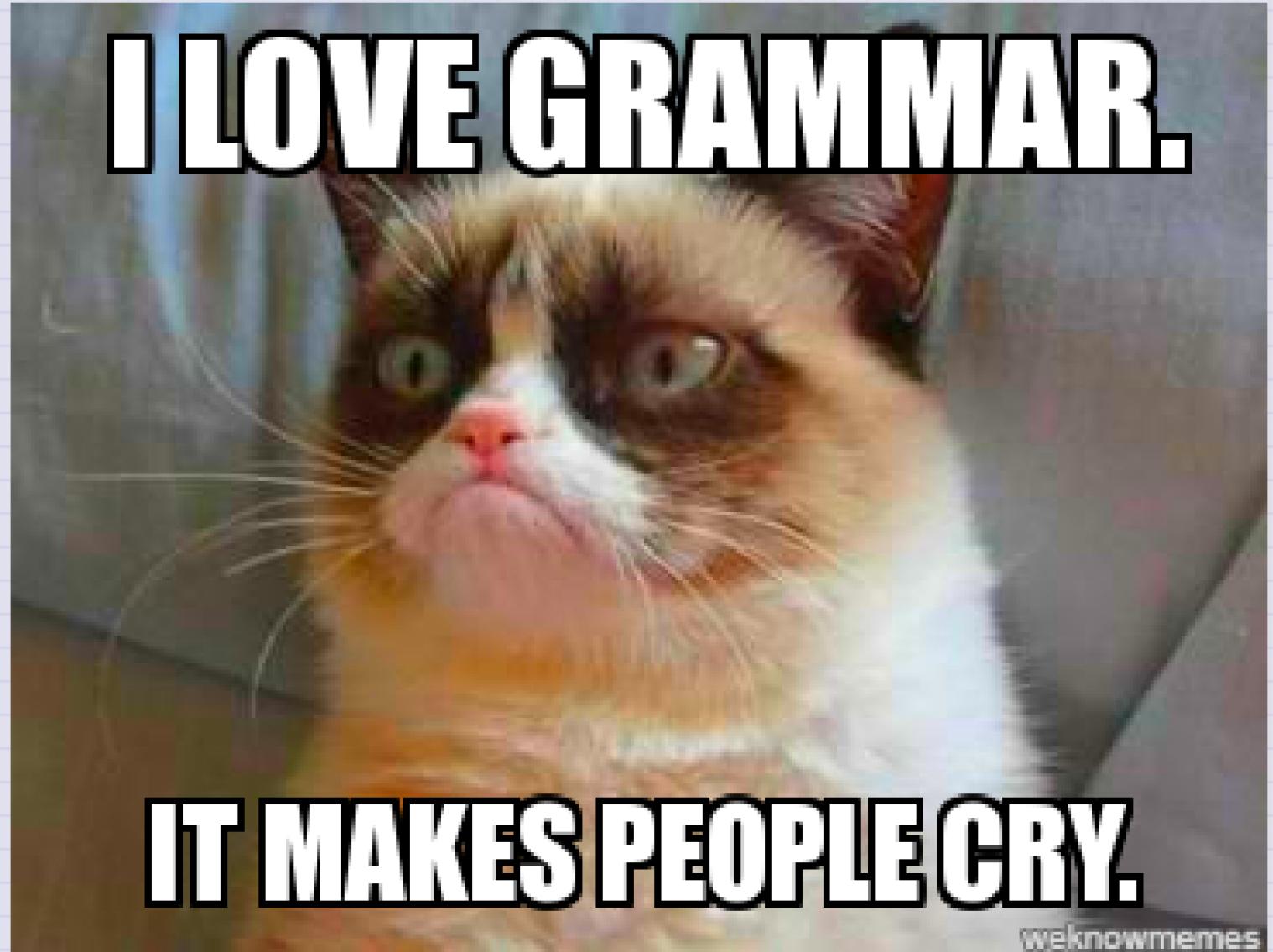
A photograph showing a person's arm and hand in a dark suit sleeve, holding a black fountain pen over a white clipboard. The clipboard has some illegible handwritten notes. In the background, another person's hand is visible, also in a suit, pointing towards the clipboard. The scene is set on a wooden desk with some papers and a gold paperclip nearby.

工作經歷



英文講師

臺北科技大學  
應用英文系



博雅博士後  
研究員

北京大學  
中文系

外文系轉職AI工程師之路

The screenshot shows the homepage of the Peking University Journals Platform. The header features the university's logo and name in Chinese and English, along with navigation links for Home, Journals, Articles, and About, and options for Login/Registration and language selection. A large banner image of autumn foliage and a traditional building is prominent. Below the banner, the title 'Essays on linguistics' is displayed. A search bar allows users to search by Title or enter a search term. Below the search bar are links for Journal homepage, Introduction, Instructions for Authors, and Back Issues. The main content area displays an article titled 'Person Systems in Puyuma, an Austronesian Language of Taiwan'. The article details include the author(s) (江豪文), affiliation (北京大学中国语言学研究中心/北京大学中文系 100871), and citation information. To the right of the article summary are download and sharing options, including a PDF link (1712 KB), a 'Read' button (285 reads), and an 'add' button. A 'Related articles' section is also visible.

北京大学 期刊网  
Peking University Journals Platform

Home Journals Articles About  
Login | Registration | 中文(简体)

Essays on linguistics

Title Please enter search term Search in journals Advanced

Journal homepage Introduction Instructions for Authors Back Issues

Essays on linguistics > 2023 , (04) : 20-40.

