

LIONBRIDGE



LIONBRIDGE 2023 年 機器翻譯報告

Lionbridge 專家精闢分析機器翻譯的趨勢，就可能的發展提出深入見解，並預測技術進步對業務蓬勃發展有何助益。

執行摘要

過去幾年，有越來越多公司企業開始採用機器翻譯 (MT) 技術。即便如此，MT 品質在 2022 年卻停滯不前，沒有多大進展。Lionbridge 機器翻譯追蹤工具是市面上最悠久的主流 MT 引擎評估工具，它的分析也證實了這一點：在 2022 年，前五大引擎的品質沒有任何顯著的改善。

這些結果可說是種前兆，預告了 2023 年會有新的機器翻譯典範出現，而隨著這個新典範逐漸站穩腳步，MT 品質也可望大幅躍進。越來越多公司開始應用機器翻譯技術來提升

客戶體驗，這可說是其來有自，因為與人工翻譯相比，MT 是快速又有效的另一種選擇，能讓公司企業即時與全球客戶交流。

在 COVID-19 疫情肆虐期間，機器翻譯也協助許多公司排除溝通障礙，提升不同語言和地理區域的員工生產力，協助品牌為其員工提供更優異的工作體驗。

過去幾年，機器翻譯的商用使用案例已明顯地成長。

以下是不同產業的公司如何運用機器翻譯獲致業務上的好處：

電子商務和零售業

零售商可以運用 MT 快速而有效率地翻譯產品說明和其他行銷宣傳品，進而向多個市場推出產品。這個技術已讓這些公司企業的跨境貿易得以極快的速度成長。

旅遊和飯店業

旅遊相關公司除了可以使用 MT 翻譯目的地場所的說明，例如飯店、租賃資產、餐廳或其他觀光景點等，也可以翻譯顧客評論。這項技術有助公司提升客戶體驗，並更快地與潛在客群建立關係。

醫療保健

醫療相關組織可以使用 MT 翻譯醫學研究和臨床試驗資料。這項技術能讓他們更快地將資訊提供給大眾，進而改善治療成效和病人結果。

法律與金融服務

這些服務供應商是最早採用 MT 的其中一個產業，主要是應用在電子事證開示程序以及市場研究活動上。MT 讓他們得以處理極為大量的多語言內容。

公共部門

政府機關可以利用 MT 這個具成本效益的方式克服語言障礙，以在地社群的語言提供他們在地市政服務。

任何產業的全球公司

對於那些工作團隊遍及全球的公司，無論是交談軟體還是簡單的文件交流，都可以在內部運用 MT 增進內部溝通成效。

他們也可以善用這項技術，協助支援團隊拓展服務範疇並縮短解決事件的所需時間。這個技術已讓一些全球公司提高辦公室的生產力。



在這個連結緊密複雜的經濟體中，全球公司若想要蓬勃發展，就得要擁抱並充分善用機器翻譯這項技術。

儘管有這些進展以及越來越多的使用案例，MT 仍舊有其侷限存在。它長久以來的一些品質問題包括：無法一致地生成正確的正式度、語氣或否定句處理。這些侷限妨礙了它的進一步成長。而對大型語言模型 (LLM) 的深入研究及應用，則可望能解決這些問題，讓機器翻譯得以在技術上再次大步躍進。

科技巨擘也紛紛投資 LLM 技術，例如 Microsoft 便大手筆地對開發 ChatGPT、GPT-3 和 GPT-4 模型的公司 OpenAI 投資了 100 億美元。這些大型投資，也加快了這項技術的開發速度，並進一步推動了自然語言處理 (NLP) 領域的發展。而這些進展，無可避免地將嚴重衝擊翻譯與本地化產業，改變公司企業製作與翻譯內容的作法。

突飛猛進的 NLP，尤其是 LLM，將會徹底改變內容製作與本地化的方式。這些進展將有助翻譯人員處理更為龐大的內容量，進而在生產力與速度上獲致大幅的提升。

而精通並能在其內容引擎中善用 AI 的公司，更可在這個日漸數位化的經濟體系中，取得極大的競爭優勢。

機器翻譯的未來展望為何？

Lionbridge 追蹤主流 MT 引擎的表現已有多年，同時也擅長運用新興技術，因此可說是分析 2023 年及之後技術發展的不二人選。我們認為，MT 現有的神經機器翻譯 (NMT) 典範將會走下舞台，而像 ChatGPT 這樣奠基於大型語言模型 (LLM) 的，可能是取而代之的新典範。事實上，GPT-4 的推出以及 LLM 的成長，便已經對業務應用產生顯著的影響。

您會看到以下的發展：

- MT 品質的大幅躍進，包括工作流程自動化
- 內容產出量增加
- 一流譯者會供不應求
- 機器翻譯採用率增加
- 機器翻譯成為提升客戶體驗 (CX) 的一種方式

在這個連結緊密複雜的經濟體中，全球公司若想要蓬勃發展，就得要擁抱並充分善用機器翻譯這個技術。歡迎繼續閱讀我們的分析，了解這項技術在 2022 年的發展及不足之處，以及它們對 2023 年以及未來數年又有何意義。

從 2022 年機器翻譯的品質表現來看，它的演進情況如何？

機器翻譯 (MT) 目前可說是炙手可熱。公司企業越來越能體認到使用 MT 技術的諸多好處，連例如生命科學這類原本比較抗拒使用 MT 技術的產業也是。

但從輸出品質來看，2022 年只是持平而沒有什麼進步，這也讓人開始期望 MT 能有令人振奮的新變革出現。

除了 Microsoft 在 2022 年年末以及 Amazon 所展現的小幅進步，主流 MT 引擎在品質上的表現其實都不相上下。此外，與前幾年的品質躍升相比，這些小幅進步實在相形失色許多。

我們在統計式 MT 典範結束前也看到類似的指標：品質進展持平，沒有什麼變化，而且所有頂尖引擎的 MT 輸出品質都趨於一致。這使我們得出一個結論：目前的神經機器翻譯 (NMT) 典範正走向終點。

而具備多模態及多語言等龐大內容量的大型語言模型 (LLM)，將會在未來扮演很重要的角色。

而精通並能在其內容引擎中善用 AI 的公司，更可在這個日漸數位化的經濟體系中，取得極大的競爭優勢。



機器翻譯入門簡介

想要充分善用機器翻譯，享有它所帶來的深遠好處，並協助公司企業實現 Localize everything™ (無譯不做、溝通無礙) 的最終願景，就必須先對這個技術的演進有基本的了解。

什麼是 MT？是什麼使它廣受全球各地採用？它有什麼主要優點，使用時又該避開哪些地雷？而它一路發展下來的背景又是什麼？

人工智慧

以最淺顯的話來說，機器翻譯運用的是人工智慧 (AI)。所謂人工智慧，是指機器在執行任務時能展現「智慧」，通常是人類本身具備的思維能力，像是學習或解決問題等。在機器翻譯這個案例中，則是使用 AI 來進行翻譯。近年來，隨著電腦運算能力日漸強大，AI 也有了大幅進展。效能更加強大的電腦不但能在執行任務期間進行更密集的運算處理，也能展現更先進的「機器學習」能力，也就是透過執行工作獲得 AI 應用所需的知識。

機器學習

機器學習 (ML) 是資訊科學的一個分支，是運用龐大的資料來教電腦執行某些任務。機器學習會檢查與某個特定任務相關的資料，從中找出模式並將資料連結在一起，接著運用這些新學到的知識，規劃電腦應該如何執行這項任務。

如果經過這樣的分析後，電腦執行這個任務的表現有所改善，就代表機器學習到東西了。由於我們擁有非常龐大的語言和本地化資料，因此人們也正運用機器學習來改善電腦在各個方面的效能，例如氣象預報、自動挑選股票，以及機器翻譯等。



機器學習會檢查與某個特定任務相關的資料，從中找出模式並將資料連結在一起，接著運用這些新學到的知識，規劃電腦應該如何執行這項任務。



什麼是機器翻譯技術？

機器翻譯是自動化的翻譯，也就是您將某個語言的來源資料輸入電腦，電腦便會以另一種語言輸出翻譯。MT 技術並不完美，但若想更有效率地產出可堪使用的翻譯，這會是最強大的工具之一。

過去數十年來，MT 在輸出的品質以及支援的語言數目上都有所改善。從只能做簡單字詞代換的早期 MT 系統、根據文法和辭典明確編碼的規則式 MT、以大量數據分析模式的統計式 MT，一直到運用深度學習和類神經網路的神經 MT，機器翻譯一路走來的發展，可說亦步亦趨地反映了我們對電腦日漸精細複雜的使用。

機器翻譯有著什麼樣的演進？

規則式機器翻譯

早期的機器翻譯使用的是規則式方法。顧名思義，規則式機器翻譯 (RBMT) 是仰賴人們或字典制訂的規則來執行翻譯。這些系統是以字彙為架構。

雖然這種方式能產出正確的翻譯，但規則式機器翻譯系統還是最適合處理那些使用簡單文法及小規模字典的簡化語言。

複雜程度越低，機器翻譯引擎執行的效果也就越好。因此在翻譯特殊領域、使用慣用語或是有歧義的文本時，規則式方法的成效就沒那麼好。自 RBMT 問世以來，語言技術也有了長足的演進。

○ MT 技術並不完美，但若想更有效率地產出可堪使用的翻譯，這會是最強大的工具之一。



學習演算法接著會建立語言模型，計算特定詞彙與片語在目標語言中比鄰出現的機率。

統計模型

統計式機器翻譯 (SMT) 會針對來源語句生成大量的候選翻譯，再根據目標語言中字詞與片語同時出現的機率，從中挑選出最好的翻譯。

SMT 是透過所謂的「N 連詞」(n-gram) 來學習翻譯，後者是指在來源與目標語言中會同時出現的一小組字詞。我們會將訓練材料輸入至 SMT 系統，也就是非常、非常多的來源語言句子，以及這些句子在目標語言中的相應翻譯。

學習演算法會把來源與目標語言的句子拆解為 N 連詞，然後找出當來源語言中特定的 N 連詞出現在句子中時，目標語言的翻譯中通常會出現哪些 N 連詞。

學習演算法接著會建立語言模型，計算特定詞彙與片語在目標語言中比鄰出現的機率。等到要翻譯新素材的時候，SMT 系統會將新的來源句子拆解成 N 連詞、找出高度相關的目標語言 N 連詞，然後開始生成候選句。

最後選定的翻譯，則會是目標語言 N 連詞與來源語言 N 連詞相關程度最高，同時其目標語言字詞在目標語言中同時出現機率最大的句子。儘管 SMT 系統並沒有將任何語言學的因子納入考量，但 SMT 的翻譯成果卻出奇地好。的確，這個系統考量的只是 N 連詞，而不是完整的句子。

混合式機器翻譯

公司企業接著便開始嘗試混合式機器翻譯 (HMT) 這種作法，也就是將統計式機器翻譯與規則式機器翻譯系統的輸出結合在一起。這些進展帶動了機器翻譯技術的普及，更將這項技術推廣到全球各地。接下來的另一個技術大躍進，則來自於更加新穎的一種 MT 方法：神經機器翻譯。

神經機器翻譯

神經機器翻譯 (NMT) 克服了 SMT 最大的缺點：後者對 N 連詞分析的依賴。NMT 能賦予機器能力，就像 SMT 一樣，這個系統也要用大量的資料來訓練，但兩者之間有個根本的不同點：NMT 系統收到訓練資料後，會自行決定如何從這些資料盡可能地學習。

NMT 系統會為每個來源句子建立資訊向量，並根據詞彙前後的字詞，為每個詞彙建立關聯資訊。有些系統可以為每個字建立數百個資訊片段，因此能產出深度的準確性。

透過深度學習，NMT 系統可以為每個字詞及來源句擷取到大量的資訊，然後再運用一種注意力模型 (Attention Model) 以及這些大量資料串流的分析，來精進之前所學習到的關鍵特徵，這對翻譯流程至關重要。

