

本地化的未來



AI Business
電子書系列



了解如何運用語言和文化等工具，
有效推動您的數位轉型

合作機構：

LIONBRIDGE

2

拓展業務的核心關鍵

本地化的宗旨，在於根據在地市場調整產品、溝通、服務與其他事項，使內容能更加符合特定地點客戶的期待。

在這個四海一家的全球化世界裡，本地化能讓公司和企業進軍國際市場，進而提高營收，並在過程中獲得規模經濟的優勢。而正因為如此，對於打算將在地產品推向全球的業務策略而言，本地化都會是核心關鍵。即使某些公司在品牌調性上是與特定地區緊密連結，例如 Apple 保證產品都是由位於加州的 Apple 所設計（「Designed by Apple in California」），他們仍必須運用本地化策略贏得其他國家/地區客戶的青睞。

當然，翻譯仍然是本地化的核心要素。正如這份電子書之後會討論到，這項作業有時會由語言專家進行，有些會由先進的機器翻譯 (MT) 工具處理，更有些會採用兩者兼具的混合式翻譯方法，稱為加強式翻譯（參見第 5 頁）。

但真正的本地化不只是翻譯而已。如果想將某個市場的服務或產品體驗，成功移植到另一個市場，就必須深入了解各個地點之間的細微差異。

NI (National Instruments) 的語言服務資深經理 Inés Rubio 便向 AI Business 表示：「我們基本上是以另一個語言重新打造體驗。重點並不是語言，而是文化。」

這份電子書由翻譯與本地化產業專家 Lionbridge 贊助與製作，探討拓展國際業務時會遇到的挑戰，以及成功做到本地化的優勢。除了專家訪談外，書中還有來自包括 Zynga、Royal Caribbean、Volvo 和維基媒體基金會 (Wikimedia Foundation) 等組織的觀點，希望能提供實用資訊，協助讀者打造一流的全球產品與服務。

Luke Dormehl | 副主編 | AI Business

電子書合作機構：

LIONBRIDGE

本地化的未來 | www.lionbridge.com



3

現代本地化服務簡介
語言和文化

4

市場概況
本地化躍為主流

5

發揮 AI 的功用
什麼是加強式翻譯？

7

個案研究
Lionbridge 協助 Royal Caribbean 全面
檢討其本地化業務

8

服務全球客群
Zynga 運用本地化擄獲數百萬玩家的心

10

在地思維
運用語言服務提升 Wikipedia
(維基百科) 品質

12

機器翻譯的進展
轉捩點

14

個案研究
Lionbridge 協助 Volvo Cars 建立體驗
一致的線上市集

15

未來展望
語言技術的未來

17

學術觀點
人工與機器，不分軒輊

語言與文化

重點並不是語言，而是文化。好比說，如果以西班牙文為例，我們的目標對象在哪裡？是全球的西班牙文使用者嗎？還是歐洲的西班牙文使用者？我們又該如何贏得目標客群的認同？因此本地化絕對不只是翻譯字句，而是比純翻譯意義更深遠的過程。

翻譯固然重要，但了解目標客群更重要

本地化通常是指將網站、說明手冊和軟體應用程式等等資產的語言加以翻譯，方便使用對象閱讀並理解其含義。

Lionbridge 的產品主管暨策略師 Kajetan Malinowski 向 AI Business 表示：「本地化的目的是要讓產品或服務真正在地化，而熟知消費者的語言，是其中很重要的一環。」然而語言只是這整個流程的其中一個環節，想要打入全球獲利最高的某些市場，還需要具備文化上的專業能力。

不只是翻譯

您是否曾聽人說過「說了什麼並不重要，而是怎麼說才重要」？之所以不直接將本地化稱為翻譯，是有原因的。本地化是指將在地溝通的許多面向都納入考量，所以語言是其中的一環，地區差異同樣也是其中的一環。度量衡單位就是個簡單的例子，不同地方使用的標準都不盡相同。至於法規或技術方面的事務，甚至是「語氣」或語域等，也是其他會隨地點而有差異的元素。

NI (National Instruments) 的語言服務資深經理 Inés Rubio 便向 AI Business 表示：「我們基本上是以另一個語言重新打造體驗。重點並不是語言，而是文化。以西班牙文為例吧，我們的目標對象在哪裡？是要用全球通用的西班牙文？還是歐洲西班牙文？我們又該如何贏得目標客群的認同？這絕對不只是翻譯字句而已，而是遠比翻譯意義更深遠的過程。」

為了迎合目標客群的偏好，本地化也可能要處理例如介面圖示等圖片的翻譯。但這類圖示不一定有全球通用或普遍一致的意義。舉例來說，美國或英國這些市場常常會用燈泡代表想法或點子，但這個符號對其他地方的人來說並沒有這個意涵。又比如「拇指朝上」的表情符號，在美國或英國雖有正面的意義，卻會冒犯到拉丁美洲、中東和西非的人。同一種顏色在不同地方代表的意義也不一樣，例如橘色在印度等同於愛與勇氣，在中東文化中卻常與失去和哀悼有關。

成功的本地化可以帶來非常多好處；而如果沒做好或未能了解在地習慣的重要性，則可能會招來嚴重傷害。因此，對任何現代企業來說，本地化都是他們在 2020 年代不可或缺的重要工具。



本地化躍為主流

「一開始你根本不會想到本地化。」她接著說：「等到最後一刻，你才會想起本地化，它根本不會是初期作業的考量。但隨著公司與整個團隊慢慢從錯誤中學習，大家才會漸漸發現事後再做這類改變的代價有多大，然後才會開始將本地化納入整個流程。」

了解本地化市場的規模，及其成長背後的驅動因素

現代的本地化 (localization) 誕生於 1980 年代。它通常簡稱為「l10n」，其中 10 代表了英文字母 l 與 n 之間的字母數量。本地化出現的時機並不是偶然，當時，正值電腦軟體逐漸成為全球辦公與居家不可或缺的工具，為了把握這個日漸成長的市場商機，公司和企業必然得運用本地化這項利器，才能支援新的字元集、為特定市場增添或修改功能等。儘管最初都是組織內部自行處理本地化業務，但隨著作業的規模與重要性與日俱增，本地化也因此迅速成為一項成熟的產業。

規模遠勝以往

在過去十年內，由諮詢、翻譯及更廣泛的本地化業務所構成的語言服務市場，其規模幾乎翻漲了一倍。語言服務產業領導出版商 Sclator 近期的報告便指出，2020 年翻譯、本地化與口譯的潛在市場總值便高達 238 億美元，其中科技、遊戲、媒體及生命科學是成長速度最快的客戶垂直市場，並預估整個產業在 2021 年可望成長 10%。

