

針對數學與科學教育領域之電腦輔助英中試題翻譯系統

呂明欣
國立政治大學資訊科學系
94753007@nccu.edu.tw
劉昭麟
國立政治大學資訊科學系
chaolin@nccu.edu.tw

高照明
國立台灣大學外國語言學系
zmgao@ntu.edu.tw
張俊彥
國立台灣師範大學地球科學系
changcy@ntnu.edu.tw

摘要

由國際教育學習成就調查委員會(IEA)統一命題之國際數學與科學教育成就趨勢調查測驗(TIMSS)，為便於台灣中小學生施測與理解，英文原文試題內容需要經過許多人工討論及翻譯時間。為了增進翻譯內容一致性及其效率，我們設計一套符合測驗試題的輔助翻譯系統，將不同格式的試題文件，經執行語法分析式的片語擷取和字典查詢，透過使用者介面，選擇合適的片語詞彙翻譯選項和詞序調整，以及提供目前常用之線上翻譯服務、回顧翻譯類似句、以及加減詞彙等功能。為了能提昇翻譯詞彙的選擇正確性，我們記錄翻譯者選詞動作，讓翻譯者能回顧過去曾處理過的翻譯類似句，並且按照系統提供之選詞頻率資訊、科學領域的期刊語料之詞頻統計，以及利用統計式中英詞彙對列和語言模型，更改選詞的優先順序。我們嘗試以過去試題為實驗對象，按年級及學科區分 6 大試題類別，搭配 4 種選詞策略，透過 BLEU 及 NIST 之翻譯評估指標比較線上翻譯系統和本系統，實驗結果顯示在各實驗組的評估上均有優於線上翻譯系統的效果。

關鍵詞：自然語言處理，電腦輔助教學，受限語言，試題翻譯，機器翻譯，TIMSS

一、緒論

TIMSS(Trends in International Mathematics and Science Study) [16]，由國科會委託國立台灣師範大學科學教育中心（以下稱師大科教中心）負責，針對國小四年級及國中二年級學生執行數學與科學領域之試題翻譯及測驗工作。主要工作流程包含：從 IEA 取得英文試題內容，師大科教中心決議進行翻譯工作分配、中文試題交換審稿校正及翻譯問題討論，最後將中文翻譯試題定稿。其中完成上述說明之工作流程，依照 TIMSS 官方網頁歷年統計至少需要一個月以上的準備時間。主要原因在於試題翻譯的原則上須避免在翻譯過程中由於個人主觀的因素（如：為具體表達題意，加入不適當詞彙，或者是同一題目中，主要詞彙因在文章前後翻譯方式不一致而產生的語意混淆）或不同的翻譯者因不同的翻譯標準使得翻譯內容有所出入，間接影響試題在題意表達上的品質。故為了能夠完整表述且不影響題意的正確性，過去投入了不少的人力資源及時間成本。因此，若有一個好的翻譯輔助系統[3]協助翻譯工作的話，除了可以提供在字詞翻譯上一致性的標準以外，利用電腦及機器翻譯的技術取代人工檢查字詞翻譯亦有可能提升翻譯的準確及效率。

另外，TIMSS 試題在內容上，依題型種類區分為選擇題及問答題，句型結構多半以直述句和問句構成問題的題幹，加上誘答選項，因此語法結構和主題較一般翻譯文句明確簡短，亦不需考慮太過複雜的修辭方式。

有鑒於試題內容較一般文章結構精簡的特性，在機器翻譯的領域中有所謂子語言(sublanguage)的翻譯類型。即翻譯需求限制於某一特定領域上，減少翻譯過程所發生的

歧義、修飾文法及實作的複雜度。現今的研究多採用受限語言(controlled-language)的方式，限定專業領域的語彙和語法結構的範圍，減少詞彙歧義[4]和複雜性，以及加入人機互動式的機器翻譯模式。因此，我們在系統設計重心上，以尋找專有名詞，將試題內容經由斷詞及文法分段，利用查詢詞典，輔以統計式翻譯模型[1]找出所有可能適合原文詞彙的翻譯結果，並提供使用者介面，以人工編輯或半自動的方式，透過中文句型與英文句型的對應找到合理的詞序關係。

二、TIMSS 試題翻譯系統

基於緒論的中英試題翻譯之問題定義及系統設計重心，我們完成的試題翻譯系統，在執行的流程主要區分為 3 部分。

- i. 將 TIMSS 英文試題透過文件轉檔程式轉成文件檔
- ii. 從系統翻譯模組選擇線上翻譯模式，或字典翻譯模式，配合修正翻譯模組產生中文翻譯結果
- iii. 將翻譯結果輸出至使用者介面

首先，試題資料的檔案格式可能為純文字檔、Microsoft Word 的 DOC 檔或 Adobe Systems 制定的 PDF 檔等非純文字檔格式，故要能正確讀取文字內容之前必須先進行轉檔動作。我們呼叫 JACOB PROJECT[13]的應用程式套件進行純文字檔轉檔，目前，利用轉檔的缺點在於當 PDF 檔的內容為中文內容，且原先文書檔之存檔方式是以特殊字型的方式存檔（如利用 LaTeX 文書編輯工具定義之字型），這時中文檔案會由於編碼還原問題產生亂碼的情形，而英文內容之檔案則無此問題。接著，存入暫存檔之後再讀取其文字內容，依題號、試題內容及誘答選項，並輸出題目選單於題目選擇及新增修改面板供翻譯者選擇欲翻譯之試題內容。

系統將 TIMSS 試題檔轉換成純文字檔後，接著視翻譯者選擇可切換線上翻譯模式及字典翻譯模式。

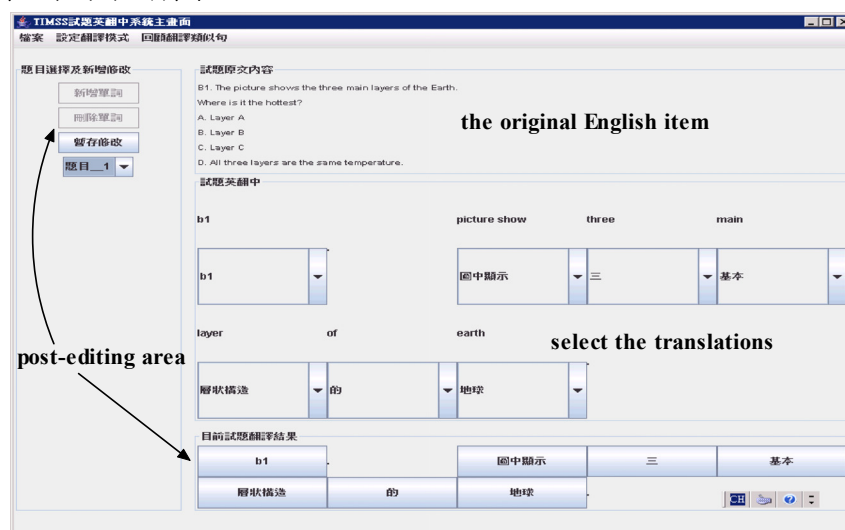
- i. **線上翻譯模式**：我們考慮到目前谷歌所推出的 Google Translate[12]以及雅虎線上翻譯系統[19]，實作 Http Connection Interface，輸入翻譯句子後傳回譯文的 html 文字檔，經過字串剖析，將 html 語法去除，逐行檢查並判斷該行是否存在具有翻譯結果字串之左右相鄰識別文字，若符合則取出左右相鄰識別文字中間之翻譯結果字串，再透過介面顯示如圖一之翻譯訊息，系統會經翻譯者確認，自動暫存英文原文和中文翻譯的結果，待最後儲存動作完成後存入檔案中。



