

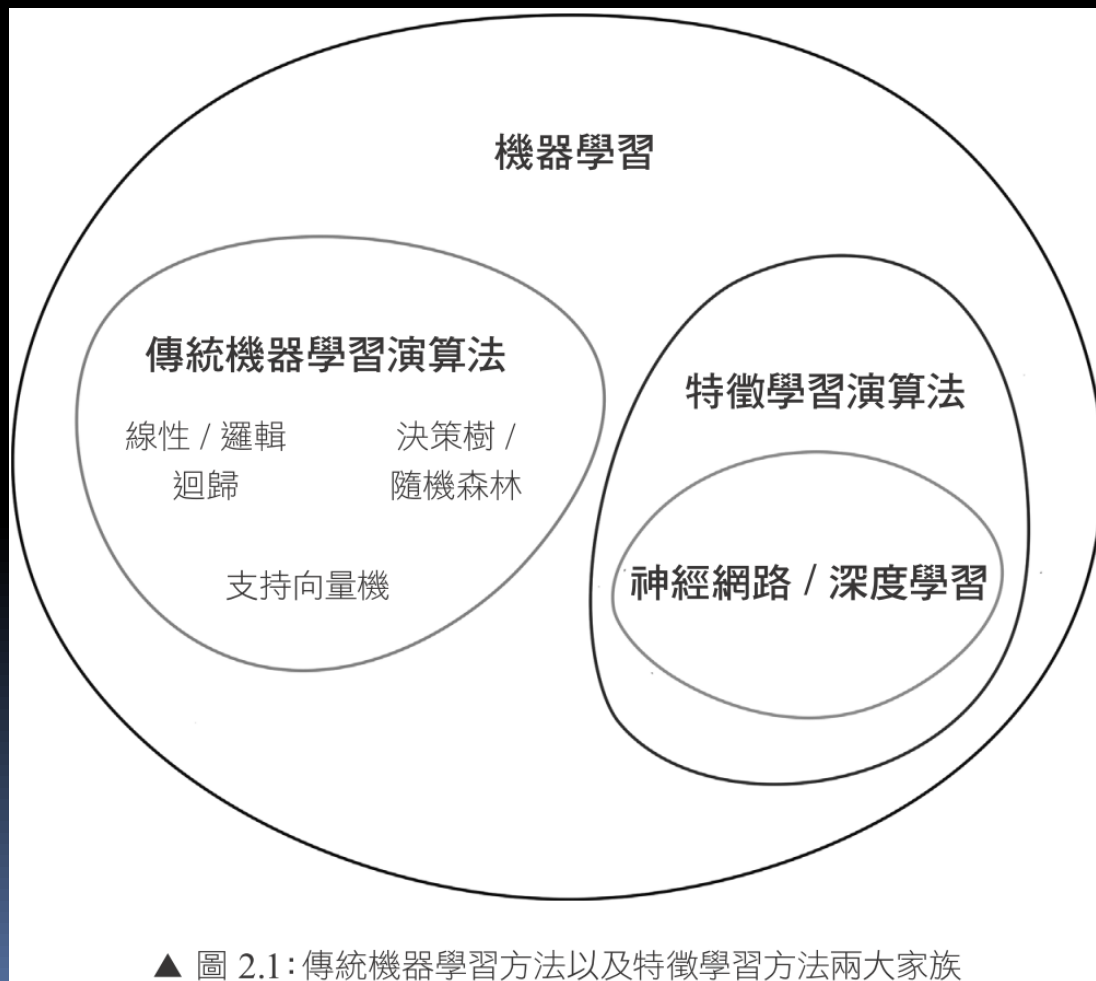


CHAPTER 2

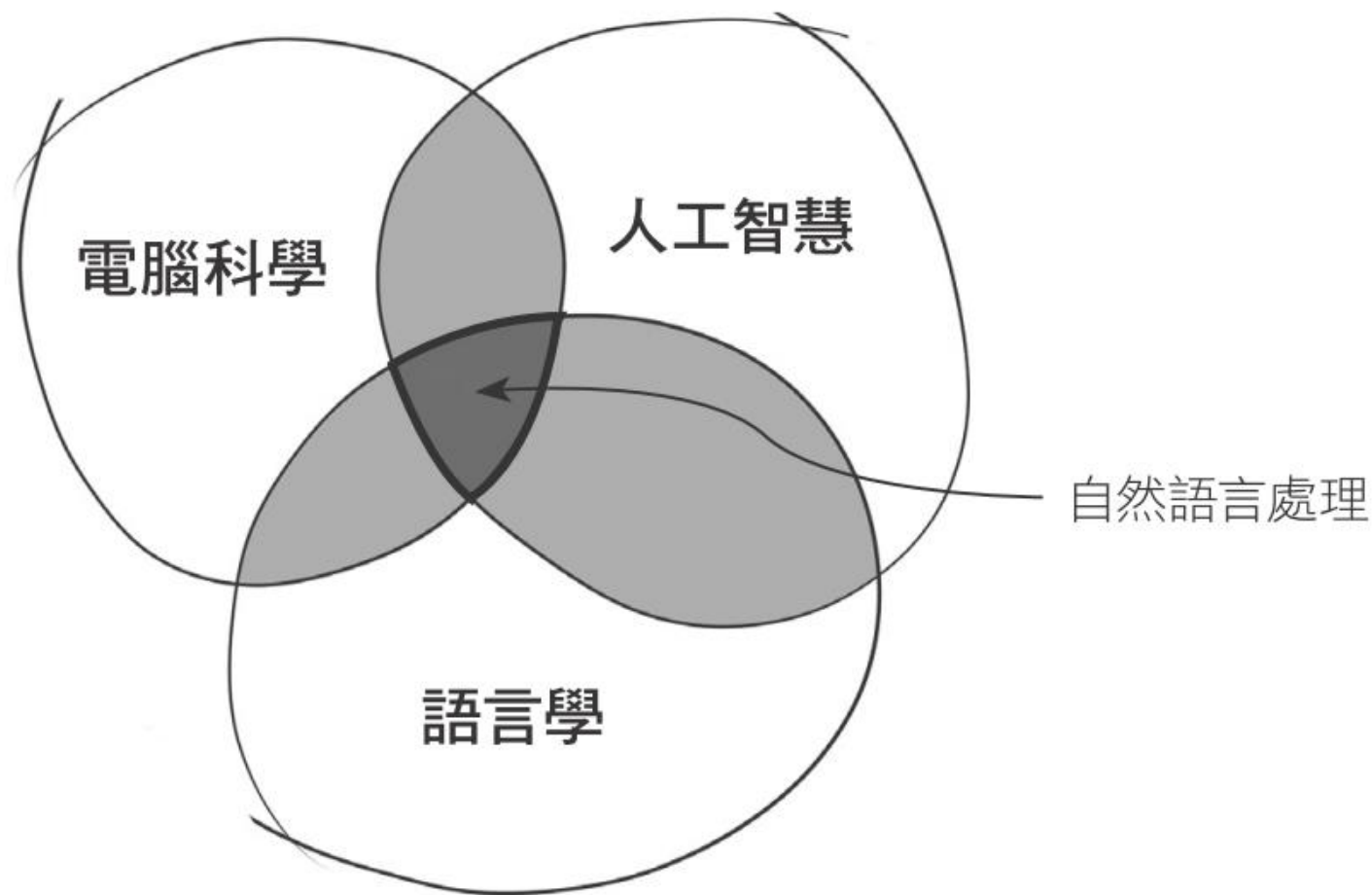
用機器處理自然語言

2.1 深度學習 + 自然語言處理

■ 2.1.1 用深度學習自動學習特徵




2.1.2 自然語言處理 (Natural Language Processing)