卑南语的人称标记

Person Systems in Puyuma, an Austronesian Language of Taiwan

Author(s) : 江豪文<sup>1</sup>

Affiliation(s) : 1 北京大学中国语言学研究中心/北京大学中文系 100871

Cite this article

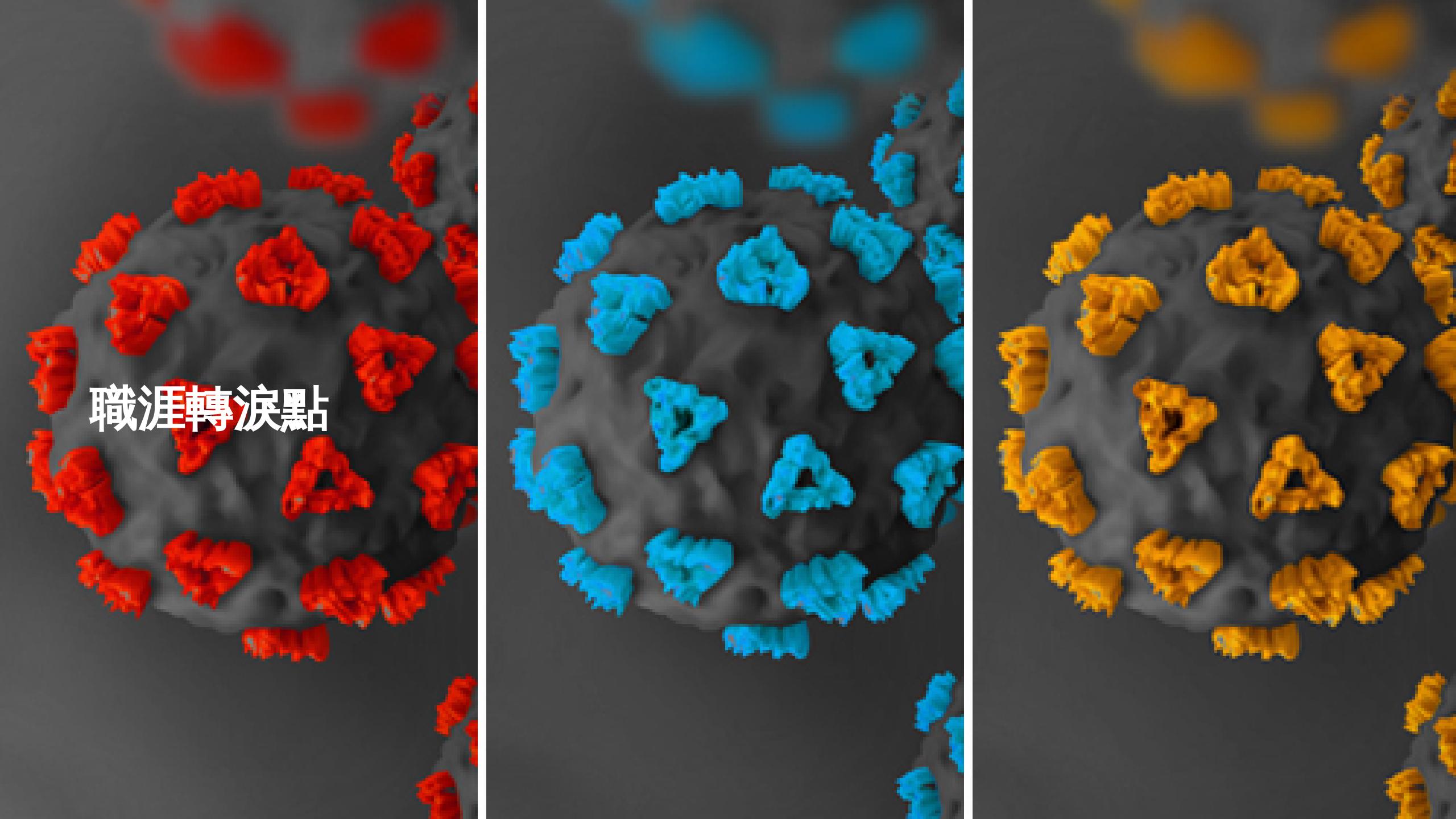
PDF(1712 KB)