最後所生成的翻譯在流暢度上有顯著的改善，這也代表電腦產出的翻譯開始越來越自然流暢。神經機器翻譯解決了機器翻譯長久以來的一些缺點，像是不通順的自動翻譯，以及難以相容於特定語言的問題。

這個進展使得機器翻譯能產出可接受的品質，不但可用來理解大量的文件或掌握其重點，也可應用在非任務關鍵的一般商務文件上。而這三種方法中，顯然以 NMT 最為優異。它的出現提高了 MT 的使用率，加快了生產流程的腳步，可說徹底改變了機器翻譯的應用。

想像神經機器翻譯的運作方式

雖然 NMT 和統計式機器翻譯一樣，也是使用翻譯記憶庫 (TM) 加以訓練，卻能運用深度學習以及往往更為龐大的訓練資料，來建置人工神經網路。

不妨將神經機器翻譯想像成是在「培養」一個神經系統。就像我們練習彈鋼琴一樣，如果彈錯了，就會回到先前的部分再彈一次，然後這樣反覆練習直到彈對為止。神經 MT 系統也會像這樣運用其神經網路來學習。

提供訓練內容給深度學習演算法時，您不一定得要告知演算法該怎麼做。您可以讓系統自行找出模式，例如與來源句子相關的脈絡跡象。然而，就許多方面而言，具體流程仍舊是個謎。

許多處理作業會於複雜資料的「隱藏層」中完成，這表示我們難以看清神經網路的決策方式，因此也難以建置 NMT 流程的具體心像。**我們只能提交訓練內容、讓演算法完成相關作業，然後在翻譯不準確時著手調整訓練內容。**

一旦系統收到訓練素材後，會自行決定如何從這些資料盡可能地學習。

機器翻譯的歷史沿革為何？

自 1950 年代問世以來，機器翻譯已經有著長足的進步。
以下是它的一些重要里程碑。

- 1954** ○ 喬治城大學研究人員進行了史上第一次的 MT 初期系統公開展示。
- 1962** ○ 機器翻譯與計算語言學學會 (Association for Machine Translation and Computational Linguistics) 於美國成立。
- 1964** ○ 美國國家科學院 (National Academy of Sciences) 成立了專責研究 MT 的委員會 (ALPAC)。
- 1970** ○ 法國紡織研究所 (French Textile Institute) 開始使用一套 MT 系統翻譯摘要。
- 1978** ○ Systran 開始翻譯技術手冊。
- 1989** ○ Trados 率先開發與行銷翻譯記憶庫技術。
- 1991** ○ 烏克蘭的哈爾科夫大學 (Kharkov State University) 研發了第一個商用 MT 系統，可進行烏克蘭文對俄文、英文、德文之間的翻譯。
- 1996** ○ Systran 與 Babelfish 在網路上提供簡短文字的免費翻譯服務。
- 2002** ○ Lionbridge 使用旗下的規則式 MT 引擎執行首個商用 MT 專案。
- 2000 年代中期** ○ 統計式 MT 系統正式向大眾公開。Google 翻譯於 2006 年推出，而 Microsoft Live Translator 則於 2007 年問世。
- 2012** ○ Google 宣布 Google 翻譯每天翻譯的文字數量約等於一百萬本書籍。
- 2016** ○ Google 和 Microsoft 皆啟用神經機器翻譯 (NMT)，不僅有助於減少語序錯誤，更大幅改善了詞彙與文法。
- 2020** ○ 截至當年 10 月為止，Google 神經機器翻譯 (GNMT) 可支援 109 種語言。
- 2022** ○ ChatGPT，一個可以根據脈絡生成有如人類撰寫之文本的大型語言模型 (LLM)，在 11 月一夕爆紅，並對機器翻譯產生巨大的衝擊。
- 2023** ○ 隨著某一類 LLM 持續發展並導致 MT 產生劇烈改變，可能會出現重大的 MT 典範轉移。

機器翻譯有哪些好處？

機器翻譯之所以深具吸引力，主要有兩個原因：

- 它的速度飛快，能在數秒鐘內完成大量的翻譯
- 它極具成本效益，能讓公司企業翻譯超乎想像的大量內容

速度 + 成本效益 = 更多本地化的內容以及更優異的客戶體驗

自動翻譯讓公司企業能夠以同樣或較少的預算，達成 Localize Everything (無譯不做，溝通無礙) 的願景，並滿足更快向市場推送更多內容這個不斷增長的需求。

企業為何應該採用機器翻譯？

除了省時又可減少本地化成本，使用 MT 處理非關鍵任務的內容，也可讓公司企業得以：

- 將更多素材本地化，進而得以根據翻譯的語言數以及進軍的市場數拓展觸及範圍
- 改善術語和品牌調性的一致性
- 即時本地化與客戶的互動，進而提高客戶滿意度
- 消弭溝通障礙，推動全球團隊的生產力更上一層樓

機器翻譯有助公司減少翻譯上的預算，進而能夠分配更多資源和預算在內容創作上。

這樣的預算重新分配，有助公司加快內容產出速度，吸引更多讀者，加速業務成長。對數位優先的公司企業來說，這個優點尤其實際。

平均而言，Lionbridge 的客戶在工作流程中使用機器翻譯時，可以獲致以下這些成果。



更高的翻譯效率

最多可減少 40% 的成本。



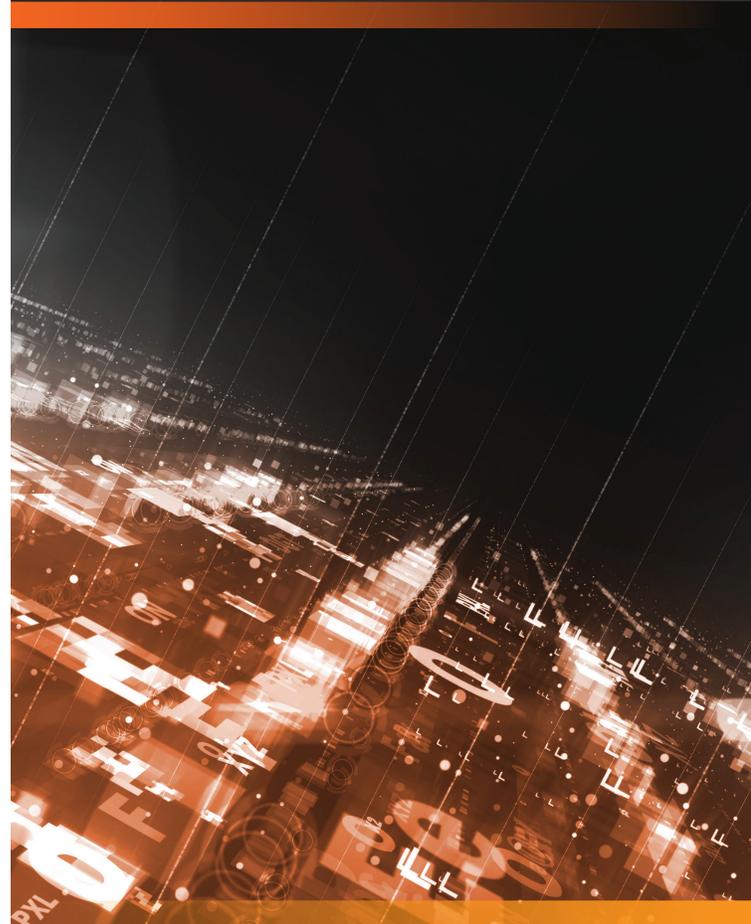
加快交付時程

最多可將交付時間縮短 60%。



提升客戶體驗

無論客戶身在何處，都能以他們的母語與之溝通。



機器翻譯最主要的缺點為何？

儘管 MT 越來越精密複雜，輸出品質也持續不斷改善，這些引擎仍有很長的一段路要走，才能像人類一樣理解細微差異、語調、諷刺、幽默、意涵等諸多語言特性。

如果完全依賴機器進行翻譯，如機器人般生硬的譯文語調可能無法適當地傳達您的訊息，反而會讓客戶覺得您提供的服務品質不良或不可靠。

公關內容的翻譯必須要流暢自然，並符合目標市場的文化脈絡，才能有效地傳遞訊息。想要製作出高品質的在地內容，多花一些錢請母語人士審閱 MT 輸出往往能帶來卓越成效，這也正是使用機器翻譯譯後編修 (MTPE) 的最佳時機。

機器翻譯加譯後編修如何能提供最佳的 MT 與人工翻譯？

機器翻譯譯後編修 (MTPE) 是將 MT 與傳統人工翻譯相結合，顧名思義，是指由人類針對機器翻譯的結果進行譯後編修。首先，資料會經過軟體處理，產生初始翻譯，接著再由譯者審閱譯文是否正確、清楚、流暢並能引起在地共鳴。這個作法不但具有 MT 飛快的速度，同時也兼具譯者注重細節與感性的優點。

機器負責快速地完成大量的翻譯工作，而譯者則檢查機器翻譯的成果，並添加只有人類才能提供的翻譯品質。如果您想要以量身打造的方法翻譯專為客戶建置的內容，並盡可能獲致最高的品質與最短的產出時間，那麼這種混合式方法會是很不錯的選擇，是取得高品質成果最快也最經濟實惠的一種作法。

機器翻譯還有哪些其他風險？

儘管 MT 有著長足的進步，但這個技術離完美還有段距離。因此，採用這個技術除了會獲得如機器人翻譯般生硬的內容外，您還得承受一些其他風險，尤其是如果不採取行動減輕 MT 缺點的話。

到底可能會有什麼問題？

使用機器翻譯時，可能會發生以下情況：

- 您的文案可能包容性不足，或無法反映您原本想展現的語言正式度，這是因為引擎並不知道什麼時候該使用中立的字詞，也不知道要配合您的目標對象而使用適當的正式度。
- 您文案的翻譯內容可能會有程度不一的錯誤。

雖然這些問題可能會造成不良的後果，但錯誤也可能會對貴公司帶來非常嚴重的傷害。



機器無法像人類一樣依脈絡判斷。

MT 風險：錯誤

MT 造成的錯誤類型

MT 引擎主要會產生兩種錯誤類型：

一般錯誤

是目標語言中跟內容的語言特徵有關的錯誤。一般錯誤包括了文法、拼寫或標點符號錯誤。儘管母語人士通常會注意到這種錯誤的存在，但它們很少會為公司招致災難性的負面影響。

重大錯誤

MT 重大錯誤往往比語言學上的錯誤來得嚴重，是指引擎的輸出內容沒有正確地傳達原文的意圖。這類錯誤導致的錯誤資訊或誤解，可能會為公司企業帶來聲譽、財務或法律上的影響，甚至對公共安全或公共衛生產生不良後果。

MT 引擎為何會產生重大錯誤

重大錯誤可以視為是 MT 引擎功能失常的結果，當引擎無法理解文本，例如一詞多義或是原文內有錯字時，就可能產生重大錯誤。如果沒有妥善訓練引擎，或是使用的詞彙表有問題，進而導致同樣的錯誤一再重複，也會產生重大錯誤。機器翻譯的引擎儘管非常精密複雜，但並不完美，因此可能會產生重大錯誤。畢竟機器無法像人類一樣依脈絡判斷。

重大錯誤一般會發生於三種主要的脈絡情況：

重要實體的誤譯

這種錯誤是指專有名詞(個人或組織)、重要數字或度量衡單位的翻譯錯誤。

否定和相反的意思

這種錯誤是指目標語言中的錯誤會導致意思與原文的意圖相反。

機器翻譯幻覺

這種錯誤是指 MT 置入了原文中根本不存在的內容。引擎可能會在特定情況下，生成冒犯、粗俗、挑釁或極為敏感的字詞。發生這種類型的重大錯誤，代表 MT 引擎的軟體本身有問題。