根據 Statista，儘管歐洲人口不到全球的 10%，卻佔全球語言服務市場約一半之多，主要是因為該地區的語言種類非常多樣。若以市場規模來看，仍舊是美國獨占鰲頭。

這樣的差異並不只存在於國家/地區之間：本地化在某些產業雖然已是非常成熟的領域，但對其他產業 (特別是新興產業) 和公司而言，其重要程度仍在持續發展中。

NI 語言服務資深經理 Inés Rubio 認為，對本地化方面的投入程度，可以視為業務成熟與否的表徵。「一開始你根本不會想到本地化。」她接著說：「等到最後一刻，你才會想起本地化，它根本不會是初期作業的考量。但隨著公司與整個團隊慢慢從錯誤中學習，大家才會漸漸發現事後再做這類改變的代價有多大，然後才會開始將本地化納入整個流程。」

拜本地化相關技術進步之賜，推動產業前行的工具越來越優異。諸如神經機器翻譯 (NMT) 等創新技術雖然出現不過短短幾年，卻已展現出非常大的影響力。對客戶以及最終使用者來說，這些技術上的突破，將會持續改變本地化這片天地。





什麼是加強式翻譯？

與 Lionbridge 的 Will Rowlands-Rees 及 Jaime Punishill 一同探究人類與機器的關係

Lionbridge 產品長 Will Rowlands-Rees 向 AI Business 表示：「機器翻譯是人工智慧 (AI) 的其中一個基本應用。機器翻譯最早出現於 1950 年代初期，即使在那時，人們就預言未來可能再也不會需要人類譯者，這樣的預言當然是有點過頭了。」

將近 75 年後，人類在翻譯上仍舊扮演了非常重要的角色，連那些運用了機器翻譯的工作上也是如此。儘管在特定情境下，免費的公用翻譯引擎可以發揮一定的效用，但還是難以應用於商業上。

Lionbridge 的行銷長 Jaime Punishill 接受 AI Business 訪問時表示：「我們也都看過，公開的機器翻譯引擎會產出什麼奇怪的譯文來。那些公開引擎都沒有經過訓練，而商用系統的環境雖然好一些，可是兩種語言之間的翻譯，還是會出現很詭異的語言問題。從英文翻譯至韓文已經不容易，從德文翻到韓文更困難。機器翻譯的結果有時就是很奇怪。」

高品質翻譯的重要性

如果只是想大略了解一篇新聞的內容，那麼機器翻譯的結果或許「可以接受」。但對其他某情況來說，「可以接受」的翻譯其實是不夠的。比如說投資產品的說明書、醫學研究的標籤或是法律內容等，都是必須精準翻譯的例子。

Rowlands-Rees 說：「假設我要為某個頂尖跑車製造商製作手冊，萬一相關說明有錯誤，就可能發生健康安全上的問題。或者，想想臨床研究或其他任何攸關生死的領域。而從公司的角度來看，任何事情都會影響到品牌，所以自然也要確認翻譯正確無誤，否則可能會給品牌帶來負面的後果。」

這個流程習慣上稱為譯後編修，亦即先用機器翻譯工具進行初步的翻譯，再將這些未經編修的 MT 段落交由語言專家編輯，修正電腦所犯的任何錯誤。這種做法理論上可以節省不少時間，但也有其侷限，比如說它只在流程的最後才加入譯者的人工處理。

解方：加強式翻譯

加強式翻譯 (Augmented Translation) 這個詞最早是由 CSA Research 所提出，是指運用多種技術來協助語言專家更快速、更一致且更正確地產出高品質的翻譯。加強式翻譯能藉由人工與技術間更優異的搭配，進一步提升翻譯的體驗。其中的許多做法其實早已經過廣泛的探究，但將不同方法連結組合在一起，就能打造出更為完善的架構。這種方式能將許多一般視為低價值但又佔用許多時間的工作自動化，讓語言專家得以充分發揮他們的專業能力。➤

6

►加強式翻譯涵蓋七個主要領域：

翻譯記憶庫

適應性神經機器翻譯

品質估算

自動化內容充實 (ACE)

術語管理

全天候專案管理

翻譯管理系統 (TMS)

所有這些工作都是相輔相成的。舉例來說，翻譯記憶庫對重複的翻譯工作會有很大的助益。Punishill 就說：「我們每年都可能要翻譯同樣的產品手冊，畢竟多數產品手冊的內容都跟前一年的幾乎一樣。但在過去沒有數位化的類比世界中，公司和企業等於是要花錢再翻一次或更多次。而翻譯記憶庫的設計宗旨，就是要檢查來源文字，比對看看有沒有符合的既有譯文。」

如果有完全符合的字句，這段文字就可直接略過，不需要再做任何修改；另一方面，如果有不一樣的地方，系統就會將這些文字標記為需要由譯者更新，讓譯者得以檢查前後的字句，確認新增添的元素不會影響到內文的意涵。

加強式翻譯的另一個特點，就是可以自動判斷最適合某個特定翻譯工作使用的人工和機器資源：可以是將工作轉交由譯者處理，又或者是找出哪一個機器翻譯引擎能產出最佳結果。

發揮所長

Punishill 接著表示：「在這個領域你會發現一件事，那就是不同的機器翻譯引擎，所擅長的事項也不相同。某個引擎英文到德文翻得好，另一個則是西班牙文到韓文；某個引擎製藥內容翻得好，另一個則是行銷內容。所以你得要組合搭配不同的 MT 引擎，結合多種能力讓你的翻譯系統更強大。」

「Lionbridge 能協助公司和企業設置相關事務，並為他們執行系統，其中包括所有客戶需要做的事，從訓練引擎，一直到維護品牌和調性指南、術語辭典以及所有運作所需的語言資產等，我們都能提供相關服務，確保系統順暢運作。這意味著我們不只是能提供高品質翻譯，更能根據客戶要求的調性翻譯，例如以 Apple 或是 Microsoft 的調性翻譯。」

Rowlands-Rees 指出：「我們可以明確地告訴使用者，『因為如此這般，所以這份文件最好採用人工翻譯』，或是『我們建議可以交由 MT 處理，而 MT

可以產出不錯結果的原因是如此』。透過經驗，我們知道為什麼有時人工翻譯會比較好，為什麼有時又是 MT 會比較好。客戶提供內容後，我們根據流程便能了解與辨別適用的作法，並接著內容轉交給最合適的資源處理，不但能提高我們的效率，更重要的是也有助提升客戶的效率。」

加強式翻譯也能以其他方式支援翻譯工作，例如運用預測分析來評估哪些內容的翻譯效果會比較好，進而改善前後期的內容創製作業。此外，也可以針對例如其他語言的搜尋引擎最佳化 (SEO) 等領域進行全面性的評估，這對提高線上可見度會有很大的助益。