285 Read add

Related articles

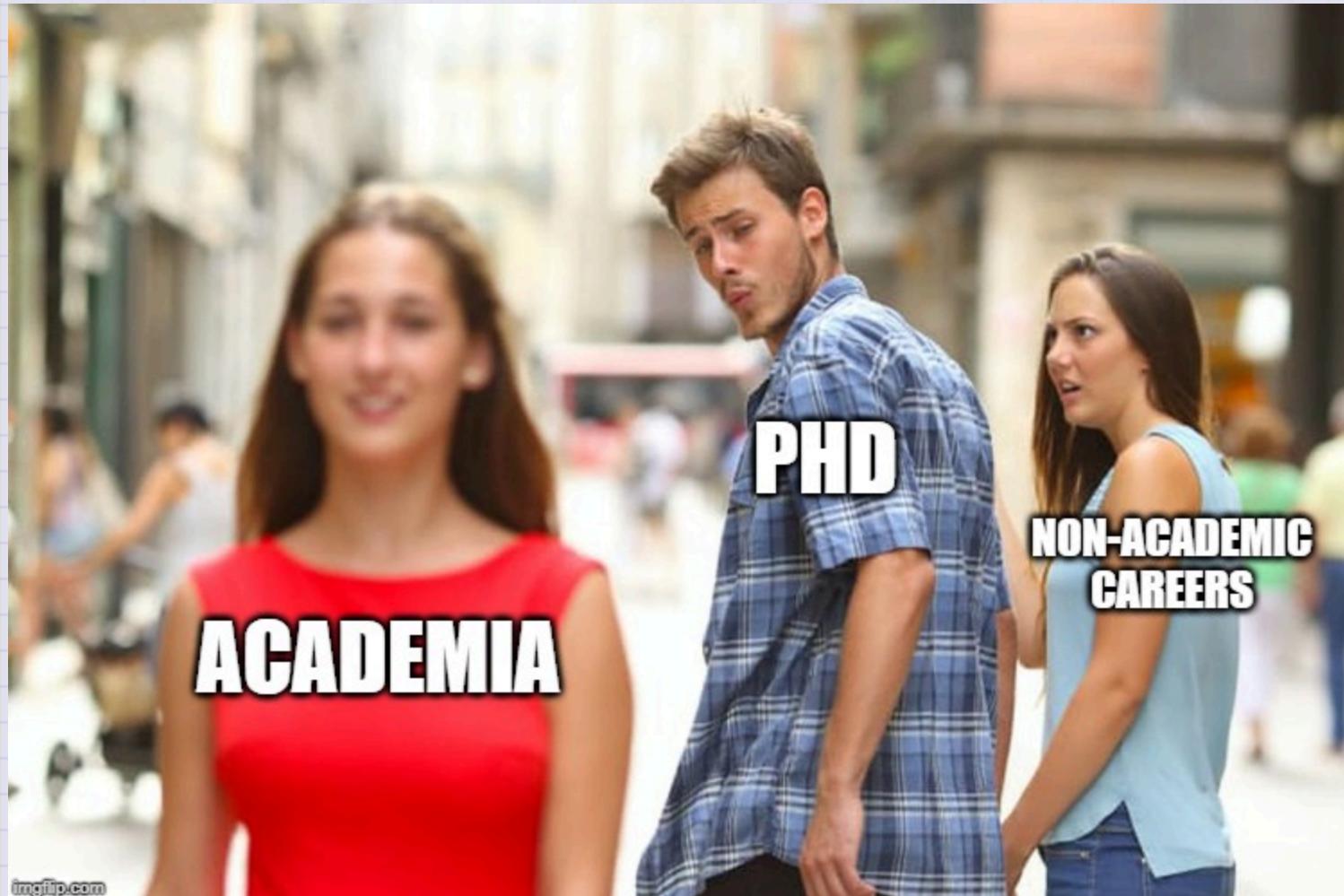
[1] 吴语人称代词复数标记来源的类型学考察

[2] 指人名词人称分裂的语言共性研究

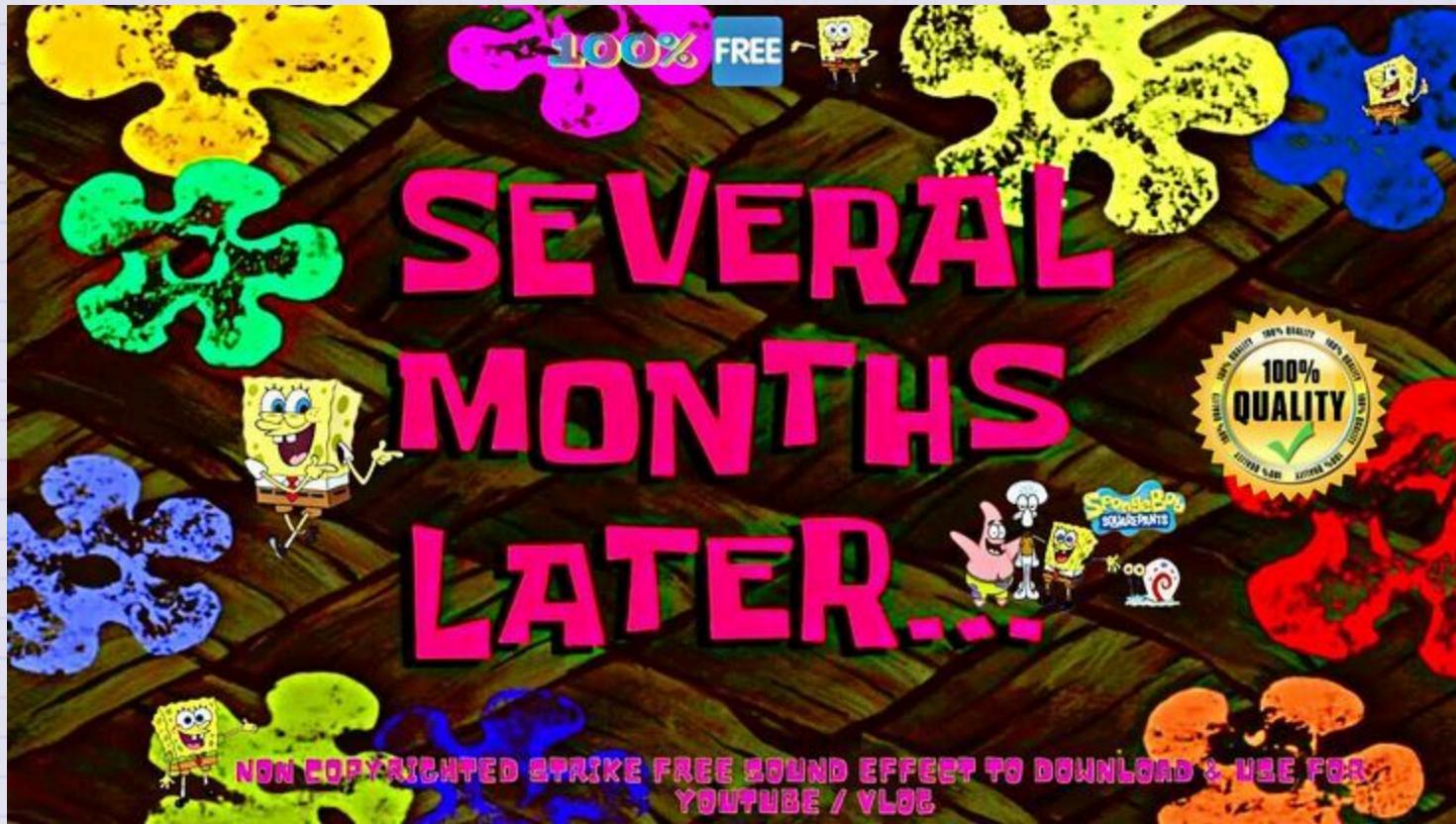


職涯轉捩點

- 心態



- 自學





AI 工程師

哈瑪星科技  
創新研發中心

- 業界初體驗



哈瑪星科技  
HAMASTAR TECHNOLOGY CO., LTD.



AI 工程師

香港商慧科訊業  
AI Lab

• 跨區合作

成立于 **2014** 年  
在香港、台北、北京设有研发中心

香港      台北      北京

The three icons represent the company's research centers: Hong Kong (a stylized skyscraper), Taipei (a traditional pagoda), and Beijing (the Tiananmen Gate Tower).