▲ 圖 2.2: NLP 位於電腦科學、語言學和人工智慧領域的交界處

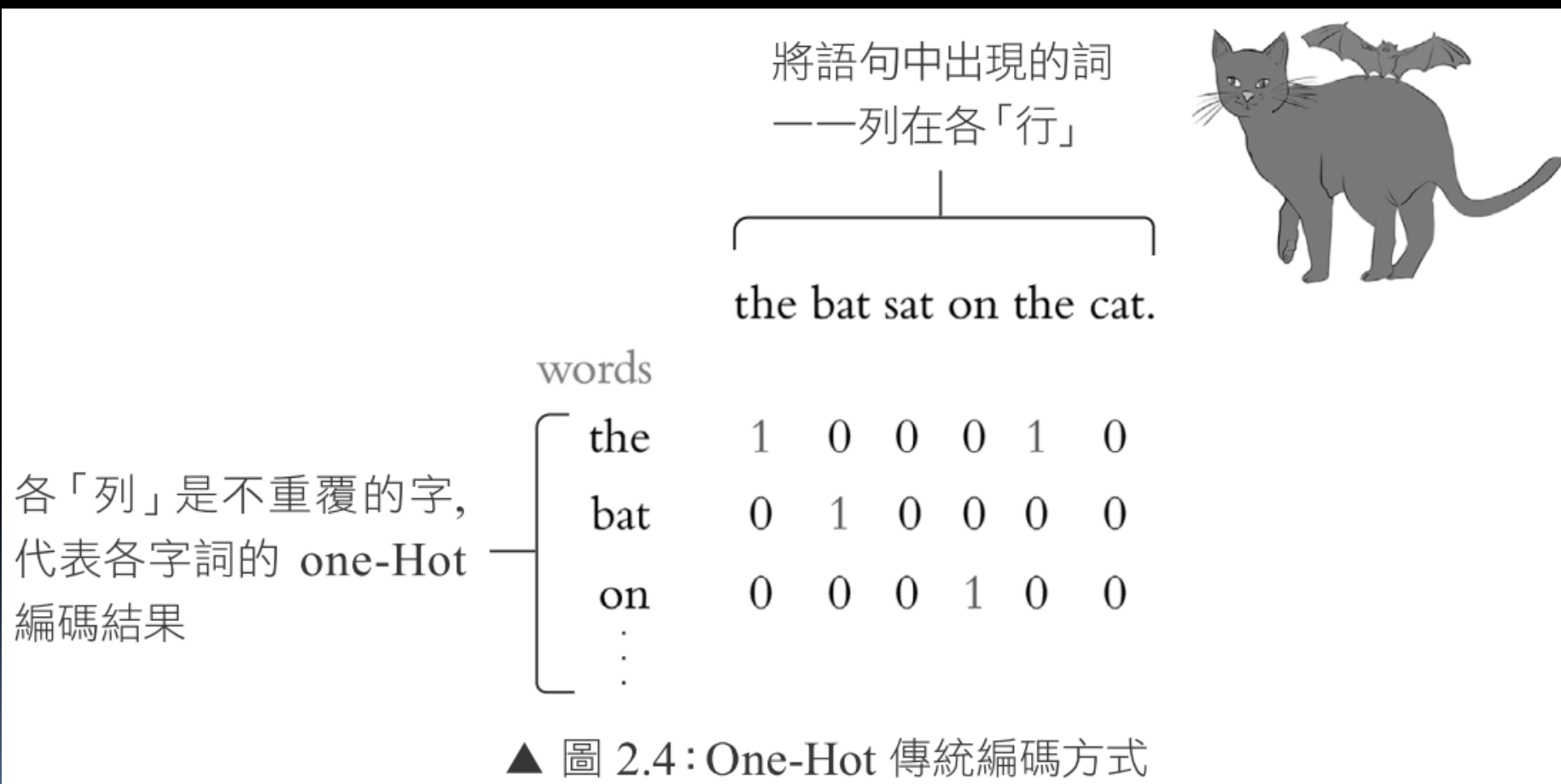


NLP 的應用

- 文件分類
 - 機器翻譯
 - 搜尋引擎
 - 語音辨識
 - 聊天機器人
