

自然語言處理（Nature Language Processing），演講中以文字翻譯、問答與意見探勘三者為例，簡單說明自然語言處理與日常生活中的關係。

文字翻譯須能夠識別不同的語言以及每種語言的詞彙與語句結構。以 Google Translate 來說，講者每年觀察同樣的文句在 Google Translate 上翻譯的結果，在早期翻譯功能較不完善的狀況下，經常出現詞彙使用錯誤，甚至連語句結構都有過完全不對的案例，但自從 Google Translate 導入 Deep Learning 的神經網路架構後，很明顯突破了翻譯貧頸，近幾年的翻譯結果較為準確，詞彙使用較為精準、語句結構也比較貼近一般人的語法。

問答的要點則有三個：聽題、知識庫查找與答題。以演講中提到的 Watson 機器人來說，他參加了問答節目，所以首先需要「聽懂」主持人問了什麼問題，再來才能在知識庫中搜尋答案。在知識庫中搜尋答案也有可能遇到一些困難，例如代名詞指涉的問題，在一段文章中出現了很多個「他」，然而這些他可能分別代指不同的人，知識庫檢索的過程中要能將這些區分開來。又或者會遇到詞彙的同義詞問題，像是某些狀況下「建設」與「製造」所指涉的意思可能相同，知識庫檢索時必須要能辨別。

最後一個意見探勘，演講中舉例讓機器人去辯論是否要投資更多錢在太空探索上，而這會涉及到文本分析的技術，NLP 要能夠分析一段文章中的論點與立論基礎，甚至要能分辨一篇文章中多個不同（甚至相悖）的觀點，因此難度也不小。由於資料量眾多，必要時也需要做數據清理與標準化，讓資料能夠有效的被解讀以及使用。

講完三個例子後，講者開始詳細闡述 NLP 會遇到的問題，並與前面提到過的例子相呼應。例如：談到歧義，也就是說一個詞可能有許多意思，在不同狀況下有不同或相同的解讀，因此需要能識別同義詞與歧義詞間的關係。

<https://www.youtube.com/watch?v=pWqm9K9Aio>