2016年荣获香港政府 **853** 万创新科技基金



2018年荣获国际自然语言处理与中文计算大会“单文本自动摘要”技术评测 **第一名**



**10+** 项国际发明专利



广东省智能创新协会 **理事单位**

数据分析  
襄理

新光人壽  
數位資訊部

- 資安法規



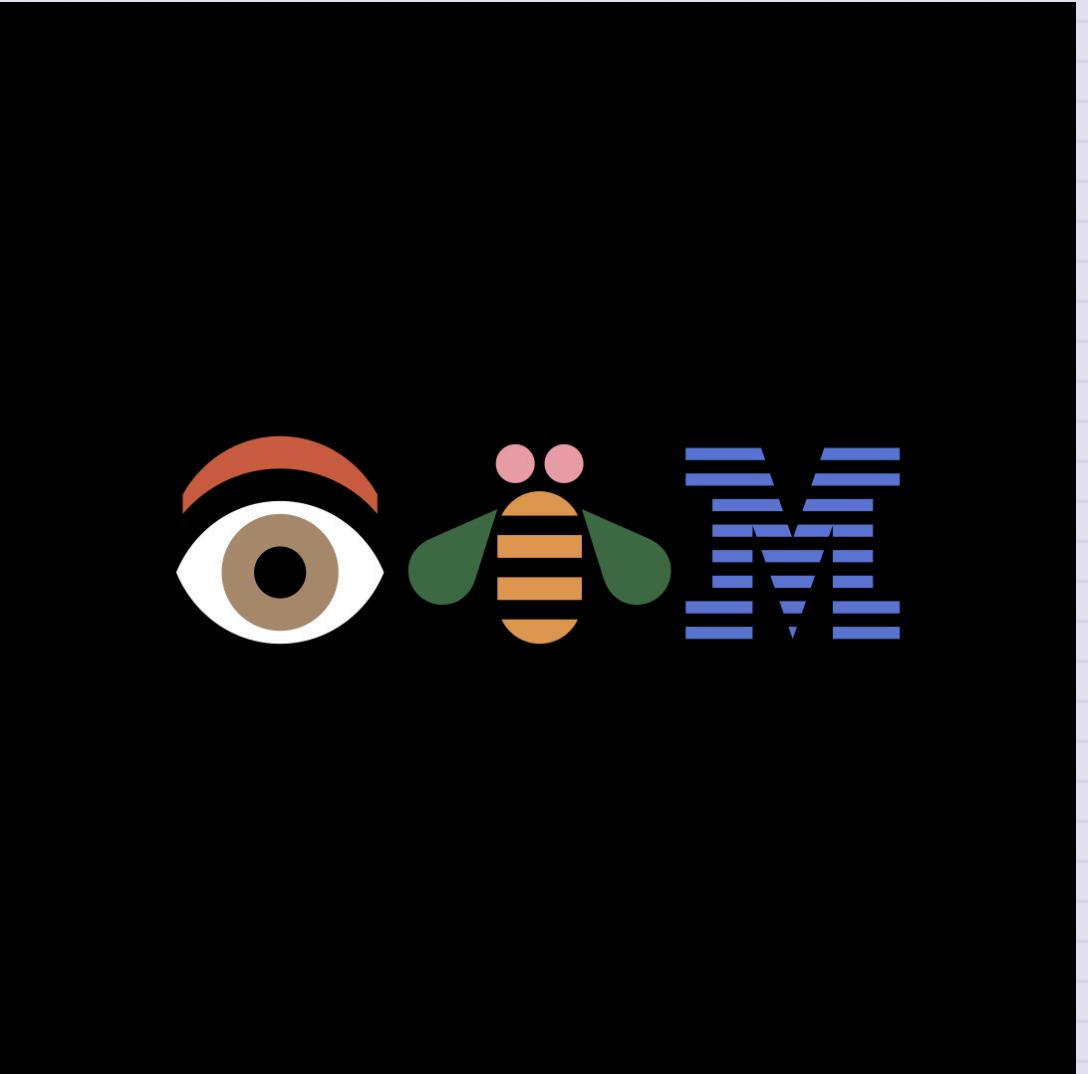
新光金控

新光人壽  
Shin Kong Life

AI 暨資料解決  
方案技術工程師

IBM  
科技事業部

- 美商文化



# 自學歷程



## 學習進程

01 Python 程式語言

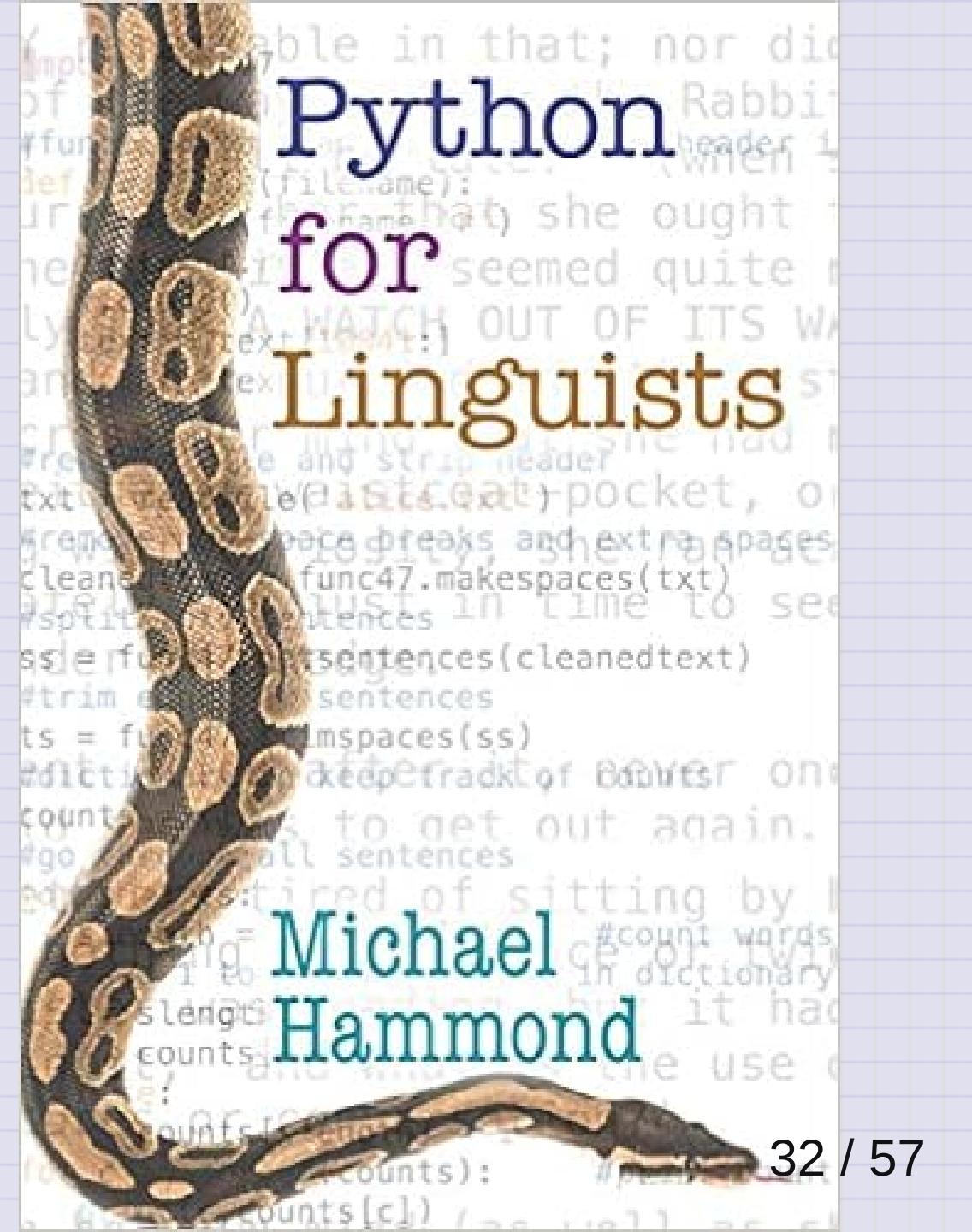
02 網路爬蟲

03 自然語言處理

04 機器學習

# Python 是蟒蛇也是語言

## 01 Python 程式語言



## 程式語言相比自然語言的優勢

- 沒有人是母語者
- 不需練聽說，只需讀寫

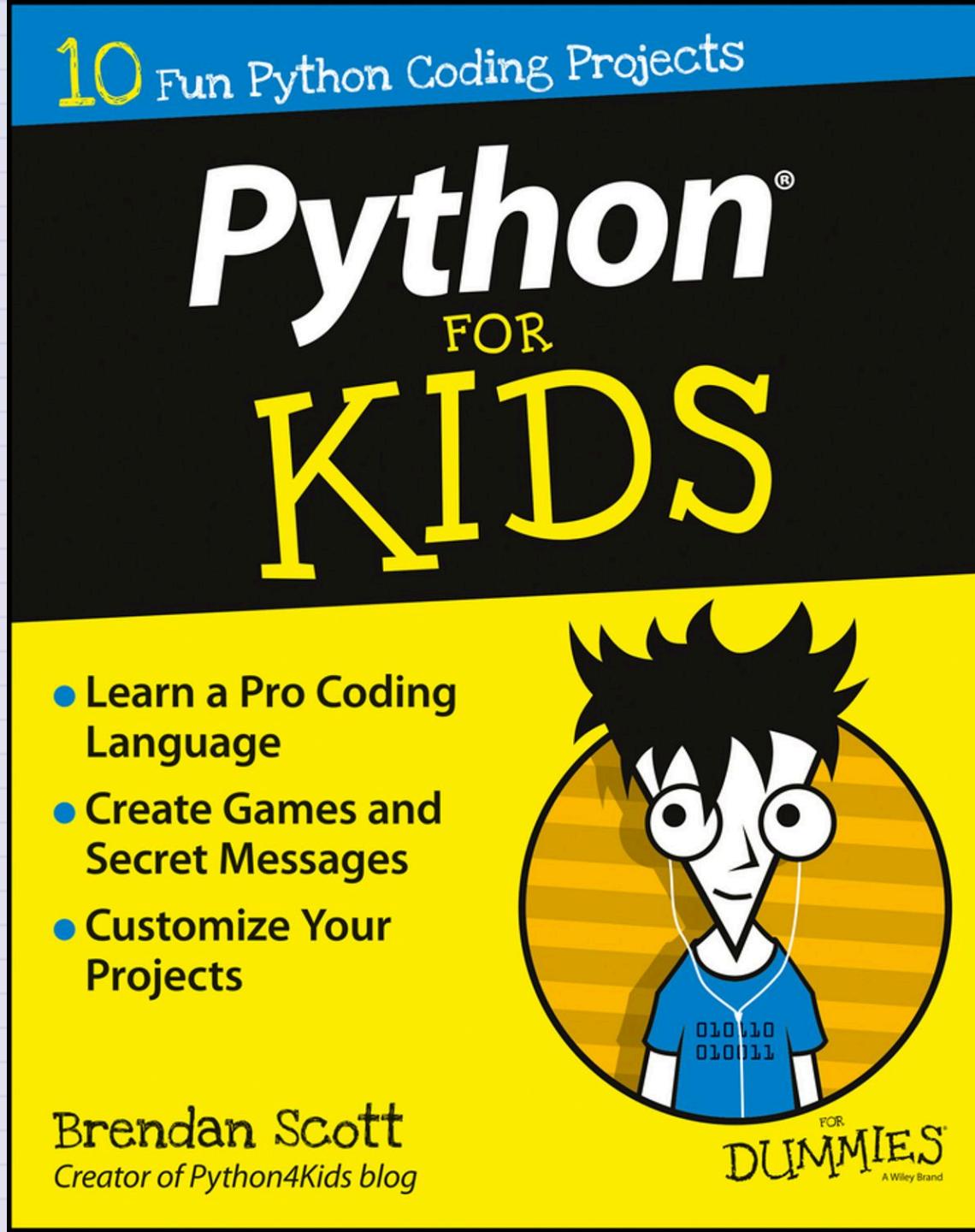
"the best way to learn a language  
is to speak to natives"  
the guy learning Python:



