【2021 全國科學探究競賽-這樣教我就懂】

高中（職）組

題目名稱： Google Translate 你懂我嗎？

一、摘要：

本研究以日常生活中最常見的 AI 翻譯軟體「Google Translate」為研究對象，探究 Google Translate 是否能辨識中文歧義句（即，一句話有兩種以上的語意）的所有語意，並正確翻譯成英文與日文。此研究主要分為兩個主題：句構歧義句（指一個句子因為結構的不同而導致兩種可能的語意）、詞彙歧義句（指一個詞彙在句子中有兩種或兩種以上可能的語意）。而研究結果顯示，在分辨句構歧義句的正確率，中翻日高於中翻英；在分辨詞彙歧義句的正確率，中翻英傾向高於中翻日。但這兩大主題的中翻英和中翻日的正確率都沒有到達 100%，由此可知，Google Translate 並不能準確及完整的翻譯出句構歧義句和詞彙歧義句。

二、探究題目與動機

現代社會的科技日新月異，從電腦到手機，再到所謂的智慧「AI」，現在科技能替我們做的工作越來越多也越來越複雜，很多事情我們不需要透過書面資料的查找或是擁有專業的姿勢就能夠解決，例如—語言的翻譯。現在，如果你需要將中文轉換成英文，只要到網路瀏覽器上去輸入關鍵字，你需要的那個單詞就會出現，像是 Google Translate、牛津字典，又或者是語音助理 Siri，這些工具皆能夠幫助你達成你的目的。

目前以廣大數據庫為依據的 Google Translate 應屬網路上最常被使用的翻譯工具。然而，我們觀察到 Google Translate 有時候翻譯出來的句子和我們平常在書寫、口說用語差得天差地遠，因此我們想要了解 Google Translate 的翻譯程度能夠準確到什麼程度，在什麼樣的組織結構上會開始出現明顯的錯誤。

在日常生活中我們發現中文時常會有歧義的問題。歧義是指一個單詞、片語、句子等擁有一個以上的解釋。常常會因為歧義的關係，有可能導致說話者與聆聽者對於這個句子產生不同的闡述，甚至會因不同的解釋，造成不必要的誤會。因此，我們的研究題目是：當需要翻譯的中文句子是歧義句時，Google Translate 是否能準確翻譯出所有語意，如果不能，又會出現什麼樣的問題？我們希望從這份研究可以瞭解 Google Translate 的翻譯模式與翻譯困境，進一步看見目前 AI 翻譯的進程。

三、探究目的與假設

為了瞭解 Google Translate 在處理中文歧義句時，是否能準確翻譯出所有語意，我們必須從兩個歧義的種類分析：「句構歧義」與「詞彙歧義」。「句構歧義」是指一個句子因為結構的不同而導致兩種可能的語意，例如 ：「我想出來了」有兩種意思：第一種是「我想出一個點子」，第二種是「我想踏出一個空間」。「詞彙歧義」是指一個詞彙在句子中有兩種或兩種以上可能的語意，例如 :「用毒毒毒蛇，毒蛇會不會被毒毒死」，在這個例句當中，「毒」具有三種語意: 第一個是「名詞的毒，意思對生物體有害的物質」第二個是「動詞的毒，意

思為傷害」，第三個是「形容詞的毒，意思為有害的性質或東西」。

生活中時常被使用的 Google Translate 是否能將這類具有歧異的句子完整翻譯出每種解釋，是我們想要去探討的。由於英文是國際語言，而日文是除了英文以外，在台灣也很多人會學習及貿易往來使用的語言，因此，我們希望能以「中翻英」和「中翻日」來測試 Google Translate 處理中文歧義句的情況。我們的研究問題有二：

（一）Google Translate 是否能在「中翻英」時，正確翻譯含有「句構歧義」與「詞彙歧義」的中文歧義句？

（二）Google Translate 是否能在「中翻日」時，正確翻譯含有「句構歧義」與「詞彙歧義」的中文歧義句？

四、探究方法與驗證步驟

(一)研究方向 ：本研究探討 Google Translate 在「中翻英」和「中翻日」時，是否能正確翻譯含有「句構歧義」與「詞彙歧義」的中文歧義句？

(二)研究對象 ：Google Translate

(三)使用材料 :我們與朋友間的嬉鬧時發現，在我們說話的過程中只要稍微變換語調或者劇中字詞有兩個以上的可能語意，整句話的理解就會變得不同。有些網站跟 Youtubers 也會探討這些問題，因此，我們便從那些網站以及 Youtubers 的影片裡的例句來當我們做研究的參考依據。原本找了 30 餘句左右，其中包含了句構歧義、詞彙歧義以及文章，但之後考量有些文句的語意分解起來實在太過繁雜或者原本中文的意思就非常艱澀難懂，Google Translate 可能無法判讀，最後只留下 18 句（句構歧義與詞彙歧義各 9 句），如表一和表二所示。