加強式翻譯是非常尖端的一種機器翻譯，可以大幅提高產出結果的品質。透過將例如翻譯與品質估算 (亦即針對機器翻譯的品質有多高進行統計估算) 等技術，與最新的適應性神經機器翻譯工具相結合，加強式翻譯將會是任何本地化團隊裝備庫中的強大工具，相信日後也會更加普遍應用。



Lionbridge 協助 Royal Caribbean 全面檢討其本地化作業能力

一流遊輪營運商希望提高對全球客群的吸引力

一個在全球擁有眾多航線的遊輪公司，自然是間國際企業。全球每年約有 530 萬旅客搭乘 Royal Caribbean 遊輪欣賞異國美景，為了確保旅客能如願享受精彩體驗，並維持一致的 Royal Caribbean 品牌調性，該公司必須要以高品質的翻譯成果呈現其網站 royalcaribbean.com，讓使用者得以便利地預訂假期，並盡情探索 Royal Caribbean 提供的各式遊輪、活動與目的地。

所欲解決的問題

對 Royal Caribbean 這樣的知名品牌來說，繁複又龐雜的內容翻譯工作流程，可能會對其內容的正確性及溝通有效性產生不良影響。因此，該公司決定要更新其內容平台，並向本地化產業中聲譽卓著的專家尋求協助。Royal Caribbean 最後選擇了 Lionbridge，希望借重後者在語言服務上超過 25 年的深厚經驗。該公司的目標，是要運用集中的內容網站來提高網頁流量與銷售量、改善翻譯工作流程，並提供更優異的使用者體驗。

採用的解決方案

最後的解決方案採用了由兩種工具組成的服務：Lionbridge 的 Connector™ 與 Adobe Experience Manager™。這個解決方案可以提供 Royal Caribbean 所需的高效工作流程和多語言專業能力，確保他們可以妥善地更新其內容，亦能針對在地搜尋進行最佳化。我們成功地在可接受的時程內，以合理的成本和高度能力的語言推出相關內容。在短短四個月內，Royal Caribbean 便將超過 80% 的現有內容（合計約 1,500 頁）翻譯至七種語言。

Royal Caribbean Cruise Lines 產品內容最佳化經理 Ariadna Castro 表示：「Lionbridge 的 Adobe Experience Manager 連線程式能將網頁翻譯流程自動化，因此我們可以輕鬆翻譯數

以千計的網頁。這項由 Lionbridge 提供的技術，協助我們實施高效率的翻譯工作流程。他們的客戶團隊也非常優秀，隨時都準備好支援我們達成目標。」

本地化專案最重要的任務，應該是要協助客戶提供更好的使用者體驗。

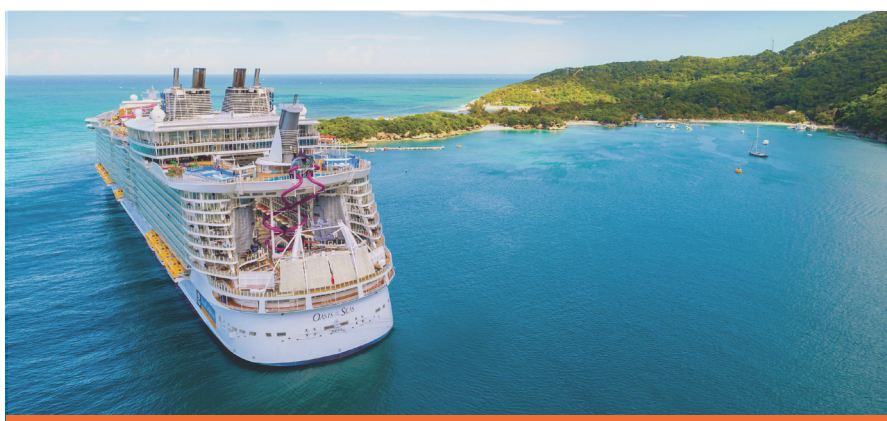
然而，這些專案也必須要能轉化為實際的成果，證明這個投資對公司和企業甚有助益。以 Royal Caribbean 為例，很快就看到立竿見影的成效。在採用解決方案的頭幾年內，該公司的自

然點閱數就成長了 35%，自然搜尋曝光數更是增加了 37%，全球點閱數則是令人驚艷地成長了三倍。

這次優異的本地化作業能力檢討，順利為全球旅客打造了更為個人化的體驗，使每一位相關人士都得以享受成功的果實。

1,500

頁的內容，由 Lionbridge 在 4 個月內翻譯至 7 種語言



Zynga 運用本地化擄獲數百萬 玩家的心

“ Zynga 發行部門總裁 Bernard Kim 表示：「我們儘量不用本地化一詞，而會以『文化在地化』稱之。也就是說，我們除了在語言上改變產品呈現文字內容的方式，在遊戲方式以及互動交流功能上，更會儘量靈活地配合在地市場進行調適。」

發行部門總裁 Bernard Kim 及策略合作關係主管 Vineet Rajosi Sharma 分享他們對開發能吸引全球客群之遊戲的看法

遊戲一向廣受世人喜愛，是能跨越地理藩籬的娛樂活動。而想在目前激烈的遊戲市場蓬勃發展，即使是大型遊戲開發商也得效法最新趨勢。創立於 2007 年的 Zynga，是全球一流的社群遊戲開發商，每一天全球都有數百萬玩家玩該公司的遊戲。他們最著名的遊戲《FarmVille》於 2009 年 6 月在 Facebook 上推出後，短短數週內便累積了多達 1 千萬的日活躍使用者。十多年後，大部分業務已移轉至行動裝置的 Zynga，依舊在超過 150 個國家/地區擁有數以千萬計的每月玩家，旗下熱門遊戲包括《Zynga Poker》、《Words With Friends 2》、《Game of Thrones Slot Casino》和《Harry Potter Puzzles and Spells》等。

他們之所能創造出這麼多成功的「長青遊戲」(該公司這麼稱呼他們歷久不衰的作品)，其中一個關鍵就是特別注重依全球不同的市場來調整其產品。畢竟，以英文為母語或第二語言的人，只佔全球人口的十二分之一不到，因此只有英文的遊戲體驗，必定會錯失大部分的潛在客群。所以從商業的角度來看，以多種語言呈現遊戲其實是非常合理的決策。但正如 Zynga 主管對 AI Business 所言，成功的本地化，絕不只是翻譯遊戲內的文字那麼簡單。

所謂的「文化在地化」

Zynga 發行部門總裁 Bernard Kim 表示：「我們儘量不用本地化一詞，而會以『文化在地化』稱之。也就是說，我們除了在語言上改變產品呈現文字內容的方式，在遊戲方式以及互動交流功能上，更會儘量靈活地配合在地市場進行調適。舉例來說，韓國的玩家一天內經常會互動交流許多次，並特別投入我們的某一款遊戲，所以我們也會根據在地市場的文化，特別為這些玩家推出一些優惠和挑戰。」