- 

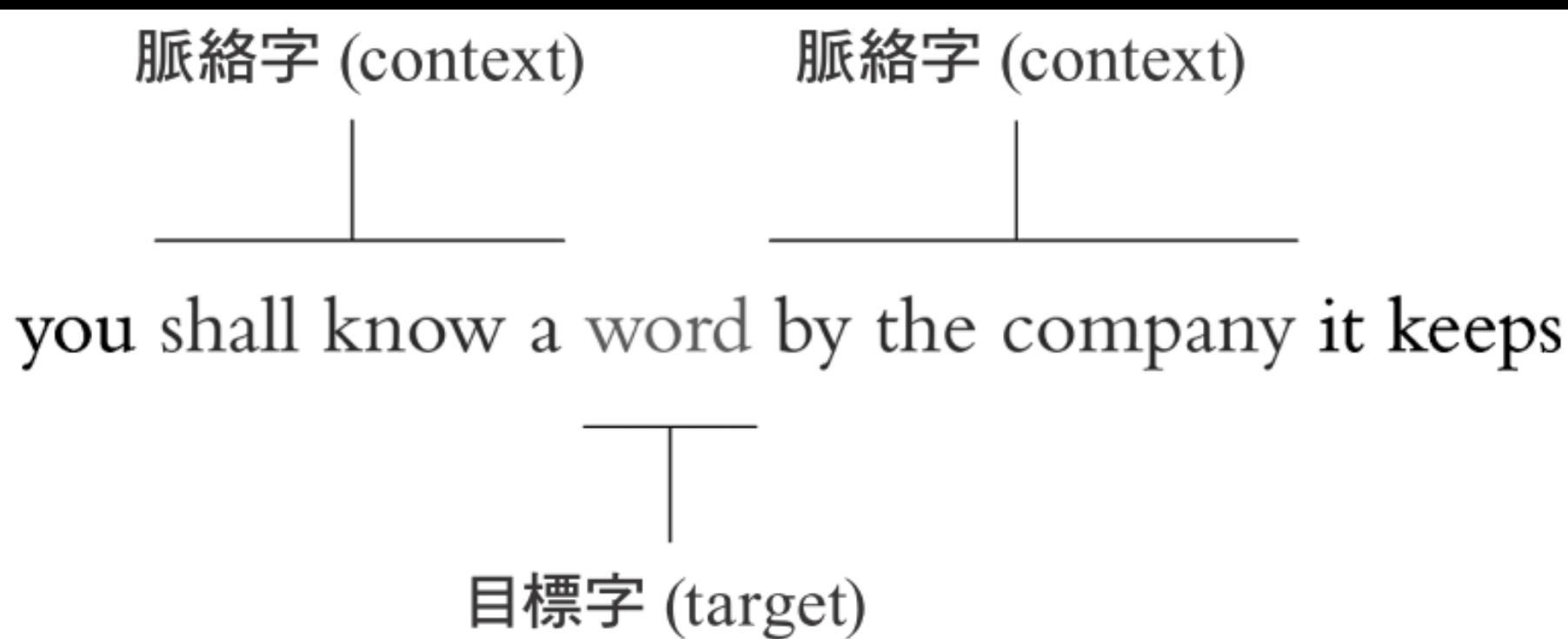
2.2 將語言量化

■ 2.2.1 字詞的 One-Hot 編碼



2.2.2 詞向量 (word vector) 與詞向量空間 (vector space)

