電腦輔助中學程度漢英翻譯習作環境之建置

賴敏華　　劉昭麟
國立政治大學資訊科學系
{g9523, chaolin}@cs.nccu.edu.tw

摘要

電腦輔助教學系統主要在於幫助教師教學與學生自學。本研究提供能協助學生學習中翻英的優良環境，透過提供適當的參考資料，如相似文法與相似句型等，可增進學生對單字與文法的熟悉度，並提升學生學習英文的興趣及能力。本研究主要針對中學生程度設計，利用自然語言處理技術，藉由輸入中文句或中英文混合句，經文法、詞性及結構樹分析後，能提供相似的中英文句子，給予適當的英文翻譯建議，讓使用者有更多方向的參考。

關鍵詞：電腦輔助語文教學、句型搜尋、例句式教學、電腦輔助翻譯

1. 緒論

英文在現今社會上是非常普遍的語言，是地球村中不同母語背景的人用來相互溝通以及傳達訊息、信念不可或缺的工具。近年來，政府的教育政策以及家長對於小孩的期待，開始學習英文的年齡層下降，國小中高年級已排入英語課程、教育部國民教育司公佈國中常用兩千字字彙以及基礎一千字字彙[20]、各大專院校為學生英文程度把關，紛紛訂定畢業門檻等，由此可看出學習好英文、提升英文能力是未來生活重要的一環。

為了協助學生在課餘時間自學，各式各樣多元的輔助教學系統如雨後春筍出現，電腦輔助教學系統 (computer assisted tutoring system) 可以協助學生在課餘時間自我學習。目前許多學科，如：語言、數學、自然科學等教學，皆可以利用電腦技術與應用軟體設計出生動的互動教學軟體，也可以利用電腦輔助出題系統協助老師出題、批改。

本系統的中英文語料來源為從網路上收集，適合中學生或全民英檢中級與中高級程度的文件，包括教育部委託宜蘭縣建置語文學習領域國中教科書補充資料題庫[19]、旋元佑文法[16]、基礎英文 1200 句[17]、國民中學學習資源網[18] 的評量題庫及資源手冊等，經由人工擷取，建構中英文對照語料庫，以及利用現有的自然語言處理工具，中研院的中文斷詞系統將語料庫中的字彙標記其詞性以及中文句結構樹檢索系統建立中文句的結構樹，來建立標記化語料庫 (tagged corpus) 。這個標記化語料庫即是中英翻譯推薦句的來源。

本篇論文的組織架構如下：第二節為文獻探討；第三節為詳細描述本系統建立標記化語料庫的步驟；第四節則介紹本系統所提供的功能；第五節為系統評估，並於第六節提出簡單的結語。

2. 文獻探討

隨著電腦的普及化，電腦輔助教學系統在學生自我學習與教師教學上逐漸扮演著重要的角色。電腦輔助教學系統不但可以協助教師出題、閱卷、統計分數及對於學生學習效果作審慎的評估；亦可協助學生在沒有老師的陪同下自我學習，透過電腦輔助教學系統的回饋機制，系統可以依據學生所需，提供適合的教材，進而提升自我學習的效果。

有關電腦輔助出題系統，Mitkov和Ha[8]在2003年提出採用自然語言技術來自動產生閱讀測驗的試題。用來產生試題的句子皆具有「主詞+動詞+受詞」或「主詞+動詞」結構，而選擇為考題答案的通常是特定領域的詞彙。利用計算詞頻數以及WordNet[7]查詢詞的定義，來擷取英文句子中的關鍵詞作為考題答案。使用WordNet找尋語意上近似的觀念作關鍵詞的誘答選項，或從語料庫找尋語意不相似，但具有部分相同關鍵詞的片語或複合詞作為誘答選項。

Liu等學者[5]在2005年提出電腦輔助英文字彙出題系統之研究，利用詞性標記、字彙頻率的統計資料以及selectional preferences[6]的技術，配合搭配詞 (collocation) 概念與機器學習，從訓練資料歸納出法則，來輔助產生題目中的誘答選項。此研究只單純的利用搭配詞的訊息並沒有深入到語意層面的分析，故無法確保所產生的試題語義的完整性，仍須經由人為篩選作最後的確認。

陳佳吟等學者[21]在2005年提出電腦輔助英文文法出題系統，FAST (Free Assessment of Structural Tests) ，對於各式文法考題依照題型作分析，將考題分為九大類型，利用Tagger撰寫分類規則，使用Wikipedia網頁上收集有意義的句子為試題的主要來源。產生的題型為傳統四選一的單選文法試題。基於題型不同，會有不同出題方式的誘答選項，即不存在某種適用於全部題型的誘答選項產生方式，所以根據不同題型的誘答選項都是針對其特性去撰寫規則來產生合適的誘答選項。

林仁祥等學者[15]在2007年提出國小國語科測驗卷出題輔助系統，包含四聲辨識、中文克漏詞、改錯字、量詞等試題。其中中文克漏詞部分，使用HowNet[2]找尋相同義原的詞彙以及中研院現代漢語語料庫一詞泛讀[14]的學習工具將近義詞回傳，給予出題教師更多誘答選項的選擇。改錯字部分，提供同音字、相似字兩種功能，同音字利用新酷音輸入法的詞庫檔，給予相同發音的國字視為誘答選項參考；相似字利用倉頡碼建構構字式檔案，提供具有部分相同偏旁的中文字當作誘答選項參考。

除了電腦輔助出題系統外，另有電腦輔助自我學習的系統，其主要目的是在沒有老師的伴隨下可以輔助學生自我學習各個學科（如：語言、數學、科學等）；而語言學習上可分為聽、說、讀、寫四大部分，目前針對「讀」的方面，有較多相關的電腦輔助系

統；在「寫」的方面則較少。對於閱讀的輔助，多為網頁式 (Web-based) 即時的輔助翻譯，提供單字的解釋，進而為整個網頁的翻譯。

Weir和Lepouras[10]在2001年提出一個Web-based的自動註解英語的資訊為希臘文及中文。由於語言學習者對於新的單字或是不熟悉的字會有文字誤用、不同的定義、聯想錯誤或是對於上下文有所誤解，甚至於語言學習者並不熟悉語言上的特殊用法（如：專業術語、俚語或是慣用語）而導致閱讀上的困難。論文中將相關的對應文字在離線狀態 (off-line) 先行建立，再透過動態的連結來查詢英文單字的希臘文或中文的定義及解釋。由於希臘文與英文皆由羅馬字母構成，故中文翻譯比希臘文翻譯要來的複雜、效果也沒有希臘文來得好。

