

## O. Language

# 人類智慧的最終形是數學

## - Paradigm Shift 曲範轉移

由地心說→日心說，所有的科學體系都得根據新體系重建

## - 主流 AI 發展及逆思

First 1950~1960：符号逻辑，金上层规则 tons of if...else → fail

Second 1980 ~ 1990 : 壓縮系統 限定場境 → fail

## Third · role~now · ; Machine · Learning

## - 現代語言學

1957 ~ : 有限的規則 for all human's language.

## 一、這三學典範轉手

Before it, we thought babies' brain like white paper.

After it, we know brain containing something, because a white paper can't do anything.

一、結論

## - Conclude

「Classic」：用資料(真相)寫條件邏輯

Modern : 用結構(遞式)寫條件邏輯

語言和數學沒有太多差別

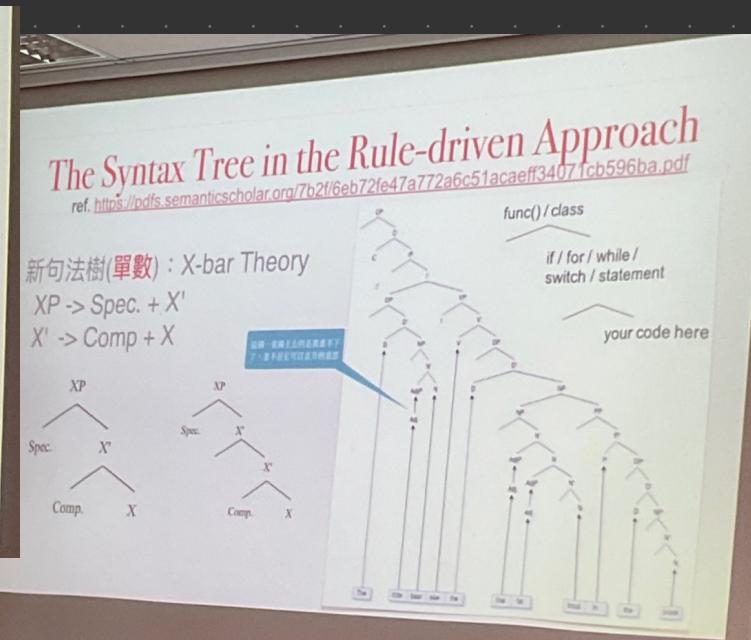
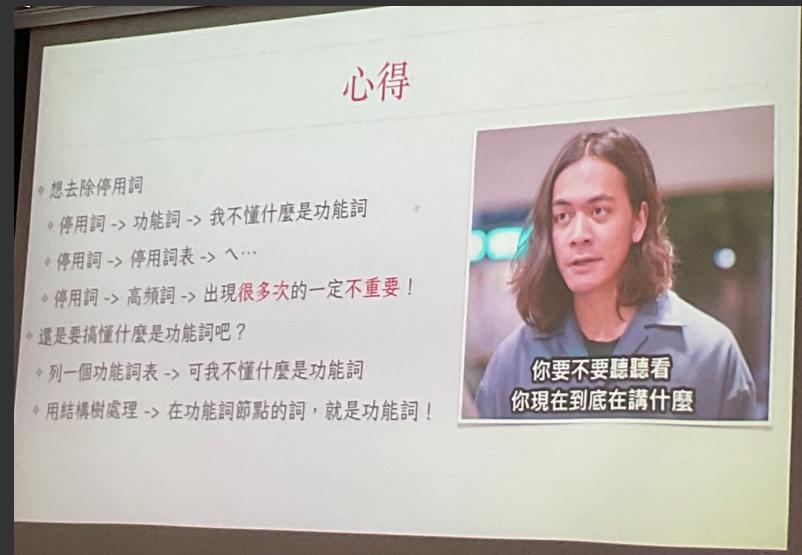
空格 ... (a lots of vocabulary)

Rule-based 我昨天買了一隻很會

貞小説

## Rule-driven

## - 豆瓣清洗



### - 停用詞 stop words.

[on, in, at, ...] such like that. 例如介系詞之類的，不具詞義

了，以

### - 將高頻詞列入停用字的問題

$2+8=10 \Rightarrow 2\ 8\ 10$  because +, = are high frequency word

### - 將敏感詞列入停用表

錯認引入停用詞，將導致模型訓練失真

⇒ 「詞性」、「語意」是兩回事

- 人類語言組成

- 百科知識

- I/O 能力

- 內建的語言處理系統

- 詞典/本位斷詞法

- Challenge

紅匯這成本、領域差異、左右生詞

- 統計/機率模型

已知詞用字典、未知詞用機率

Ex py module jieba.

- Challenge

紅匯這成本、領域差異、資料量

jieba 用 5000 萬詞處理箇中，箇中沒有如此字典且實際詞量更大

同一詞在箇中已有差異。

「非指比」在香港沒負面，在中國為正面

- 框架學習

不同於人類思維

## • LUIS : Microsoft Language Model

- It requires some human assistance.

語意相當於程式函數