圖一．線上翻譯模式介面

- ii. **字典翻譯模式**：翻譯者按題號選擇所需翻譯內容後，系統會讀取該題目，在取得欲翻譯之英文句子後，會先行判斷目前系統翻譯模組的選擇，若翻譯者選擇字典翻譯模式，則會經由字典查詢單字的方式，將所有單字和詞組的所有可能的中文翻譯，輸出至使用者介面，讓使用者進行後續之調整翻譯詞序和中文翻譯選詞

等動作，如圖二所示。



圖二．字典翻譯模式介面

然而，翻譯詞序和中文翻譯選詞的動作，就翻譯者本身來說仍然需要許多人工檢查的時間，因此，為增進選詞和調整詞序的效率，我們需要足夠多的中英試題翻譯資訊，利用機器學習的技術，找出中英試題在翻譯和文法結構的關連性。然而 IEA 每隔四年舉行一次 TIMSS 測驗，語料來源上，我們目前僅有 1999 年及 2003 年所舉辦的 TIMSS 中英對照測驗題目（在之後章節，我們分別以 TIMSS1999 及 TIMSS2003 來表示 1999 年及 2003 年所出題之測驗題目），在試題資料部分並不足以能利用機率統計的方式，依照語境準確選出適當表達語意的中文詞組，並還原中文翻譯的詞序。因此，需要透過其他外部資訊或中英語料庫的輔助，做為替代的方案。在第三章和第四章，我們將會介紹如何做到翻譯的一致性，詳細說明中文詞組翻譯及選詞順序調整的策略。

三、翻譯的一致性

一篇好的中文翻譯內容，能掌握翻譯前後間內容的一致性，即完整陳述原文的意思，在語彙上使用正確詞義表達，表達的文法形式也要能和原文一致，並且流暢易讀。以 TIMSS 試題來說，要能正確翻譯出試題陳述部分，包含內容出現的問句片語、專有名詞，和提示考生在做題目時常用詞彙，以適切引導考生閱讀試題內容。此外，問句主體在翻譯內容前後亦不可出現不同的翻譯結果。

為避免不同翻譯者會有不同的翻譯習慣，翻譯者經常需要檢討需要統一翻譯的詞彙。以提示考生的常用詞彙來說，如 check one box 會統一翻譯成“勾選一項”、as shown below 統一翻譯成“如下所示”，故針對上述現象，系統亦需提供翻譯者可隨時動態調整其中英對照翻譯規則之彈性。

另外，傳統市售的翻譯系統，或線上翻譯系統所提供之翻譯結果通常為語料統計中最有可能的翻譯結果，但對於特定領域或一詞多義的原文內容，不能完全保證其翻譯結果之正確性，翻譯者也無法直接針對特定領域知識之原文詞彙，任意修正其翻譯結果，讓系統經過修正後，選擇最好的解釋。故系統在使用者介面的設計上，要能提供翻譯者選擇符合語境的翻譯選項及合理的詞序調整方式，並且在接受翻譯者調整翻譯後，能夠根據先前的調整，自動修正類似句之翻譯詞彙。

本系統在翻譯過程中針對需一致性翻譯的詞彙或詞組，定義以下檔案，讓翻譯者能自行定義詞組的中英翻譯規則。

- i. **規則詞典檔**：將試題翻譯小組所決議之統一翻譯詞組以人工的方式建立中英對照詞典檔。
- ii. **片語句型檔**：為常用的名詞片語、動詞片語及形容詞片語等詞組的中英翻譯對照檔。和規則詞典檔不同的是，規則詞典檔所記錄的詞組是連續的詞組，如 *in order to*（爲了）或 *build up*（建立，構成），而片語句型檔所記錄的詞組則爲非連續的詞組，如 *place~ on~*（散佈某物於某處）或 *carry~away*（載走某物）。

另外，爲了減少討論單字翻譯的正確性及適切程度的時間，讓翻譯者檢索過去已有的翻譯內容，我們提供回顧翻譯似句之功能，將檢索的翻譯結果以中英文字對齊（*concordance*）的方式呈現於使用者介面。以下爲主要之檢索執行流程。

- i. **從中英對照詞組檔建立倒轉檔（*inverted file*）**：當翻譯者利用檢索視窗輸入英文檢索單字或連續性片語，系統會從過去的倒轉檔內容，找尋所有和檢索單字或片語有關的中英對照詞組檔之存放位置，再逐一搜尋檔案之內容，檢查檔案內所有中英對照結構，其英文單字片語是否和檢索詞相同，若相同，則將目前比對位置的所有中英對照結構、英文原文句子及中文翻譯結果儲存於一緩衝區（*buffer*）。
- ii. **產生中英文字對齊（*concordance*）結果**：從緩衝區儲存的中英對照結構，找出檢索詞的中英的對照結構，接著把英文原文句子及中文翻譯結果，以檢索詞的中英的對照結構爲中心，將英文原文句子及中文翻譯結果切成左右兩部分，最後根據翻譯者所輸入的資料筆數，檢查緩衝區符合檢索詞的句子數目，以及檢查所有檢索詞中文翻譯，將英文原文句子及中文翻譯結果，依照相同中文翻譯句子最多的順序爲前的排序方式，輸出至結果視窗，如圖三所示。



圖三 中英文字對齊介面

四、中文詞組翻譯及選詞順序調整

除了利用自行定義詞組的中英翻譯規則，以及回顧翻譯句的方式，掌握部分詞組的中文翻譯之外，我們還定義一套詞組翻譯及翻譯修正流程，以下爲詳細步驟。

i. 詞性標記及 *stemming*

我們呼叫 *MXPOST*[15]詞性標記工具將原文加入標記，並根據 *Porter* 演算法[17]，實作還原各詞性（名詞、動詞、形容詞、副詞）的詞幹變化（如 *~ly*、*~ing*、*~ed* 等）。對於某些特殊詞幹變化的單字（如動詞 *break*，其過去式爲 *broke*，被動式爲 *broken*，以及名詞單複數型態）比較難以用演算法處理的部分，我們利用 *WordNet*[25]，依照詞性

做字典檔搜尋找到 stemming 後的結果。

ii. 斷詞處理及詞組翻譯

英文句子部分包含單字 (word)、詞組 (term) 及標點。其中，詞組即為數個單字的組合、片語 (phrase, 如名詞片語、動詞片語、形容詞片語) 或複合名詞。

在單字和詞組的語彙表達程度上，詞組往往能掌握比較完整的動作表達或描述。若單純只用一般字和字之間的空白做斷詞，便無法利用斷出來的單字決定詞組的意思。

因此，我們希望在翻譯上能做到英文單字的字意能完整對應中文的單字字意，以及做到詞組式的翻譯，使得文句的語意能捕捉得更完整。而第一步要先有辦法斷出正確的字詞，接下來才能找出正確的中文詞組翻譯。