關於寫作方面，在第二語言的電腦輔助自我學習系統，其他語言也有相關研究。Kakegawa等學者[3]在2000年提出電腦輔助學習第二語言（日語），日文字主要可區分成“用言 (Yougen) ”以及“体言(Taigen) ”兩種，“用言”為有語尾變化的詞性，如：動詞、形容詞；“体言”則是沒有語尾變化，如：名詞、代名詞、指示詞 (demonstratives) 。論文中提出使用LTAG (Lexicalized Tree Adjoining Grammar) 來分析日文句子，句子結構樹採用bottom-up方式配合堆疊來建立，系統可以提供相似詞意的詞以及針對有語尾變化的日文字作偵錯，對於不恰當的用法可立即更正或是提供簡單的文字描述讓學生自行修正。

Knutsson等學者[4]在2003年提到有關第二語言（瑞典語）寫作學習環境文法檢查技術。面對非母語使用者在撰寫句子常常會有不可預期的文法型態產生，為了能提供學習者組織以及修正文章，需要足夠支持的語料資源。文章中，字詞的形態錯誤可由語言工具挑出；對於句法錯誤的部分僅僅提供錯誤訊息與建議，而非給予制式的解答。

Chang 和 Schallert[1]在 2005 年提到在學生撰寫英文作文時，透過互動式鍊結文法 (link grammar) ，可以作適時的英文文法確認，進而使學生提升書寫技巧；則教師在批改作文時，也可以專心致力於學生所想表達的意念，而不需專注於英文句文法的使用。提供使用者適時的文法確認的前提為，使用者必須對文法句型結構有相當足夠的認知，否則僅僅知道文法錯誤，卻不知道該如何修正，並無法達到良好學習的效果，若系統可以提供修正的句型參考則更能提升學習效果。

劉吉軒等學者[22]在 2007 年提出利用語言資訊檢索的方法，開發了名為 SAW (Sentence Assistance for Writing) 的雛形系統。此系統提供了完全比對與部分比對的檢索功能，為了允許不同語言程度的使用者能彈性的使用此系統，以正規表示法 (regular expression) 的概念為基礎，利用特殊符號來代表部分確定而部分設定範圍的查詢條件。對於查詢選取出的例句，利用多重序列排列 (Multiple Sequence Alignment) 的技術[9]進行查詢條件與選取例句之間相關程度評估與排序。

在語言寫作學習上，輔助系統提供寫作時，可為文句中的文法或是字詞的型態錯誤作偵錯並且給予修正的建議，但初學者可能對文法或句型完全沒概念，無法下筆將句子作適當翻譯或敘述想表達的意思，則偵錯的輔助系統也沒辦法發揮最大的效益。若輔助系統需要用較多符號去輔助來設定範圍查詢，使用者無法使用直覺去查詢、不夠人性化，

所以本研究的目標是設計並且實作出一個中英翻譯寫作輔助系統。學生透過查詢系統可以取得合適的英文翻譯以及類似句參考，藉由多組類似句可以學習到正確的單字用法以及文法知識。本節主要為介紹自然語言處理應用在電腦輔助教學上，但本系統技術上仍可參考 Example-Based Translation 的相關文章。

3. 建立標記化語料庫

本研究希望能提供一個簡單方便的系統介面，讓學習者可以輸入簡單的中英文關鍵字或句子，即可查詢到相關例句；並藉由本系統提供的中英翻譯推薦句，給予學習者在練習中英翻譯時有更多的參考資料，協助學習者在寫作時能達成既定的目標，達到學習的效果。

網路上提供程度適合國、高中生或全民英檢中級、中高級程度的中英文對照句資訊為數不多，大多數以學生學習單的樣式呈現，故採人工方式將中英文對照句從學習單中取出，此時語料庫內已經包含了對應好的中英文句子組合，每一組中的中英文句都是互為翻譯的句子，所以我們並不需要對語料作句子層次以上的處理。原始中英文對應的語料可經由圖 1 所列的流程，將人工擷取的中英文對照句的中英文句子，分別利用中研院的中文斷詞系統[12]及中文句結構樹系統檢索系統[11]進行詞性標記及結構樹的建立，將其結果回傳建立可用於查詢的標記化語料庫，以下為詳細描述各步驟的細節。

斷詞 (segmentation) ：中文與英文不同之處在於英文的詞與詞之間會以空白作為區隔，而中文的詞與詞是沒有空白，所以必須對中文作斷詞，本系統語料在詞性標記及中文結構樹建立時，會作中文句的斷詞，本系統是使用中央研究院所提供的中文斷詞系統[12]，中文斷詞系統會將中文句以常用的詞彙為基礎，將句子中的詞彙區隔出。如圖 2 所示，輸入的中文句為「我們都喜歡蝴蝶」，中研院中文斷詞系統回傳斷詞後的資訊為「我們(Nh)　都(D)　喜歡(VK)　蝴蝶(Na)」，其中 Nh 為代名詞，D 為副詞，VK 為狀態句賓動詞，NA 為普通名詞。

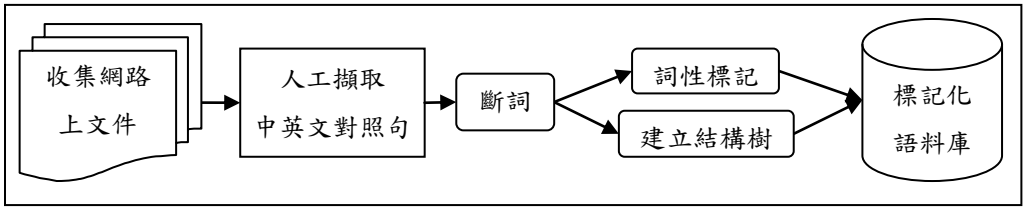


圖 1 建立標記化語料庫的流程圖

輸入：我們都喜歡蝴蝶
輸出：我們(Nh)　都(D)　喜歡(VK)　蝴蝶(Na)
詞性標記擷取： Nh D VK Na

圖 2 中研院中文斷詞系統輸出範例

詞性標記 (Part-Of-Speech tagging) ：將語料標記詞性，可以在搜尋功能中提供依照詞性來作為搜尋依據。因為擁有相同詞性順序的句子，句子相似度會越高，所以可依照詞性來給予使用者中英文類似句的推薦。本系統詞性標記是利用中央研究院所提供的中文斷詞系統[12]所回傳的斷詞結果會伴隨著詞性，將斷詞後的詞性擷取留下。如圖 2 所示，將中文斷詞系統回傳的結果中的詞性部分依序取出儲存，以“我們都喜歡蝴蝶”為例，擷取出的詞性順序為「Nh D VK Na」。