某個西班牙政府機關網站上的專有名詞誤譯，就讓大眾見識到這樣的重大錯誤。在這個實際案例中，人們在政府官方網站上看到的不是部門主管的姓名 Dolores del Campo，而是這個名字的字面翻譯：「痛苦的原野」。

如何將重大錯誤的風險降至最低

只有當電腦科學家能夠改善現有 MT 技術來避免這些錯誤的發生，我們才可以在翻譯流程中運用自動化技術，辨別可能的問題、修正有問題的字句並增進正確度。舉例來說，Lionbridge 便可以透過 **Smart MT™** 服務搭配先進的 **Smart Content™** 語言人工智慧 (AI) 技術，對翻譯文本進行特定的自動品質檢查，一方面偵測可能的錯誤，一方面保持 MT 作業速度並將由翻譯人員進行譯後編修的需求降至最低。

這些自動化方法可以偵測下列項目：

專有名詞翻譯錯誤：

藉由辨識來源文字中既是實體名稱也是常用字詞的實體，找出包括個人或組織名稱的專有名詞翻譯錯誤

冒犯、粗俗或極為敏感的字詞：

運用監督式機器學習 (ML) 演算法搭配冒犯詞語列表，偵測這類字詞

譯文與原文意思相反的情形：

辨識只出現於原文或譯文中而非兩者皆有的否定助詞 (亦即英文中「not」或其縮寫「n't」或是中文中「非」、「沒有」等否定助詞的句子)

機器翻譯幻覺：

運用字典或冒犯字詞列表，找出譯文中會冒犯他人的機器翻譯幻覺 (亦即翻譯中有但原文中沒有的字詞)

自動品質檢查無法保證能完全清除重大錯誤。這類檢查可能會漏掉錯誤，或是將正確的內容回報為錯誤。儘管如此，這還是非常有效的作法，有助我們找出問題。

採用這種方式，我們便能請專業譯者專注處理那些標記為有問題的句子，避免重做整份文件。透過提醒專業譯者哪些地方最容易找到錯誤，我們就能提高本地化流程的效率。

機器翻譯的引擎儘管非常精密複雜，但並不完美，因此才可能會產生重大錯誤。畢竟機器無法像人類一樣依脈絡判斷。





MT 風險：不相稱的語言正式度

什麼是語言正式度，它為何很重要？

語言正式度是人們表達自我意見時的正式程度。即使是講述同一件事情，我們對上司以及跟好友說話的方式通常就不一樣，會在用字遣詞以及文法上有所不同。

在商業場合或較為正式的情況下，我們往往更傾向於使用正式用語，因為正式用語意味著禮貌或尊敬。而在比較悠閒隨意的場合或是與熟識的人們說話時，則通常會使用非正式用語。

有些特定文化非常重視正式度級別，若使用了錯誤的級別，可能會讓對方有粗俗無禮的感覺。例如在韓文中，與年長者說話時絕對不能用對小孩說話的文法。韓文共有七種語階 (Speech Level)，每種語階的動詞結尾都不一樣，而其中五種屬於正式用語，其他兩種則是非正式用語。一旦使用了錯誤的正式度級別，可能造成令人難堪的情況。

此外，許多案例也顯示有時使用正式用語並不正確，倒不是因為會讓人感到冒犯，而是完全不適當並顯得很突兀。這些錯誤往往會讓目標對象開始疏遠您。舉例來說，翻譯電腦遊戲和學生活動計畫時，就需要使用較為日常、沒那麼正式的語氣。

公司企業如果願意花時間與心力，試著將翻譯內容調整為適合目標對象的風格，就能讓客戶強烈感覺自己受到重視。這些心力甚至能吸引到更多潛在顧客的關注。而有效地與目標對象建立連結，自然有助您成功達到目標。

有些特定文化非常重視正式度級別，若使用了錯誤的級別，可能會讓對方有粗俗無禮的感覺。

引擎為何難以產出正確的語言正式度

機器翻譯以及正式與非正式用語之所以可能會產生問題，是因為在語言正式度這方面可能會得到錯誤且不一致的結果。為什麼會這樣？一般而言，MT 模型會為輸入的每個分段文句傳回一種翻譯。但若是輸入的分段文句有歧義時，模型就必須在數個有效選項中擇一輸出，而不會考量目標對象是誰。

而任由模型在不同的有效選項中選擇，就可能造成翻譯不一致，或是翻譯的正式度級別有誤。

如果來源語言的正式度級別較目標語言來得少，就更難以得到正確的輸出成果。舉例來說，有些語言具有明確定義的敬稱，例如法文的 tu (你) 或 vous (您)，但英文則沒有。

解決使用 MT 時的語言正式度問題

公司可以運用規則式技巧及非規則式技巧來控管語言正式度，前者是使用規則以正確的翻譯取代不合適的風格，後者則是使用自訂的 MT 模型。

雖然多數的商用 MT 系統並不支援語言正式度或性別參數，但仍舊可以看到這些公司在這方面有一些進步。DeepL (API) 和 Amazon (主控台及 SDK) 目前就提供可調整正式度的功能。

共有三種調控選項可選：預設、正式以及非正式。預設選項不會變更神經機器翻譯 (NMT) 輸出的正式度。正式/非正式功能則可讓使用者選擇正式或非正式的語氣，更明確地說，這個功能會決定翻譯中所使用的代名詞及相關詞彙。

Lionbridge 的企業級機器翻譯解決方案 **Smairt MT**，則可以設定語言規則並套用至目標文本，以便依照所需的風格或正式度生成機器翻譯結果，甚至可以推行中性語言，避免**本地化中的偏見問題**，以英文為例，就是以 Salesperson 這個稱呼取代 Salesman 的差別。

我們的專業人員會維護最新的規則資料庫，並運用於分析 MT 輸出以調控結果。如果有足夠的資料，我們還可將規則式策略與自訂的 MT 模型相結合，取得最合宜的結果。

使用 MT 產出正確的語言正式度雖然並不容易，但也不是不可能的任務。本白皮書稍後會探索一些令人振奮、可望解決語言正式度相關問題的技術發展。

Lionbridge 的企業級機器翻譯解決方案 **Smairt MT**，可以設定語言規則並套用至目標文本，以便依照所需的風格或正式度生成機器翻譯結果。

2022 年機器翻譯年度回顧

2022 年有哪些重要的 MT 趨勢？

Lionbridge MT 專家發現，機器翻譯在 2022 年有所進步以及沒有進展的地方，都非常值得我們留意。有鑑於 MT 相關技術在過去數年一直不斷大幅進步，我們的團隊自然預期同樣會看到更多進展。然而根據我們的機器翻譯追蹤工具，MT 在這一年其實並沒有長足的進展。

除了極其少見的例外情況，主流引擎這一年幾乎沒有什麼進展。這個趨勢當然會影響到未來的發展。不過在放眼未來前，我們先來深入檢視 2022 年的成果。

頂尖 MT 引擎在 2022 年的表現如何？

對於想開始使用 MT 或改善目前 MT 使用方式的公司來說，首要之務就是要找出最適合他們特殊需求的 MT 引擎。而越深入探究主流 MT 引擎在 2022 年的表現，有件事情也就越清楚：沒有一個引擎是全能的。

根據語言比較 MT 引擎的表現

要處理西班牙文內容的公司，不妨選擇 DeepL 進行自動翻譯；但如果要翻譯日文，就得考慮改用其他更合適的選項。因為每個引擎的表現，會因為所處理的語言而有所不同。我們針對三個主流引擎，計算了它們從英文翻譯至多種語言的自動翻譯表現。在判斷品質高低時，則是根據計算所得的平均編輯更動程度而定，也就是編譯人員為了取得與人工翻譯一樣的高品質結果，而對 MT 輸出做的編輯更動次數。

數值越低，代表自動翻譯的效果越好。而從圖 1 便可以看出，這些結果的差異絕對值得公司企業留意。

根據我們的分析結果，在特定情況下：

- DeepL 在西班牙文翻譯的表現優於 Google 與 Microsoft
- Google 在日文翻譯的表現優於 DeepL
- Microsoft 在波蘭文翻譯的表現優於 DeepL
- 這三個引擎在義大利文、土耳其文和希伯來文翻譯的表現相差無幾

這些結果說明了機器翻譯本質上非常複雜且具挑戰性，必須將不同語言、文化與領域間的差異與複雜性納入考量。

因此不同 MT 引擎會有不同的表現，也就並不令人感到意外，畢竟沒有任何一種演算法或方法，能完美地適用於所有語言和內容類型。



MT 的表現為何會因語言不同而異？

造成這些成效表現差異的其中一個可能原因，是可供各個語言用於訓練的資料品質與數量都不一樣。機器翻譯模型需要有以高品質雙語資料組成的龐大語料庫，才能學習如何正確地翻譯某個語言。一旦某個特定語言缺乏這些資料，MT 引擎往往就難以產出正確的翻譯。

此外，每一個 MT 引擎的技術和語言特色也不盡相同，各個引擎都有各自的一套演算法、架構和機器翻譯作法。視引擎模型的特色與功能而定，某些引擎在特定語言組合或內容類型上的表現會比較優異。除此之外，MT 引擎最佳化及自訂的程度，也可能會影響到效能。有些 MT 引擎可能是針對特定的語言或領域最佳化與自訂，因此在這些使用案例上可以產出更加正確和有效的翻譯。

整體而言，這些因素都會影響到機器翻譯的複雜度和變異性，也凸顯出我們必須要了解不同 MT 引擎在不同語言和使用案例上的優點與侷限，才能加以妥善運用。

要針對特定使用案例選擇最合適的 MT 引擎，自然得具備機器翻譯方面的專業能力。這並不是個簡單的工作，需要審慎評估並持續進行最佳化，才能確保取得最佳的成果。

不同語言組合的 MT 引擎平均表現

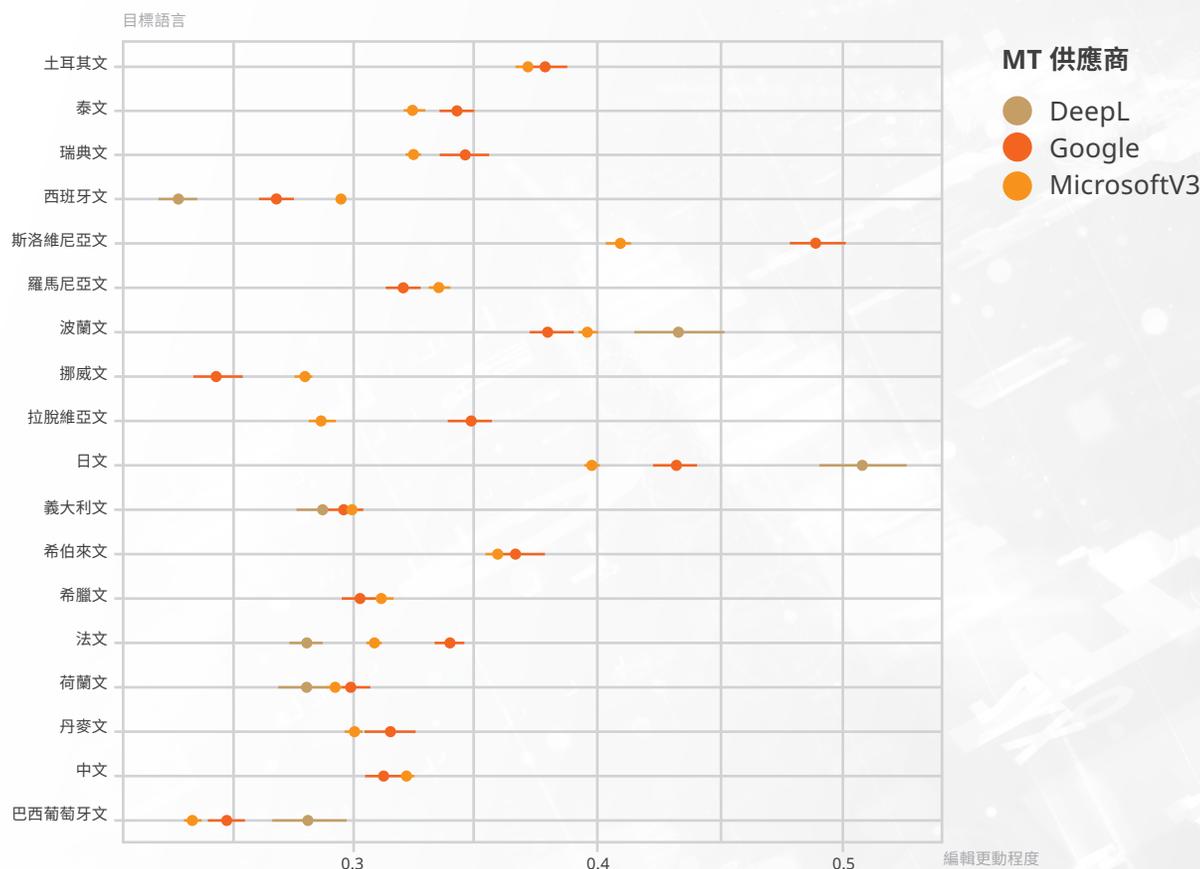


圖 1：MT 引擎在不同語言組合的平均編輯更動程度

不同領域的自動翻譯表現

內容越富創意，引擎便越難有效地翻譯。

根據領域評估而得的翻譯品質，也證實了這個論點。我們針對多個主要領域追蹤了不同引擎在處理多種內容類型時的平均編輯更動程度，也就是編譯人員為了取得與人工翻譯一樣的高品質結果，而對 MT 輸出做的編輯更動次數。