在斷詞方法上，首先將英文句子，利用字和字之間的空白做基礎斷字的動作。但這時要先考慮標點符號的情形，因為標點符號在英文文句中通常是連在單字右方，因此先檢查文句有標點符號時再插入空白字元，再利用空白斷詞。接下來進入詞組斷詞部分，我們分別利用英文剖析器 MINIPAR[14]及片語句型檔斷詞方法，將英文句子內片語或詞組找到，再用字典找出其片語翻譯，其中字典部分我們使用牛津現代英漢雙解詞典[29] (Concise Oxford English Dictionary, 收錄 39429 個詞彙)。在字典查詢完兩種尋找片語或詞組方法之翻譯後，接著進入長詞詞組篩選階段，首先，排除重覆找到的詞組翻譯，再來檢查若片語句型檔分詞或 MINIPAR 分詞之中出現片語結構類似，片語長度不同的情況，則選擇較長的片語翻譯結果，最後輸出完成片語翻譯之句子。

MINIPAR 分詞上，MINIPAR 以解析其語法解析結構，擷取最常出現之名詞片語 (即 NP 或 NPL) 及動詞片語 (VP 或 VP+PP) 及形容詞片語 (ADJP)。

片語句型檔斷詞方法上，我們以句型比對的方式，除了按照相鄰片語做搜尋外，利用 “*” 代表任一 word, “<>” 代表左右詞彙限制範圍內所有之 word, 都可以利用該演算法擷取其片語。故像要抓取 “pulse rate”、 “success rate”、 “~rate” 等複合名詞，可利用 “* rate”，代表抓取 rate 前一個 word 當做複合名詞；若需要抓取 “as ~ as possible” 或 “as ~ as” 等句型時，可利用 “as <> as possible” 及 “as <> as” 之描述方式，來抓取這些類似片語。惟須注意的是，比對的句型若有相似結構但不同長度的字串樣式，如一英文句子為 “...the diagram shown below...”，同時滿足片語句型檔內的 “the diagram shown” 和 “the diagram shown below” 片語句型，則我們會選擇長度較長的 “the diagram shown below” 而不是選擇 “the diagram shown” 加上 “below” 做為分詞的結果。

iii. stop word filtering

執行英文翻譯時，會有某些單字因無法單獨存在，需搭配其他單字出現，因而出現不需翻譯或無法翻譯詞彙之情形。例如冠詞 the、an、a，常加於單數名詞之前，當不須特別強調單數時，便不需要特別翻譯冠詞。介系詞 for、to、of，當出現於疑問詞之後，或片語之中，其意義表達助動詞如 do、does，為了不讓後續翻譯動作翻譯出上述之詞彙，我們將這些詞彙建立一集合，並針對英文句子並做出以下判斷。

- 冠詞部分，the 直接去除，a 和 an 則判斷出現位置是否為句首，並檢查 a 後面是否有其他標點符號 (如括號或句點，代表有可能是編號)，若以上條件皆不符合，則予以去除。
- 介系詞部分大致保留，惟需判斷前一字元是否為 what、how、who、when、why 等疑問詞，若條件符合，則予以去除。另外，to 出現於句首時亦予以去除。
- 助動詞部分，則判斷前一字元是否為 what、how、who、when、why 等疑問詞，同介

係詞判斷方式，若符合則予以去除。

iv. 單字和專有名詞翻譯

在斷詞處理及片語翻譯階段，我們完成了片語的中文翻譯，剩下來的單字及專有名詞的部分，我們則再次利用查詢字典及修正翻譯模組的中英對照規則詞典檔的方式，找出每個單字及專有名詞所有可能的中文翻譯結果，做為中文選詞之選項。另外，針對英文姓名轉換成中文命名部分，我們參考 20000-NAMES.COM 網站，將網頁所列的男性及女性英文姓名按性別分類，分別存成文字檔，接著將待翻譯的英文句子所切分出來的詞彙，和先前儲存的所有英文姓名加以比對，找出相同姓名後，我們利用事先建立好的中文姓名庫，按性別回傳適當名字。例如，當英文名詞出現“Bob”時，屬於男性姓名，故從中文姓名庫的男性姓名分類中，以亂數方式，任意取得“小華”當作對應的姓名翻譯結果，並記錄目前“Bob”是和“小華”相互對應，以便同題內容中再出現“Bob”姓名時可做到人名一致性的翻譯。

西元表示部分，我們則是判斷句中的詞彙是否包含西元年月日等詞彙，作為自動轉換民國表示的標準。

v. 翻譯結果修正

在完成單字及片語翻譯之後，最後就是針對各單字翻譯出來的中文翻譯選項，調整適當的選項排列順序，以及詞序調整。

在第二章已說明目前 TIMSS 試題資料並不足以利用機率統計的方式，依照語境選出適當表達語意的中文詞組，並且還原中文翻譯的詞序。因此我們利用詞頻高低決定選詞重要程度的概念，記錄翻譯者過去選詞之詞頻數，並且利用中文語料庫統計詞頻，透過中英語料庫，以統計式翻譯為基礎，建立中英詞彙對列及語言模型。以下將針對翻譯者選詞詞頻、中文語料庫詞頻及中英詞彙對列和語言模型詳細介紹。

(一) 翻譯者選詞詞頻

中文選詞順序調整部分，我們考慮試題本身的敘述部分及問句所運用到的修詞方式不多，大部分以直接翻譯為主，故常用的詞組其對照的中文詞組種類大致不會變動太大。在這邊我們使用漸進式的訓練學習方法，翻譯者可利用介面選擇中文詞組選項，藉由統計最常被翻譯者選擇的選項做為日後選詞的優先順序，其他的選項則是再按次常選擇順序依此類推調整。為了達到這個目的，我們首先將每次經翻譯者確認翻譯無誤的中英文對照詞組統一記錄至中英對照詞組檔內，如表一所示。

表一、中英對照詞組檔

<p>英語題目： L5. When male wolves place their scent on trees, they most likely are doing this in order to</p> <p>經前處理及對照翻譯後儲存的中英對照詞組檔（以 BNF 表達儲存範例）： <English-Chinese Form> := [<English Word> <English Phrase> : <Chinese Word> <Chinese Phrase>] <Symbol> </p> <p><u>When male wolves place their scent on trees .</u> [when:當][male:雄][wolf:狼][place:置於][their:他們的][scent:氣味][on:上][trees:樹木][.]</p> <p><u>they most likely are doing this in order to</u> [they:他們][most:最][likely:可能][be:是][do:做][this:此][in order to:爲了]</p>

英文句子經過翻譯後，按下儲存動作，會自動將所有中英詞彙對照翻譯的結果，依照表一之格式，儲存在中英對照詞組檔內，接著，將目前儲存於中英對照詞組檔所有中英對照的結構，和修正翻譯模組的多義索引修正檔所有過去已有記錄之中英對照結構比對，若可以找到相同的結構，則將目前結構的選詞次數欄位加 1，若無相同結構，代表

過去的記錄無該中英翻譯對照之記錄，則將該結構儲存，並將選詞次數設定為 1。完成更新多義索引修正檔次數後，在下次翻譯新的句子時，就可以依照新的選詞次數高低，適當調整翻譯詞彙的選項順序。