# Python 相比其他程式語言的優勢

Top Paying and Most Popular Programming Languages in 2020

Rank by Average Salary		Rank by Volume of Job Openings			
1.	Python	\$119,000	1.	Python	50,000
2.	JavaScript	\$117,000	2.	SQL	50,000
3.	Java	\$104,000	3.	Java	45,000
4.	C	\$103,000	4.	JavaScript	38,000
5.	C++	\$102,000	5.	C++	29,000
6.	C#	\$97,000	6.	C#	21,000
7.	PHP	\$94,000	7.	PHP	13,000
8.	SQL	\$92,000	8.	C	9,000



連小朋友都學得會

Brendan Scott  
*Creator of Python4Kids blog*

外文系轉職AI工程師之路

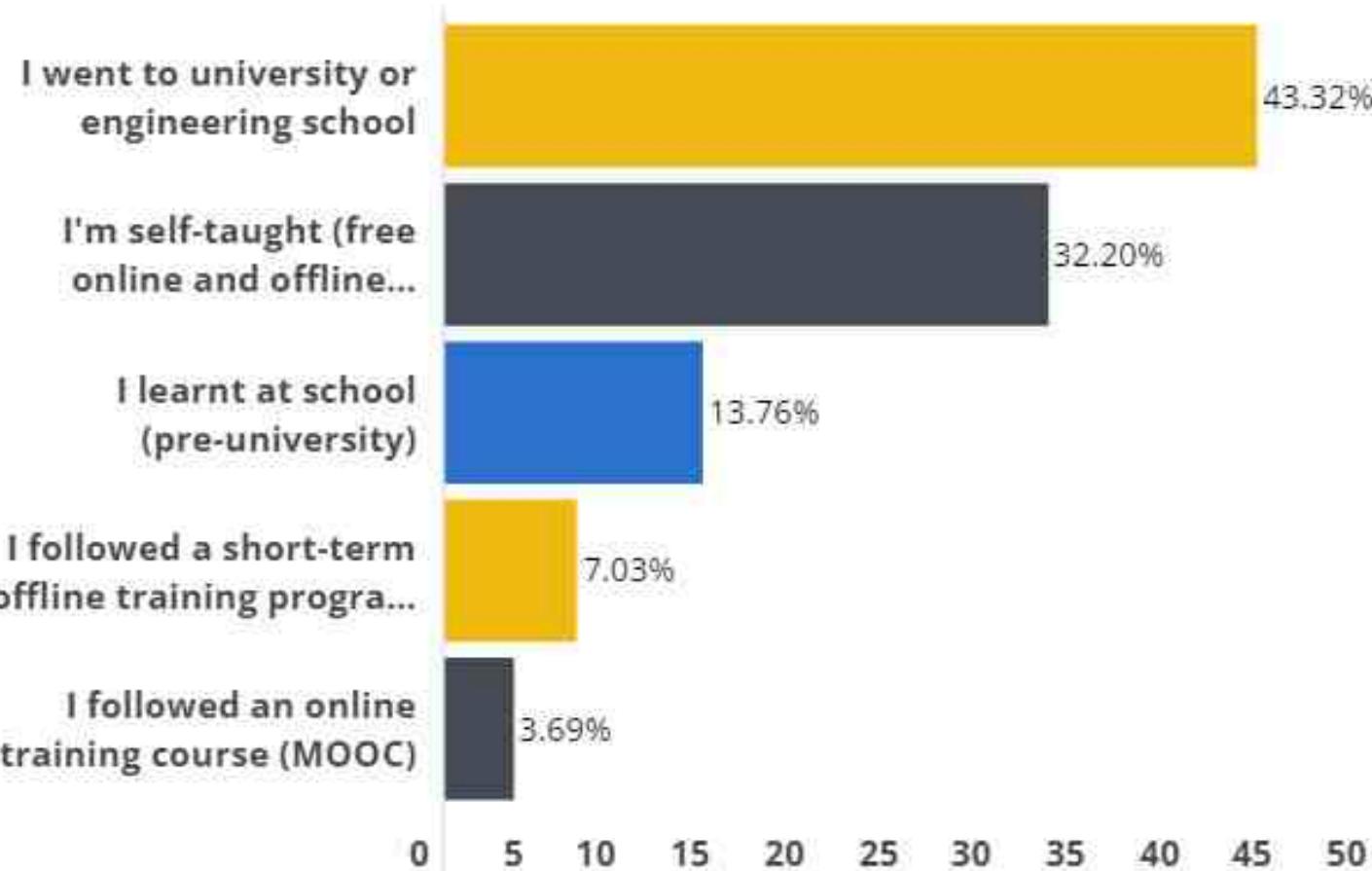
35 / 57

<b>CEFR levels</b>	<b>Study hours</b>	<i>from</i>	<i>to</i>
<b>C2</b>		<b>1,000</b>	<b>1,200</b>
<b>C1</b>		<b>700</b>	<b>800</b>
<b>B2</b>		<b>500</b>	<b>600</b>
<b>B1</b>		<b>350</b>	<b>400</b>
<b>A2</b>		<b>180</b>	<b>200</b>
<b>A1</b>		<b>90</b>	<b>100</b>

The advertisement for the IntelliPaat Python Complete Course features a dark blue background with a grid pattern. At the top left is the IntelliPaat logo. To its right is a white circular icon containing a blue and yellow Python logo. Below these, the word "PYTHON" is written in large, bold, yellow capital letters. Underneath it, "COMPLETE COURSE" is written in white capital letters. A yellow banner across the middle contains the text "BEGINNER TO ADVANCE" in black and "7 HOURS+" in white. On the right side of the banner, there is a portrait of a young man with dark hair, wearing a black t-shirt with the IntelliPaat logo. The overall design is clean and professional.

