

程式設計與基礎資料型態與 中文構詞學

台師大通識教育課程

文本分析與程式設計

授課：卓騰語言科技 _ PeterWolf

諸君，語言的基礎乃是結構！

- 人類：
 - 自然語言
 - 句子的結構 - Syntax (句法)
 - 詞彙的結構 - Morphology (構詞)
- 電腦：
 - 程式語言
 - 程式的結構 - Syntax (語法)
 - 詞彙的結構 - Reserved Keywords (保留關鍵字)、Statement (陳述)、class (類別) 和 function (函式)

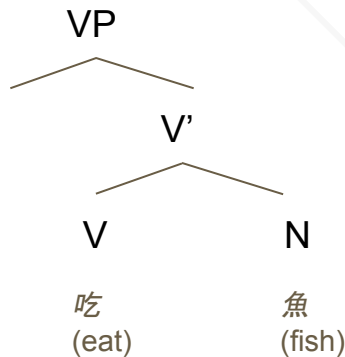
#人造的語言不是自然語言哦！

● 語言學：

- 所有的人類大腦結構是一致的
- 所有的人類都具有一樣的語言處理機制
- 人類的語言習得機制允許參數改變
- 所有語言的差異是來自參數的差異
- 參數的變化是有限的

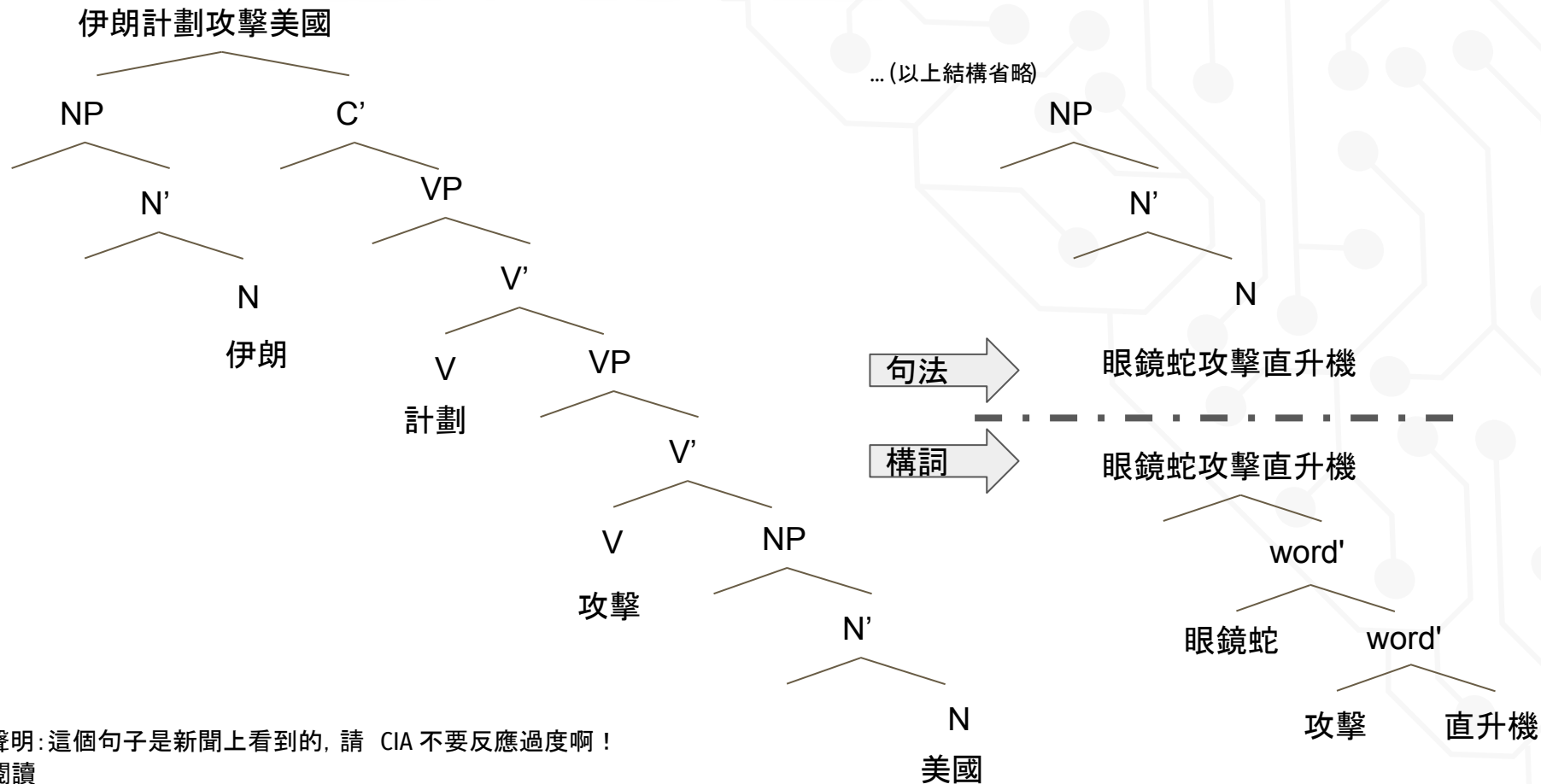
● 基本結構如下：

- X-bar theory



中文：句法和構詞的內在都有結構

#中文斷詞的時候，應該在**句法**上找「詞邊界」，而不是在**構詞**的內部找邊界！



免責聲明：這個句子是新聞上看到的，請 CIA 不要反應過度啊！

延伸閱讀

https://www.youtube.com/watch?v=12975615000577126642

中文：字的常見結構

- **並列**：「字符 + 字符」，但兩個字符的意義相近。
 - e.g., 語言、價值、嘲諷、花草