(四)資料收集 ：使用 Google Translate 做為翻譯的工具。將蒐集到的句子放入 Google Translate 的左側 (原文語言選擇繁體中文)，則右側將會跑出翻譯完成的句子 (譯文語言分別選擇英文和日文)，並將所有的題目與翻譯完成的句子整理至表格中，方便分析比較。

(五)資料分析：將 Google Translate 翻譯出的英文句子、日文句子的意思與中文句子互相比較，探討譯出的句子是否和中文歧義句的語意一致。若不一致，正確率有多高？另外，也檢視中翻英、中翻日的文法問題，探究造成翻譯問題的原因，進而分析 Google Translate 的翻譯模式與困境。為確保分析內容無誤，我們邀請一位教授英文課程的本國籍大學教師與一位取得日文檢定一級的本國籍專家共同討論。

五、結論與生活應用

(一) 句構歧義句：

研究結果顯示，9 句中文歧義句應有 20 個意思，但英文只翻出了 9 個意思，因此中翻英的正確率為 45%，而日文總共翻譯出了 11 個意思，因此中翻日的正確率為 55%。經過這次研究發現 Google Translate 在中翻英和中翻日是有著許多問題的，而我們將這些問題與發現做了以下統整 ：

1.大部分的中翻英或中翻日都只能翻出一種意思

2.以比例看來，中翻日的正確率高於中翻英

3.有時候雖然有翻譯到中文句子的意思，卻沒有到很完整以及精確，例如第 4 句

4.有時候只是湊巧翻對，例如第 6 句，假設又換成另外一個句子，那翻譯就會是錯的

而透過這些問題與發現，我們推測：

1.電腦會優先將處理中的詞彙連結到剛剛處理好的資訊，而不選擇保留給之後的資訊，或者是某些字詞在電腦的資料庫顯示特定一種意思使用的頻率較高，造成只有翻譯出一種意思。

2.日文的句構大部分和中文較像，例如「關係子句」，中文和日文都是形容詞子句都是放在名詞子句的前面，而英文的形容詞子句則是放在名詞子句的後面，才會造成中翻日的的正確率高於中翻英。

3.Google Translate 尚未有能力處理中翻英時，中文句子沒有主詞的情況（中文是可以沒有主詞的語言，但英文不行）。

我們的結論是，以現今的 Google Translate 來說，尚未有能力分辨有結構歧義的句子是否為歧義句。因篇幅有限，本報告呈現部份詳細結果，如表一：

表一　句構歧義句翻譯結果

中文歧義句與翻譯結果
1. 縣裡的通知說，讓趙鄉長本月 15 日前去匯報。 英：The notice from the county said that the mayor Zhao will report before the 15th of this month.（正確率：1/2） 日：郡からの通知は、趙市長に今月の 15 日までに報告させるように言っています。（正確率：1/2） 說明：英、日文翻譯都只翻譯出「本月 15 日前」。 推測：(1) Google Translate 尚未有能力分辨句子是否為歧義句。 (2)「前」被系統分類為時間副詞，可能是電腦會優先將處理中的詞彙連結到剛剛處理好的資訊，而不選擇保留給之後的資訊。也有可能是電腦的資料庫顯示「前」拿來當作時間副詞的頻率比動詞複合詞（如：「前」去）的頻率更高。
2. 剛才打電話來的是前幾天才認識的張洪民的哥哥。 英：The one who called just now was Zhang Hongmin's brother who only met a few days ago.（正確率：1/2） 日：ちょうど今電話をかけたのは、ほんの数日前に会った張紅民の兄でした。（正確率：2/2） 說明：1. 英文只翻譯出「前幾天才認識的」用來形容「張洪民的哥哥」，但日文的翻譯結果和中文一樣是歧義句（和中文一樣有兩個意思）。因為日文的形容詞子句（即，關係子句）和中文一樣，都放在名詞片語之前，但是，英文的形容詞子句是放在名詞片語之後。 2. 「前幾天才認識的」的英文翻譯文法有誤，缺少主詞。 3. 「張洪民」的日文翻譯有誤，變成「張紅民」。 推測：(1) Google Translate 尚未有能力處理中翻英時，中文句子沒有主詞的情況（中文是可以沒有主詞的語言，但英文不行）。 (2) 「張洪民」的日文翻譯有誤，不確定是什麼原因。
3. 今年我校夏季運動會的入場式，每班有比賽項目的男生和女生的一半要保證參加入場式。 英：For the admission ceremony of our school's summer games this year, half of the boys and girls in each class have