高比例的軟體  
工程師都是  
自學的

## How did you learn computer programming?



Share

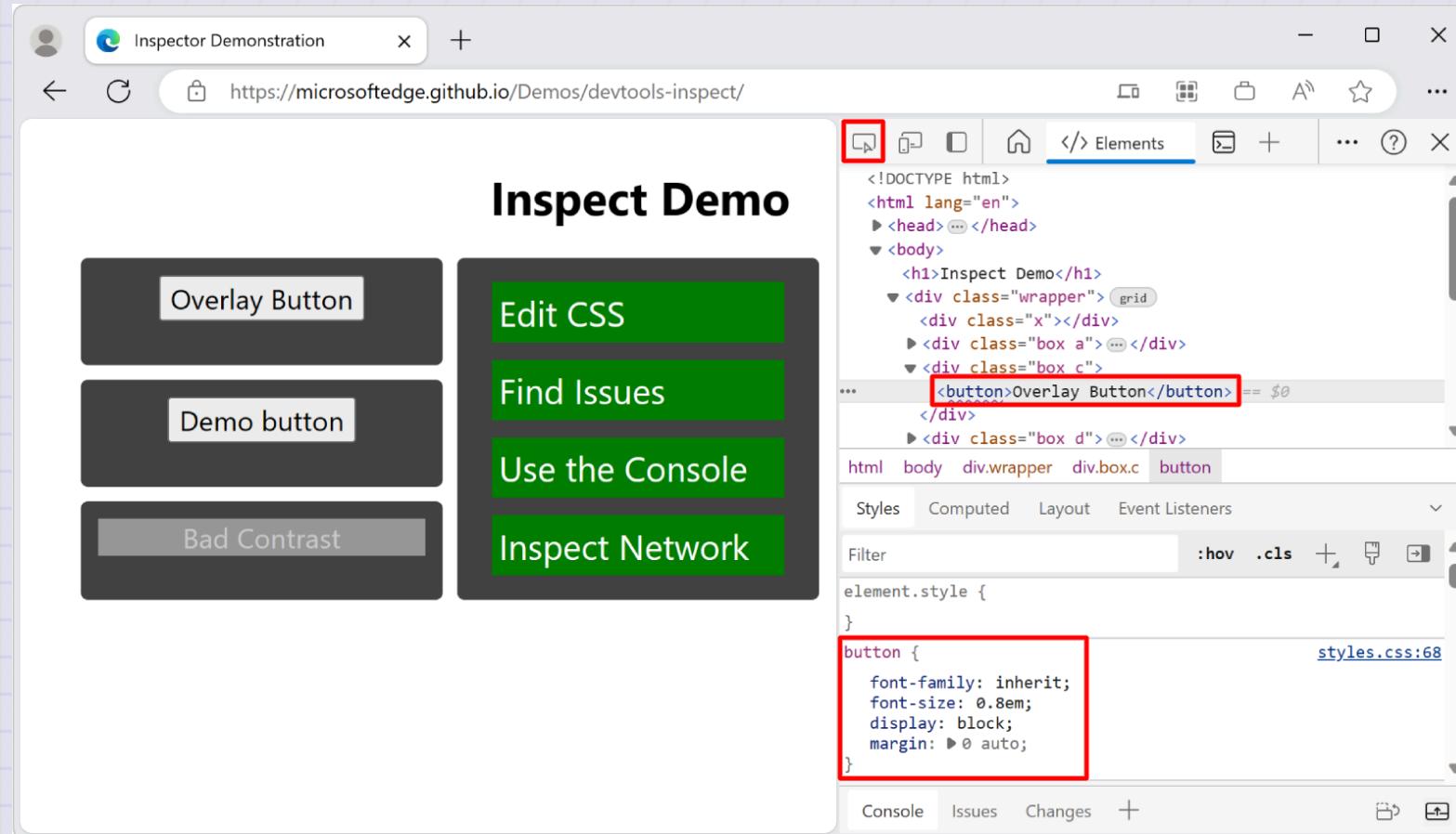
## 爬蟲能快速取得資料

### 02 網路爬蟲

Web Scraping Using Python



# 網路爬蟲 - 用 Python 解析 HTML



契機



>> 族語 E 樂園



代名詞  
指示詞



疑問詞



抽象名詞

2024-09-11@NSYSU

ezTravel 易遊網 找住宿 搜機票 瘋台灣 饗玩樂 玩世界 搭遊輪 企業服務 客製旅遊 | 追蹤 | 查訂單 | 登入

目的地或飯店名稱：高雄  
入住日期：2021年05月15日 週六 ~ 2晚  
退房日期：2021年05月17日 週一  
房間數及人數：1間房，2位成人，0位孩童

共找到 252 家高雄飯店

顯示地圖 易遊推薦 價格 好評優先 飯店星級 距離遠近

**自訂篩選**

**熱門篩選**

- 免費取消
- 即時確認
- 免費WiFi
- 免費早餐
- 很棒：評分4.7以上

**星級**

- 五星/豪華
- 四星/高級
- 三星/舒適

ZH_Name	EN_Name	Hotel_ID	City	Location	Address	Campaign	Mask	Hold_Room	Min_Price
高雄比歐緻居	Brio Hotel	22260305	高雄	None	中山一路 14-26號	None	None	2	2584
和逸飯店(高雄中山館)	Hotel Cozzi Zhongshan Kaohsiung	2985883	高雄	None	中山二路 260號30樓	None	None	19	2488
高雄喜達絲飯店		2328586	高雄	None	成功一路 286號	None	None	3	3306
高雄國賓大飯店	The Ambassador Hotel Kaohsiung	701744	高雄	None	民生二路 202號	None	None	9999	2097
高雄麗尊酒店	The Lees Hotel	701698	高雄	None	五福一路 105號	None	None	10	2208
高雄水京棧國際酒店	H2O Hotel	8194467	高雄	None	明華路 366號	None	None	1	4701

```
1 max_df.head()
```

# 爬蟲效果

# 台灣南島語爬蟲成果

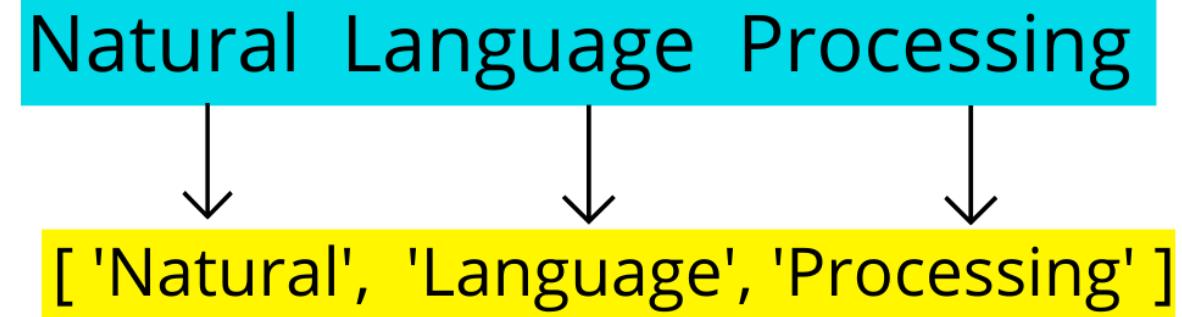
>> 台灣南島語-華語句庫資料集 APP

<b>Lang_En</b>	<b>Lang_Ch</b>	<b>Ab</b>	<b>Ch</b>	<b>From</b>
58650	Saaroa	拉阿魯哇 umacʉka maivʉvʉrʉnga tumuakia umarua sipaasʉrʉpa masʉrʉpʉ.	早上起床，都要用毛巾洗臉。	詞典
115515	Bunun	布農_郡群 Namaun a uvavazan mas hangvang tu balang.	孩子們要吃牛排。	生活會話
92691	Saisiyat	賽夏 rohaenan panrahi: ka rarahi: kayzaeh 'okay tango:i.	晚上掛蚊帳才不會被蚊子叮。	詞典
28392	Rukai	魯凱_霧台 malra patuelre kwini ya cukudhane kwi belebele.	那桌上的香蕉全部拿去。	詞典
117888	Paiwan	排灣_中 kasi nu sun a na kitulu tua pinaiwanan ?	你在哪裡學排灣話的？	九階教材
126445	Puyuma	卑南_南王 munuma tu ami kanu wadi na babayan?	你妹妹幾歲？	句型
6699	Amis	阿美_中部 O cisafakay ko felac no miso.	你的米堆摻有空殼米。	詞典
119373	Rukai	魯凱_東 kadro ka mabocoko si katha, takosiya sa'ararake .	不但吃得飽，並且都很營養。	九階教材
46746	Seediq 賽德克_德固達雅	Mkela melux ka tama mu ma; ini kela smberih.	我父親會對果樹及瓜果施咒；但不會解咒。	詞典
5051	Sakizaya 撒奇萊雅	maudad tu saupuen ku kasuy i salining.	下雨了把木材堆在走廊。	詞典