數值越低，代表翻譯的品質越好。

根據我們的分析結果，如圖 2 所示，

- 向來以創意聞名的媒體和行銷內容，對各個引擎而言都是翻譯時最具挑戰性的內容類型
- 織品和時尚相關內容往往有許多富想像力而又細膩的描述，在機器處理的難度上則排名第二
- 汽車與機械產業的內容清楚明確，因此機器翻譯處理的成果也最好

結果也正如我們所預期。如果內容直接單純，相對而言簡單易懂，同時結構明確，詞彙也沒有太過專業或技術性，常見 MT 引擎的翻譯表現便頗為亮眼。

不同領域的 MT 輸出品質

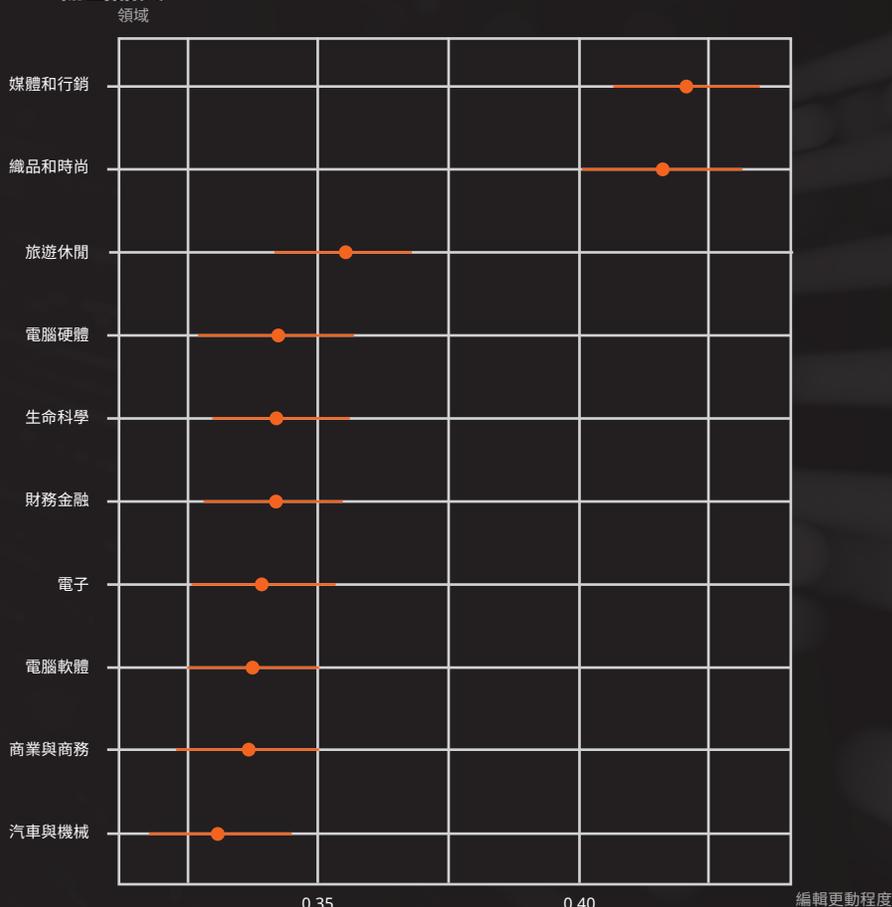


圖 2：不同領域的平均編輯更動程度。

根據領域和 MT 引擎比較 MT 引擎的表現

翻譯媒體與行銷內容的公司，最好選擇 DeepL 做為他們的 MT 供應商；而翻譯織品和時尚相關內容的公司，不妨考慮其他更合適的選項。根據所取得的資料，我們衡量了四大主流引擎處理不同產業領域內容時的表現。在判斷品質高低時，則是根據計算所得的平均編輯更動程度而定，也就是編譯人員為了取得與人工翻譯一樣的高品質結果，而對 MT 輸出做的編輯更動次數。數值越低，代表自動翻譯的效果越好。從圖 3 可以看出，在特定情況下，不同引擎在不同產業領域和內容類型上的表現也不一樣。

根據我們的分析結果，在這些特定案例中：

- **媒體與行銷**：DeepL 的表現優於 Google 與 Microsoft
- **生命科學**：DeepL 的表現優於 Microsoft 與 Google

- **財務金融**：Microsoft 和 DeepL 的表現優於 Google
- **汽車與機械**：Google 的表現略微優於 Microsoft 與 DeepL

然而要注意的是，如果要推薦最合適的系統，我們必須先針對您的特定內容進行分析，確切了解您的需求，才能找出最符合需求的引擎。

從上述的結果可以看出，MT 引擎的效能不但如我們所述，會因語言組合而異，也會受特定領域或產業所使用的術語和措辭所影響。

因此，在挑選 MT 引擎時，務必要將翻譯的語言組合以及所屬的特定領域和產業脈絡納入考量。這種作法一方面必須要具備語言和特定領域的專業能力與知識，但同時也能就特定的內容類型和產業取得更優異的翻譯品質。

MT 引擎在不同領域的平均表現

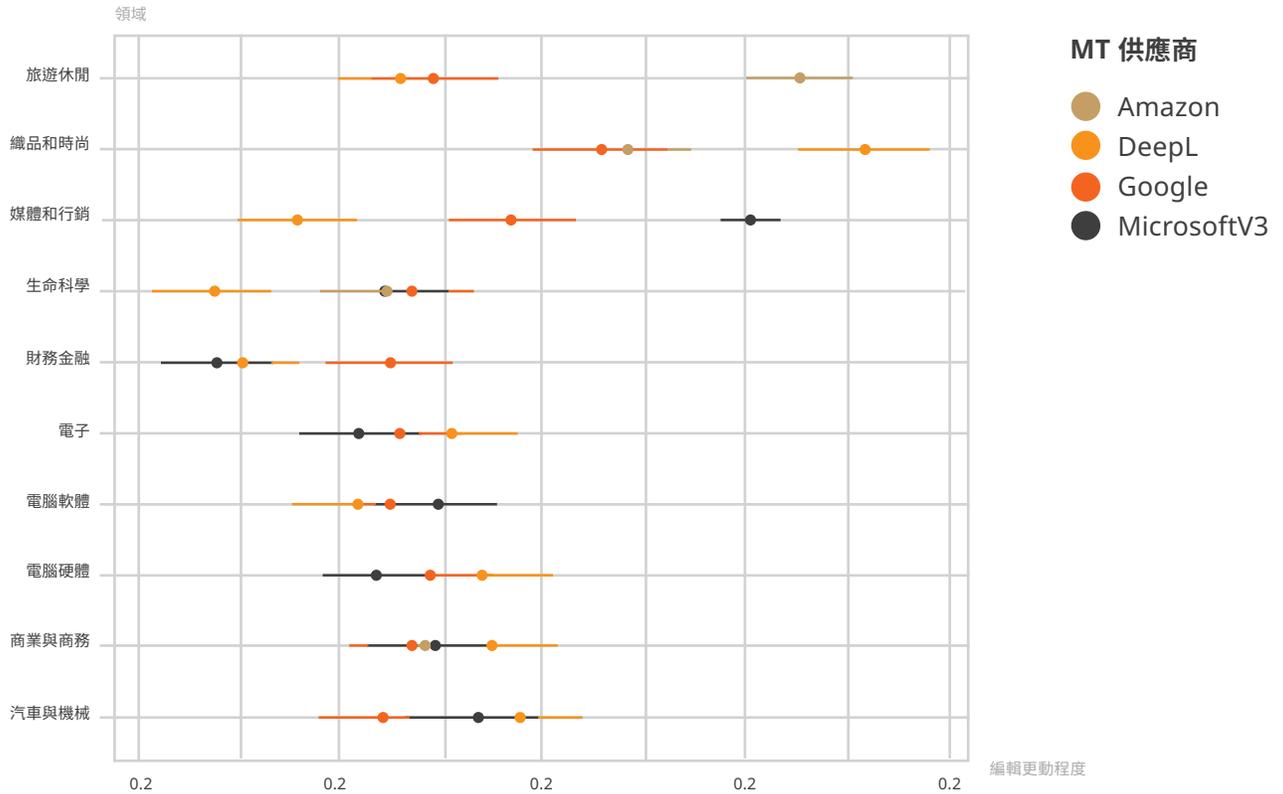


圖 3：不同領域和 MT 供應商的平均編輯更動程度

機器翻譯引擎品質之比較

這些主流引擎彼此之間的整體表現如何？在圖 4 中，我們根據從 2018 年 5 月到 2022 年 12 月這段期間的反向編輯距離，比較了五大主流引擎彼此之間在德文、西班牙文、俄文和中文的輸出品質。

編輯更動程度測量的是編譯人員為了取得與人工翻譯一樣的高品質結果，而對 MT 輸出做的編輯更動次數。所得的反向編輯距離數字越高，代表品質越好。

在圖 5 中，我們運用了趨勢線來呈現各個引擎的表現，可以從中得出非常有意思的結果。將圖 4 和圖 5 一起對照，就可以看出頂尖 MT 供應商的引擎品質越來越相近。

根據我們的分析：

- Microsoft Bing 急起直追，迎頭趕上領先者
- Amazon 和 Google 的趨勢線幾乎是完美的平行線
- 從技術面來看，機器翻譯在 2022 年並沒有大幅度的進步

到了 2023 年的年初，主流引擎間的差異已經微乎其微。如果神經機器翻譯 (NMT) 典範仍舊蔚為主流，而 MT 供應商也以相近的速度繼續投資於其中的話，我們認為 MT 引擎的成效表現最終會在這一年內趨於相同，只會在不同語言組合和領域上有些微差別。