Zynga 策略合作關係主管 Vineet Rajosi Sharma 向 AI Business 表示，讓玩家感到遊戲好像是專為他們設計似的這點非常重要：「假如我是使用者，我在跟一個應用程式互動時，應該要有這是專門為我而打造的感觉。」 >



▶ 要讓遊戲能同時贏得多個國家/地區玩家的青睞，又要給人個人化的感受，可絕對不是件容易的事。在遊戲產業有著豐富資歷的 Kim 表示，他就曾遇到過非常多 IP 難以在世界各地取得耀眼成效的挑戰。而他學到的其中一個教訓，就是在為某個市場本地化遊戲時，不能給人「臨時抱佛腳」的感覺。

「玩家都知道是否能長久地玩某個遊戲」，Kim 接著說：「感受、語言、內容的呈現方式等，這些事情如有任何一件是最後一刻才為特定市場完成，從之後的投入數據就能看得出負面影響。」

為了解決這個問題，Zynga 的任何新專案在一開始時就會將本地化（或是文化在地化）策略納入考量。重點就是要「把餅做大」，儘量擴大潛在玩家人數。

Sharma 說：「看我們那些長青遊戲就知道，它們對全球客群深具吸引力。在選擇要呈現哪些概念時，他說：「我們會將文化在地化視為非常重要的優先任務。應用程式內要呈現的所有面向，都必須要能為市場客群所理解。」

語言：很重要，但不是唯一要素

語言無疑是 Zynga 的重要考量，所以他們在世界各地廣設工作室，在美國、加拿大、印度、愛爾蘭和土耳其都設有辦事處。Sharma 大方分享了一個能讓本地化流程更簡單的訣竅：減少文字量。他說：「如果在遊戲中放入太多文字，基本上就是給自己的工作找麻煩。而且也沒有人喜歡閱讀那麼多文字。」對螢幕較小的行動裝置來說，此言更是中肯。

讓事情更具挑戰性的是，任何文本的翻譯，還必須要能傳達原版遊戲的「調性」。除了遊戲內的文本與註釋之外，這也包括虛構世界中的各種元素，像是貨幣、度量衡、日期格式等，此外還有應用程式商店的說明、行銷素材以及其他必須翻譯的相關文本。Sharma 解釋道：「不管是內部還是外包，都需要有譯者處理。我們真正要追求的是翻譯的品質。我們會進行本地化品管，我們又稱之為「國際品管」，好確保所拿到的翻譯符合原文意旨而且合情合理。」

文化在地化的範疇非常廣泛，語言的書面和口語形式，只不過是這座龐大冰山的一角。舉例來說，特定的符號、性別的處理方式以及稱謂等，都是必須費心思量的事項。跟在地常用的社群媒體平台整合也是如此，無論是為了行銷目的、認證還是社群分享等，一樣也要審慎處理。Kim 表示很多事情會因市場而異，其中一個最值得留意的細節，就是遊戲內「App 內購買方案的形式」，常會為了吸引特定地區的玩家購買而有所更改。過去該公司就曾為了確保能提供最佳的體驗，而暫緩遊戲的推出，例如亞洲特定市場的哈利波特系列遊戲。

Kim 說：「這就是我們用人工翻譯的原因，我們的譯者都非常熟悉所負責的遊戲，而且也在這些遊戲上工作了非常久的時間。」

本地化對免費遊戲的重要性

Zynga 的主管表示，本地化對免費遊戲這個領域而言更為重要。付費遊戲是指玩家以固定價格購買遊戲，但所謂的「免費增值」(freemium) 遊戲，則是靠一再吸引玩家回流來獲利。換句話說，光是表面的本地化還不夠，例如花心思推出很成功的在地行銷廣



告活動，卻只在遊戲內做小幅度的更動。Kim 表示，免費增值遊戲能確保遊戲具有一定的品質，因為在競爭激烈的市場中，玩家只會一再玩那些能夠深深吸引他們的遊戲，而這通常會是那些採用本地化方法開發的遊戲。此外根據 Statista 的統計，全球現在大概有 38 億人擁有智慧型手機，佔全球人口約 48.33% 之多，因此對開發手遊的工作室來說，本地化自然會是非常明智的投資。

Kim 表示：「整體而言，我們認為文化在地化是需要創意的作法，它囊括了整個過程，從使用者第一次聽聞我們的遊戲 - 可能是從廣告文案或是在地的行銷廣告活動得知，一直到運用客戶關係管理 (CRM) 接觸並吸引他們成為玩家。比如說要透過電子郵件還是推播通知？什麼時候發？總不能在人家睡覺的時候發送推播，也不能以錯誤的語言傳送訊息給他們吧？凡此種種，都是構成文化在地化的要素。」

運用語言服務提升 Wikipedia 品質

Albon 表示：「運用機器學習的目的... 是要使用模型讓讀者和編輯者享有更好的體驗。比如說，有的模型會去檢查並試著預測某個編輯的品質，有的則會檢查某個編輯是不是所謂的破壞性編輯。」

維基媒體基金會機器學習總監 Chris Albon 探討專案使命及在地資料的重要性。

維基百科 (Wikipedia) 擁有 300 種不同語言共超過 4 千萬個條目，每月的觀看讀者超過 5 億人。做為全球最大的使用者創作百科全書，自然應該儘可能讓越多人理解越好。對各地的使用者而言，它已經是非常寶貴的資源，但本地化服務是否能協助它更上層樓，有更優異的表現？這便是維基百科的創立者所面臨的大哉問。至於答案，他們表示：當然可以。

想要確保每一個 Wikipedia 版本都儘可能一樣的全面且完備，最直覺的作法就是把所有內容彙整在一起，然後進行自動翻譯，將例如英文的條目 (英文是規模最大的 Wiki 版本) 翻譯至使用者數目較少的其他語言。

維基媒體基金會 (Wikimedia Foundation) 採納了部分這樣的作法，使用由基金會贊助的內容翻譯工具，先將某個 Wikipedia 條目進行初步的機器翻譯，再由編輯人員審閱並視需要修改，然後才發表條目。人工編修非常重要，因為機器翻譯如果未加管理，可能會造成不少問題。以某次知名的錯誤為例，英文的「village pump」在葡萄牙文中就被翻成「轟炸村莊」之意。在流程中納入人類的參與，就可以將這類錯誤的發生降至最低。到目前為止，這個自動化工具已經協助翻譯超過五十萬個條目。

然而，正如 Wikipedia 機器學習總監 Chris Albon 向 AI Business 所說，該組織並沒有全面啟用自動翻譯，而是採用了另一種「真正能融入在地社群意見的作法」。

