**鑒往知來：子非人，安知人之語？—談自然語言處理 / 陳信希教授**

**演講摘要&重點筆記**

演講摘要：

背景：應用和挑戰

淺談自然語言處理

分類：掌握共通性

分布式表示

分散式表示

言談剖析

問答系統應用

機器翻譯應用

意見探勘應用

結語

重點筆記：

1. 三大應用：機器翻譯、問答系統、意見探勘
2. 機器翻譯代表不同語言間的文本自動翻譯。
3. 機器翻譯與圖像、語音做結合，成為了生活中很重要且貼切的應用。
4. Google翻譯的成長，就詞彙面向來看，翻譯的單詞逐漸符合人性，就結構面向來看，翻譯出來的句子中，介係詞等等的位置逐漸符合真實口語，說明了翻譯技術的進步。
5. 問答系統應用了代名詞指示的技術，從文章結構中找到代名詞所指的人物。
6. 意見探勘做正負面等等的意見分析，找出一段敘述中主觀的意見以及評論。
7. 機器人透過做意見探勘與人類進行辯論比賽，並獲得一勝一敗，說明了意見分析技術的成長以及進步。
8. 自然語言處理的一些典型問題：歧義解析、容錯力、強健性
9. 歧義性指的是單詞在文章中所代表多項詞彙、語法、語義，透過歧義解析來判斷單詞在該文章中所代表的意義。
10. 概念表示用符號來代表一項實體，這樣的方式會導致歧義性。
11. 透過詞性、語意、句法、相依、言談、意見、情感、立場來做分類掌握句義。
12. 以上方法都屬於符號計算。
13. 分布式假設表示如果兩個詞的上下文相似，則兩個詞的詞義是相似。
14. 分散式表示用上下文的詞來預測目標詞，以當下的詞來預測上下文的詞。
15. 符號的離散轉為連續代表文字轉變成向量的概念。
16. 問答系統透過相似性來選擇答案。
17. Google多語言神經機器翻譯系統透過訓練找出相關的翻譯輸入與輸出。
18. 意見探勘技術辨識出句子中正面、負面、中性的詞彙差異性。
19. 機器翻譯尚未到達人類翻譯的境界，因為缺乏人類主觀意識。