從語言資料到資料主權——夏威夷語言學院與 AI 的運用

言語データからデータ主権へ——ハワイ言語学院とAIの運用

From Language Data to Data Sovereignty: The Hawaiian Academia Language School and the Use of AI

文・圖｜李政政（東華大學原住民族研究國際博士班助理教授）

「‘O ka ‘Ōlelo ke Ka‘ā o ka Mauli」可以理解為：「語言是連結我們生命與認同各個面向的纖維。」這句格言，是理解我此次參訪夏威夷大學希洛分校（University of Hawai i at Hilo）夏威夷語言學院（Ka Haka ‘Ula o Ke‘elikōlani, Hawaiian Language College）最重要的起點，也是學院的核心理念。

本篇文章首先介紹夏威夷語言學院所建立的文化空間，接著延續文化空間的脈絡，透過 Lauleo 計畫的案例分析，探討夏威夷原住民如何透過社群參與與資料蒐集，逐步建立資料主權，進而決定夏威夷語如何進入 AI 的世界，並脈絡化 AI 是如何成為語言復振運動中工具的一環。



kūpuna（耆老與祖先們）在教室的空間與學生上課。

夏威夷語言學院的文化空間

在今年七月，透過 Kekoa Harman 與 Pelehonuamea Harman 兩位教授的介紹，我有機會拜訪夏威夷大學希洛分校的夏威夷語言學院。參訪過程中，我特別感受到學院如何有意識地將 kūpuna（耆老與祖先）、以 Ruth Ke‘elikōlani 為代表的領導者典範。她曾擔任夏威夷島總督，也是夏威夷語的堅定守護者，以及文化脈絡、諺語、地名、人與自然環境、 mo‘ōlelo（故事與歷史敘事）、 hula（夏威夷舞蹈）與 mele（聖歌、歌謠與詩歌），融入學院的建築、空間設計與



ʻōlelo Hawaiʻi 語音檔案庫 Kaniʻāina（意為「土地之聲」）。

日常環境之中，並進一步將這些核心價值融入課程設計。

因此，學院所營造的核心概念是，空間本身就是一種述說故事與傳遞價值的方式。當學生進入玄關、穿過迴廊、進入教室，在經過校園各個角落的過程中，他們所接觸的不只是文字與課程，更是在接觸一套關於土地、環境、 kūpuna（耆老與祖先們）、歷史故事與身份認同的知識體系。Kekoa 老師特別向我強調，語言學習是一種生活方式。只有當語言真正融入每天的生活，成為生活方式的一部分，語言才會進一步觸及人類的靈魂，並內化為身體與肌肉記憶的一部分。

此外， Kekoa 老師在介紹中反覆強調的一個重要觀念，即是語言復振是一項跨越世代才能完成的工作，它不可能由單一世代完成，而是由一代又一代的人共同投入。而這樣的理念，與我在學院的日常生活中相遇。我在走廊上遇到一位媽媽與兒子一起上相同的課程，也看到學院與整個社區之間密切的教育關係。Kekoa 告訴我們，高中生就可以開始修習夏威夷語言學院所開設的課程，而在正式進入大學後，學院會承認這些學分。這樣的制度安排，讓學生可以更早開始自己的學習旅程。因此，對我而言，ʻŌlelo Hawaiʻi（夏威夷語）是一個被有意識地建構於文化整體空間之中，在這個設計下，空間與語言是密不可分的關係。

從文化空間到 AI 世界

在這樣的脈絡下，我們就能理解為什麼夏威夷語言學院會投入 AI 的運用。如果只看到學院使用 AI 的相關報導，讀者可能會產生一種印象，好像 AI 是當代夏威夷語言復振運用的核心方法。但從我與學院互動的經驗來看，AI 其實是語言復振運動中的眾多工具之一。在 AI 出現之前，學院與相關社群早已投入大量心力，保存與活化夏威夷語。例如，學院將具有約 40 年歷史的夏威夷語廣播節目進行數位化，讓今天的我們可以「聆聽另一個時代」的聲音。這些工作對於文化空間的建構同樣重要，因為語言復振不只是創造文化空間，也是不斷發掘新的語言工具，用於保

Kekoa 老師特別向我強調，語言學習是一種生活方式。只有當語言真正融入每天的生活，成為生活方式的一部分，語言才會進一步觸及人類的靈魂，並內化為身體與肌肉記憶的一部分。



參與者不需要是語言專家，只需要具備基本且清楚的發音能力，並願意朗讀指定句子即可。因為，Kimura 認為如果 AI 要真正服務夏威夷語社群，它就不能只理解某一種標準的語言使用方式，而必須能夠收納不同年齡、不同背景與不同語言程度使用者的聲音。



存並重現過去的聲音，讓下一代仍然能夠聽見 kūpuna（耆老與祖先）的話語。

在這個過程中，Larry Kimura 教授是一位非常重要的人物。他長期投入夏威夷語，及原住民族研究，也是 ʻōlelo Hawaiʻi 語音檔案庫的重要推動者。根據夏威夷大學的報導，Kimura 提到，「在建立 Kaniʻāina（意為「土地之聲」）數位資料庫的過程中，光是轉錄線上語音資料庫中的