本地化的機器學習

Wikipedia 對機器學習模型的使用，比許多使用者以為的還來得廣泛。Albon 表示：「運用機器學習的目的..... 是要使用模型讓讀者和編輯者享有更好的體驗。比如說，有的模型會去檢查並試著預測某個編輯的品質，有的則會檢查某個編輯是不是所謂的破壞性編輯。」

為了維持頁面順暢易讀，Wikipedia 開發了數以百計複雜程度不一的「機器人」，用以執行從條目建議到取締破壞行為等各式各樣的工作。雖然 Albon 及其同事 (他們本身不會為 Wikipedia 撰寫任何內容) 的一切行事都是為了這個免費百科全書著想，他們也同時想避免將自身的偏見植入所建置的 AI 模型內。

對 Wikipedia 來說，什麼叫做偏見？簡單地說，關於 Wikipedia 需要什麼或怎樣才算好的 Wikipedia 條目，一個住在舊金山的使用者無論對此有什麼看法，都不能自動套用到住在肯亞、文化與習慣均相異的 Wikipedia 使用者身上。 >





Albon 解釋道：「到目前為止，我看最多的顯然是英文版的 Wikipedia。所以對於何謂優秀條目、條目應該要有哪些特定主題、條目該有什麼樣的架構，或是撰寫條目時應該採用什麼語氣等，我對所有這些事情的看法，全都來自我在 Wikipedia 專案中單一個社群的經驗。但這些想法不能完全套用到所有其他不同的社群上。每個參與 Wikipedia 專案的人都有廣泛的共識，但對於何謂高品質的編輯、什麼算是破壞性編輯，或是該怎麼處理特定主題，這些看法都會因地區而異，所以這些事情也應該由在地語言所屬的在地社群管理。」

本地化是 Wikipedia 使命的一環

Wikipedia 目前在約 30 個不同的 Wiki 中部署了大約 100 個機器學習模型，而且完全是用在地社群的資料進行訓練。Albon 表示：「本地化對我們基金會之所以如此重要，因為它真的涵蓋了整個專案」。他指的是為全球所有人以其母語提供免費的百科全書這個目標。

之所以要用地資料訓練各個模型，是因為不同國家/地區對高品質 Wikipedia 條目的構成條件，都有不同的期盼。

Albon 說：「這很重要，比如說，什麼對條目很重要或不重要、必須要連結哪些項目，或是尚未開發的主題等，斯瓦希里語社群的看法，可能就會因為在地脈絡而跟英文版 Wikipedia 不一樣。舉例來說，英文版 Wikipedia 的使用者可能不會覺得 matatu (私營的小巴士) 這類條目有什麼重要；但斯瓦希里版的 Wikipedia 使用者對這個條目就會深感興趣，因為這是東非、南非和西非人民日常生活中不可或缺的交通工具。如果我們用英文版 Wikipedia 創建模型，就會將英文版的價值觀強加在斯瓦希里版的 Wikipedia 上，這是我們所不樂見的事。」



再以 Wikipedia 使用的其中一個機器學習模型為例，這個模型的功能是針對條目內容中的詞彙提出連結至其他頁面的建議。不過，由於 Wikipedia 中大部分的名詞都有自己的條目頁面，因此這麼做可能會創建出內容中滿是超連結的條目，不但讓網頁看來雜亂無章，也容易使讀者錯過可能相關的連結。所以 Wikipedia 在這方面採取的立場，就是如果詞彙是種「常識」，也就是訪客幾乎都知道的基本知識，就不會加上超連結。但問題在於：何謂常識，往往會因地區而不同。所以部署 AI 模型執行這項作業時，就必須採用本地化作法來判斷詞彙是否為常識。

金科玉律

Albon 表示：「我們製作模型的金科玉律，就是機器學習模型日後要應用於哪個社群，就要用來自那個社群的資料訓練該模型。比如說，假設斯瓦希里版的 Wikipedia 使用者來找我們，說想要個『能偵測編輯品質的模型』，根據我們的金科玉律，我們要收集該社群認為是高品質的條目，然後用這些資料訓練模型，再將訓練好的模型提供他們使用。」

改善不同 Wikipedia 版本的品質，能提高它們對使用者的實用性，進而吸引更多使用者，而這些新使用者又能提供他們的意見反應，讓在地訓練過的機器學習模型更加聰明、更貼近在地文化。這種獨特的「良性循環」本地化作法，就是這個寶貴的線上資源可以更加優異的關鍵。

Albon 說：「所以對我們來說，讓社群參與這類事務非常重要。如果不這麼做，就會讓來自特定地區、對 Wikipedia 有特定看法的一群人，強加他們的意志在所有其他人之上。人們常常說要透過機器學習模型來消除偏見，但這正是我們不想做的事。我們希望社群能對他們使用的模型有話語權，更希望他們能參與建立與管理這些模型的流程。因為到頭來，我們是要運用這些 AI 模型，讓社群享有更優異的體驗。」



轉捩點

與 Lionbridge 的 Rafa Moral 及 Kajetan Malionowski 一同探究機器學習和語言的關係

翻譯一直是人工智慧發展的重點之一。早在 1933 年，距離人工智慧 (AI) 成為自成一格的領域還有二十多年，蘇聯科學家 Peter Troyanskii 就向蘇聯科學院 (Academy of Sciences of the USSR) 展示了「能在翻譯語言時選取和列示字詞的機器」。而 1964 年的紐約世界博覽會上，IBM 則自豪地公開了一個機器翻譯工具，可以在英文與俄文間進行初階的翻譯。

當然，我們可以將機器翻譯視為是 AI 大旗下眾多子領域中的一個，但 Lionbridge 創新部門副總裁 Rafa Moral 卻認為，機器翻譯是這個科技的核心主題。Moral 向 AI Business 表示：「AI 旨在模仿人類智慧，而沒有什麼比語言更能稱得上是人類的智慧了。假設有 AI 研究人員想做出個行為像人類的東西，那麼語言正好就符合這個概念。所以我才認為打從一開始，語言就是 AI 研究的主題，而且可能是很主要的主题。」

突飛猛進的機器翻譯

和 AI 的其他領域一樣，機器翻譯的演進也經歷過非常多變革，不但為近幾年的突飛猛進打下穩固的基礎，更使得這項技術從前途看好的研發 (R&D) 工作，晉升為可充分獨當一面的工具。

其中一個突破性的發展，就是詞向量 (word embedding，又稱為詞嵌入) 這個概念的出現。詞向量隸屬於機器學習與深入學習的工具組，是一種詞彙表示的方法，是指意義相近的字詞可以用相近的方式表示。詞向量這種處理方式，能自然地擷取字詞的意義，進而大幅提高翻譯處理過程的品質。它和之前作法的不同之處，在於之前不同的字詞是以各自完全不同的方式表示，即使意義相近的字詞亦是如此。