到了 2023 年的年初，主流引擎間的差異已經微乎其微。

MT 引擎的表現

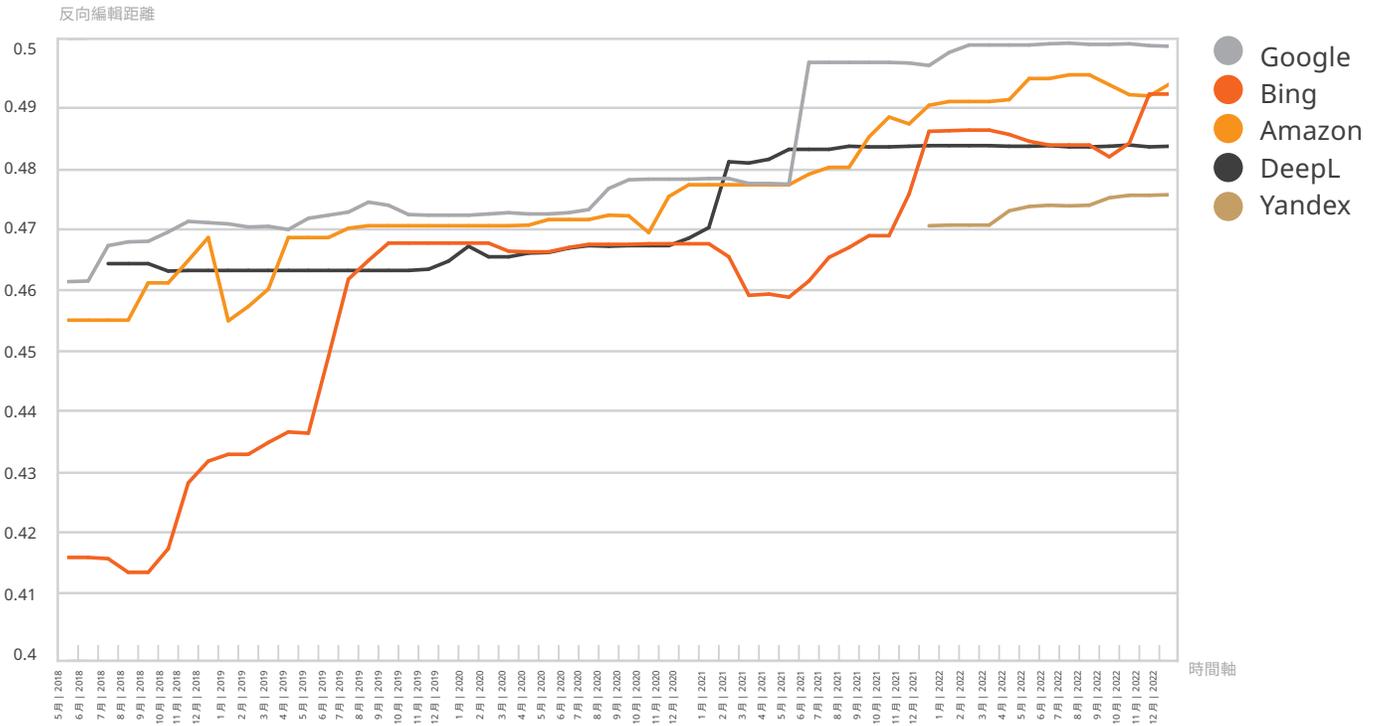


圖 4. 根據反向編輯距離比較 MT 整體品質

MT 引擎的表現：時間軸

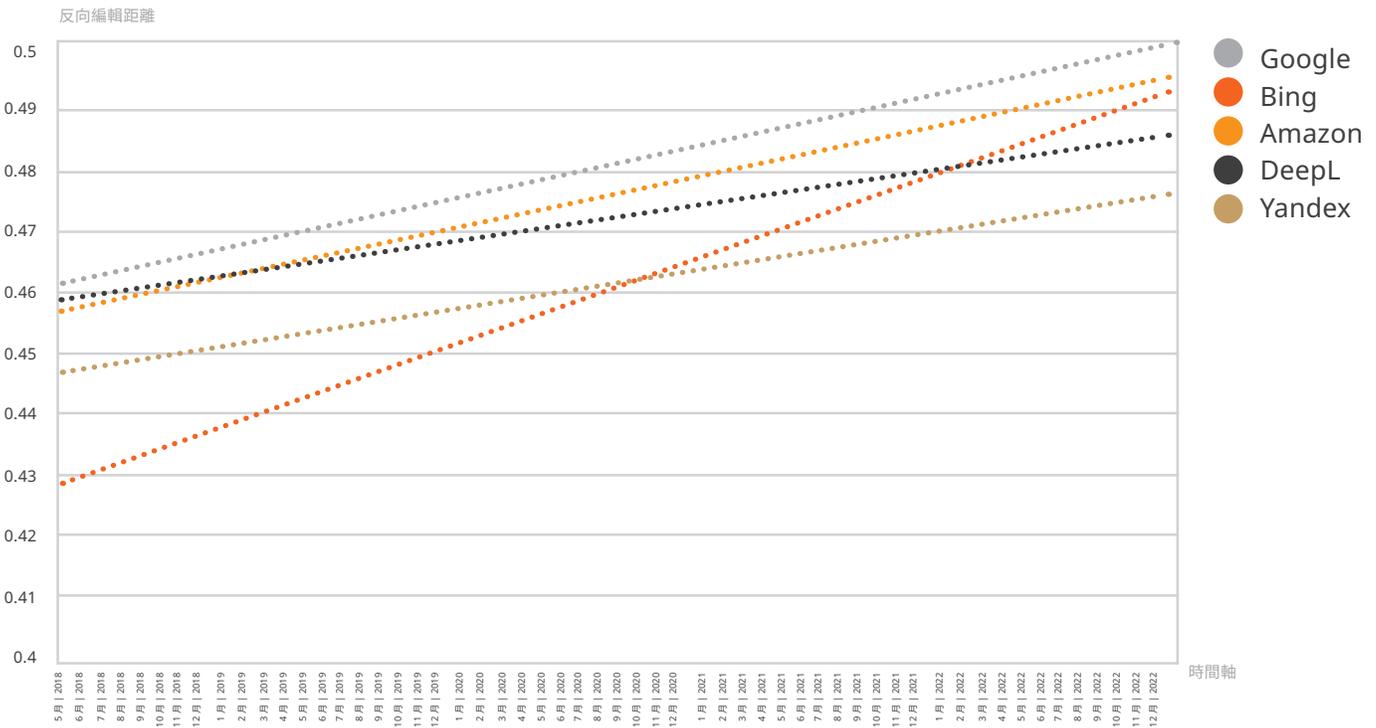


圖 5：五大主流 MT 引擎表現的趨勢線

不同語言之機器翻譯引擎品質的比較

在 2022 年，這些主流引擎彼此之間的表現如何？更明確地說，它們在德文、西班牙文、俄文和中文翻譯上的表現如何？我們根據反向編輯距離，衡量了這些引擎的品質。

編輯更動程度評量的是編譯人員為了取得與人工翻譯一樣的高品質結果，而對 MT 輸出做的編輯更動次數。

所得的反向編輯距離數字越高，代表品質越好。如圖 6 所示：

- 從顯示反向編輯距離的比例尺即可看出，整體來說 MT 只有些微的改善
- 在 10 月/11 月時，Microsoft Bing 的德文、西班牙文和中文翻譯有著小幅的進步
- 數據證明 2022 年是表現沒有什麼起伏的一年

我們可以就此得出結論：神經機器翻譯已經到達頂點，必須要有新的典範出現，MT 才有可能在品質上再度大幅突破。

依反向編輯距離評估機器翻譯引擎在特定語言的表現

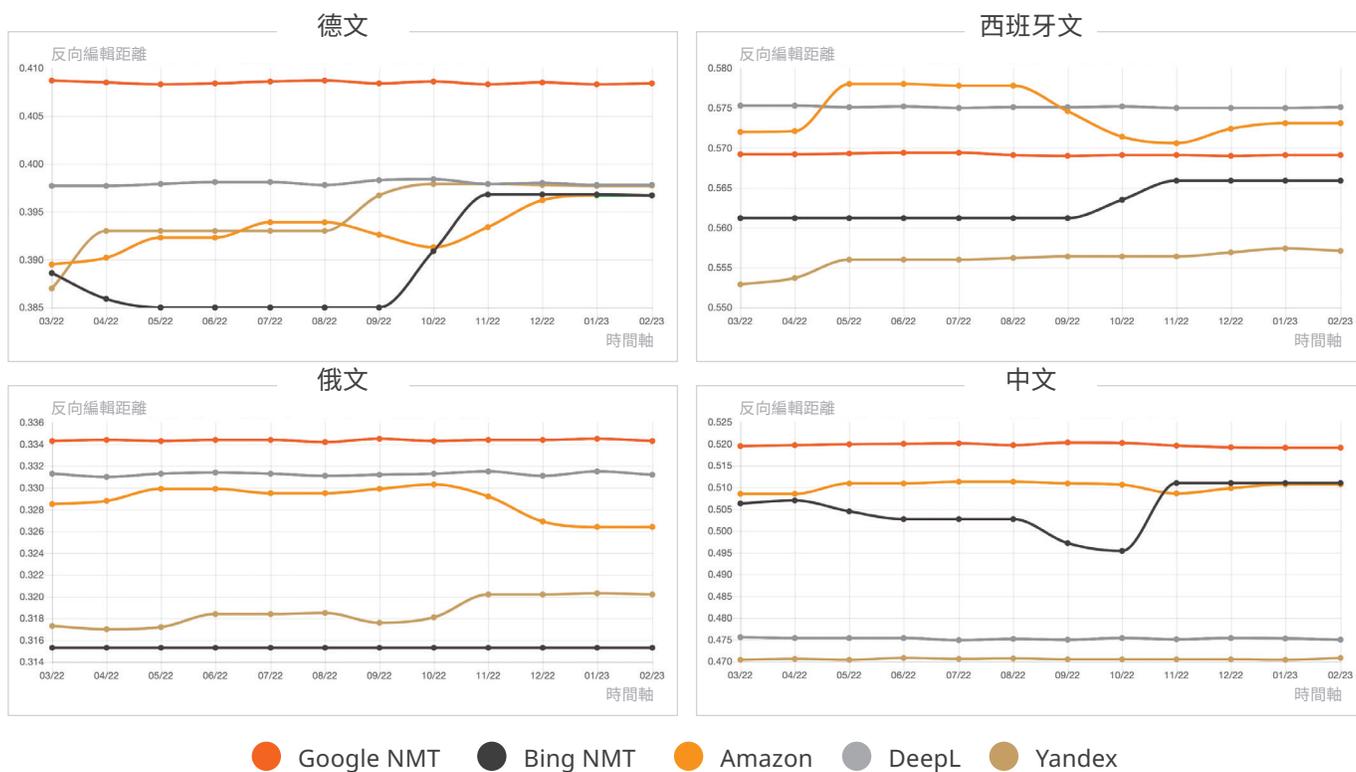


圖 6：根據反向編輯距離比較不同語言的 MT 品質

○ 我們可以就此得出結論：神經機器翻譯已經到達頂點。

不同領域之機器翻譯引擎品質的比較

在 2022 年，這些主流引擎彼此之間在特定領域的表現如何？我們根據反向編輯距離，衡量了這些引擎的品質，結果如圖 7 所示。

編輯更動程度測量的是編譯人員為了取得與人工翻譯一樣的高品質結果，而對 MT 輸出做的編輯更動次數。所得的反向編輯距離數字越高，代表品質越好。

根據我們的分析：

- **機器翻譯**在程序化內容的表現往往優於較具創意的內容，因為程序化內容通常比較直接單純，更易於機器處理。因此，MT 在汽車和電腦軟體產業上的品質表現，自然會優於在零售、行銷和旅遊觀光業上的表現，因為後者的內容全都更為複雜。這些結果不但很類似我們根據不同領域及平均編輯更動程度 (見圖 2) 評量自動翻譯表現所得的結果，而且更加明顯。