結構樹：透過結構樹可以知道中文句法、語意關係，兩個句子若結構樹的結構相似或是相同，可以猜測為語法相似。由結構樹的結構可以提供語法相似的例句，雖然使用的單字可能相差甚遠，但可以參考相同結構的句子，對於文法的搜尋有很大的幫助。本系統是利用中研院中文句結構樹資料庫檢索系統[11]作為取得結構樹的依據。如圖 3 所示，輸入的中文句為「她不覺得自己幸運」，接收中研院中文句結構樹系統回傳結構樹的資訊，並將結構樹依照分層取出。

我們將樹根定義為第 0 層，樹根的子樹為第 1 層，越往下層數字越大，故葉子節點為一個中文詞。在擷取各分層的結構樹時，樹根層（第 0 層）僅一個節點，不作記錄；依序將每一分層取出，假設某一中文句的結構樹深度為 i，而其一分支子樹最深深度為 j，則在第 j+1, j+2, …, i 層的詞性仍記錄第 j 層的詞性。將中研院中文句結構樹系統回傳結構樹的範例句資訊擷取各分層資料，如圖 3 所示，第一層結構樹為「NP Dc VK1 S」；第二層結構樹為「Nhaa Dc VK1 NP VH11」，第二層結構樹因為中文詞「不」與「覺得」的深度只有 1，所以在第二層結構樹詞性紀錄是記錄深度為 1 時的詞性「Dc」與「VK1」；第三層結構樹，中文詞「她」與「幸運」的深度只有 2，是記錄深度為 2 時的詞性「Nhaa」與「VH11」，中文詞「不」與「覺得」的深度為 1，詞性是記錄深度為 1 時的詞性「Dc」與「VK1」，故第三層結構樹為「Nhaa Dc VK1 Nhab VH11」。依照結構樹的分層取得各