除了利用多義索引修正檔調整翻譯詞彙的選項順序，針對目前翻譯句子中待調整選詞順序之單字，取出該單字所有中英翻譯選項及目前該單字前後詞彙之中英翻譯，即緊鄰詞彙，接著，從所有中英對照詞組檔中找出所有同為目前單字中英翻譯選項的前後詞彙，比較目前翻譯句子找到的緊鄰詞彙和過去留存在中英對照詞組檔的緊鄰詞彙是否相同，若相同或包含，則以當時所對應的中文翻譯選項為優先。如不相同，則取該字的中文翻譯中選擇次數最多的詞組為優先。例如：我們假設在“wolves place their scent”中，place 需要調整翻譯選項，我們知道 place 可以翻譯成“置於”或“地方”，並且確定 place 的緊鄰詞彙：wolf 翻譯成“狼”，their 翻譯成“他們的”。接著，從所有中英對照詞組檔找尋過去 place 的所有緊鄰詞彙中，是否存在 wolf 翻譯成“狼”和 their 翻譯成“他們的”記錄，假使發現存在，並且當時 place 翻譯成“置於”我們便決定 place 的翻譯即以“置於”為優先。

（二）中文語料庫詞頻

為了彌補目前資料量過少的語料來源，我們透過尋找有關數理、科學教育領域的中文語料，以補足語料庫的不足。之後，系統依此統計資料，計算中文詞組之詞頻高低，檢查英文詞彙所對應的所有中文翻譯，若該中文翻譯在語料統計資料上有相對應的詞頻高低值，則依詞頻調整其翻譯詞彙順序。

然而，英漢字典的中文翻譯詞彙，並不必然存在於語料庫之中，使得英漢字典詞彙，無法順利找出相對應語料庫詞彙之詞頻（如：科學領域專有詞彙），加上字典詞彙在描述上和語料庫詞彙可能不盡相同，可能具有意思相近或同義（synonymous）關係。例如：在字典詞彙中，[test：測試]表示英文在翻譯 test 一詞的中文詞彙為“測試”，但在語料庫句子中，有可能不使用“測試”一詞，而使用和“測試”相似的詞彙，例如“試驗”、“實驗”、“試”、“嘗試”等詞彙，這樣的情況下，就無法直接找出並統計“測試”一詞在語料庫上的詞頻數，而必須考慮和“測試”相似的詞彙在語料庫上統計的詞頻數，間接反映“測試”在使用上的重要程度。故我們先將語料庫所有中文詞彙找出其相關之近義詞彙集合，之後，針對無法直接比對字典詞彙和語料庫詞彙找出其詞頻高低的詞彙，可檢查字典詞彙是否存在於語料庫詞彙之近義詞彙集合，若存在，代表可間接找出和該語料庫詞彙對應的詞頻，最後，建立字典詞彙對應語料庫詞彙之詞頻關係。

我們的中文語料庫來源參照國立科學教育館的期刊語料[9]（40 卷第 7 期至 46 卷第 2 期共 111 篇）以及師大科教中心教學教育月刊語料[10]（240 到 285 期共 218 篇），共 329 篇。首先，將上述語料經過中研院 CKIP 斷詞系統輔助，獲得約 73 萬 4 千個中文詞彙，接著統計重覆出現的中文詞彙次數，產生語料庫詞彙詞頻表，共有 23422 個中文詞彙，平均每篇詞彙個數約 2800 字，字典來源為牛津現代英漢雙解詞典[11]共 39429 個詞彙。

為了取得語料庫與字典詞彙的近義關係，我們首先將斷詞後的語料庫詞彙，分別輸入至 HowNet[8]近義詞尋找方法，及中研院一詞泛讀系統[2]，接著將這兩種方法所找到的近義詞集合，取交集篩選，最後輸出為經過篩選過後之近義詞詞彙集合。

經過上述步驟，可建立字典和語料庫相似字對應檔和透過原先經過斷詞統計詞頻之語料庫詞彙詞頻表，建立字典和語料庫相似字對應檔和語料庫詞彙詞頻表之目的在於建

立字典詞彙對應語料庫詞彙之詞頻關係。建立流程如下。

- i. 直接比對所有字典詞彙與語料庫詞彙詞頻表所記錄的語料庫詞彙，若相同，則記錄目前字典詞彙及所對應之詞頻數，若無，則記錄目前字典詞彙及其詞頻數為 0。
- ii. 針對目前字典詞彙及其詞頻數為 0 的情況，利用字典和語料庫相似字對應檔，檢查字典詞彙是否出現於近義詞集合，若有，代表目前字典詞彙可間接對應近義詞集合所指向之語料庫詞彙，並找出其對應詞頻數，若無，則記錄目前字典詞彙及其詞頻數為 0。

針對上述流程 2 部分，我們以圖四為例，假設語料庫內有“發揮”一詞，查詢相似字對應檔發現其近義詞集合為“表達，達”，若英漢字典不存在“發揮”中文詞，而有“表達”或“達”之詞彙，則根據近義詞對應關係，可以間接指向語料庫詞彙詞頻表之“發揮”所對應之詞頻（24 次）。因此，透過查找相似字的方式，找出所有相似字的詞頻，建立屬於字典詞彙的詞頻統計表，可做為判斷中文詞彙重要程度的選詞參數。

解析幾何法=[解析幾何法] 到處=[到處，處處，在在，四面八方，四方，萬方] 火=[火] 品味=[嘗試，品味] 測試=[試驗，實驗，試，嘗試，測試，測試，考驗，試用] 蜂鳴器=[蜂鳴器] 沒有=[不及，沒有，小，低，差，不如，亞] 交流電力=[交流電力] 發揮=[表達，發揮，達]	解析幾何法 : 2 到處 : 1 火 : 15 品味 : 1 測試 : 25 蜂鳴器 : 7 沒有 : 223 交流電力 : 1 發揮 : 24
--	--

圖四、字典和語料庫相似字對應檔(左)及語料庫詞彙詞頻表(右)

(三) 中英詞彙對列和語言模型

由於語料庫統計詞頻和使用者選詞次數的方式調整選詞順序，並未考慮前後文之間的搭配關係，因此我們利用 word alignment 的技術，以科學人雜誌之中英對照雙語平行語料庫為來源，以 GIZA++[6]為分析工具，找出雙語詞組或詞彙的對照關係及其機率值，並透過以下機率模型計算以下的條件機率。

$$\Pr(c_i | c_{i-1}, e_i) \cong \Pr(c_i | e_i) * \Pr(c_i | c_{i-1}) \quad (1)$$

公式(1)中定義 c 為中文翻譯詞彙， e 為英文詞彙，利用前一個中文翻譯選詞的結果，即 $\Pr(c_i | c_{i-1})$ ，找出和目前中文翻譯詞 c_i 共現的機率，以及中英詞彙對列機率，找出兩者相乘之最大機率分數，以近似 $\Pr(c_i | c_{i-1}, e_i)$ 的數值，做為選擇中文翻譯詞 c_i 的可能值。

不過，根據公式(1)所計算出來的機率分數，僅決定 bi-gram 的翻譯機率，若需要計算整句中文翻譯的機率分數，即考慮中文句子長度為 N 時，相乘所有 bi-gram 的翻譯機率，結果如下。

$$\prod_{i=1}^N \Pr(c_i | c_{i-1}, e_i) \cong \prod_{i=1}^N \Pr(c_i | e_i) * \Pr(c_i | c_{i-1}) \quad (2)$$