- **Google MT** 和 **DeepL** 在這一一年中的表現，比 **Amazon** 和 **Bing** 來得穩定。

- **Amazon** 在我們分析的這段期間並沒有顯著的改善，只有一些上下震盪起伏，充其量也只是在評量結束時又回到原點。但有兩個領域則是例外：法律與法規，以及媒體、廣告和行銷。在這兩個產業中，我們觀察到了正面的演進，在法律與法規方面，Amazon 的表現稍微優於其他競爭對手，而在媒體、廣告和行銷這個領域中，則是與 **Google** 一同名列前茅。

- **Bing** 顯然在兩個領域中獨占鰲頭，而且本質上都與其母公司 **Microsoft** 有關：電腦軟體和財務金融領域。在這一一年中，這個 MT 引擎展現了大幅的進步，排名從第三躍升至第一名。沒有任何其他 MT 引擎能在任何領域中，有跟它足以匹敵的進步幅度。

- **Yandex** 是我們分析的 MT 供應商中，唯一一個沒有在任何領域領先的引擎。在 2022 年末時，它終於在兩個領域中以較為優秀的表現超越了 **DeepL**：汽車以及電腦軟體。



機器翻譯引擎在特定領域的表現

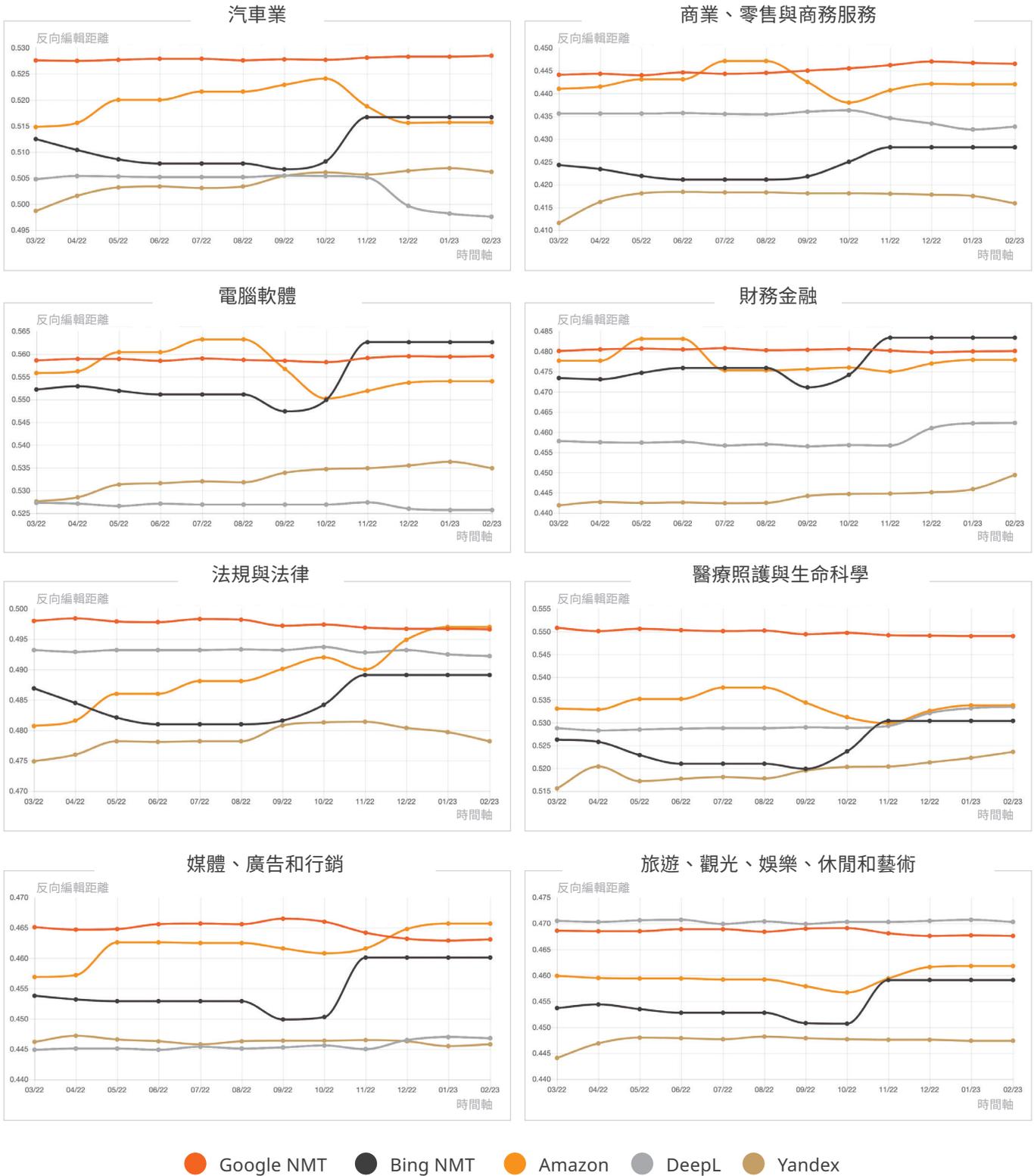


圖 7：根據反向編輯距離比較不同領域的 MT 品質

如何充分發揮機器翻譯的功效？

想要充分發揮 MT 技術的價值，不妨採取以下這些步驟：

- 找出哪一種 MT 引擎可更輕易翻譯特定的語言組合，這又稱為語言的**機器可譯性**或 M 可譯性。這麼做可協助您擬定 MT 策略，並決定要向哪些市場進軍，而我們已經幫您把麻煩的分析工作做好了。
- 有效地運用術語來改善所有領域的輸出品質，包括 MT 通常難以妥善處理的領域。
- 思考執行 MT 自訂或 MT 訓練的時機。這兩種做法雖然都能改善 MT 輸出品質，但彼此無法互相替代。

機器可譯性

了解不同引擎處理特定語言組合的難易程度，您在規劃各個語言的翻譯費用時，就能更妥善地分配預算。您會更清楚知道哪些語言組合需要投入更多翻譯心力。因此，對語言的複雜性有深入了解，不但能協助您做出更明智的商務決策，也有助您回答以下的問題：

- 是否應該將大部分的預算，投入較複雜語言組合的譯後編修？
- 在預算吃緊的情況下，小幅譯後編修，或是只針對內容中重要部分進行譯後編修的重點譯後編修，對某些語言是否便已足夠？哪些語言應該使用這些譯後編修作法？
- 在思考預算的最佳分配方式，特別是低預算的專案時，公司是否應該將語言排名納入商務和文化因素的考量？如果某個文化可以接受較低的品質，公司是否應該將內容翻譯至 M 可譯性排名較低的語言？

依照可譯性來排序語言不是個簡單直覺的流程，然而，我們可以使用不同的指標來進行評估。編輯更動程度，也就是譯後編修人員為確保最終譯文擁有與人工翻譯相同的品質，而對機器翻譯之內容所做的更動次數，就有助於我們了解各個語言組合的 MT 複雜性和可譯性。為了協助您比較這些語言，我們列出了從英文譯出時，前 28 種目標語言的機器可譯性 (亦即 M 可譯性) 排名。

如表 1 所示，從英文譯出時，多數羅曼語系的語言，例如葡萄牙文、西班牙文、法文和義大利文等，通常只要稍做修改就能得到高品質的譯文。我們發現這些目標語言是機器最能輕鬆處理的語言，它們也名列我們 M 可譯性排名的前四名。

匈牙利文和芬蘭文這兩個烏拉爾語系的語言，則屬於較為複雜的語言，因而在我們的可譯性排名中墊底，分別是第 27 和第 28 名。另一個同一語系的愛沙尼亞文，也屬於較為複雜的語言。這些根據 Lionbridge 處理過的數百萬個句子分析所得的結果，可以看出語系對 MT 產出結果的重要性。

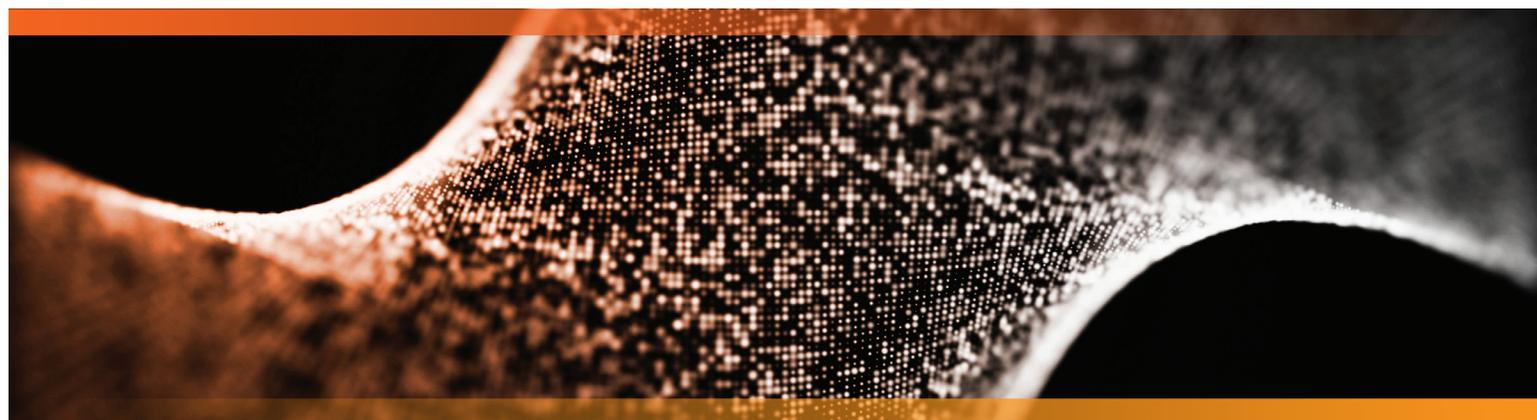
雖然語言比較仍有其侷限，但這樣的排名能提供一些有趣的深入見解，協助您更有效地管理多語言專案。



表 1：語言 M 可譯性排名

排名	語言 (由英文譯出)
1	葡萄牙文
2	西班牙文
3	法文
4	義大利文
5	中文 (簡體)
6	荷蘭文
7	丹麥文
8	日文
9	希臘文
10	羅馬尼亞文
11	泰文
12	挪威文
13	德文
14	瑞典文

排名	語言 (由英文譯出)
15	土耳其文
16	斯洛伐克文
17	希伯來文
18	拉脫維亞文
19	波蘭文
20	中文 (繁體)
21	立陶宛文
22	捷克文
23	阿拉伯文
24	愛沙尼亞文
25	韓文
26	俄文
27	匈牙利文
28	芬蘭文





訓練 MT 引擎、客製化詞彙表，再加上制訂前處理及後處理的規則，將有助確保 MT 輸出使用正確的術語，並展現與客戶文件相近的風格。

運用術語改善專業領域的翻譯表現

如前所述，常見 MT 引擎輸出的翻譯有時會有不少錯誤，尤其是應用在特定領域上時，在術語方面可能會產生不如人意的結果，如果是醫療或法律領域，更有可能因此造成嚴重的傷害。無論是處理什麼主題，有效地使用術語都有助改善 MT 品質，生成正確而一致的翻譯。

請務必使用含有特定領域專門術語的雙語文本，來訓練客製化的 MT 系統。但是，如果在訓練引擎用的專門文本中，使用的術語並不一致，那麼也還是無法保證能產出正確的翻譯。這個領域的相關研究建議，可以透過註解 (Annotation) 的方式將語言學資訊輸入神經機器翻譯 (NMT) 系統中。至於要以人工還是半自動的方式進行註解，則取決於可用資源 (例如詞彙表) 以及限制條件 (例如時間、成本及可用註解人員等) 而定。

Lionbridge 的 **Smart MT** 可以設定語言規則並套用至來源與目標文字，也可以將「請勿翻譯」(DNT) 及詞彙表清單加入特定的設定檔，再據此進行術語管理，解決**機器翻譯術語**的問題。我們會協助客戶建立與維護詞彙表，並定期調整以加入新的相關詞彙和汰除過時的術語。此外，只要在 **Smart MT** 建立詞彙表，就能將它們使用於所有的 MT 引擎，進一步節省時間與金錢。