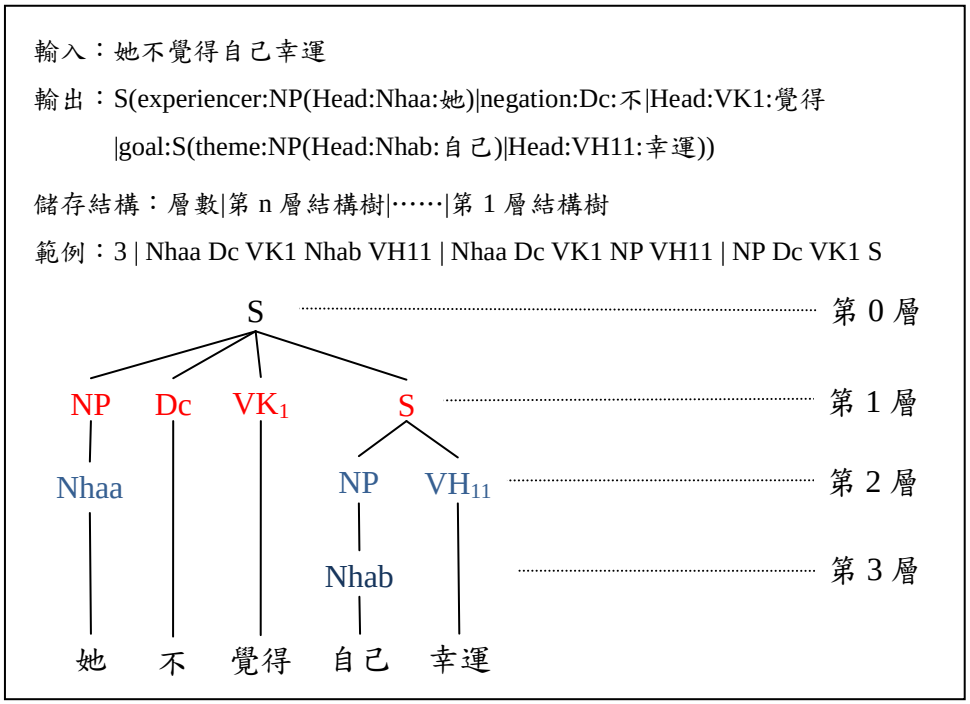


圖 3 中研院中文句結構樹輸出範例

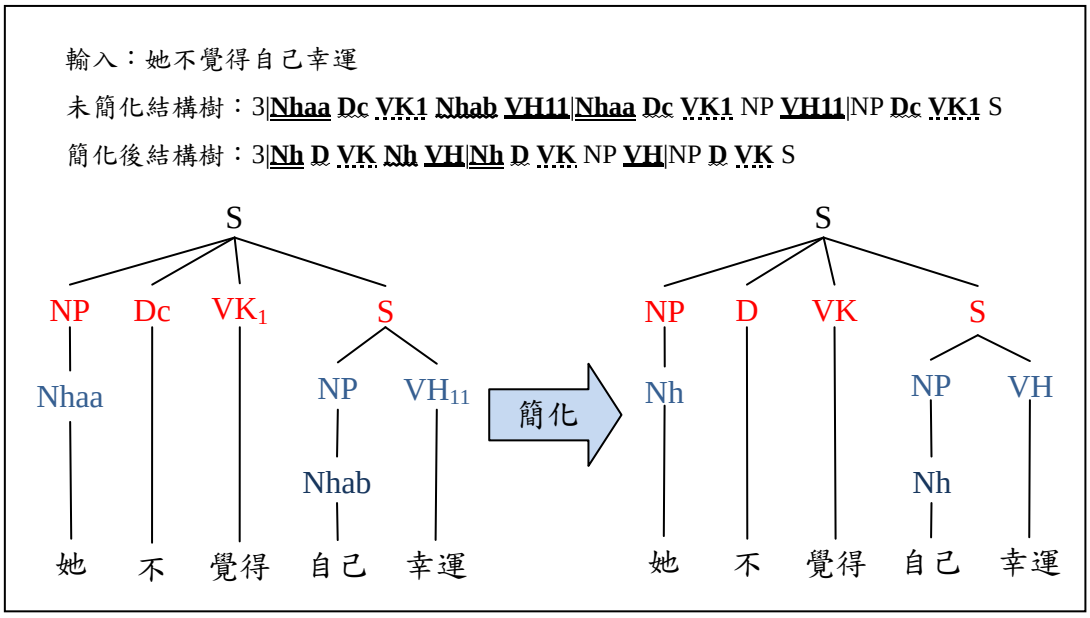


圖 4 結構樹詞類簡化範例

表 1 平衡語料庫詞類標記集中四個簡化詞類對應表

簡化標記	對應的 CKIP 詞類標記	附註
Nh	Nhaa, Nhab, Nhac, Nhb, Nhc	代名詞
D	Dab, Dbaa, Dbab, Dbb, Dbc, Dc, Dd, Dg, Dh, Dj	副詞
VK	VK1,2	狀態句賓動詞
VH	VH11,12,13,14,15,17,VH21	狀態不及物動詞

分層的結構樹，儲存的結構為「層數|第 n 層結構樹|……|第 1 層結構樹」，而最後記錄在標記化語料庫資訊為「3|Nhaa Dc VK1 Nhab VH11|Nhaa Dc VK1 NP VH11|NP Dc VK1 S」。

由於結構樹回傳的詞性分類很細，共計有 115 種詞類，為了提升搜尋時提供更多的例句給予參考，我們根據中研院資訊科學所詞庫小組所編列的中研院平衡語料庫詞類標記集[13]，將標記化語料庫中的例句詞類由 115 種全面簡化成 46 種。以“她不覺得自己幸運”為例，如圖 4 所示，依據表 1 平衡語料庫的標記集（僅附例句中使用的詞類對應表）中得知，可將例句中的四個詞類作簡化，分別為代名詞、副詞、狀態賓動詞及狀態不及物動詞，圖 4 中將此四個詞類作相對應的底線粗體標示，並有簡化前後的樹狀結構表示。

4. 提供搜尋功能

在標記化語料庫建立完成後，可以透過自然語言處理的技術提供一些加值的服務。對於搜尋中文句的類似句，我們的系統利用中文詞、中文詞的詞性以及中文句結構樹的方法來分析處理並提供參考例句，本系統亦提供中英文混合的搜尋。

4.1 以中文詞為搜尋依據

當使用者輸入查詢句後，透過中文斷詞系統取得斷詞後的結果，利用斷詞結果在標記化語料庫搜尋；或是將斷詞的結果透過 HowNet 辭典[2]或是中研院現代漢語語料庫一詞泛讀[14]的學習工具取得與查詢句詞彙中相似的詞，來增加搜尋時中文詞彙的多樣性。

HowNet 辭典[2]是一個以中文和英文的詞所代表的概念為描述對象，以概念與概念之間以及概念所具有的屬性之間的關係作為基本內容的常識知識庫。在 HowNet 辭典中，每個詞彙多個欄位去記錄特殊資訊，如：中文詞條、中文詞性、英文詞條、英文詞性、義原關係等，其中一個欄位名稱為 DEF，描述著此詞彙的義原關係。在 HowNet 中，義原 (sememe) 是描述一個概念最小意義的單位，定義中英雙語知識詞典中的每個詞彙，並且建有描述各個義原之間關係的分類樹。例如：「讀書」一詞是由「從事」、「學」及「教育」三個義原定義而成，所以我們定義 S 為「讀書」義原的集合，S={從事，學，教育}，我們尋找 HowNet 中所有的詞彙，並將詞彙的義原與集合 S 作比對，與集合 S 有交集的詞彙都找出來，「讀書」一詞經過 HowNet 搜尋得到的結果為「攻，攻讀，苦讀，留，留美，念書，旁聽，求學，死記硬背，聽講，同窗，習作，修業，學到，學好，學習，學以致用，專攻，專修，自修，走讀」等，我們的系統會將這些詞彙視為與查詢句詞彙相似的詞，並加入搜尋時詞彙比對的依據。本研究中 HowNet 辭典是採用 1999 年版本。

中研院現代漢語語料庫一詞泛讀的學習工具[14]是利用電腦所收集的文本，針對一個詞彙，閱讀該詞彙出現的許多句子，記錄了各種該詞彙和其他詞彙共同出現的情形，作整理後所得的資料，可讓使用者藉由查詢更能掌握該詞彙的用法。我們的系統會將所要查詢的詞彙自動連到中研院一詞泛讀的網頁，網頁會回傳有關該詞彙的近義詞，本系統會將回傳的近義詞，視為搜尋時詞彙比對的依據。以「讀書」一詞經過一詞泛讀系統回傳所得到的近義詞詞彙為「學習，上，學，讀，念，修，讀書，就讀，念書，上學，入學，求學，攻，攻讀，就學，習，深造，修業，向學」。

如圖 5 所示，使用者輸入查詢句 T 後，經由中文斷詞系統回傳斷詞結果 $t_1, t_2, \ldots, t_n$，

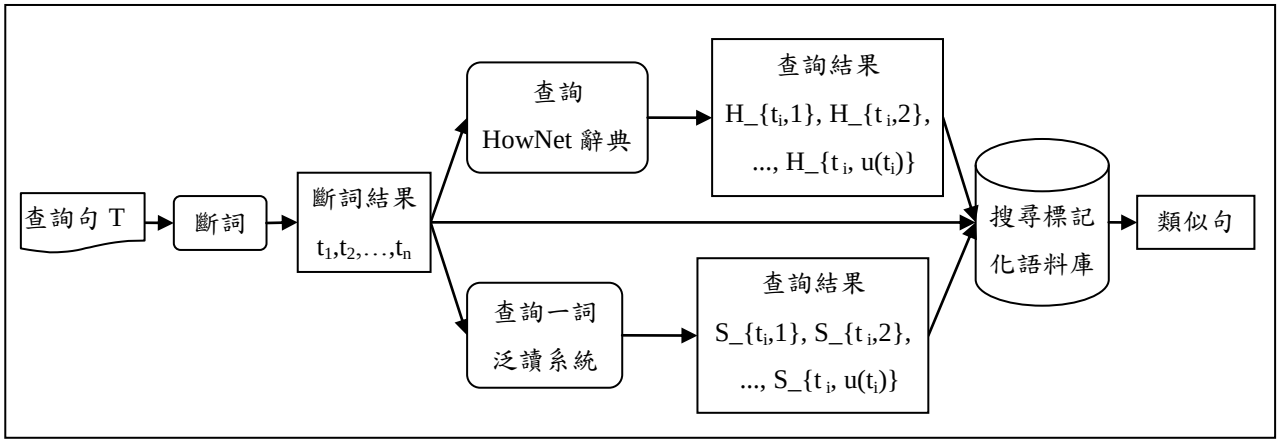


圖 5 使用中文斷詞結果搜尋

輸入：我有時上學遲到

中文斷詞輸出：我(N)　有時(ADV)　上學(Vi)　遲到(Vi)