綜合語料庫統計詞頻資訊、使用者選詞次數，以及利用雙語語料統計的機率值，可以將這些衡量詞彙選詞重要程度的方法當做參數，定義字典原文詞彙 e 的所有中文翻譯選項為 w_1, w_2, \dots, w_i ，以及定義語料庫統計詞頻資訊為 $W_f(w_i)$ 、使用者選詞次數 $G_f(w_i)$ 以及根據公式(2)利用雙語語料統計的機率值 $P(w_i)$ ，作為原文詞彙 e 之中某一翻譯選項

w_i 輸入至三種方法的函數名稱，其相加的分數定義 $S(w_i)$ 為選詞計分函數的計算結果，如下表示。

$$S(w_i) = \alpha W_f(w_i) + \beta G_f(w_i) + \gamma P(w_i) \quad (3)$$

其中， $S(w_i)$ 所定義的 α 、 β 、 γ 為各別決定 $W_f(w_i)$ 、 $G_f(w_i)$ 和 $P(w_i)$ 計算之權重，由於目前研究仍無法決定 α 、 β 、 γ 的最佳值，因此設定 $\alpha=\beta=\gamma=1$ ，代表三者的重要性相同。

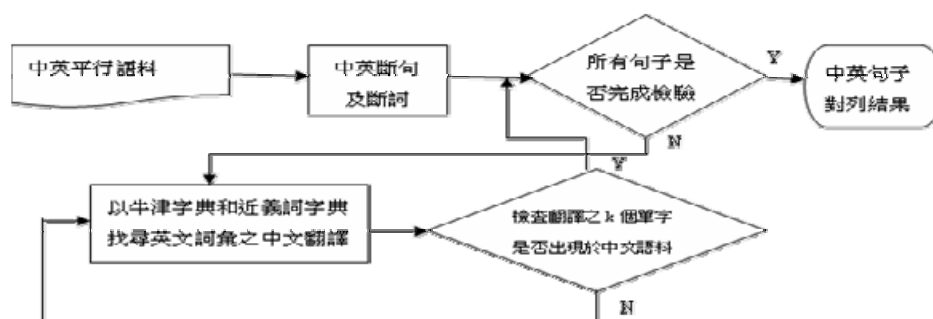
接下來，我們將利用 GIZA++[12] 及 mkcls[11] 等工具，介紹如何將利用科學人雜誌提供的中英雙語語料，經過簡單的中英語句對列 (sentence alignment) 技術，接著將中文語料部分，利用中研院 CKIP 斷詞系統加以斷詞，英文語料部分，則是經過英大小寫轉換及利用字和字之間空白斷詞，再來利用 word clustering 技術，建立詞彙分群，並加以編號，將文字語料轉換成純文字資訊，最後輸入至 GIZA++ 產生詞彙對列結果，並評估其實驗效果。

我們抽樣科學人雜誌 2002 年 3 月至 2006 年 12 月共 110 篇文章，句對數、中英詞彙數及中英總詞彙個數如表二所示。

表二、實驗語料來源統計

句對數	詞彙數	總詞彙個數(tokens)
中文	9313 字	301747 個
英文	10608 字	390020 個
2780 句		

首先，由於科學人雜誌文章內容格式為 html 網頁格式，加上其中英對照網頁的內容僅利用分欄的方式，將中英內容各別置入左右兩欄內，而未標註其中英句子相互對照的位置，以致於不能符合訓練資料需滿足句子對列 (sentence alignment) 的條件。然而，以人工方式針對中英語句判斷是否句子對列，需要大量的時間和精神，故我們依照圖五之簡易句子對列流程，找出可能的中英句子對列結果。



圖五 簡易句子對列流程

GIZA++ 計算 alignment 的機率，為目標語言 t 在長度 m 以及來源語言 s 在長度 n 條件下語彙對列的條件機率，即計算 $\Pr(t_1^m | s_1^n) = \prod \Pr(t_j | s_{a_j})$ 。然而，實際檢查語料庫，根據語料庫提供的統計資料進行公式計算時，會因為語料庫所提供的部分資訊不足，使得條件機率的值為 0 的情形，即所謂的稀疏資料 (sparse data) 問題。因此，為了能提昇 alignment 的品質，增加日後計算詞彙翻譯機率的正確性，Och 等人在 1999 年於 “An Efficient Method for Determining Bilingual Word Classes”[5] 提出一種 word clustering 方法，將分佈稀疏的語料資訊進行分配 (partition)，使得經常同時出現的單一語言詞彙 (monolingual word) 或雙語詞彙 (bilingual word) 能在同一類，即使來源詞彙和目標詞彙的對照資訊不足，也可改以檢查來源詞彙所屬的類別和目標詞彙的對照資訊，來增加其機率值。同時，Och 也提出了一套最佳化機制，在調整目標詞彙類別 T^* (target word class) 或來源詞彙類別 S^* (source word class) 時，檢查其詞彙翻譯條件機率值是否符合

$\arg \max \Pr(s_1^n | S^*) \Pr(t_1^m | s_1^n, S^*, T^*)$ 條件，確保 clustering 結果之正確性。

^{S*,T*}以 word clustering 概念為基礎，我們利用 Och 所開發 mkcls 工具[11]，設定需要執行最佳化 (optimize) 的次數執行 5 次及 word class 個數設定為 80 個，之後輸入中英語料給 GIZA++ 執行，透過 EM algorithm、HMM model 及 IBM-model 1~5 之訓練過程，反覆訓練 5 次，輸出中英詞彙對列結果及中英詞彙對照翻譯機率表。

在 2780 句中英句對中，我們發現在中英詞彙對照翻譯機率表裡，針對中文 9313 個詞彙對應至英文 10608 個詞彙共出現 18181 種不同詞彙對列組合。接著，我們將中英詞彙對照翻譯機率表裡出現過 18181 個英文詞彙，扣除相同英文詞彙對應多個中文詞彙的次數，可得到 5224 個不同英文詞彙的接受配對個數，和原先 10608 個英文詞彙比較起來，發現 5384 個為 GIZA++ 判定為英文詞彙對應至 NULL，即語料中無任何合適中文詞彙可供配對。若扣除 5384 個未出現於中英詞彙對照翻譯機率表中的詞彙，接著，以人工檢查的方式，檢查中英詞彙對照翻譯機率表內正確配對的中英詞對共有 3372 筆記錄，扣除掉相同英文詞彙對應多個中文詞彙的情況，尚有 2761 個不同英文詞彙是配對正確的中文詞彙解釋。

接著，根據公式(1)，計算中文字的組合影響到最後中文詞彙選擇的機率，因此需考慮中英詞彙對列機率及中文雙連語言模型，即 $\Pr(c_i | c_{i-1})$ 。

其中，我們建立中文雙連語言模型的方式為利用 SRI Speech Technology and Research Laboratory 所開發的自然語言工具 SRILM (網址為 <http://www.speech.sri.com/projects/srilm/>)，輸入中文語料檔後，根據統計各中文詞彙及下一個中文詞彙的出現次數，計算其出現機率，最後輸出中文語言模型機率表，如圖六所示。

-2.150164	運動 消耗
-2.139987	運動 能力
-2.153002	運動 健將
-2.146435	運動 模式
-2.142739	運動 模型
-2.153002	運動 模擬儀
-2.153002	運動 器官
-2.144583	運動 聯盟
-1.341907	運算 ，