to participate in the admission ceremony. （正確率：0/4）

日：今年の当校の夏季オリンピックの入学式では、各クラスの男子と女子の半数が入学式に参加する必要があります。（正確率：0/4）

說明：中文的句子有四個語意，但是英、日文都沒有正確翻譯出任何一個語意，都沒有翻譯到「有比賽項目的」。

推測：Google Translate 似乎對於複雜的句構，如：**形容詞子句**（「有比賽項目的」）+ **含有連接詞的名詞片語**（「男生和女生」）+ **量詞**（「的一半」），有處理困難。

4. 我看見張原扶著一位老人走下車來，手裡提著一個黑色皮包。

英：I saw Zhang Yuan helping an old man get out of the car with a black leather bag in his hand. （正確率：2/2）

日：張元が黒い革のバッグを手に持って車から降りるのを手伝っているのを見た。（正確率：2/2）

說明：1. 英文的翻譯和中文一樣形成句構歧義句，恰好兩種語意都有。

2. 日文的翻譯兩種語意大略有翻到：第一句是說老人（の）拿黑色皮包，第二句是說張原拿黑色皮包。

3. 日文的翻譯不夠完整：「我」沒有翻譯到，「張原」翻成「張元」。

推測：(1) Google Translate 似乎知道日文時常省略主詞。

(2)「張原」的日文翻譯有誤，不確定是什麼原因。

5. 小明想出來了。

英：Xiao Ming figured it out. （正確率：1/2）

日：シャオミンはそれを理解しました。（正確率：1/2）

說明：英、日文都只翻譯「想出來什麼東西」，而不是「想要出來」。

推測：Google Translate 把「想出來」看做是一個動詞片語，而不是兩個動詞。有可能是電腦的資料庫顯示「想出來」拿來當作動詞片語的頻率比雙動詞的頻率更高。

6. 他誰都不認識。

英：He doesn't know anyone. （正確率：1/2）

日：彼は誰も知らない。（正確率：2/2）

說明：1. 英文沒有主題化的語法，所以只翻譯了一種語意。

2. 日文和中文恰巧都有「主題化」的語法，在此例中，翻譯恰好相同，都形成歧義句。

3. 但是如果中文句子是「飯你吃了嗎？」，日文的翻譯就有問題。

推測：Google Translate 沒有中文、日文主題化的相關知識。

7. 哥哥看到妹妹拿刀子怕她傷到自己。（分析結果略）

8. 衣服嘛，冬天能穿多少就穿多少，夏天能穿多少就穿多少。（分析結果略）

9. 以前，我不懂什麼叫做愛 。 現在，我不懂什麼叫做愛。（分析結果略）

(二) 詞彙歧義句：

本研究結果發現，9 句中文詞彙歧義句應有 27 個意思，英文翻出了 12 個意思，因此中翻英的正確率約為 44.4%，但日文只翻譯出了 11 個意思，因此中翻日的正確率約為 40.7%。

換言之，Google Translate 在中翻英和中翻日是有著許多問題的，而我們將這些問題與發現做了以下統整 ：

1.大部分的中翻英中翻日都只能翻出整個句子中有多重意思且相鄰的的第一個字，後面就無法完整翻出，但還是有例外，例如第 8 句，就是只翻譯出第三個「毒」字的意思。

2.遇到人名可能無法辨認，例如第 7 句，無法辨認出「過兒」為人名。

3.在中文有的詞彙，英文和日文反而沒有類似的解釋，導致翻譯完全正確，例如第 6 句。

4.以比例來說，中翻英的正確率高於中翻日，但因測試句只有 9 句，恐不達統計上的顯著差異。

而透過這些問題與發現，我們推測 ：

1. Google Translate 對於連續三個同形異義詞尚無法處理。
2. Google Translate 在面對含有形容詞子句、較為複雜的句子時，會忽略處理形容詞子句。
3. Google Translate 尚無法擁有目標語言的文化知識，所以無法翻譯出較有文化特色的事物。
4. Google Translate 對於中文詞彙的詞義掌握仍不齊全，無法完全翻譯正確。