擷取中文詞部分：我 有時 上學 遲到

利用一詞泛讀查詢近似詞彙：

我 身 個人 人家 本人 予 吾 余 咱 俺 儂 咱家 洒家 有時 學習 上 學 讀 念

修 讀書 就讀 念書 上學 入學 求學 攻 攻讀 就學 習 深造 修業 向學 遲到

輸出：

你上學常常遲到嗎?={Are you often late for school?}

我花了十分鐘走路上學。={It took me ten minutes to walk to school.}

我以前讀國中時，我都走路上學。={When I studied in junior high school, I used to walk to school.}

我每天騎腳踏車上學。={I go to school by bike every morning.}

前天他上學遲到了。={He went to school late the day before yesterday.}

上學讀書以前，他原本是個小頑童。={Before he was in school, he used to be a naughty child.}

我有時上學遲到。={I go to school late sometimes.}

我昨天和同學在圖書館讀書。={I studied with my classmates in the library yesterday.}

圖 6 使用中文詞一詞泛讀搜尋的範例

n 為詞彙個數，將 $t_1, t_2, \ldots, t_n$ 依序查詢 HowNet 辭典，取得查詢結果 H_{t_i,1}[†], H_{t_i,2}, ..., H_{t_i, $u(t_i)$}，其中 $u(t_i)$代表透過 HowNet 辭典所查到的第 i 個中文詞的近義詞的數量；或將 $t_1, t_2, \ldots, t_n$ 依序查詢中研院現代漢語語料庫一詞泛讀的學習工具，取得查詢結果 S_{t_i,1}, S_{t_i,2}, ..., S_{t_i, $u(t_i)$}，其中 $v(t_i)$代表透過一詞泛讀系統所查到的第 i 個中文詞的近義詞的數量。本系統會將查詢到的相似詞結果{H_{t_i,1}, H_{t_i,2}, ..., H_{t_i, $u(t_i)$}}或{S_{t_i,1}, S_{t_i,2}, ..., S_{t_i, $u(t_i)$}}與斷詞結果{$t_1, t_2, \ldots, t_n$ }作聯集，並將聯集結果於標記化語料庫作搜尋，提供類似句供使用者參考。

圖 6 為中文詞使用一詞泛讀搜尋的範例，我們輸入查詢句「我有時上學遲到」，經過中研院中文斷詞系統，得到斷詞後的結果為「我(N)　有時(ADV)　上學(Vi)　遲到(Vi)」，將所斷詞後的結果，擷取其中文詞，並利用一詞泛讀系統查詢得到每一個詞彙的近義詞後，與標記化語料庫中的中文句子作比對，若與語料庫中詞彙符合的句子，則會輸出視為類似句給予使用者作為參考例句。

4.2 以詞性為搜尋依據

在使用者輸入查詢的中文句後，本系統透過中文斷詞系統[12]取得斷詞後的結果，從斷詞的結果中把詞性依序擷取出來，接著透過搜尋標記化語料庫裡中文對照句的詞性，在僅考慮詞性出現順序，若相同即視為類似句。查詢句的詞性長度與標記化語料庫中英對照句所標記的詞性長度可能不相同，所以分為查詢句的詞性長度大於中英對照句詞性長度，以及查詢句的詞性長度小於或等於中英對照句詞性長度，兩種情形。

[†] X_Y 表示 Y 為 X 的下標，此底線為 LaTeX 語法

輸入：那個包包很好看

中文斷詞輸出：那(DET)　個(M)　包包(N)　很(ADV)　好看(Vi)

擷取詞性部分：**DET M N ADV Vi**

利用詞性搜尋：

我們這個週末要去露營了。={We are going camping this weekend.}

={N **DET M N ADV** ADV **Vi** ASP}

這個學生正在鐵路附近溜冰。={The student is roller-skating near the railroad.}

={**DET M N ADV** N N **Vi**}

那隻狗很凶惡。={The dog was mean.}={**DET M N ADV Vi**}

那個包包很好看。={That purse looks pretty.}={**DET M N ADV Vi**}

這個問題好像很容易。={The question seems easy.}={**DET M N ADV** ADV **Vi**}

我覺得這個包包很漂亮。={I find the purse pretty.}={N Vt **DET M N ADV Vi**}

那個島可以搭飛機或搭船去。={You can go to the island by plane or by boat.}

={**DET M N ADV** Vt N C **Vi** ADV}

圖 7 依照詞性搜尋範例

使用者輸入查詢句，經由中文斷詞系統回傳斷詞結果，並從中取得其相對應的詞性，透過與標記化語料庫中所記錄的中英文對照句的詞性標記資訊作比對，可以給予類似句當作參考依據。當查詢句的中文詞詞類個數小於或等於中英文對照句的詞類個數時，查詢句的中文詞詞類都依序出現在對照句的詞類序列裡面，則我們將此中文對照句視為查詢句的類似句，反之則不為查詢句的類似句；當查詢句的中文詞詞類個數大於中英文對照句的詞類個數時，對照句的詞類序列裡都依序出現在中文詞的詞類中，則我們將此中文對照句視為查詢句的類似句，反之則不為查詢句的類似句。如圖 7 所示，輸入的查詢句為「那個包包很好看」，輸出詞性標記順序為「DET M N ADV Vi」；搜尋後所得的部分結果如圖 7 所示，詞性比對成功部分以粗體底線標示之。