自然語言處理能透過程式  
處理大量文字

### 03 自然語言處理

#### Tokenization



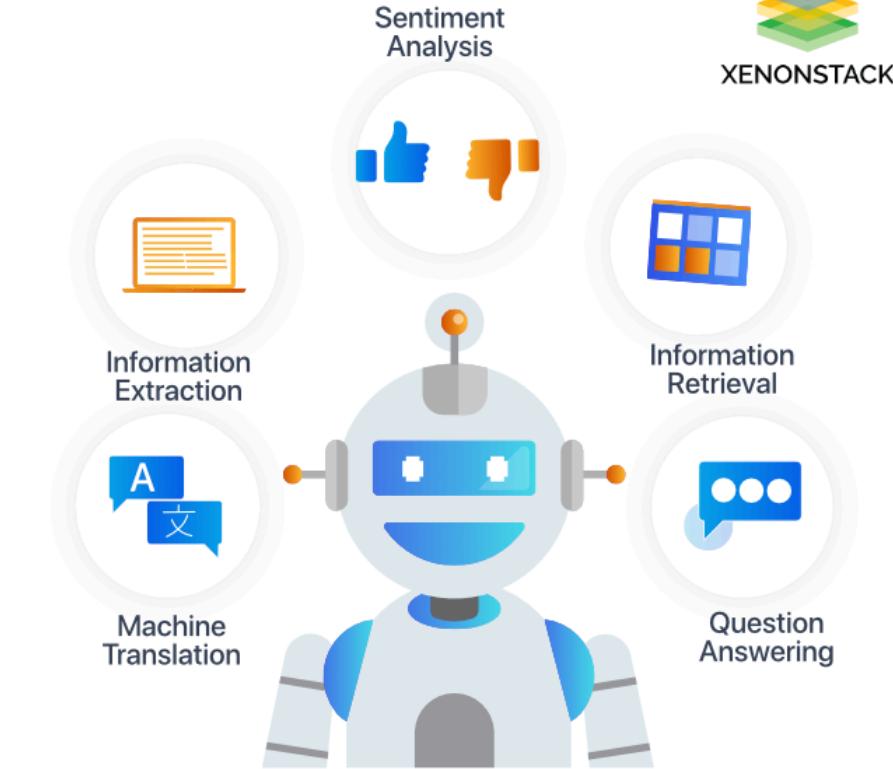
## 自然語言處理的定義

Natural language processing (NLP) is a subfield of **linguistics**, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to **program computers to process and analyze large amounts of natural language data**.

[>> Wikipedia](#)