我們的結論是，以現今的 Google Translate 來說，無法完整辨認句子中所出現具有多重意思的字詞。因篇幅有限，本報告呈現部份詳細結果，如表二：

表二　詞彙歧義句翻譯結果

中文歧義句與翻譯結果
1. 女孩給男朋友打電話： 如果你到了，我還沒到，「你就等著吧」；如果我到了，你還沒到，「你就等著吧」。 英：The girl calls her boyfriend: If you arrive, I haven't arrived yet, "Just wait." If I arrive and you haven't arrived yet, "Just wait." (You die.) （正確率：1/2） 日：女の子は彼氏に電話します。あなたが到着した場合、私はまだ到着していません。「ちょっと待ってください」。私が到着し、あなたがまだ到着していない場合は、「待ってください」。（正確率：1/2） 說明：英、日文都只翻譯「等待」的意思，沒有「警告意味」的翻譯。 推測：Google Translate 只收錄了語意資訊，但尚未收錄足夠的語用資訊。
2. 遇見她之前，「我沒想過跟別人結婚」。遇見她之後，「我沒想過跟別人結婚」。 英：Before I met her, "I never thought about marrying someone else". (anyone) After meeting her, "I never thought about marrying someone else".（正確率：1/2） 日：彼女に会う前に、「私は他の誰かと結婚することを考えたことはありませんでした」。彼女に会った後、「私は他の誰かと結婚することを考えたことはありませんでした」。（正確率：1/2） 說明：英、日文的翻譯相同，只正確翻譯了第二個句子。 推測：Google Translate 把「別人」看作「除了某人之外的其他人」，而沒有「除了說話者以外的其他人」的意思。
3. 多虧跑了兩步，差點沒上 上 上海的車。 英：Thanks to two steps, I almost missed the car in Shanghai. (to) （正確率：2/3）

日：2 つのステップのおかげで、私は上海で車をほとんど逃しました。（almost, miss 直翻，但日文不這樣講）（以都市為核心：「上」大城市、「下」鄉下）（正確率：1/3）

說明：1. 英文的翻譯只翻了第一個「上」（missed）和第三個「上」（Shanghai），而第二的「上」則翻錯。

2. 日文的翻譯只翻了第三個「上」（上海），而第一個「上」（逃しました）翻錯（應該用「乗る」），第二個「上」也翻錯（應該用「へ」の）。

推測：Google Translate 對於連續三個同形異義詞尚無法處理。

4. 明明 明明 明白 白白 白，就是不承認。

英：Obviously, I just didn't admit it. (Ming Ming clearly understands that Bai Bai has fair skin,) （正確率：明：1/3；白：0/3）

日：明らかに、私はそれを認めませんでした。（和英文翻譯一樣）（正確率：明：1/3；白：0/3）

說明：英、日文的翻譯相同，只翻了第三、四個「明明」（obviously,明らかに），而第一、二個「明明」（人名）、「明白」、第二、三個「白白」（人名）和第四個「白」，通通沒有翻譯。

推測：Google Translate 對於連續三個同形異義詞尚無法處理。

5. 校長說:衣服上除了校徽別 別 別的。

英：The principal said: There is nothing else on the clothes except the school badge. (Do not wear anything)（正確率：1/3）

日：校長は言った：学校のバッジ以外に服には何もありません。（正確率：1/3）

說明：英、日文的翻譯相同，只正確翻譯出第三個「別」。

推測：Google Translate 對於連續三個同形異義詞尚無法處理。

6. 雨天騎腳踏車，車輪打滑，還好我反應快，一把 把 把手穩住了。

英：Riding a bicycle on a rainy day, the wheels slipped, but fortunately, I reacted quickly, holding the handle firmly.（正確率：3/3）

日：雨の日に自転車に乗ると車輪が滑ってしまいましたが、幸いハンドルをしっかり握って素早く反応しました。（正確率：3/3）

說明：雖然英、日文只翻譯了一個「把」（第二個「把」），但因為這兩個語言沒有對應中文的「一把」這種詞彙與「把東西做何處置」的介詞用法，所以，翻譯算完全正確。

推測：Google Translate 應該對於連續三個同形異義詞尚無法處理，只是恰巧英、日文沒有「一把」和「把東西做何處置」的介詞用法，以至於此句雖沒有翻譯卻仍屬翻譯正確。

7. 來到楊過曾經生活過的地方，小龍女動情的說:「我也想過過 過兒 過過的生活」。（分析結果略）

8. 用毒 毒 毒 蛇，毒蛇會不會被毒毒死。（分析結果略）

9. 小明，今晚有思修課，你去不去? " 我去 ！ 我不去 ！！"（分析結果略）

上述研究結果告訴我們不能完全相信翻譯軟體，軟體翻譯結果僅能用作參考。

參考資料

McGregor, W. B. (2015). *Linguistics: An introduction (2nd Edition)*. UK: Bloomsbury Academic.