4.3 以結構樹為搜尋依據

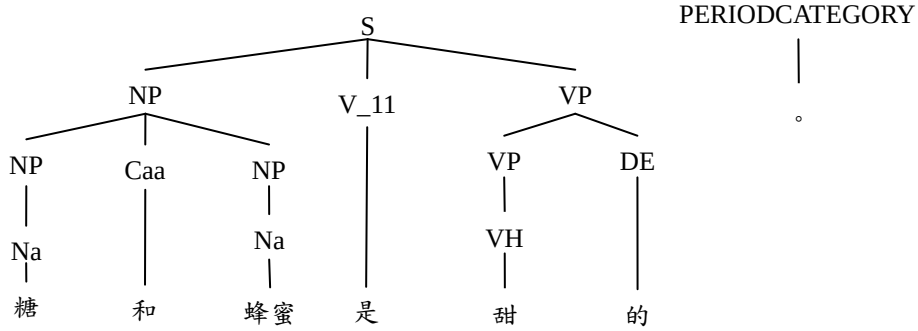
使用者輸入查詢句後，本系統可以透過中文結構樹取得查詢句的結構樹，為了查詢比對時能提供更多類似句供使用者參考，將結構樹中的詞性作簡化。依照結構樹階層分層取出，第 0 層（樹根層）不作記錄及搜尋，針對每一階層的結構樹，在語料庫中作結構樹比對搜尋，並將擁有相同結構樹的句子視為類似句給予使用者作參考。為了保有結構樹的原始結構，此功能只會針對每一階層去作完全比對以提供參考的類似句。若某一查詢句，經由中研院中文結構樹、詞性化簡以及擷取各分層的結構樹後，將各分層的結構樹與標記化語料庫中所記錄的結構樹作比對，其分層結構樹與標記化語料庫中的中英文對照句的結構樹相同時，則我們將此中文對照句視為查詢句的類似句，否則則不為查詢句的類似句。計算從最高層結構樹至第一層結構樹所查詢到的類似句，若較高層結構樹的類似句句數已達到 20 句，則不會再搜尋較低層結構樹的查詢句。

輸入：糖和蜂蜜是甜的。

中文結構樹輸出：S(theme:NP(DUMMY1:NP(Head:Naa:糖)|Head:Caa:|DUMMY2:NP(Head:Naa:蜂蜜))|Head:V_11:是|range:V(head:VP(Head:VH11:甜)|Head:DE:的))
%(PERIODCATEGORY:。)

簡化後結構樹：S(theme:NP(DUMMY1:NP(Head:Na:糖)|Head:Caa:和|DUMMY2:NP(Head:Na:蜂蜜))|Head:V_11:是|range:VP(head:VP(Head:VH:甜)|Head:DE:的))
%(PERIODCATEGORY:。)

結構樹：



分層結構樹： **Na Caa Na V_11 VH DE PERIODCATEGORY**
NP Caa NP V_11 VP DE PERIODCATEGORY
NP V_11 VP PERIODCATEGORY

利用結構樹搜尋：**Na Caa Na V_11 VH DE PERIODCATEGORY**

結果：糖和蜂蜜是甜的。={Sugar and honey are sweet.}
茶和咖啡是苦的。={Tea and coffee are bitter.}
柳橙和檸檬是不同的。={An orange and a lemon are different.}

利用結構樹搜尋：**NP Caa NP V_11 VP DE PERIODCATEGORY**

結果：糖和蜂蜜是甜的。={Sugar and honey are sweet.}
茶和咖啡是苦的。={Tea and coffee are bitter.}
柳橙和檸檬是不同的。={An orange and a lemon are different.}
布朗教授和太太是友善的。={Prof. and Mrs. Brown're friendly.}

利用結構樹搜尋：**NP V_11 VP PERIODCATEGORY**

結果：大衛是年輕而且強壯的。={David is young and strong.}
台灣是溫暖和潮濕的。={Taiwan is warm and humid.}
橋是方便的。={Bridges are convenient.}
誰是富有而且慷慨的？={Who's rich and generous?}
布朗教授和太太是友善的。={Prof. and Mrs. Brown're friendly.}
她的姐妹是害羞的。={Her sister is shy.}
台灣的天氣是溫暖和潮濕的。={The weather of Taiwan is warm and humid.}
她的洋裝的材料是進口的。={The material of her dress is imported.}

圖 8 依照結構樹搜尋範例

如圖 8 所示，輸入的查詢句為「糖和蜂蜜是甜的。」，透過中研院中文結構樹分析、詞性化簡以及擷取各分層的結構樹，得到「3|Na Caa Na V_11 VH DE PERIODCATEGORY|NP Caa NP V_11 VP DE PERIODCATEGORY|NP V_11 VP PERIODCATEGORY」的結果，利用每一階層結構樹去標記化語料庫中搜尋，取得與每一階層結構樹相同結構樹的句子。查詢句的第 3 層結構樹 (Na Caa Na V_11 VH DE PERIODCATEGORY) 比第 2 層結構樹 (NP Caa NP V_11 VP DE PERIODCATEGORY) 多搜尋到一句類似句，「布朗教授和太太是友善的。」，此類似句的分層結構樹為「3|Nb Na Caa Na V_11 VH DE PERIODCATEGORY|NP Caa NP V_11 VP DE PERIODCATEGORY|NP V_11 VP PERIODCATEGORY」，此類似句的各分層結構樹與查詢句的第 3 層結構樹不相符，所以在標記化語料庫中搜尋查詢句的第 3 層結構樹時，「布朗教授和太太是友善的。」並沒有被列為類似句給予推薦；而在標記化語料庫中搜尋查詢句的第 2 層分層結構樹時，「布朗教授和太太是友善的。」的分層結構樹中的第 2 層結構樹與查詢句的第 2 層結構樹相符，所以在此分層結構樹搜尋，「布朗教授和太太是友善的。」會被列為類似句推薦給使用者作參考例句。查詢句的第 1 層結構樹 (NP V_11 VP PERIODCATEGORY) ，因為搜尋到類似句句數較多，故僅附上部分結果。

4.4 中英文混合搜尋

本系統也提供中英文混合的搜尋，對於一道中翻英的題目，學生可能知道其中幾個中文詞的英文翻譯，而沒有英文文法的概念，或是可利用辭典查詢到各個中文詞彙的相對應英文字，卻因為不了解英文文法無法將查到的單字組織成一句完整的英文句子。透過輸入中英文混合的搜尋關鍵字，本系統可以幫忙尋找類似句，透過提供的類似句可以了解此句型結構。如：「宜蘭的空氣非常新鮮」，假設學生知道「空氣」與「非常」的英文字分別為「air」與「very」，則學生可以輸入「宜蘭的 air very 新鮮」，雖然輸入的查詢句