- **偏正**：「字符 + 字符」，但意義由後者決定。
 - e.g., 皮鞋、鐵路、火車、微笑
- **述補**：「字符 + 字符」，但意義由前者決定。
 - e.g., 擊敗、推翻、摔破、改善
- **述賓**：「字符 + 字符」，第一個「字符」通常可做獨立的動詞「字」，第二個字符則是前者的受詞。
 - e.g., 爬山、解題、傷心、賣命
- **主謂**：「字符 + 字符」，第一個字符通常可做獨立的名詞「字」，第二個字符則為獨立的動詞或形容 / 副詞。
 - e.g., 頭痛、臉紅、腦殘、

中文：詞的常見結構

- **並列**：「字 + 字」，但兩個字符的意義相近。
花園城堡、機關團體、
- **偏正**：「字 + 字」，但意義由後者決定。
攻擊直升機、電腦硬碟、消費者
- **述補**：「字 + 字」，但意義由前者決定。
吃不起、笑死人、吃壞肚子
- ~~**述賓**：「字 + 字」~~
- ~~**主謂**：「字 + 字」~~

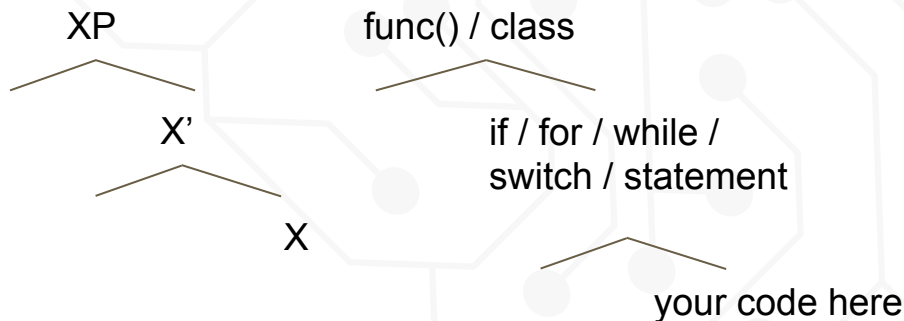
這兩者已進入到句法的層次了。

Recap! 諸君，語言的基礎乃是結構！

- 人類：
 - 自然語言
 - 句子的結構 - Syntax (句法)
 - 詞彙的結構 - Morphology (構詞)
- 電腦：
 - 程式語言
 - 程式的結構 - Syntax (語法)
 - 詞彙的結構 - Reserved Keywords (保留關鍵字)、Statement (陳述)、class (類別) 和 function (函式)