Ruth Ke elikōlani 是夏威夷與學院的代表的領導者典範，她曾擔任夏威夷島總督，也是夏威夷語的堅定守護者。



由社群共同打造、Lauleo 網站，為社群服務的 ʻŌlelo Hawaiʻi（夏威夷語）AI 工具。

錄音，每一小時的語音，大約就需要花費 30 小時，才能完成多個版本的轉錄與校訂，之後才能上傳分享。」這個例子讓我看到保存工作中的一個實際困難，即是整理與保存既有的資料庫本身就是一項龐大的工程，而透過人工進行這項工作，所需要投入的時間成本極高。因此，AI 在這裡的價值，是協助社群處理龐大的語音資料，降低整理與轉錄所需的時間成本，使原本需要長時間投入的工作能夠更有效率地完成。

Lauleo 計劃：由社群建立的 AI 語料庫

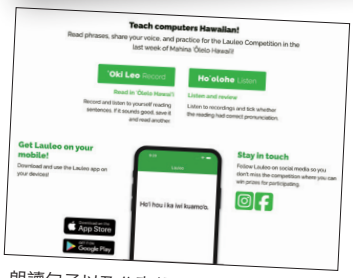
Lauleo 計畫正是在這樣的背景下孕育而生。這是一項由夏威夷語言學院、Kanaeokana（由超過 70 所學校與組織組成，目標在深化 lāhui 夏威夷族群／國族意識）、Awaiaulu（原住民人才培育機構）以及 Aotearoa（紐西蘭）的 Te Hiku Media 共同合作的跨島嶼、跨國家的計畫。Te Hiku Media 是一個以毛利文化為主體的機構，投入於 te reo Māori（毛利語）的語言資料保存、檔案建立與相關人才培訓工作。

Lauleo 的目標，是蒐集夏威夷語使用者的語音資料，建立能夠辨識 ʻŌlelo Hawaiʻi 的人工智慧模型，並進一步將語音轉換成文字。也就是說，它不是單純利用 AI 翻譯夏威夷語，或利用 AI 生成夏威夷語文字，而是先由夏威夷語社群自主提供語音資料，再建立能夠理解這套語言的 AI 模型。這些資料未來會被運用在不同的語言工具，例如夏威夷語文字訊息、夏威夷語版本的 Siri，以及協助語言學習者改善發音的應用程式。更重要的

是，這項計畫強調：「讓我們的聲音為語言匯聚：由社群共同打造、為社群服務的 ʻŌlelo Hawaiʻi 人工智慧工具。」因此，這項計畫也開放不同程度的夏威夷與使用者參與。就此，Lauleo 舉辦了一項語音資料蒐集競賽，參與者只需要在手機下載應用程式，並以組隊的方式，廣邀不同背景、語言程度的參與者加入。根據網站資料，目前已有約一千位參與者提供總計 413 小時的語音內容。這項競賽，回應了Kimura所說的，Lauleo 並非讓 AI 只學習單一或標準化的語言使用方式，而是透過社群聲音，使 AI 更能反映夏威夷語社群實際使用語言的多樣性。



穿過迴廊、進入教室。在經過校園各個角落的過程中都有滿滿的文化知識。



朗讀句子以及分享聲音，並舉辦比賽刺激大家的參與。

從語言資料到資料主權

Lauleo 案例真正啟發我思考的地方，其實不只是 AI 的運用，而是「誰擁有資料？誰管理資料？以及資料最終要服務誰？」等核心問題。這些問題，也是我在夏威夷大學求學期間，反覆與自己、教授、社群以及研究夥伴討論的倫理與關係性課題。這些蒐集而來的錄音，並不是一般意義上的「data」，而是社群交付給他們保管的文化資產。他們以守護者的角色保存這些資料，並承諾讓社群最終能夠從這些資料所發展出的應用程式與工具中受益。這樣的做法，與 Tahu Kukutai 與 John Taylor 在《Indigenous Data Sovereignty》一書中所提出的概念相呼應，即是資料主權（data sovereignty）。在這個脈絡下，資料主權可以被理解為原住民族自決權在 AI 世界中的延伸。

值得一提的是，Lauleo 的工作建立在更早以前的努力之上，即是 kūpuna（耆老與祖先）留下語言與智慧，社群保存錄音，學院建立文化空間，家庭參與傳承，社群建立廣播與文化活動，研究者則進行資料整理與轉錄。而現在，新一代的計畫開始使用 AI，將這些累積多年的語言資料帶入 AI 的世界。因此，AI 並不是語言復振的起點，真正的起朗讀句子以及分享聲音，並舉辦比賽刺激大家的參與全貌。

最後，這也讓我重新理解開頭那句話：「ʻO ka ʻŌlelo ke Kaʻā o ka Mauli」。語言本身就是連結生命、土地、祖先、家庭、社群與身分認同的纖維。而今天的課題，是如何讓這條纖維同時連結過去與現在、土地與認同，也能延伸到 AI 所構成的世界。◆



李政政

高雄市三民區人。1990年生。夏威夷大學政治學博士。現任東華大學原住民族研究國際博士班助理教授。曾於中央研究院民族學研究所博士後研究。由於父親在高雄港的油槽工作，從小在港口看著形形色色的貨櫃船進出，就此成為與大洋洲世界連結的起點。目前研究範圍包括大洋洲政治、原住民族外交，以及台灣、帛琉與區域之間的關係。