Moral 說：「在詞向量這種方法出現之前，基本上是将字詞視為文字字串。而詞向量的概念引進了向量的觀念，定義了如何將字詞轉為數字，同時這些數字還可以具備語義意義。詞向量的重點是要根據上下文，將數字或向量與某個字詞相連結。」

神經機器翻譯：明顯的轉捩點

近年最重要的其中一個突破，便是神經機器翻譯的出現，Moral 更稱之為是業界「明顯的轉捩點」。神經機器翻譯 (NMT) 與之前的規則式 MT (RbMT) 和統計式 MT (SMT) 都不相同，是仰賴神經網路這個以人腦為靈感、也是現代機器學習技術主要推手的工具，以完全自動化的方法進行機器翻譯。NMT 並不是單獨分析每一個詞彙，而會將字詞的上下文納入考量，因此能產出更優異、更正確的翻譯。也就是說，這個技術看的不只是詞彙，更會去了解詞彙間的關係。

諸如 NMT 等技術，使得 Lionbridge 等企業得以將其納入裝備庫，藉由更高品質的文件和其他資料翻譯能力，推出突飛猛進的本地化服務。➤

► Lionbridge 產品主管暨策略師

Kajetan Malinowski 向 AI Business 表示：「現在的內容之多，比全世界所有譯者能處理的量還要多上數百倍，早已遠超出我們所能翻譯的量。從這點來看，機器翻譯正在推動我們這個產業革新，協助我們管理更多內容，讓人們得以專注在他們可以增添更多價值的事物上。」

跟 Moral 一樣，Malinowski 也表示諸如 NMT 等方法的出現，對自動化翻譯非常重要。但高品質翻譯的其中一個重點，就在於知道什麼時候該使用什麼方法，畢竟有些方法更適合特定類型的內容。

選擇最合適的方法，取得最佳成果

Malinowski 說：「內容的類型、用途各不相同。比如說有的內容是暫時性的而且能見度低，這種內容不需要極為精確的翻譯就能使用；又比如說其他類型的內容，像是行銷文案、廣告或非常特殊的法律、醫療保健或法規文件等，就需要極為高品質的翻譯。而有了機器翻譯，我們便能處理更多內容，並將人力運用在那些「附加價值」更高的工作上。」

以 Lionbridge 為例，該公司在開始著手進行翻譯之前，會先在 AI 工具的輔助下，決定能產出最佳成果的方法。

Malinowski 表示：「我們採用的混合式流程，除了能充分善用各個方法的優點，也會試著辨別下列所有這些問題：詞彙密度與多樣性、複雜度和變異性、品質等。這就是我們的出發點。一旦清楚了解客戶內容的特性後，我們就能接著找出最適合這個內容的處理作法，也就是工作流程，包括最合適的翻譯記憶庫、最適合所屬語言或領域的機器翻譯引擎，以及最適用於這個內容類型的工具組合。」



一旦清楚了解客戶內容的特性後，我們就能接著找出最適合這個內容的處理作法，也就是工作流程，包括最合適的翻譯記憶庫、最適合所屬語言或領域的機器翻譯引擎，以及最適用於這個內容類型的工具組合。

人類的參與

正如本電子書關於加強式翻譯 (第 5 頁) 的敘述，在許多情況下，最好的作法通常是將機器翻譯與人工專業能力相結合。

Malinowski 說：「機器翻譯引擎已經足以所有能見度低、只需要一般水準的內容類型。如果需要稍微好一點的品質，可以由人類介入，協助改善這些內容。這樣一來，我們人類便可以將精力專注在那些高價值、更需要投入心力的內容，比如說需要更通順好讀、需要有特定風格，或需要能在目標客群中引起特定情感等目標。機器翻譯做不到這些事情，至少現在還不能。」

雖然語言的主觀性，以及對何謂優質翻譯莫衷一是，使得即便是最優異的機器翻譯工具，仍持續受到人們質疑，但這個領域的進步對現代翻譯還是大有益處。Malinowski 便說：「人類產出的內容量呈級數般暴增，遠超過我們過去一兩年前所產出的量，而且還在持續增加。機器翻譯以及我們手上的 AI 和機器學習工具，都能協助我們創製更多內容、與更多目標對象溝通，並更妥善地處理這些內容。」

Lionbridge 協助 Volvo Cars 建立體驗一致的線上市集

我們必須解決目前網站外觀及風格各自為政的問題，建立一個內容與風格趨於一致的單一品牌平台。

該公司原本要維護約 100 個不同的網站

Volvo Cars 決定要在 2030 年前轉型為純電動車品牌，並全面採用數位線上銷售模式。因此，他們與 Lionbridge 以及商業服務及雲端解決方案供應商 Avanade 攜手合作，進行任何必要的改革，包括全面檢討該公司的線上能見度，並以數位市集而非行銷工具的角度重新規劃其網站。為了實現此目標，就必須打造能為全球潛在客戶提供實用資訊、令他們感到滿意的使用者體驗。

待解決的問題

在進行全面檢討之前，Volvo Cars 的網站能見度雖然很高，但卻極為分散瑣碎。該公司原本要維護約 100 個網站，語言總數高達 45 種。這些網頁之前是由一個中央團隊創製，然後再交由在地團隊進行翻譯。這種方式嚴重影響了整體的品牌一致性，因為不同市場的網站設置與銷售策略都不盡相同。

採用的解決方案

為了改善翻譯流程效率，Volvo Cars 使用 Smartling 翻譯管理系統，並因此省下大約 1 千小時的翻譯任務管理時間。之所有能有這樣的成效，部分原因在於將原本由個別市場各自管理的翻譯記憶庫，改為集中管理與使用。該公司也重新規劃了翻譯工作流程，多加入了一個步驟，確保內容針對在地搜尋引擎最佳化 (SEO) 做最適切的調整。

集中作業與本地化：不容小覷的力量

多虧這些變革，現在 Volvo Cars 的網站在不同市場也能呈現一致的觀感。無論顧客來自英國、德國還是比利時，雖然瀏覽產品頁面時看到的是各自語言的高品質翻譯，但都能體驗一致的使用者歷程。這也讓 Volvo Cars 更容易跨越多個疆域推出廣告活動和產品。

Ernby 表示：「我們同時推出了為數眾多的廣告活動。如果採用先前的系統，這根本是不可能的任務。」

此個案研究是個很重要的實際案例，說明了本地化與集中作業如何能相輔相成，既能提供全球一致的使用者體驗，也能保留任何在地顧客要求的優勢。

1,000

採用新系統後減少的工作小時數

Volvo Cars 希望透過網站集中化，在本地化作業方面享有由上而下更全面的控制權，既能由中央監管翻譯事務，又能直接與執行這個極其重要工作的人員直接溝通。Volvo Cars 全球線上數位化部門生產作業主管