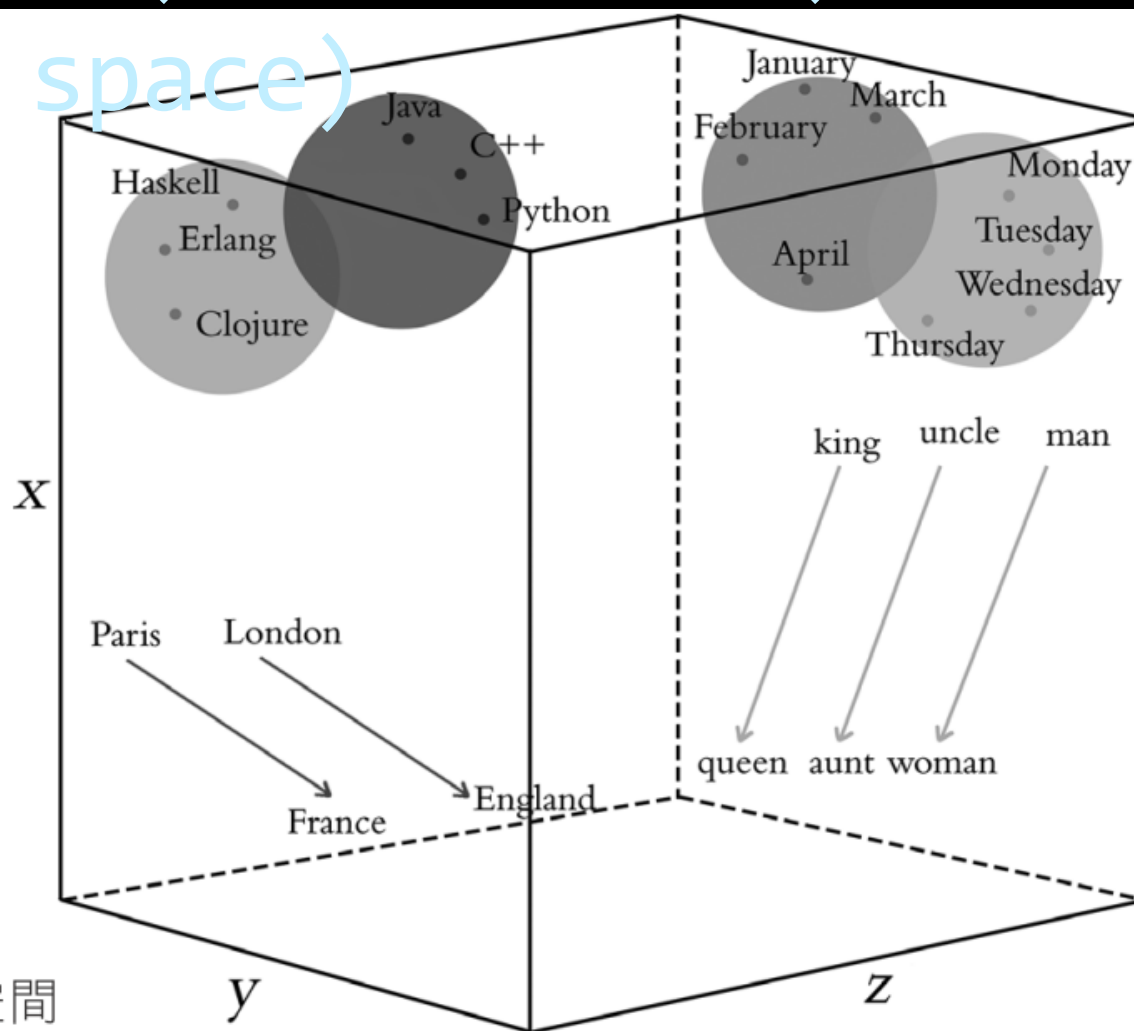
- 詞向量空間 (vector space) 的概念



▲ 圖 2.5：準備將語料庫的各單字轉換成詞向量

2.2.2 詞向量 (word vector) 與詞向量空間 (vector space)