在 MT 專案中使用詞彙表，遠比表面看來還要複雜。詞彙表如果使用不當，反而會對機器翻譯的整體品質造成負面影響。如果希望 MT 生成的翻譯能使用所提供的術語，最好的作法就是透過 MT 訓練。訓練 MT 引擎、客製化詞彙表，再加上制訂前處理及後處理的規則，將有助確保 MT 使用正確的術語，並展現與客戶文件相近的風格。

應該使用 MT 自訂還是 MT 訓練

MT 自訂和 MT 訓練能協助您發揮 MT 輸出的最大效益，但您必須特別留意應用這些方法的時機。表 2 大概說明了**機器翻譯自訂與機器翻譯訓練的差異**，並就評量各個方式時可考量的因素提供一些建議。

表 2：機器翻譯自訂與機器翻譯訓練

	MT 自訂	MT 訓練
意義及運作方式	使用詞彙表及「請勿翻譯」(DNT) 清單來調整既有的機器翻譯引擎，藉此改善機器生成翻譯的正確性	使用來自語料庫及翻譯記憶庫 (TM) 的龐大雙語資料，來建置與訓練 MT 引擎，藉此改善機器生成翻譯的正確性
功用	改善 MT 的翻譯建議，產出更正確的輸出並減少對譯後編修的需求	改善 MT 的翻譯建議，產出更正確的輸出並減少對譯後編修的需求
特有好處	能讓公司保持其品牌名稱、遵循所應使用的術語，並顧及地區性差異	能讓公司展現特有的品牌調性、語氣和風格，並顧及地區性差異
使用的風險	如果執行不當，MT 可能會產出品質不佳的翻譯建議，對整體品質帶來負面影響	如果沒有足夠的資料來訓練引擎，MT 訓練可能無法對輸出有任何影響。如果編寫人員經驗不足，使用太多不必要的術語，可能會導致 MT 產出品質不佳的翻譯建議，對整體品質帶來負面影響
使用時機	非常適合技術性以及講究細節的內容，以及有以下要求的任何內容： <ul style="list-style-type: none"> • 正確的術語翻譯 • 地區性差異，但缺乏足夠的資料進行 MT 訓練 	非常適合高度專業的內容、行銷與創意內容，以及有以下要求的任何內容： <ul style="list-style-type: none"> • 特有的品牌調性、語氣或風格 • 地區性差異，且有足夠的資料進行 MT 訓練
成功要素	由經驗豐富的 MT 專家，為您成功管理輸入與輸出正規化規則、詞彙表和 DNT	要有最低 15,000 個獨特不重複的句段，才足以訓練引擎
成本考量	更新 MT 所用的設定檔會有個一次性支付的成本，日後也需要隨時間維護詞彙表。若將潛在好處納入考量，成本相對而言並不昂貴，而且通常比 MT 訓練的成本來得低	成本包括初次訓練及後續訓練的成本，如果 MT 表現監測顯示還有改進的空間，成本亦會隨時間增加。若將潛在好處納入考量，MT 訓練對特定案例會是值回票價的投資

● 結論以及機器翻譯的展望

從 2022 年的資料，以及這一年品質表現停滯不前這令人驚訝的結果，我們可以對機器翻譯的情勢做出什麼結論呢？那就是：這項技術不但已臻成熟，而且一再證明是深具價值的企業級技術，因而也將持續獲得廣泛的採用。

無論採用人工干預與否還是混合作法，人們已經體認到這個技術對幾乎任何翻譯案例而言都很實用。的確，根據 Global Market Insights，從 2022 年到 2030 年，翻譯市場估計會以 30% 的年複合成長率 (CAGR) 成長。公司企業會日漸擁抱 MT 技術，包括那些以往極為抗拒 MT 的遊戲和生命科學等領域。如果能具備充分發揮這個技術效益的能力，再搭配使用採用 AI 的技術來自動化工作流程與挑選合適譯者，公司企業將可加快內容產出速度、產出總是符合品牌特色而又生動有趣的多語言內容、擴大市場範疇，進而得以在競爭極其激烈的數位市場中蓬勃發展。

機器翻譯的未來會如何？

2022 年機器翻譯的分析結果，使我們開始質疑目前的神經機器翻譯典範。

- NMT 這個典範是否已經達到頂點？
- 如果這些引擎無法再有長足的進展，我們是否要改採新的做法？是否要有新的典範轉移？
- 新的典範又會是什麼？

我們認為，具備多模態及多語言等龐大內容量的大型語言模型 (LLM)，可能會與未來的新典範有關。

為什麼這麼說？因為我們做了一個比較 ChatGPT 與 MT 引擎翻譯表現的開創性分析，所得的結果讓我們有了上述這樣的想法。

OpenAI 的 ChatGPT 翻譯表現雖然不如選定的 MT 引擎，但其實差距並沒有很大。它的表現可說頗為卓越。在某個例子以及某個語言組合上，GPT-4 的結果甚至還勝過某個主流神經機器翻譯引擎。這些結果，無疑地會影響到機器翻譯的未來。

為何可能會出現新的機器翻譯典範？

目前 MT 引擎的趨勢，給人似曾相識的感覺。在統計式機器翻譯時代式微、但尚未被 NMT 取代之際，MT 輸出品質幾乎沒什麼變化；此外，不同 MT 引擎的輸出品質也漸趨相同。現在的情況也正是如此。

儘管 NMT 還不會馬上被取代，但如果我們相信指數成長和加速回報理論，再考慮到規則式 MT 為期 30 年和統計式 MT 盛行十數年的壽命，以 NMT 問世到目前進入第六年來，新的典範轉移應該也不遠了。



下一個機器翻譯典範可能會是誰？

LLM 在 2022 年的重要進展，使得這項技術有望在 2023 年涉足 MT 領域。LLM 是訓練用來執行許多事情的通用模型。然而，我們也在 2022 年結束之際，觀察到有些專用或精心調整過的 LLM，在某些特定領域有著非常重要的進步。從這些發展來看，這項技術可能可以在做過額外一些訓練後，用以執行翻譯作業。

以 ChatGPT 為例，OpenAI 便精心調整了這個模型，讓它可以進行問答式的對話，同時又可執行通用 LLM 可以做的任何事。該公司在他們的 GPT-4 模型上甚至還做了更多改善。可以預期，日後應該會有更多 LLM 會精心調整以使用於翻譯。

如何精心調整大型語言模型以使用於處理翻譯作業？

如果能用語言種類更為均衡的語料庫來訓練機器，那麼運用 LLM 執行翻譯的可能性會更高。GPT-3 的訓練用語料庫 93% 是英文，只有 7% 是來自所有其他語言的語料庫。我們猜想 GPT-4 的訓練資料可能含有更多語言的更多語料，所以它的輸出品質才會更加優異。以語言種類更均衡的語料庫為基礎，就有可能建置精心調整的 LLM，進而專門用於處理翻譯。

這個以 LLM 為基礎的新 MT 典範假設，還有另一個值得注意的面向，就是多模態趨勢。我們可以用語言以及例如影像和視訊等其他訓練資料來訓練 LLM，這類訓練可以提供更多有關真實世界的知識，進而提高翻譯品質。

大型語言模型會是神經機器翻譯典範之外的另一個好選擇嗎？

為了評估 LLM 取代 NMT 典範的潛力，我們將 ChatGPT 和 GPT-4 的翻譯品質，與我們機器翻譯追蹤工具中使用的五大主流 MT 引擎的翻譯表現進行了比較。如圖 8 所示，ChatGPT 的成果非常接近專業化引擎。而如圖 9 所示，GPT-4 在英文譯至簡中這個語言組合上，表現略微勝過 Yandex 的輸出。

我們是如何評估 ChatGPT 跟常見 MT 引擎的品質？

為了評估引擎的品質層級，我們使用多個參考翻譯來計算反向編輯距離 (Inverse Edit Distance)。編輯更動程度 (Edit Distance) 評量的是編譯人員為了取得與人工翻譯一樣的高品質結果，而對 MT 輸出做的編輯更動次數。在計算時，我們是將原始的 MT 輸出，與 10 個 (而非僅只一個) 不同的人工翻譯 (也就是多個參考翻譯) 相比較，所得的反向編輯數字距離越高，代表品質越好。