Cecilia Ernby 表示：「改採線上銷售後，我們需要重新取得品牌訊息的主導權。我們必須解決目前外觀及風格各自為政的問題，建立一個內容與風格趨於一致的單一品牌平台。」

語言技術的未來

探究穿戴式裝置、擴增實境以及超龐大語言模型的潛能

關於機器翻譯的侷限，有個經典但可能是杜撰的故事，大意是這樣的：有人將馬可福音中的一句英文「The spirit is willing, but the flesh is weak」（心靈即便充滿鬥志，肉體卻軟弱無力，即「心有餘而力不足」之意）輸入早期的某個 MT 工具翻譯至俄文，所得到的結果卻是「威士忌很濃烈，但肉卻腐爛了」。

與產業早期相比，機器翻譯和語言技術已經有了長足的進步。正如語言技術擁護者所一直深信，這個技術已逐漸成為我們生活中不可或缺的一環，但它仍舊有很長的路要走。

新一代語言技術一些特別令人印象深刻的表現，都來自於硬體。這些創新通常跟快速成長的穿戴式技術有關，在後者的推波助瀾下，可望讓使用者隨時隨地順暢地使用翻譯，甚至不會意識到自己正在使用這類服務。我們距離科幻小說中設想的「宇宙翻譯機」，也就是能立即理解任何語言的裝置，可說相去不遠了。

通往萬用翻譯裝置之路

這項技術其中一個應用範例，是語言翻譯耳機。這種裝置能將語音處理與機器翻譯相結合，就語音文句提供即時翻譯，像 Pilot Smart Ear Buds（也提供對話的逐字記錄功能）、Google Pixel Buds 以及 WT2 Plus AI 實時翻譯耳機等都是些知名的產品。

而這個創新光譜的另一端，則是視覺上的應用：Google 翻譯的擴增實境 (AR) 功能。有了這個稱為 Google 智慧鏡頭 (Google Lens) 的功能，使用者只要將智慧型手機的鏡頭對準另一種語言的標誌、菜單或其他文本片段，就可以在手機螢幕上看到疊覆顯示的翻譯，而且字體通常會和原文一樣。這些全都是由手機上安裝的機器學習工具執行，也就是說智慧鏡頭不需要網際網路連線也可以運作。該公司在 2021 年 1 月另外推出了單獨的 Google 智慧鏡頭應用程式，在 Google Play 應用程式市集的下載次數已經超過 5 億次。Google 翻譯目前的可用語言有 100 多種，Google 智慧鏡頭也支援其中大多數的語言。

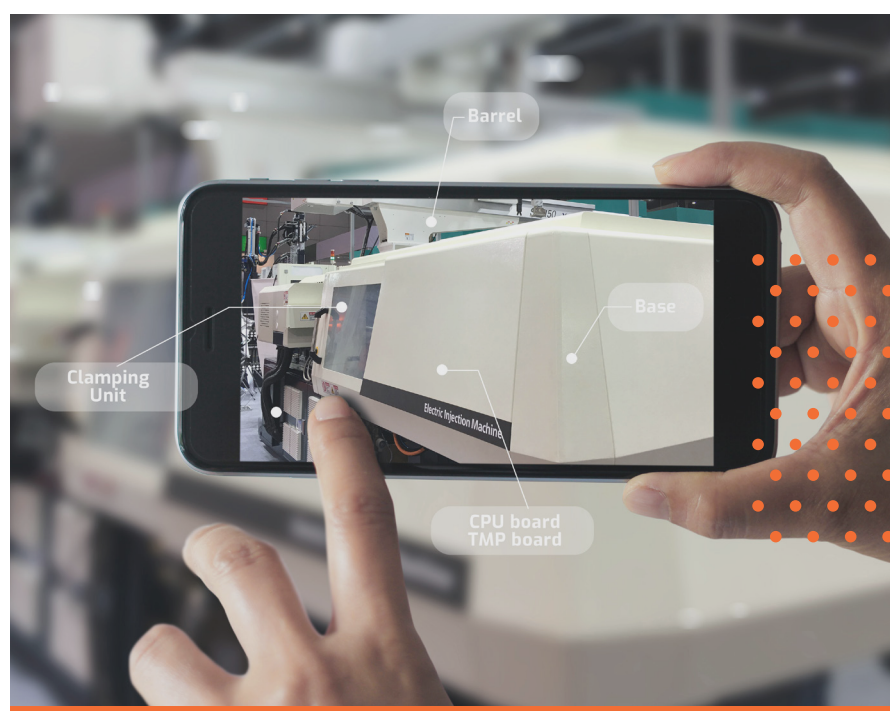
Google 智慧鏡頭雖然看似非常先進，卻是第一代採用 AR 技術的翻譯工具。智慧鏡頭最初是於 2017 年底推

出，一直到後來才獲得較為廣泛的支援。但智慧型眼鏡等抬頭顯示器的問世與廣為流行，可能會大幅扭轉這個勢態。這類穿戴式裝置將打破 AR 僅能在手機和平板電腦上使用的侷限性，讓穿戴者可更順暢地隨時隨地體驗周遭的擴增實境。從本地化的觀點來看，這也代表翻譯在未來更能輕易顯示。

檯面下的進展

公開展示當然令人印象深刻，能提醒世人語言技術的快速進展。但對需要高品質翻譯的人而言，真正的創新仍潛藏在檯面下，深深烙印在軟體中，而不是在智慧型耳機或智慧眼鏡等硬體上。

正如我們在「什麼是加強式翻譯？」(第 5 頁) 中的說明，在不久的將來，翻譯技術必然會採用混合且加 >



16

►強式的作法，將 AI 與翻譯人員相結合。Absolute Scientific Translation (ASTV) 是一家專精於智慧財產權、法律與生命科學領域的語言服務供應商，他們的執行長 Domenico Lombardini 便認為，人們經常將人類與機器視為相抗衡的兩方，這其實是個錯誤的概念。

Lombardini 向 AI Business 表示：「雖然新技術在某種程度上能抑制或減少人工干預，但它們同樣創造了新的專業角色和新的工作風格。現在有越來越多的譯者擔任機器翻譯輸出的譯後編修人員，但他們的貢獻卻非常之重要，因為機器翻譯還是很容易出錯，所以語言專家的訓練，是確保翻譯最終品質所不可或缺的條件。」

即使人類仍舊是決定翻譯品質優劣的最後仲裁者，AI 工具對加強式翻譯工作來說還是非常之實用。例如，我們可以建立推薦系統，根據譯者各自的技能組合，針對特定專案推薦合適的人選。此外，與其將整個翻譯工作交由演算法處理，專業人員可以運用 AI 來分析文件的內容，找出關鍵字與風格要求，也可以標記與分類文本，然後再交叉比對這些資訊。