程式語言的結構是什麼？

```
week02.py
(bottom)
1  #!/usr/bin/env python3
2  # -*- coding: utf-8 -*-
3
4
5  def main(inputSTR):
6      """
7      這支程式的主要函式：(「函式」就是「功能」的意思！)
8      """
9      print("Hello {}, ".format(inputSTR))
10
11     messageSTR = """
12     「程式設計與基礎資料型態與中文構詞學」
13     整堂課的資訊量爆炸，在知識的海洋裡衝浪
14     超過癮的啊啊啊啊！
15     """
16     print(messageSTR)
17
18
19
20
21
22  # 程式進入點！ week02.py 這支程式，從這裡開始「執行」！
23  if __name__ == "__main__":
24      nameSTR = "PeterWolf"
25      main(nameSTR)
```



原來程式語言和自然語言這麼像呀？



Noam Chomsky (諾姆 喬姆斯基) (a.k.a. 杭士基、荷姆斯基)

計算語言之父

■ Noam Chomsky

- "a language to be a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements"
 - 語言是句子的集合 (有限或無限) · 每個句子的長度都是有限的 · 而且從有限的元素組合
- "the structure of language is biologically determined"
 - 語言結構是被生理結構確定
- "that humans are born with an innate linguistic ability that constitutes a Universal Grammar"
 - 人類生來就有天生的語言能力 · 這個能力包含了一種通用語法
- Syntactic Structures 句法結構



<https://www.slideshare.net/secret/loBqz9B6wpPcej>

<https://linguistics.mit.edu/user/chomsky/>

自然語言 in 程式語言

```
xINT = 1
```

```
xFLOAT = 1.5
```

```
xSTR = "w"
```

```
xSTR = "word"
```

```
xSTR = "the word is fine"
```

```
xLIST = xSTR.split(" ")
```

```
xSubSTR = xSTR[3]
```

整數 (integer): 程式語言的「整數」值

浮點數 (float): 程式語言的「小數」值

字串 (string): 程式語言裡的「字符串」

符記 (token): 自定切分規格後的結果

詞彙 (lexicon): 字典中列出的獨立項目

詞條 (entry): 資料庫中列出的獨立項目

自然語言 in 程式語言

```
week02.py
(bottom)

1  #!/usr/bin/env python3
2  # -*- coding: utf-8 -*-
3
4
5  def main(inputSTR):
6      """
7      ... 這支程式的主要函式 (「函式」就是「功能」的意思!)
8      ...
9      ... print("Hello {}".format(inputSTR))
10
11     messageSTR = """
12     ... 「程式設計與基礎資料型態與中文構詞學」
13     ... 整堂課的資訊量爆炸，在知識的海洋裡衝浪
14     ... 超過癮的啊啊啊啊！
15     ...
16     ... print(messageSTR)
17
18
19
20
21
22  # 程式進入點！ week02.py 這支程式，從這裡開始「執行」！
23  if __name__ == "__main__":
24      nameSTR = "PeterWolf"
25      main(nameSTR)
```

func() / class

if / for / while /
switch / statement

your code here

```
xINT = 1
xFLOAT = 1.5
xSTR = "w"
xSTR = "word"
xSTR = "the word is fine"
xLIST = xSTR.split(" ")
xSubSTR = xSTR[3]
```

Quiz:

<http://www.differencebetween.net/language/difference-between-grammar-and-syntax/>

回想課程中提到了自然語言中的「句法 (syntax)」和程式語言中的「語法 (syntax)」，試回答以下問題：

1. syntax 和 grammar 是否是一樣的東西？
 - a. 若一樣，為什麼？
 - b. 若不一樣，為什麼？
2. 英文有沒有文法？
3. 中文有沒有文法？
4. 程式語言有沒有「文法」？

Assignment:

1. 從課程 github repo 中把課程中提供的 week02.py 載下來。
2. 把 week02.py 改名為 week02_你的學號.py
3. 在 week02_你的學號.py 中：
 - a. 新增一個字串, 內容為:"姓名 學號"
 - b. 利用 substring index 切出前述字串中你的姓名和學號
 - c. 利用 split() 切出前述字串中你的姓名和學號
 - d. 在畫面中利用 print() **分別**印出你的姓名和學號