以語言種類更均衡的語料庫為基礎，就有可能建置精心調整的 LLM，進而專門用於處理翻譯。



根據反向編輯距離進行多個參考資料評估

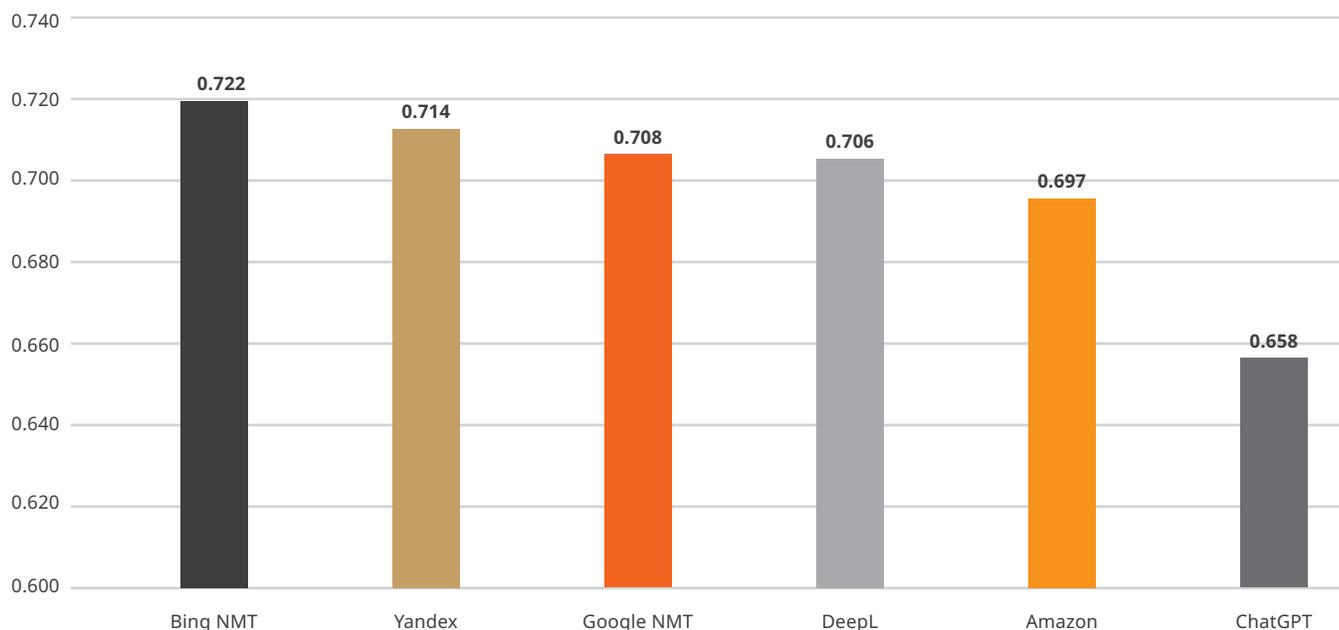


圖 8. 根據英翻西這個語言組合的多個參考翻譯計算反向編輯距離，比較 ChatGPT 與主流機器翻譯引擎的自動翻譯品質。

英譯簡中翻譯品質

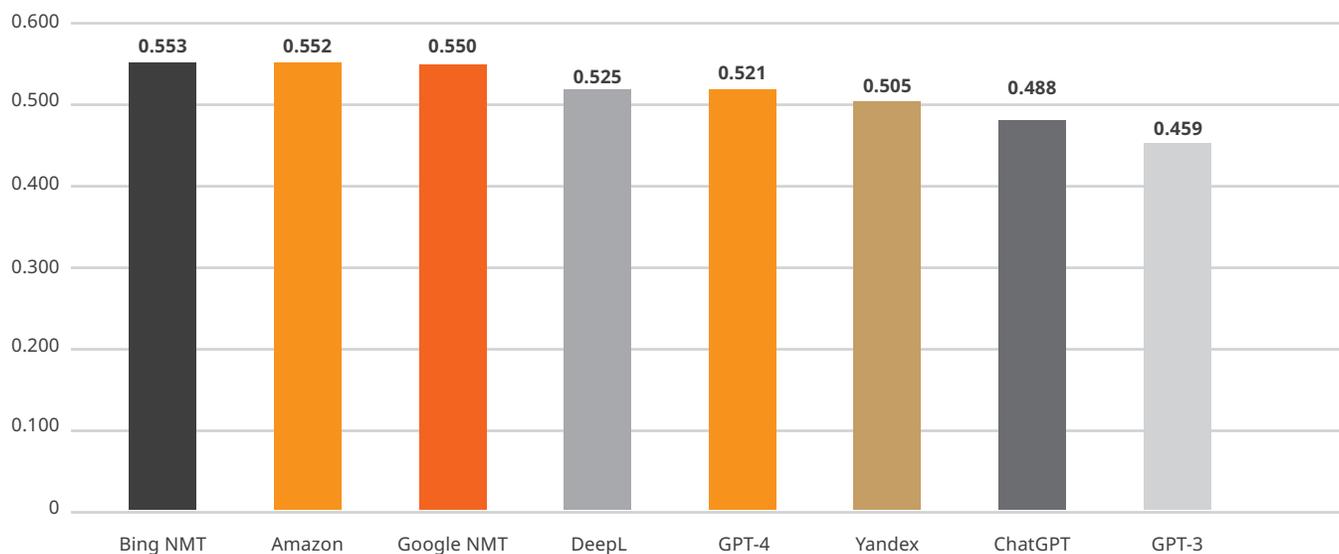


圖 9. 根據英譯簡中這個語言組合的多個參考翻譯計算反向編輯距離，比較 GPT 模型與五大主流神經網路 MT 引擎的自動翻譯品質。



大型語言「通用」模型的一個好處，是它們可以執行很多不同的任務，而且多數時候都能提供卓越的品質。

LLM 翻譯成果為何值得我們注意？

我們這個比較分析所得的結果非常值得注意，因為這個通用模型是訓練用以執行許多不同的自然語言處理 (NLP) 任務，而不是像 MT 引擎那樣，訓練的目的只有翻譯這一種 NLP 任務。即使 GPT 並沒有特別針對翻譯進行訓練，它的翻譯品質已經非常優異。

大型語言模型會促使機器翻譯有什麼樣的發展？

從大眾的關注與科技公司對這個技術的大量投資，都可看出 LLM 的成長，因此我們應該很快就能知道，MT 是否將開始採用新的 LLM 典範。

MT 或許會以 LLM 為基礎，但針對機器翻譯微調這項技術。這會很像 OpenAI 及其他 LLM 公司目前的作法，也就是針對特定使用案例去改進他們的通用模型，例如讓機器能以對話聊天的方式與人類溝通。至於專業化則可進一步讓所執行的任務更加準確。

總的來說，大型語言模型的未來有何展望？

大型語言「通用」模型的一個好處，是它們可以執行很多不同的任務，而且多數時候都能提供卓越的品質。舉例來說，另一個通用智慧模型，DeepMind 的 GATO，便曾做過超過 600 種任務的測試，並在其中 400 種取得最先進 (SOTA) 結果。

這樣看來，將有兩個開發路線會繼續存在，一個是像 GPT 和 GATO 這類通用模型，另一個則是以通用模型為基礎再針對特定用途加以專業化的模型。通用模型對推進「通用人工智慧」(AGI) 非常重要，以長程來看甚至可望催生更令人期待的發展。專業化模型則能在短期內於特定領域中發揮實際功用。

而 LLM 令人讚嘆的其中一點，就是它可以同時參與這兩個路線的發展。

機器翻譯若發生典範轉移，會有什麼影響？

隨著目前神經機器翻譯技術典範的發展逐漸到達極限，加上極具優勢、且很可能是以 LLM 為基礎的新機器翻譯技術典範浮上檯面，我們預期 MT 領域也會出現一些改變。大多數的效應會有益於公司企業，但對想實行人工翻譯的公司來說，應該會面臨額外的一些挑戰。

以下是我們對未來的看法：

品質提升

隨著技術進步，長久以來的一些問題，例如語言正式度以及其他跟語氣相關的品質問題，都可望獲得解決，因此機器翻譯的品質將會大幅躍進。LLM 甚至可能可以解決 MT 引擎最大的問題：對真實世界知識的缺乏。透過多模態訓練，便有可能達到這個目的。

除了龐大的文本外，技術人員還可以用影像跟視訊來訓練現代的 LLM。這類訓練能讓 LLM 擁有更多相關連的知識，進而有助機器解讀文本的意義。

內容產出量會提高，但一流譯者會供不應求

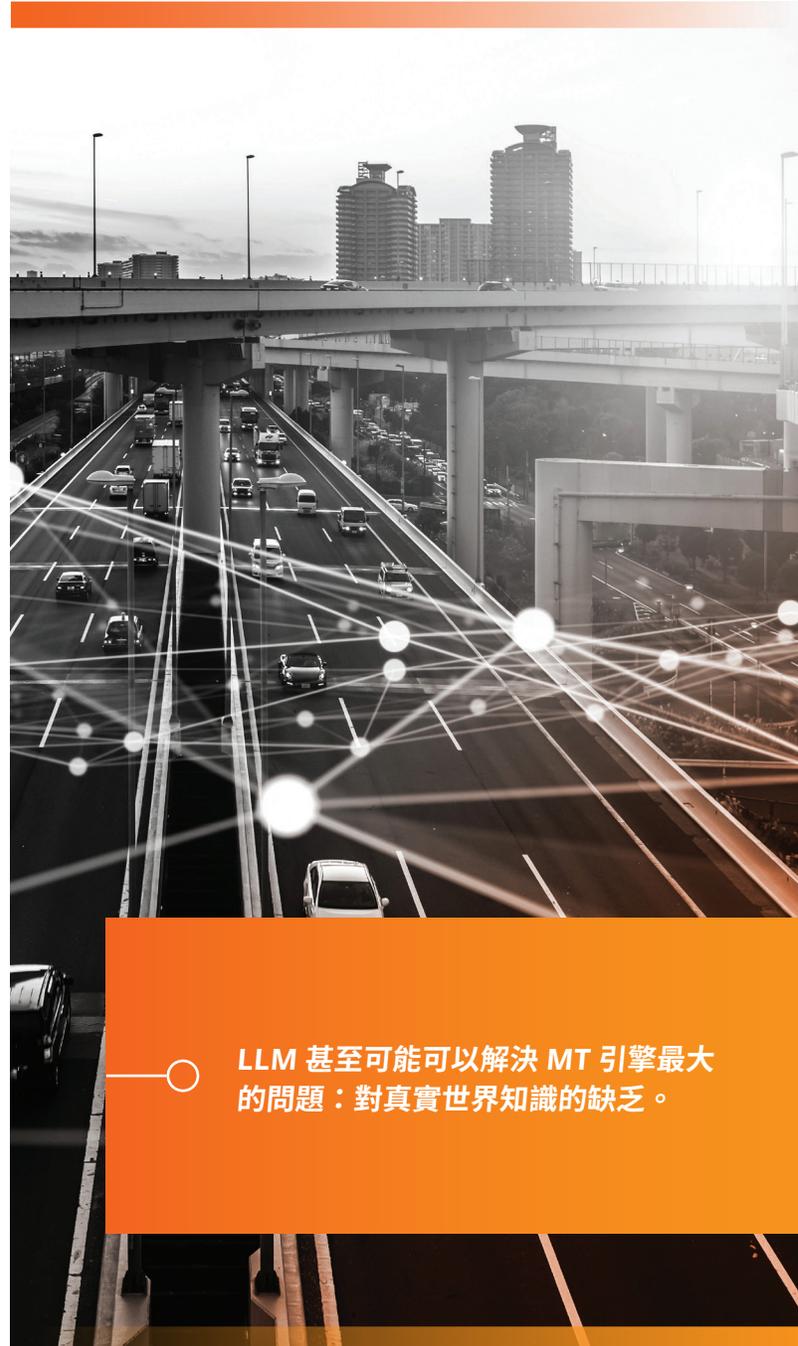
公司企業將能以更快的速度製作更多內容，但能翻譯這些內容的譯者資源成長速度，會趕不上內容創作的速度。即使 MT 有所改進且譯者生產力也會提高，翻譯社群仍舊將難以滿足這樣的翻譯需求。

機器翻譯採用率增加

隨著新技術典範的出現，以及機器翻譯品質的改善，對翻譯服務的需求將持續成長，會有更多情況及使用案例開始採用這個技術。

運用機器翻譯提升客戶體驗

MT 品質的改進，加上需要打造更貼近個人的客戶體驗，公司企業將會更頻繁地運用 MT 來增進全球客戶的數位體驗，建立更穩固的關係。



LLM 甚至可能可以解決 MT 引擎最大的問題：對真實世界知識的缺乏。

總結

許多科技公司都對 LLM 技術展現極高的興趣。Microsoft 就對 OpenAI 投資了 100 億美元，而 Nvidia、Google 和其他公司也大幅投資在 LLM 及 AI 技術上。想充分了解 GPT 與本地化以及大型語言模型的真正潛能，以便為語言產業帶來價值，我們必須要：

- 以大量的實際案例進行測試，針對各個本地化與編修工作類型，評估其錯誤率
- 針對本地化價值鏈中的宏觀與微觀使用者歷程進行詳細分析，找出這類文本自動化作業可能會對哪些地方造成衝擊
- 了解如何提示和提供 GPT 和其他 LLM 大量的相關脈絡，並記錄可能的錯誤及最佳實務做法。
- 研發新的自動化工作流程以及人機迴圈 (human-in-the-loop) 編輯工作流程，並以納入這個嶄新 AI 為前提，發想日後該怎麼進行譯後編修與品管 (QA)
- 針對每個可能改善的地方，為本地化公司及客戶設計新的自動化與使用者體驗 (UX) 互動場景
- 確認授權、部署費用及維護等相關經濟條件對我們的業務而言是合理的

對於未來，我們非常期待，也將會繼續評估 LLM 並發表結果，協助您掌握這個令人興奮演進的最新狀況，進而能從中獲益。



○ 許多科技公司都對 LLM 技術展現極高的興趣。

◎ LIONBRIDGE 機器翻譯專家



Rafa Moral

創新副總裁

Rafa 負責監管和語言及翻譯相關的研發 (R&D) 事務，包括跟機器翻譯、內容剖析與分析、術語探勘及語言品質管與控管的相關事務。



Yolanda Martin

MT 專員

Yolanda 負責建立客製化翻譯模型，以及用以細微調整這些模型的品質分析與策略研發。她同時也會與研發 (R&D) 團隊協同合作，開發新的語言工具和資源。



Thomas McCarthy

MT 業務分析師

Thomas 的職責是要確保 Lionbridge 客戶與利益相關人士能充分獲益於 MT 相關技術、服務及諮詢。



Lionbridge 專家能協助客戶輕鬆且有效地進行機器翻譯並增進業務執行。有鑑於機器翻譯領域正不斷快速變動，Lionbridge 也會密切評估新興技術，協助客戶進一步獲益於技術發展。

如要深入了解 Lionbridge 如何能協助您充分發揮自動翻譯的效益，歡迎立即與我們的團隊聯絡。



關於 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。超過 25 年以來，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 24 個國家/地區設有解決方案中心。



如需深入了解，歡迎造訪

[LIONBRIDGE.COM](https://www.lionbridge.com)



LIONBRIDGE

© 2023 Lionbridge. 著作權所有，並保留一切權利。