正如新一代的文法或拼字檢查，機器在翻譯工作上也將能為語言專家提供更多協助，可以是修正錯字這類簡單的工作，也可以是更為複雜的工作，例如防止譯者在風格上有不一致或異常的表現、改善 SEO 分數表現，或就語言專家（及僱用他們的公司企業）應注意的文化敏感度問題發出警示等。

龐大的語言模型

運用 AI 語言模型處理其他工作，也有助於我們達成本地化的要求。2020 年最重要的 AI 進展，就是由 OpenAI 建置與開發的龐大語言模型 GPT-3。GPT-3 擁有多達 1750 億個參數，模型本身更是以來自網際網路的大量內容、高達 45 兆位元組的資料訓練。由於吸收了如此大量的資料，在預測某個字詞統計上通常會接續什麼字詞（例如「橘子」後面常常接「汁」），GPT-3 的正確性可說是令人驚艷。

善用本身極為龐大的語料訓練資料，GPT-3 可以應用於非常廣泛的用途，像是可以在不同「語氣」間轉換的翻譯工作，例如從法律文本轉換為淺白英文，反之亦然。

諸如 GPT-3 這類工具，也為語言技術開啟了新的可能性：它們不但能翻譯現有的文本，甚至還能生成全新的文稿。Lionbridge 產品主管暨策略師 Kajetan Malinowski 說：「在語言方面，除了翻譯，AI 還可以協助我們做很多事。比如說，它也能協助我們撰寫內容。我們一直在測試最近的 GPT-3 模型，它有許多使用案例，其中最值得注意的就是內容生成。」

全球的使用者無不希望能看到更細膩、更個人化的訊息，因此像大型語言模式這樣的創新技術，不但能夠針對特定國家或地區進行文本的本地化，更有可能針對個別使用者進行本地化。



語言技術的未來

機器模仿人類語言的能力到底有多好？

在 2020 年知名的研究論文〈A Set of Recommendations for Assessing Human-Machine Parity in Language Translation〉(評估語言翻譯中人機均勢狀況的一些建議) ([網頁連結](#)) 中，作者群認為「機器翻譯的品質在過去幾年有了長足的進步，在一些實證研究中跟專業人工翻譯之間甚至已經難以分辨。」

這篇文章其中兩位共同作者，來自蘇黎世大學 (University of Zurich) 的 Samuel Lübli 及都柏林城市大學 (Dublin City University) 的 Sheila Castilho 接受了 AI Business 的訪問，討論機器翻譯的能力與潛力。

AIB: 兩位在論文中指出，機器翻譯在過去幾年已經有了非常大的進步。主要是什麼原因造成這樣長足的進展？

Samuel Lübli (SL): 我認為有三個主要因素：資源的大眾與普及化、更優異的電腦運算能力，以及商業利益的提高。多虧 Paracrawl 等倡議行動，任何對建置機器翻譯系統有興趣的人，基本上都能取得大量的資料進行訓練。而採用神經網路方法的機器翻譯，概念上會比統計式等過往作法來得簡單，而且又有多種可免費使用的實作可用，因此不僅是學術應用，就算是要建置商業用的機器翻譯系統，也遠比以往來得更加簡單。

以大量的資料訓練神經模型，也需要耗用很多電腦運算能力。不過以同樣的金額，目前可以買到的設備其性能也遠勝於幾前年能買到的設備。因此即使不是 Google、Microsoft 或 Amazon 這等企業巨擘，組織或個人也有能力投入其中。此外，還有商業利益

的誘因。之前的技術早已證明能在翻譯流程上節省許多時間，這也就等於能節省更多金錢。而神經模型更是首次能產出真正通順句子的技術，優異的成果讓一般大眾都感到讚歎不已，更別提其中還有許多決策者。

AIB: 什麼是評估翻譯正確性的最佳方法？

Sheila Castilho (SC): 這可以說是個無法回答的問題。目前有數以千計關於翻譯評估的作法，而對於什麼方法好或甚至是翻譯品質的定義，許多作者都難以取得一致的意見。對我來說，評量翻譯品質最好的方法，就是先針對所要評量的翻譯文本定義出對品質的要求，然後再據此設計評量的方法。

SL: 我同意 Sheila 的看法。如果某個翻譯的目的，是要讓使用者能夠解決筆電的問題，那麼請專業譯者檢查與分類這份譯文中的語言錯誤，可能就不是很實際的作法。不過我還想再指出一點：不管做什麼事，都要在某個時候加入人類干預。如果說翻譯品質的提升有什麼意義，那就是評估翻譯品質的自動化方式，例如 BLEU ([網頁連結](#)) 等，已經越來越不可靠。

AIB: 目前有什麼最令人振奮、最具前景的翻譯技術？

SC: 我認為任何混合不同方式的作法都值得我們關注，比如說互動式譯後編修和語音。

SL: 除了採用混合式方法的翻譯，我認為未來還有兩個很值得注意的改變。首先，機器翻譯系統將能處理更多文本脈絡，也就是可以翻譯整份文件，而不是像以前一樣得先將文件切成句子、分別翻譯各個句子，然後再把這些句子組合連接起來。在研究環境下，這個問題基本上已經獲得解決 ([網頁連結](#))，而一旦商業化之後，將能提供更優異的一致性。

其次，未來我們對機器翻譯系統的產出內容將有更高的掌控能力。機器翻譯已經能夠產出令人讚嘆的高品質翻譯，但例如風格等決定，卻不是使用者所能置喙。可是在未來，你不但能夠要求系統就文本產出優雅而非口語化的翻譯 ([網頁連結](#))，更可以控制詳細程度或性別化等許多其他特色。

本訪談內容經過編輯，以求清楚簡明。





AI Business
電子書系列



LIONBRIDGE

Lionbridge 與眾多品牌攜手合作，致力在全球各地達成「通情達意，暢行無阻」的使命。過去 25 年來，我們已經為許多客戶與員工遍布全球的公司，提供超過 350 種語言的翻譯與本地化解決方案，協助他們順利獲得成功。Lionbridge 運用世界級作業平台，以及由全球各地充滿熱忱的專家所組成的網路，與各大知名品牌聯手打造豐富的文化體驗。我們憑藉著對語言的無窮熱忱，運用尖端的人類與機器智慧，打造出能讓客戶與其客戶之間產生共鳴、增進認同，並提高相互理解的優質內容。Lionbridge 總部位於美國麻州沃爾瑟姆市，在全球 23 個國家/地區設有解決方案中心。

如需深入了解，歡迎造訪 www.lionbridge.com

如需更多資訊