# NLP 的應用

# Applications of **Natural Language Processing** in Different Domains



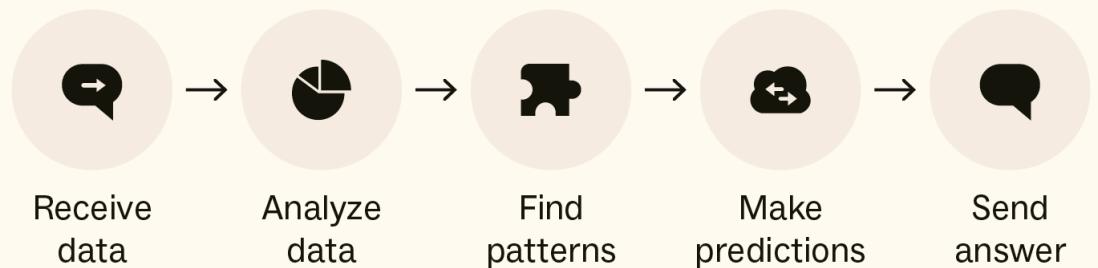
# NLP 的市值



# 機器學習能透過已知預測 未知

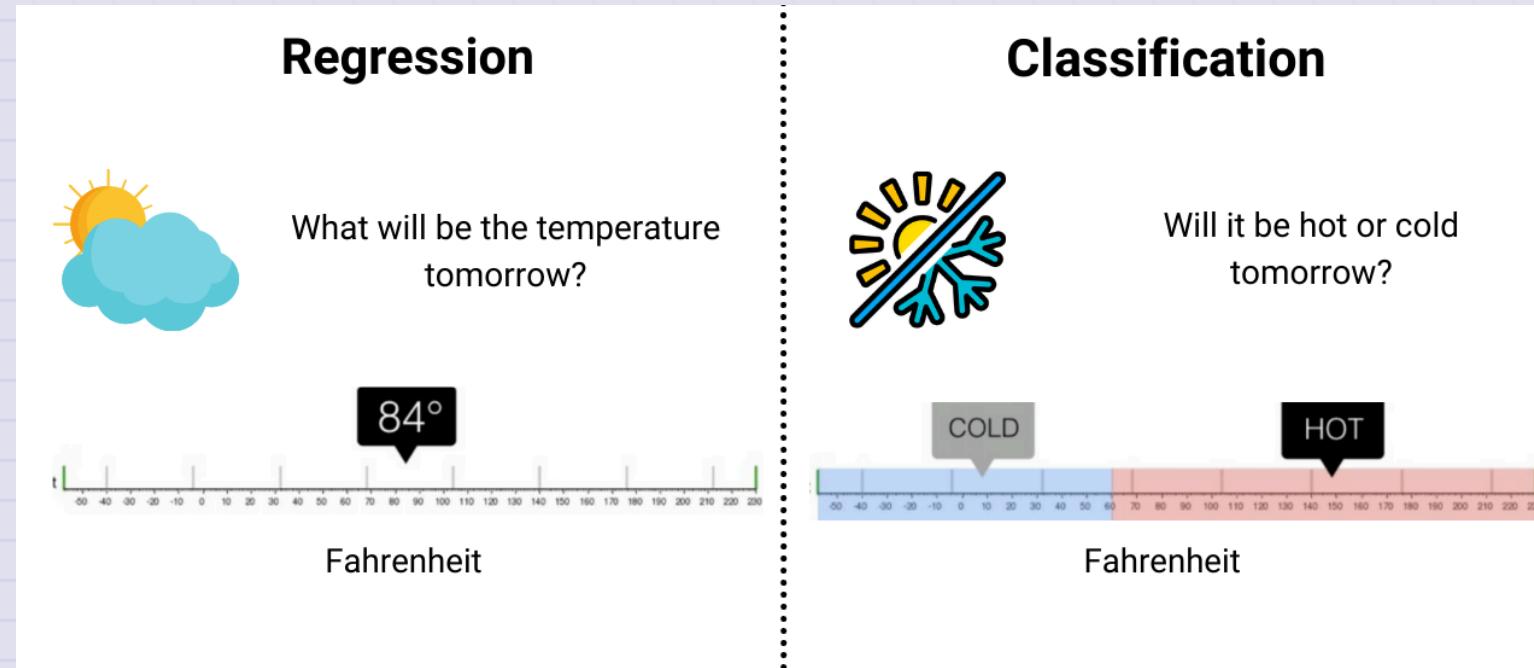
## 04 機器學習

### The machine learning process



## 兩類機器學習模型

- 回歸預測數值
- 分類預測類別



## 學習 4B

01Basics

02Badges

03Bootstrapping

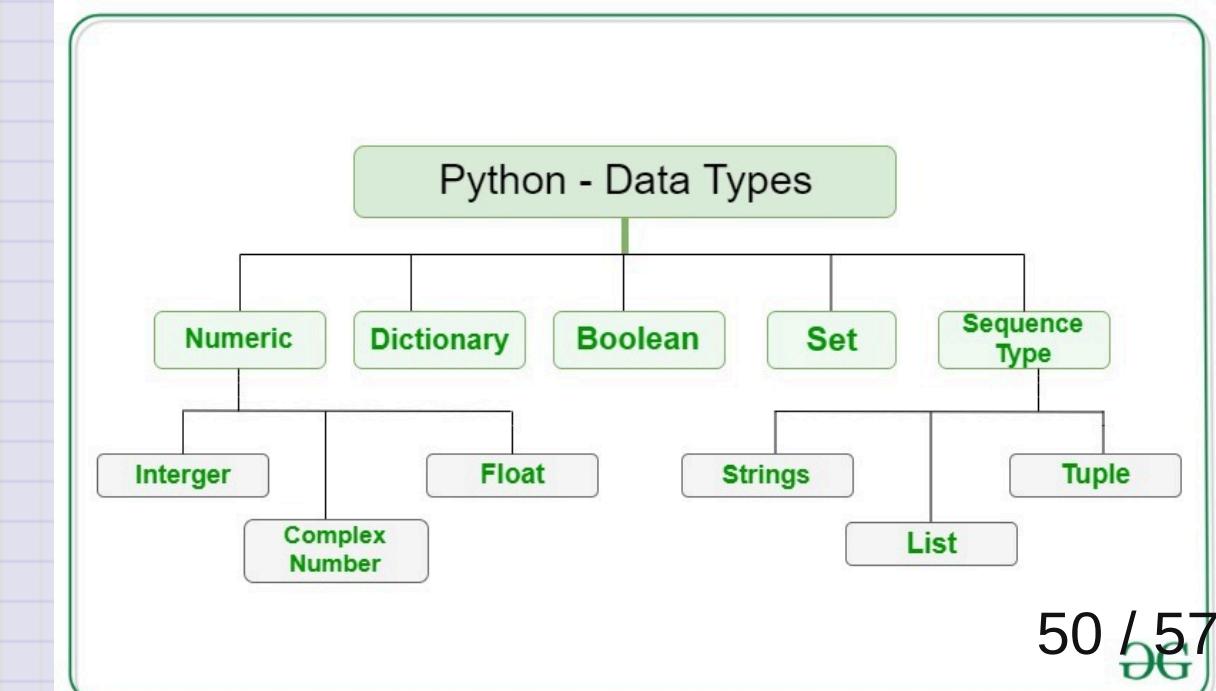
04Blogging

## 利用geeksforgeeks 一步一步練基本功

### 01Basics

# GeeksforGeeks

A computer science portal for geeks



生成式 AI  
帶來了許多  
重大改變

[>> source](#)

★ Member-only story

# English is the most powerful programming language — even for data science: Introduction to prompt engineering

What prompt engineering is, which are the steps involved in it, and how it changes the way we solve problems with ML.



Facundo Santiago · [Follow](#)

10 min read · May 9, 2023

利用cognitiveclass.ai  
獲取技術證照

## 02Badges



COGNITIVE  
CLASS.ai



# 利用GitHub 借力使力開發應用程式

## 03 Bootstrapping

- AI 模型輔助語言學習



Welcome to Streamlit!

Edit [/streamlit\\_app.py](#) to customize this app to your heart's desire ❤️

If you have any questions, checkout our [documentation](#) and [community forums](#).

In the meantime, below is an example of what you can do with just a few lines of code:

Number of points in spiral

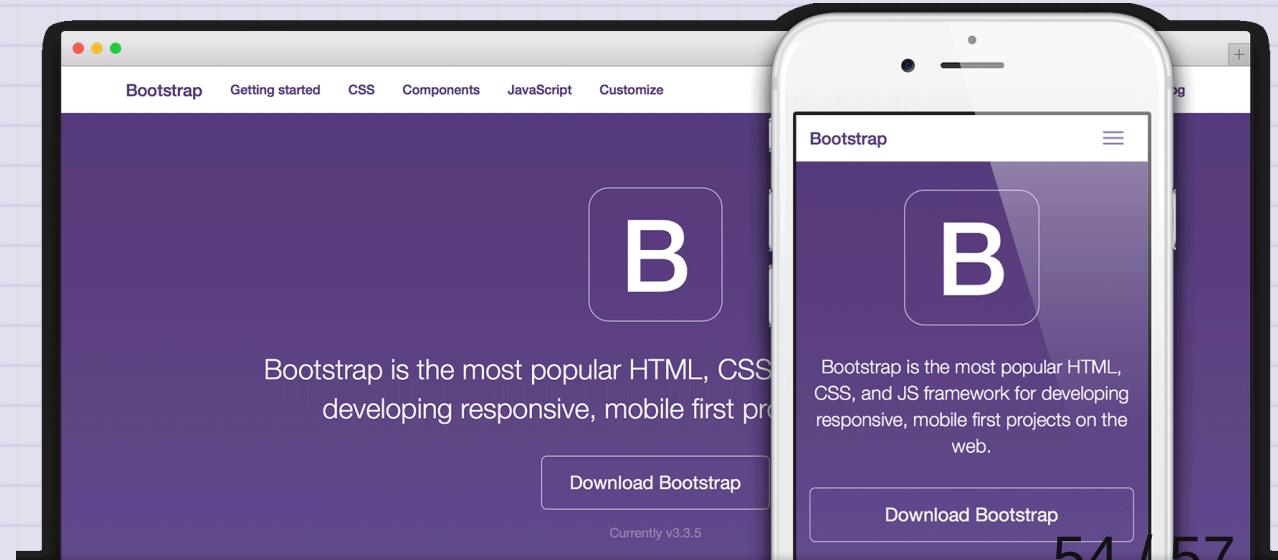
Number of turns in spiral

53 / 57

Share Rerun Settings Print Record a screencast View all apps About Developer options Clear cache

# 利用GitHub Pages 撰寫部落格文章分享所學

## 04Blogging



## 結語：有產出的學習更有說服力

- 獲取技術證照
- 開發應用程式
- 撰寫部落格文章



問答時間？



WHAT IF I TOLD YOU

I'M ALLERGIC TO QUESTIONS

MEMEFUL.COM

## 聯絡方式



- Email:  
**howard.haowen@gmail.com**
- 網站:  
**<https://howard-haowen.github.io>**