- 詞向量 (vector) 的例子

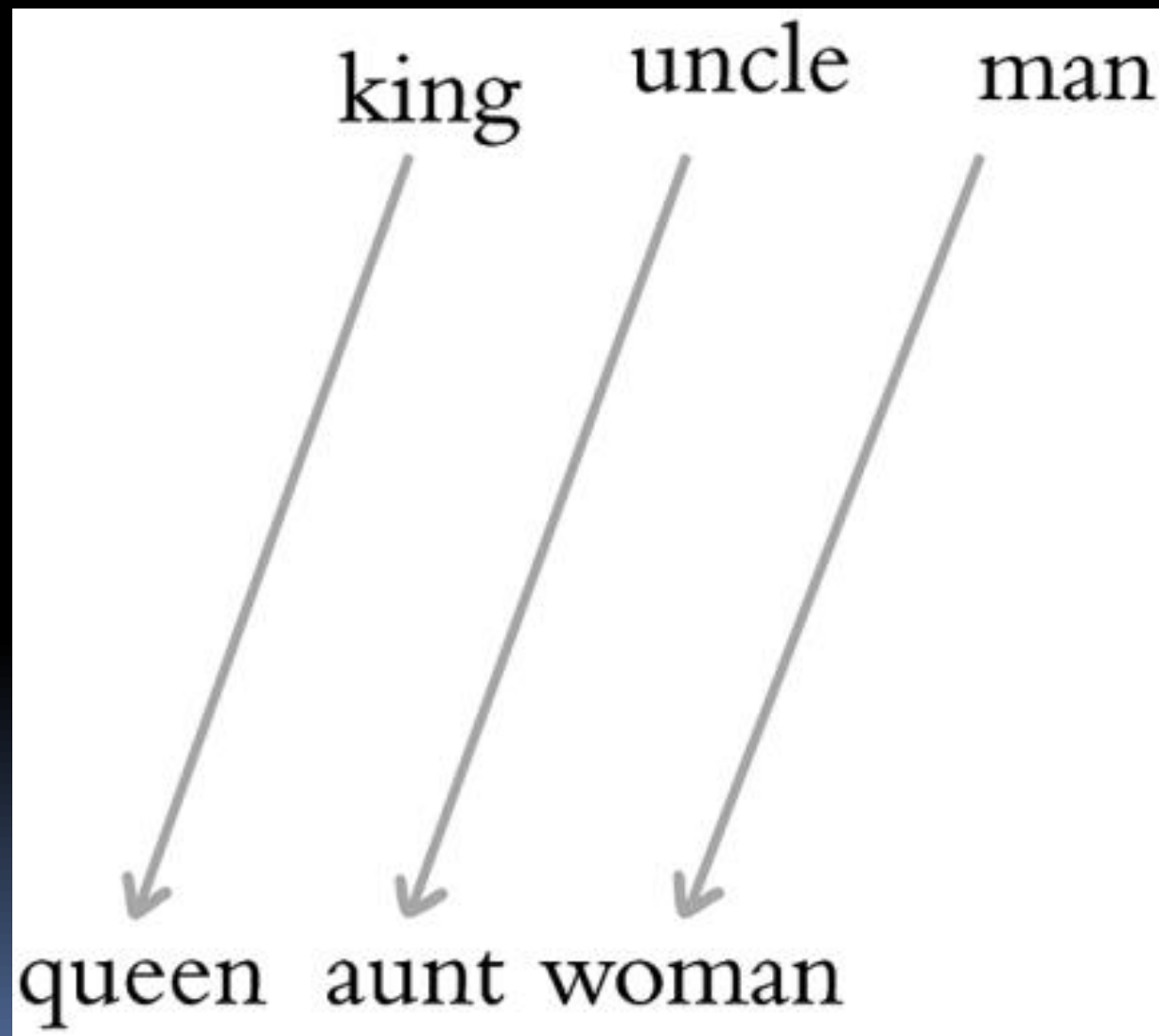


► 圖 2.6:3 維空間
中每個字詞的位置

3 維詞向量空間

2.2.3 詞向量的數學運算

- 從任何代表男性的詞向量出發，最終將停在與該男性詞向量對應的女性詞向量

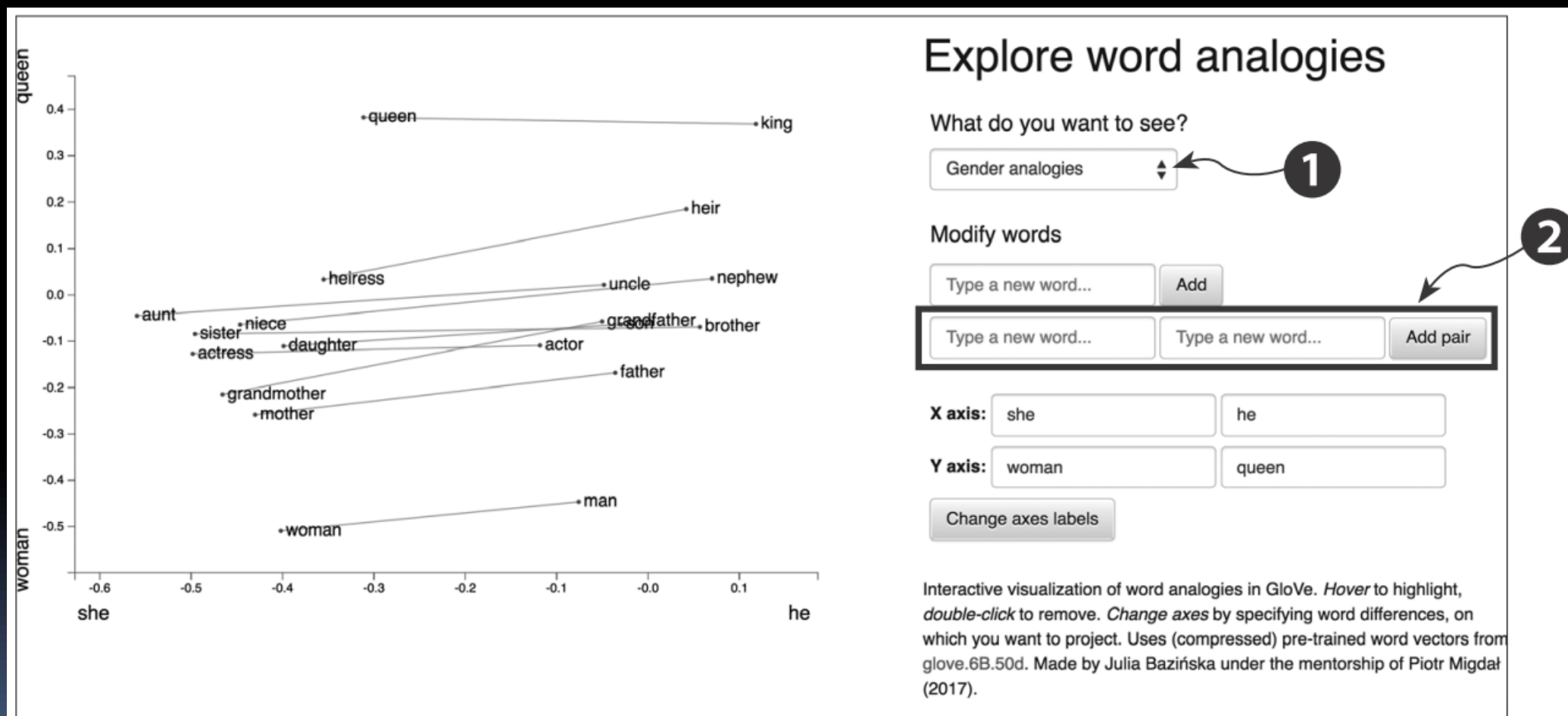


2.2.3 詞向量的數學運算

$$\begin{array}{rclcl} V_{\text{king}} & - & V_{\text{man}} & + & V_{\text{woman}} = V_{\text{queen}} \\ V_{\text{bezos}} & - & V_{\text{amazon}} & + & V_{\text{tesla}} = V_{\text{musk}} \\ V_{\text{windows}} & - & V_{\text{microsoft}} & + & V_{\text{google}} = V_{\text{android}} \end{array}$$