圖六、中文語言模型機率表

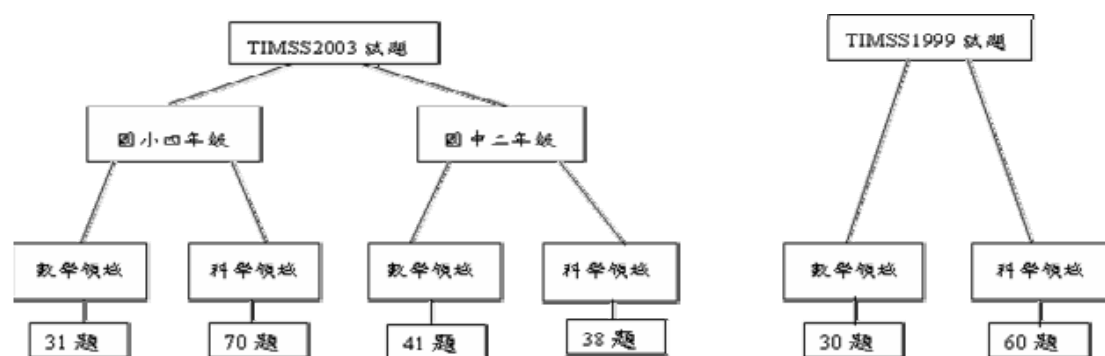
由圖三可觀察到，第一個欄位和第二個欄位分別表示為各 bi-gram 詞彙下對應的 log probability (以 \log_{10} 為基底)。因此，我們以 “sports leagues” 為例，依照公式(1)改寫的公式，計算當目前英文為 “sports leagues” 時，我們確定 “sport” 翻譯為 “運動”，而 “leagues” 翻譯成 “聯盟” 的機率為。

$$\Pr(\text{聯盟} | \text{運動}, \text{leagues}) \cong \Pr(\text{聯盟} | \text{leagues}) * \Pr(\text{聯盟} | \text{運動})$$

其中， $\Pr(\text{聯盟} | \text{leagues})$ 可經由中英詞彙對照翻譯機率表查出 leagues 機率最大的中文翻譯， $\Pr(\text{聯盟} | \text{運動})$ 則是經由圖一中文語言模型機率表，找出中文 bi-gram 共現的條件機率值。

五、實驗和評估

我們實驗主要的資料來源為 TIMSS1999 及 TIMSS2003 的試題，在試題分類上，依照 TIMSS 師大科教中心網頁，說明在 1999 年所出題之試題內容，主要區分數學及科學類別，並且以國中二年級為考試對象。而在 2003 年所出題的題型內容上，除了依照 1999 年的試題類別外，在對象上則是以國小四年級及國中二年級為考試對象。因此，我們按照上述之試題分類，將網頁所提供的中英試題內容，下載並轉檔為純文字檔，接著將試題分類及題目個數統計如下。



圖七、TIMSS 試題分類及題數統計樹狀圖

從圖七來看，由於試題依照數學及科學領域的範圍內容，以及不同年級分別，在難易程度上會比照年級有所不同。例如國小四年級的數學會著重於數的認識、單位比較或基本四則運算，而國中二年級則除了前述之基本觀念外，還增加代數、分數四則運算等概念。在科學領域部分也有類似的學習進度分配，例如從國小四年級針對地科及生物等觀念，到國中二年級則進入理化等觀念的配合。另外，在專有名詞的使用上，以及題目內陳述的內容長短，國中二年級的內容較國小四年級的用詞上較為艱深。

除了衡量以試題內容為翻譯對象，測試線上翻譯系統及本系統在數學及科學領域翻譯程度的好壞，同時也應該考慮不同年級的試題內容，驗證翻譯系統在翻譯能力上，是否按照愈低年級其翻譯正確愈高的趨勢。故實驗的比較上按照年級別、試題種類分類(以數學領域為 A 代號，科學領域為 B 代號)及年分可大致區分如表三之組別表，以實驗國中二年級試題和國小四年級試題對於翻譯效果的影響程度。

表三、TIMSS 試題實驗組別表

中二 A 組	中二 B 組	小四 A 組	小四 B 組	中二 AB 組	小四 AB 組
TIMSS1999 及 TIMSS2003 國中數學領域 試題	TIMSS1999 及 TIMSS2003 國中科學領域 試題	TIMSS2003 國小數學領域 試題	TIMSS2003 國小科學領域 試題	TIMSS1999 TIMSS2003 數學及科學領域 試題	TIMSS2003 數學及科學領域 試題

我們實驗主要比較的翻譯系統為谷歌(以下稱 Google Translate)和雅虎線上翻譯系統(以下稱 Yahoo!)，以及本系統。其中本系統根據公式(3) $S(w_i)$ 之選詞計分公式調整，計有利用語料庫統計詞頻資訊為 $W_f(w_i)$ 、使用者選詞次數 $G_f(w_i)$ 以及根據公式(2) 利用雙語語料統計的機率值 $P(w_i)$ 等選詞策略。我們的實驗目的，在於評估本系統在不同選詞策略之下，經由和線上翻譯系統比較的結果，找出各別選詞策略，或不同選詞策略的搭配之下，其翻譯水準的差異性，並藉由各選詞策略實驗出來的結果，以排名的方式，試圖找出對本系統最有利的選詞策略組合。

有鑒於此，我們考慮以下不同選詞策略的組合，系統依照這些組合共可區分為。

- i. 隨機選詞：即系統不考慮任何選詞策略，以任意選詞的方式，輸出其選詞結果，

之後，藉由此方式評估並實證系統在最差情況的翻譯水準。

- ii. 考慮翻譯修正階段選詞 $W_f(w_i)$ 模式：即考慮翻譯者選詞次數，來決定選詞順序。由於翻譯者選詞次數需要經由翻譯者長時間使用系統，才可進行實測，故在本研究中，無法針對 $W_f(w_i)$ 部分進行實證。
- iii. 考慮中文語料庫詞頻選詞 $G_f(w_i)$ ：即透過收集來的語料庫詞頻高低，決定選詞順序。
- iv. 考慮中英詞彙對列機率及語言模型 $P(w_i)$ ：即利用 GIZA++ 及 SRILM 工具完成的中英詞彙對照翻譯機率表和中文語言模型機率表等資訊，經由公式(2)計算最佳選詞路徑。
- v. 同時考慮 $G_f(w_i) + P(w_i)$ 模式：在無法得知 $W_f(w_i)$ 資訊之情況下，計算 $S(w_i) = G_f(w_i) + P(w_i)$ 的選詞計分結果。

在評估翻譯效果的標準上，我們採用目前以 n-gram 匹確正確率為基礎的 BLEU 和 NIST 評分策略為標準，其評估方式如下。

i. BLEU 部分：

BP: 長度懲罰因子
 L_{ref} : 參考翻譯句子長度
 L_{sys} : 系統翻譯句子長度
 $\delta_n(e_s, r_s)$ 系統翻譯句子 e_s 與參考翻譯句子 r_s 匹配的 n-gram 個數
 $c_n(e_s, r_s)$ 系統翻譯句子 e_s 中 n-gram 的個數

$$\text{score} = \text{BP} * \exp\left(\sum_{n=1}^N W_n \log p_n\right) \quad (4)$$

$$\text{BP} = \min\left\{1, \exp\left(1 - \frac{L_{ref}}{L_{sys}}\right)\right\} \quad (5)$$

$$p_n = \frac{\sum_{s=1}^S \delta_n(e_s, r_s)}{\sum_{s=1}^S c_n(e_s, r_s)} \quad (6)$$

ii. NIST 部分：

$$\text{Info}(w_1 \dots w_n) = \log_2\left(\frac{\text{number of occurrences of } w_1 \dots w_{n-1}}{\text{number of occurrences of } w_1 \dots w_n}\right) \quad (7)$$

$$\text{score} = \sum_{n=1}^N \left\{ \sum_{\substack{\text{all } w_1 \dots w_n \\ \text{that co-occur}}} \text{Info}(w_1 \dots w_n) / \sum_{\substack{\text{all } w_1 \dots w_n \\ \text{in sys output}}} \max(c_n(e_s, r_s), 1) \right\} * \exp\left\{\beta \log^2\left[\min\left(\frac{L_{sys}}{L_{ref}}, 1\right)\right]\right\} \quad (8)$$