輸入：上學 bus
結果：我搭公車**上學**。={I go to school by **bus**.}

輸入：幾點 go school
結果：你**幾點**上學？={What time do you **go** to **school**?}
你**幾點**去上學？={What time do you **go** to **school**?}

輸入：You 很少 late
結果：你**很少**遲到。={**You** are seldom **late**.}

輸入：昨天 打 basketball
結果：我**昨天**跟麥克**打**籃球。={I played **basketball** with Mike yesterday.}
我**昨天**晚上沒有**打**籃球。={I did not play **basketball** last night.}

圖 9 中英文混合的搜尋範例

是不符合一般英文文法，但本系統可以透過搜尋，去找到類似句法的中英文對照句，進而協助學生練習中翻英。

對於使用者輸入的中英混合查詢句，本系統會依據使用者所輸入的中、英文在標記語料庫中分別搜尋中文句與英文句，若所輸入的詞彙在標記化語料庫中有被搜尋到，則本系統會將此中英文對照句子視為查詢句的類似句，輸出給予使用者作為參考。圖 9 為中英文混合的搜尋範例，可以輸入任意的中、英文，利用空白鍵作為分隔，本系統會根據所輸入的詞彙在語料庫中搜尋，輸出符合搜尋資訊的例句。假設學生不知道頻率副詞“很少”應該使用何字，則輸入「You 很少 late」來查詢，透過本系統查詢給予的建議例句可知道應使用「seldom」並且得知建議例句「你很少遲到。」與相對應的英文句「You are seldom late. 」；若輸入「昨天 打 basketball」來查詢，則可以得到兩句建議例句，分別為「我昨天跟麥克打籃球。={I played basketball with Mike yesterday.} 」以及「我昨天晚上沒有打籃球。={I did not play basketball last night.} 」。

5. 系統效率評估

本系統的中英文語料來源為從網路上收集，適合中學生程度的文件，包括教育部委託宜蘭縣建置語文學習領域國中教科書補充資料題庫[19]、旋元佑文法[16]、基礎英文 1200 句[17]、國民中學學習資源網[18] 的評量題庫及資源手冊等，標記化語料庫共計有七千多句中英文互為翻譯的語料。測試的資料類型為參考旋元佑文法[16]中，所提到的英文五大基本句型，分別為，句型一：主詞+動詞、句型二：主詞+動詞+受詞、句型三：主詞+動詞+補語、句型四：主詞+動詞+受詞+受詞，以及句型五：主詞+動詞+受詞+補語。

測試的資料來源為，賴世雄所編著的文法從頭學[23]、旋元佑文法[16]以及基礎 1200 句[17]中所挑選出來符合五大基本句型的例句，每一句型使用四句測試資料進行測試，透過本系統提供的以中文詞彙、詞性以及結構樹搜尋功能作查詢，並提供使用者類似句做為參考依據。

表 2 為利用五大句型例句使用本系統的搜尋功能所得到的類似句句數。以中文詞為搜尋依據，所得到的類似句以句型一（主詞＋動詞）句數較多，因為搜尋句本身字數較少，相對與標記化語料庫裡中文句詞彙數完全符合的機會較高，故所得到的類似句就會較多。句型三（主詞+動詞+補語）中，其中三句例句類似句句數也高達百句，其推測原因為在中文詞彙的近義詞詞數較多，所以在搜尋標記化語料庫中的中英文對照句，會有為數較多的類似句提供參考。

以詞性搜尋依據，搜尋句的詞性若與標記化語料庫中，中英文對照句的詞性順序相符，則會視為類似句給予推薦。句型較為簡單的句型一（主詞＋動詞）會得到較多的類似句，甚至高達上千句；而較複雜的句型的類似句，則會較少。平均來說以詞性為搜尋依據所得到的類似句大於以中文詞與結構樹為搜尋依據的句數，推測其原因為詞性比對始採用較寬鬆的詞性順序只要依序出現即視為類似句，而不是採用較嚴格的完全比對，故使用詞性比對搜尋會得到較多的參考類似句。

表 2 五大句型例句使用搜尋功能查詢所得到的類似句句數

	中文例句	以中文詞為搜尋依據所得類似句句數	以詞性為搜尋依據所得類似句句數	以結構樹為搜尋依據所得類似句句數
句型一：主詞+動詞	我走路。	97	2655	6/17
	有事發生了。	232	64	0
	他過世了。	1189	1045	13/29
	你跑。	32	2655	6/17
句型二：主詞+動詞+受詞	我愛她。	13	3877	1/4
	你是健康的。	38	658	18/21/141
	貓抓了一隻老鼠。	6	88	1/28
	他寫了一本書。	87	88	3/28
句型三：主詞+動詞+補語	這個問題好像很容易。	1	69	4/23
	他繼續保持單身。	307	249	1/1/7
	他是個大英雄。	297	130	0/0/0/312
	他看起來很慈祥。	379	443	1/23
句型四：主詞+動詞+受詞+受詞	老闆覺得你的提議很刺激。	1	31	2/2/2/33
	他餵貓吃罐頭。	42	611	1/1/39
	他給我一本書。	37	146	2/8
	大衛寫給蘇珊一封信。	1	146	1/8
句型五：主詞+動詞+受詞+補語	他們覺得新房子很舒服。	1	27	1/2/33
	他把鑰匙留在那裡。	108	35	1/1/6
	他叫瑪麗擦窗戶。	44	611	0/3
	我聽到門被關了起來。	7	0	42

以結構樹為搜尋依據，所得到的類似句句數，不一定只有一個數值。因為結構樹會利用各分層結構樹在標記化語料庫中搜尋，所以表格中記錄了各分層結構樹搜尋標記化語料庫所得到的句數，記錄格式為「第 n 層結構樹查詢標記化語料庫所得類似句句數/……/第 1 層結構樹查詢標記化語料庫所得類似句句數」，若搜尋句僅只有一層結構樹，則只會有一個數值。由於結構樹層數越小，結構樹的結構較為簡單，所得到的類似句會越多；相反的，層數越大，所得到的類似句雖較少，但是類似句的句法會與查詢句較為類似。