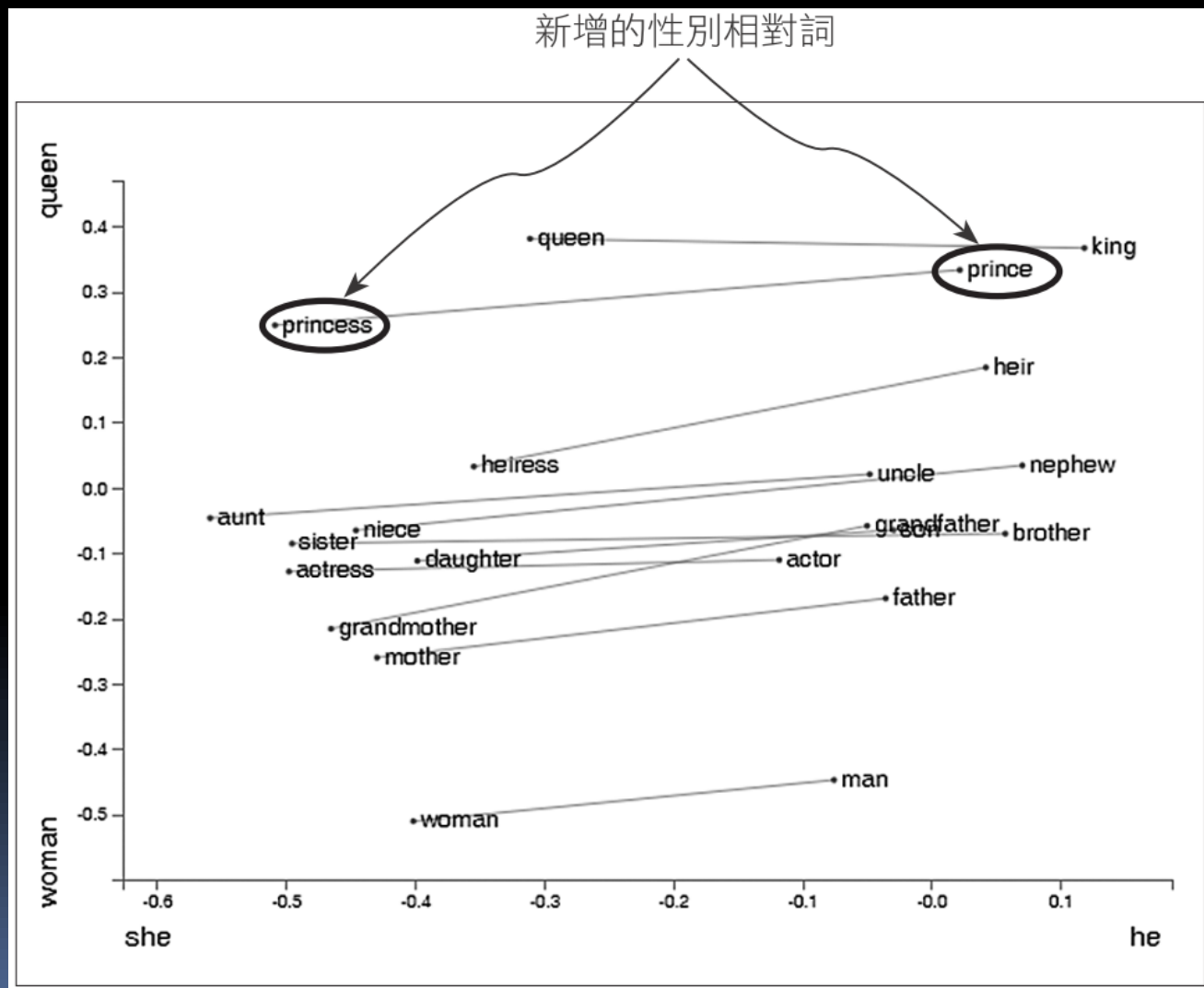
▲ 圖 2.7: 詞向量的運算

2.2.4 上 word2viz 體驗 2 維詞向量空間



▲ 圖 2.8 : word2viz 的網頁畫面

2.2.4 上 word2viz 體驗 2 維詞向量空間



2.3 Google Duplex 的自然語言功力

- Duplex 結合了 Tacotron (bit.ly/tacotron) 和 WaveNet (bit.ly/waveNet) 的波形合成技術、以及較為傳統的「拼接式 (concatenative)」文字到語音合成引擎, 而這正是它能越過恐怖谷 (uncanny valley) 的關鍵



2.4 總結

- 深度學習能自動從資料中萃取關鍵特徵的優勢，將字詞轉成詞向量，嵌入詞向量空間的概念，記錄字詞的脈絡資訊，提供給NLP 模型使用。
- 