首先，我們針對系統依上述隨機選詞、 $G_f(w_i)$ 、 $P(w_i)$ 以及 $G_f(w_i) + P(w_i)$ 等選詞策略下，依照 cumulative 4-gram scoring 之 BLEU 和 NIST[7] 評分標準，將表三 6 類的英文原文試題，連同師大科教中心所提供的 6 類原文試題的系統標準翻譯，以及本系統根據 6 類英文原文翻譯的系統建議翻譯，輸入至美國國家標準與技術局 (NIST) 所開發的 mteval-v10 之 BLEU 和 NIST 評分工具(網址為 <http://www.nist.gov/speech/tests/mt/scoring/index.htm>)，再將計分結果輸出至檔案，如表四所示。

表四、本系統各選詞策略之 NIST 及 BLEU 值比較表

組別	中二 A 組		中二 B 組		小四 A 組		小四 B 組		中二 AB 組		小四 AB 組	
	NIST	BLEU	NIST	BLEU	NIST	BLEU	NIST	BLEU	NIST	BLEU	NIST	BLEU
隨機選詞	1.6573	0.0201	1.2879	0.0000	0.8693	0.0000	0.6351	0.0096	1.4902	0.0000	0.7191	0.0093
$G_f(w_i)$	4.6169	0.2130	3.8632	0.1505	3.4706	0.2099	4.5275	0.2229	4.4453	0.1866	4.6349	0.2202
$P(w_i)$	2.6173	0.0662	2.2518	0.0613	2.0501	0.0848	0.9885	0.0223	2.4986	0.1330	1.5621	0.0501
$G_f(w_i) + P(w_i)$	3.0430	0.0786	2.1371	0.0962	3.3754	0.1991	1.0977	0.0290	4.4453	0.1866	2.7722	0.1317

從表四針對各實驗組以 NIST 和 BLEU 評測的結果，可以發現隨機選詞和其他系統比較上明顯要低，尤其在中二 B 組、小四 A 組及中二 AB 組上的 BLEU 分數出現零分的情形，代表系統在無任何選詞方法調整之下，其翻譯效率是相當低的。在 $G_f(w_i)$ 語料庫選詞部分，和隨機選詞、 $P(w_i)$ 和 $G_f(w_i) + P(w_i)$ 的系統相比，還是明顯要優於其他選詞策略。

我們接著以目前本系統評測最佳的 $G_f(w_i)$ 選詞策略之 BLEU 和 NIST 分數，和 Google Translate 和 Yahoo! 線上翻譯系統加以比較，其結果如表五所示。

表五、本系統與線上翻譯系統之 NIST 及 BLEU 值比較表

組別	中二 A 組		中二 B 組		小四 A 組		小四 B 組		中二 AB 組		小四 AB 組	
	NIST	BLEU	NIST	BLEU	NIST	BLEU	NIST	BLEU	NIST	BLEU	NIST	BLEU
Google Translate	2.8204	0.0964	2.2163	0.0677	2.1460	0.0993	1.1145	0.0171	2.5948	0.0834	1.6255	0.0523
Yahoo!	3.3721	0.1051	2.7817	0.0842	2.1312	0.1111	1.3517	0.0332	3.1779	0.0962	1.8199	0.0705
$G_f(w_i)$	4.6169	0.2130	3.8632	0.1505	3.4706	0.2099	4.5275	0.2229	4.4453	0.1866	4.6349	0.2202

從表五可以觀察得到，系統在 $G_f(w_i)$ 選詞策略之下，各組的 BLEU 值和 NIST 值皆比 Google Translate 和 Yahoo! 線上翻譯系統要高，顯示建立一致性翻譯的詞彙或詞組，以及科學與數學相關之中文語料詞頻的確能提高選詞的正確性。也同時反應 Yahoo! 和 Google Translate 翻譯系統設計的目的是將廣泛領域知識 (domain general) 的翻譯內容，以語料統計中最可能的選詞方式，採用直接翻譯或 phrase-based 翻譯將內容輸出。因此，選詞策略上較無法以限制特定領域知識 (domain specific) 範圍的做法，即加入更多數學及科學領域的片語句型或專有名詞的規則轉換條件，使得反應在系統建議翻譯的翻譯詞彙與 TIMSS 試題的參考翻譯的結果有所差距，由於參考的翻譯僅有師大科教中心針對 TIMSS1999 及 2003 年版本的翻譯內容，使得 Google Translate 或 Yahoo! 爲了增加翻譯結果流暢度所加入之詞彙，可能因爲缺乏其他參考對象，而影響最後的翻譯的品質或正確性。

比較有趣的是，從表四和表五可發現考慮中英詞彙對列機率及語言模型 $P(w_i)$ 在分數上十分接近 Google Translate 的表現，並且在 $G_f(w_i) + P(w_i)$ 的選詞分數上，亦與 $G_f(w_i)$ 的分數相差不多。不過從這兩點的比較上，皆發現加入 $P(w_i)$ 的結果並不能使 BLEU 與 NIST 的分數提升，反而在分數上較 Google Translate 以及 $G_f(w_i)$ 略低。由於 $P(w_i)$ 以科學人雜誌爲雙語語料統計之機率選詞模型，在語料類型上是接近 TIMSS 試題的內容，屬於科學與數學範疇，因此理論上在 BLEU 和 NIST 分數上會較同爲統計式翻譯爲基礎，但語料來源不完全針對科學領域統計的 Google Translate 要好，但仍比僅考慮個別中文翻譯選項之詞頻高低的 $G_f(w_i)$ 之分數來得低，主要原因可歸納幾點。

- i. 統計語料在數量上的不足：由於以 GIZA++ 詞彙對列工具所找出之正確中英詞彙對列個數僅有 2761 個，因此在預測適合之中文翻譯選項時明顯無法找出牛津字典 39429 個英文詞彙中所有中文翻譯選項的可能對列機率，因此可能會有選詞上的誤差。
- ii. 系統產生的中文翻譯句子未經任何詞序上的調整：由於系統翻譯的模式是按照英文詞彙原有的詞彙順序，將單字與片語詞組切分後查詢牛津字典之中文翻譯結果，因此中文詞序上和一般中文句子不同，而 $P(w_i)$ 在語言模型的機率統計上，是依照一般中文詞序統計，因此實際將語言模型之統計結果，套用至本系統時，會因爲許多中文 bi-gram 在 $P(w_i)$ 的語言模型未出現過，使得機率值爲 0 的情形，雖然在做法上，我們利用乘上極小數（我們預設以 10^{-20} ）的方式，避免因機率值爲 0，影響中英詞