針對不同的搜尋依據給予使用者類似句作為參考，會因為句型較為簡單而得到過多的類似句，或因為查詢句與標記化語料庫中的句型相似度不高，所以無法查詢到使用者期望的例句。我們期望能利用更多的語料或是針對查詢後提供過多的參考類似句，給予較嚴格的搜尋限制，將類似句句數降低，並期望本系統能達到輔助使用者學習英文、提升使用者對於英文文法的認知及熟悉度。

6. 結語

我們利用人工收集中學生英語學習單以及網路文件中，符合中學生程度的中英文對照句，並將語料作斷詞、詞性擷取以及結構樹建立等前處理，建立標記化語料庫。本系統提供以詞彙為單位、詞性、結構樹來搜尋相似推薦句，亦可透過輸入中英文混合字查詢推薦句，可以協助學生經由簡易的關鍵字搜尋到相似的中英文對照句給予建議，並且期望學生透過本系統所提供的多句相似文法或類似句型的中英文對照句，得以學習到正確的文法以及字彙的用法，並增進學生學習外語的興趣與能力。

目前本系統評估僅使用英文參考書中例句來測試本系統可提供的類似句句數，後續會設計使用者介面並請受測者來實際使用本系統，並且評估本系統效能以及是否達到輔助的效果。本系統可讓使用者利用中文輸入或中英文混合輸入，搜尋相關英文例句，適時給予英語學習者，在寫作時提供參考例句；我們期望若依照相同概念去分析並建構英文句結構樹，亦能讓外國人利用英文輸入或中英文混合輸入法，查詢到相關的中文建議例句，讓外國人也能利用本系統學習到中文句型或是中文文法概念。

致謝

本研究承蒙國科會研究計畫 NSC-95-2221-E-004-013-MY2 的部分補助謹此致謝。我們感謝匿名評審對於本文初稿的各項指正與指導。雖然我們已經在從事相關的部分研究議題，不過限於篇幅因此不能在本文中全面交代相關細節。

參考文獻

[1] Y.-F. Chang and D. L. Schallert, The Design for a Collaborative System of English as Foreign Language Composition Writing of Senior High School Students in Taiwan. *Proceedings of the Fifth IEEE International Conference on Advance Learning Technologies*, 774-775, 2005.

[2] Z. Dong and Q. Dong, HowNet, 2000. http://www.keenage.com [Accessed: Jun. 26, 2008]

[3] J. Kakegawa, H. Kanda, E. Fujioka, M. Itami and K. Itoh, Diagnostic Processing of Japanese for Computer-Assisted Second Language Learning. *Proceedings of the Thirty Eighth Annual Meeting on Association for Computational Linguistics*, 537-546, 2000.

[4] O. Knutsson, T. C. Pargman and K. S. Eklundh, Transforming Grammar Checking Technology into a Learning Environment for Second Language Writing. *Proceedings of the HLT-NAACL 2003 Workshop on Building Educational Applications Using Natural Language Processing*, Volume 2, 38-45, 2003.

[5] C.-L. Liu, C.-H. Wang, and Z.-M. Gao, Using Lexical Constraints to Enhance the Quality of Computer-Generated Multiple-Choice Cloze Items. *International Journal of Computational Linguistics and Chinese Language Processing*, Volume 10, Number 3, 303-328, 2005.

[6] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*, the MIT Press, 1999.

[7] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross and K. Miller, Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, Volume 3, Number 4, 235-244, 1990. http://wordnet.princeton.edu/doc/ [Accessed: Jun. 26, 2008]

[8] R. Mitkov and L. A. Ha, Computer-Aided Generation of Multiple-Choice Tests. *Proceedings of the HLT-NAACL 2003 Workshop on Building Educational Applications Using Natural Language Processing*, Volume2, 17-22, 2003.

[9] S. B. Needleman and C. D. Wunsh, A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins, *Journal of Molecular Biology*, Volume 48, Number 3, 443-453, 1970.

[10] G.R.S. Weir and G. Lepouras, English Assistant: A Support Strategy for On-Line Second Language Learning. *Proceedings of the Second IEEE International Conference on Advance Learning Technologies*, 125-126, 2001.

[11] 中研院中文句結構樹資料庫檢索系統，http://turing.iis.sinica.edu.tw/treesearch/ [Accessed: Jun. 24, 2008]

[12] 中研院中文斷詞系統，http://ckipsvr.iis.sinica.edu.tw/ [Accessed: Jun. 24, 2008]

[13] 中研院平衡語料庫詞類標記集，http://ckipsvr.iis.sinica.edu.tw/category_list.doc [Accessed: Jun. 24, 2008]

[14] 中研院現代漢語語料庫一詞泛讀，http://140.109.150.65/cwordframe.html [Accessed: Jun. 24, 2008]

[15] 林仁祥及劉昭麟。國小國語科測驗卷出題輔助系統，*2007 台灣網際網路研討會論文集*，論文光碟。台灣，台北，2007。

[16] 旋元佑文法，http://tw.myblog.yahoo.com/jw!GFGhGimWHxN4wRWXG1UDIL_XSA--/ [Accessed: Jun. 24, 2008]

[17] 基礎英文 1200 句，http://hk.geocities.com/cnlyhhp/eng.htm [Accessed: Jun. 24, 2008]

[18] 國民中學學習資源網，http://140.111.34.172/teacool/new_page_2.htm [Accessed: Jun. 24, 2008]

[19] 教育部委託宜蘭縣發展九年一貫課程建置語文學習領域（英語）國中教科書補充資料暨題庫建置計畫，http://140.111.66.37/english/ [Accessed: Jun. 24, 2008]

[20] 教育部國民教育司，http://www.edu.tw/EJE [Accessed: Jun. 24, 2008]

[21] 陳佳吟、柯明憲、吳紫葦及張俊盛，電腦輔助英文文法出題系統，*第十七屆自然語言與語音處理研討會論文集*。台灣，台南，2005。

[22] 劉吉軒、洪培鈞及李金瑛，以英語寫作輔助為目的之語料庫語句檢索方法，*第十九屆自然語言與語音處理研討會論文集*，5-19。台灣，台北，2007。

[23] 賴世雄，*文法從頭學*，長春藤有聲出版有限公司。2007。