

突破低資源語言困境～達悟語 AI 翻譯與數位典藏新里程



靜宜大學寰宇外語教育學士學位學程張惠環助理教授分享達悟語相關知能（語文教育及編譯研究中心提供）

【語文教育及編譯研究中心 周一銘】

在各種科技爆炸成長、AI 語言模型高度發展的今日，全球主流語言在數位世界大步邁進；然而數位資源極度匱乏的弱勢語言，卻面臨被科技邊緣化的「數位孤立」困境。本院語文教育及編譯研究中心於 2026 年 6 月 12 日舉辦「低資源語言翻譯系統建置與數位教材開發」學術演講，特別邀請國立中正大學語言學研究所何德華教授與靜宜大學寰宇外語教育學士學位學程張惠環助理教授，共同分享將 AI 科技導入達悟語復振與族語教學的創新成果，剖析如何透過跨學科的合作，為低資源語言建構一套可永續發展的數位知識轉型路徑。

兩位教授指出，達悟語作為臺灣唯一兼具海洋文化特色的南島語，目前面臨流利使用者高齡化、年輕世代逐漸轉向中文的現況，更棘手的是，達悟語內部存在著明顯的「南北方言變異」，這種方言結構與拼寫多樣性，在缺乏數位工具的時代，大大增加了族語教學與教材編製的難度；然而，羅馬不是一天造成的，語言復振更需要長期的累積。研究團隊自 1994 年起，在語言學主持人何德華教授、資訊建置核心張惠環老

師、資訊科學夥伴楊孟蓓老師，以及達悟語母語顧問董瑪女 (si Somapni) 女士的共同努力下，展開長達 20 多年的默默耕耘，陸續完成《達悟語文化詞典》與《達悟語聖經詞典》，累積了近三萬筆珍貴的雙語平行語料與音檔，這些數位典藏經由人工清理與逐句對齊，成功轉化為 AI 機器翻譯系統的基礎大腦。

演講的另一個亮點，在於兩位教授誠實剖析了第一代翻譯模型所面臨的「美麗錯誤」，由於低資源語言的資料有限，初期 AI 極易出現「泛化不足」與「語意偏離」的現象，例如輸入中文「你不要吃魚」，AI 竟死記固定結構而誤譯為「你不要來」，或者當方言拼寫變異出現時，AI 因無法理解語境，而將同一個字誤判為完全無關的意思。為了解決低資源語言的訓練瓶頸，團隊在最新研究中引入了 Meta AI 的多語言預訓練模型，並採用「代理來源語設定 (Proxy Source Setting)」，她們大膽選擇與達悟語同屬菲律賓型語言、語法結構高度相似的「他加祿語 (Tagalog)」作為參照，透過隱性跨語言遷移學習 (Transfer Learning)，讓 AI 能自動借力相近語言的語法知識，將微調後的自動化測試評分 (BLEU) 顯著提升，大幅改善了動詞時態、疑問句及方言變異的翻譯準確度。當翻譯系統進一步與教學現場整合，應用場景也隨之擴展，團隊為了降低 AI「幻覺」，特別設計讓 AI 依據權威教材回答，開發出網頁式語法問答機器人「TAOGPT」，以及可在手機上隨時查詢的 LINE 聊天機器人 (LINE TAOChatBot)。這些工具讓醫療與教育資訊一樣，不再只是分散的檔案，而是能夠被理解、串聯並加以運用的知識體系，有效支援族語教師備課與學生自學。

從方言變異的整理、到遷移學習的模型建模，再到族語教學現場的數位教材共創，這場演講描繪出一條跨領域整合的發展路徑，這艘由臺灣出發的 AI 科技獨木舟，不僅成功跨越了低資源語言的限制，更為本土語言的數位轉型與傳承，提供值得深思的方向。