彙對列機率的結果，但系統本身的中文詞序，仍然會造成 $P(w_i)$ 對照語言模型的機率計算上出現不小的誤差。

針對 2 的推論，我們再度針對 $P(w_i)$ 公式的結構，區分新的實驗組別，進一步探討 $P(w_i)$ 公式各部分對於整體 BLEU 和 NIST 值的變化，並驗證 2 推論的正確性。由前述之公式(2)可得知 $P(w_i)$ 為

$$\prod_{i=1}^N \Pr(c_i | c_{i-1}, e_i) \cong \prod_{i=1}^N \Pr(c_i | e_i) * \Pr(c_i | c_{i-1})$$

從公式(2)右半部可得知 $\Pr(c_i | e_i)$ 為中英詞彙對列機率模型， $\Pr(c_i | c_{i-1})$ 為中文雙連語言模型，而表四完成的 $P(w_i)$ 是經由公式(2)計算的選詞結果，因此我們分別比較單獨利用中英詞彙對列機率模型（以 **Align** 取名）和中文雙連語言模型（以 **Bigram** 取名），和 $P(w_i)$ 結果上的差異性，如表六所示。

表六、 $P(w_i)$ 與各別機率模型之 NIST 及 BLEU 值比較表

組別	中二 A 組		中二 B 組		小四 A 組		小四 B 組		中二 AB 組		小四 AB 組	
	NIST	BLEU	NIST	BLEU	NIST	BLEU	NIST	BLEU	NIST	BLEU	NIST	BLEU
P(w_i)	2.6173	0.0662	2.2518	0.0613	2.0501	0.0848	0.9885	0.0223	2.4986	0.1330	1.5621	0.0501
Align	2.6940	0.0686	2.6165	0.0926	3.3518	0.1923	1.0719	0.0242	3.1006	0.1142	2.7707	0.1239
Bigram	2.3934	0.0588	2.4677	0.0895	2.4126	0.0988	0.9760	0.0229	2.8480	0.0957	2.2474	0.0824

從表六各組別 NIST 和 BLEU 值可觀察出，由於 $P(w_i)$ 的分數是按公式(2)由 **Align** 和 **Bigram** 的機率分數相乘而得，受到 **Bigram** 分數較低的影響，和 **Align** 計算的機率分數相乘會使得最後 $P(w_i)$ 的機率分數更低，因此其 NIST 和 BLEU 分數也相形較低。並且，從 **Align** 的大部分分數（除中二 AB 組）高於 $P(w_i)$ 的情形來看，可說明 **Align** 和 **Bigram** 相結合的 $P(w_i)$ 之後其 NIST 和 BLEU 分數的確會有減分的效果，亦證實 2 的推論為正確。

六、結論

在本研究中，我們為了減少師大科教中心以人工方式進行 TIMSS 試題翻譯、校稿及討論的時間，我們嘗試實作一輔助試題翻譯系統，在評估翻譯效率實驗過程當中，我們透過 BLEU 及 NIST 為評估指標，進行 TIMSS 試題翻譯的結果，發現即使在某些選詞為任意選詞策略之下，系統詞彙翻譯之正確度上能亦能有接近線上翻譯之成果。

針對翻譯結果的流暢度上，雖然實作出利用使用者介面，進行新增詞彙及刪除詞彙等動作，來達到翻譯結果符合語法之描述，但未處理有關自動加減詞彙之機制，因此，可透過統計式翻譯模型的輔助，找出翻譯詞彙中最大相鄰之可能詞彙，或者找出某些詞彙因相鄰關係而使翻譯詞彙產生減少詞彙之關係，來增進並改善翻譯的流暢程度。

在翻譯詞序的調整部分，可利用語法剖析的方式，利用中英剖析樹的對應，找出類似的結構對應關係，在判斷新的翻譯句子時，利用其對應關係的相似程度，找出最大可能之詞序組合，其中，需要克服的條件，在於語法剖析樹在結構的種類上，視描述方法不同可能會有許多不同的組合，因此需以分類及相似度比對的方式，先找出最大相似之子樹結構，之後在判斷未知句子結構時，才能有效減少比對相似之維度，同時，在中英詞性種類的對應上，由於詞性判斷的標準不一，例如“美麗 (Beautiful)”一詞，英文詞彙為形容詞，而在中文詞性判斷上，不同語境之條件下可能其中文詞彙可能會判讀成

副詞或形容詞，因此，要先能解決詞性判別之歧義問題，才可進行上述分類及相似度比對之方法，在這同時，翻譯系統執行時如何有效率執行結構的比對，也是進一步思考的範圍之一。

參考文獻

- [1] Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin, A Statistical Approach to Machine Translation, *Computational Linguistics*, 16:2, 79-85, 1990.
- [2] Cheng, Chin-Chuan, Word-focused Extensive Reading with Guidance, *Selected Papers from the Thirteenth International Symposium on English Teaching*, 24-32, 2004.
- [3] Fons Van de Vijver and Ronald K. Hambleton, Translating Tests: Some Practical Guidelines, *European Psychologist*, 1, 89-99, 1996.
- [4] Ide Nancy, Véronis Jane and Word Sense Disambiguation: The State of the Art, *Computational Linguistics*, 24:1, 1-40, 1998.
- [5] Franz Josef Och, An Efficient Method for Determining Bilingual Word Classes, *Proceedings of European Chapter of the Association for Computational Linguistics*, 71-76, 1999.
- [6] Franz Josef Och, Hermann Ney, Improved Statistical Alignment Models, *Proceedings of the Thirty-eighth Annual Meeting of the Association for Computational Linguistics*, 440-447, 2000.
- [7] Papineni, K., Roukos, S., Ward, T., and Zhu, W. J., BLEU: a method for automatic evaluation of machine translation, *Fourtyth Annual meeting of the Association for Computational Linguistics*, 311-318, 2002.
- [8] Yang Xiaofeng and Li Tangqiu, A Study of Semantic Disambiguation Based on HowNet, *International Journal of Computational Linguistics & Chinese Language Processing*, 7:1, 47-78, 2002.
- [9] 國立科學教育館科學研習月刊，<http://www.ntsec.gov.tw/publish/pdf.asp>
- [10] 師大科教中心科學教育月刊，[http://140.122.147.172/journal/\(new\)journal.htm](http://140.122.147.172/journal/(new)journal.htm)
- [11] Concise Oxford English Dictionary，http://stardict.sourceforge.net/Dictionaries_zh_TW.php
- [12] Google Translate, http://www.google.com/translate_t
- [13] JACOB PROJECT, <http://danadler.com/jacob/>
- [14] MINIPAR HOME PAGE, <http://www.cs.ualberta.ca/~lindek/minipar.htm>
- [15] MXPOST, <http://ccc.kmit.edu.tw/wiki/advanceprogramming/index.htm>
- [16] TIMSS 中文版官方網頁, <http://timss.sec.ntnu.edu.tw/timss2007/news.asp>
- [17] The Porter Stemming Algorithm, <http://www.tartarus.org/martin/PorterStemmer/>
- [18] WordNet API, <http://nlp.stanford.edu/nlp/javadoc/wn/>
- [19] YAHOO! 雅虎線上翻譯, <http://tw.search.yahoo.com/language/>