

# 國小國語科測驗卷出題輔助系統

林仁祥 劉昭麟  
政治大學資訊科學研究所  
{94753029, chaolin}@nccu.edu.tw

## 摘要

我們利用自然語言處理技術，配合網際網路所提供的資訊與傳播管道，建構協助教師建立基礎中文試題的環境。目前我們的系統可以協助教師編輯包含基礎聽力、克漏詞、找錯字與量詞填充共四大類型試題的題庫與試卷。教師所編輯的試卷可以透過網際網路施測，測驗結果可以立即評分，並且儲存於資料庫中讓教師能夠進行測後的分析。

## 1 前言

自然語言處理是研究讓電腦能處理及了解人類日常所使用的語言[18]，使得電腦能理解人類知識，讓人類與電腦能使用自然語言進行溝通。近年來有許多國內外學者[7][10][11][15][21]重視及研究在電腦輔助語言教學上，運用自然語言處理來輔助出題。國外學者如 Mitkov 和 Ha 採用自然語言的技術[21]輔助來產生多選項的試題；而國內學者張俊盛則應用自然語言的技術來輔助英文文法出題[10]，文法學習的目的是學習者能夠對語言的架構及詞彙順序的了解程度。劉昭麟等則利用詞義辨析來協助教師編輯英文克漏詞試題[19]。所以利用自然語言處理的技術來協助人類學習語言，已成為熱門的話題。

電腦硬體技術發展迅速，加上高頻寬的網路，利用電腦高計算量的特性，把所收集的資料，經由演算法，產生各種類型且大量的候選題目。早期教師製作教材時，都是需要到處去尋找資源，而且製作試卷時必須完全用手工式的方式來完成，尋找教材及編輯紙本試卷時需要許多心思及時間。所以本研究是運用自然語言處理技術在國語科目的應用上，輔助教師編輯國語注音、基本聽力測驗、中文克漏詞、錯別字及量詞各類型的題目[20]並儲存於題庫資料庫裡。編輯考卷檔時，可從題庫資料庫中選取適當的題目合成為一張試卷檔提供學生做測驗。本系統除了編輯題目或試卷功能之外，也提供管理測驗結果的查詢及統計的功能，教師可以根據學生的作答的紀錄，作為教學的評估依據（例如：那些種類的題型或題目常會造成學生答錯等）。

我們在第 2 節介紹本系統主要架構及資料來源，第 3 節說明使用者編輯界面及編輯流程，第 4 節詳細介紹相關文字處理技術，第 5 節為結論與未來希望發展的目標。

## 2 系統設計

本節介紹本系統架構以及本系統中所需要的資源以及來源處。

### 2.1 系統架構

圖 1 為輔助出題的系統架構圖，圖中矩形為資料庫，而圓角矩形的部分則為處理程序。本系統將基本語料庫加工後，提供給教師編輯試題，編輯好的試題儲存於題庫資料庫，且可以依題型的不同，細分成個別的資料庫，我們建立四聲辨識、改錯字、量詞及中文克漏詞試題等題庫。儲存於不同類型題庫的試題可供教師編輯試卷之用；編輯好的試卷檔，則儲存於

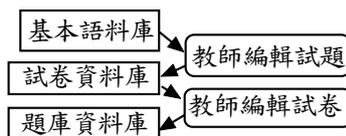


圖 1：系統主要架構

試卷資料庫。學生接受測驗時，可選擇所應受測的試卷。學生測驗後會把學生作答情形記錄在學生資料和成績檔裡。之後，教師可以根據學生資料檔，查詢特定學生的答題資料，或者是個別班級的答題資料，了解學生整體表現，做為以後教學的評量及依據。

試卷資料庫。學生接受測驗時，可選擇所應受測的試卷。學生測驗後會把學生作答情形記錄在學生資料和成績檔裡。之後，教師可以根據學生資料檔，查詢特定學生的答題資料，或者是個別班級的答題資料，了解學生整體表現，做為以後教學的評量及依據。

### 2.2 資料來源

基本語料庫是經過處理後的文字檔（例如：國小國語課本的內容，把一些標題符號刪除等）。我們收集坊間的國小國語科的參考書裡的生字表及課文內容，並且利用網路爬梳器（web crawler）下載有關國小課本內容的資料[2]。下載的檔案都為html的檔案，檔案裡頭包含了許多標記式語言，所以必須篩選出主體的文字內容的資訊，加工後成為我們所需要的格式，提供這些語料給教師來編輯試題。在產生試題中需要辨析詞彙的義原及詞性時，本系統是根據 1999 年 HowNet[16]辭典為主，而 HowNet 為一個英漢雙語的辭典，收錄約十一萬條詞條。

## 3 編輯介面

本節介紹目前本系統實作的成果。

### 3.1 四聲辨識

我們以兩個小節分別介紹四聲辨識題庫編輯及四聲辨識試卷編輯。

#### 3.1.1 四聲辨識題庫編輯

圖 2 是四聲辨識[6]題庫編輯界面，包含三個部份：編輯試題框架、題目預覽框架及編輯聲音按鈕（上方都標示注音符的按鈕）。

編輯試題框架包含編輯題目的文字方塊、題號、編輯按鈕、確認按鈕及清除按鈕。文字方塊顯示使用者所編輯的題目內容；題號的功能是將文字方塊的內容根據所選的題項號碼，來決定這個題目選項；確認按鈕的功能是根據所編輯的聲音去尋找是否有相符的聲音檔清單的內容（本系統的聲音檔是用 Windows 作業系統裡的錄音機程式所錄製，並以 wav 為副檔名儲存在硬碟，並把本系統所有聲音檔名列出清單），來決定編輯是否正確；清除按鈕的功能是將文字方塊內容清空。

本系統採用傳統的選擇題，每道題目裡，會有四個選項，而本系統在題目預覽框架裡包含兩個部份，上半部為設定考試題目，下半部則為四個選項（標示“未編輯”內容的按鈕）。我們一開始必須先編輯好四個選項，接下來，我們必須從這四個選項內容選取其中一個選項內容，成為這道考試目的答案，所以選取設定考試的按鈕（題目預覽框架

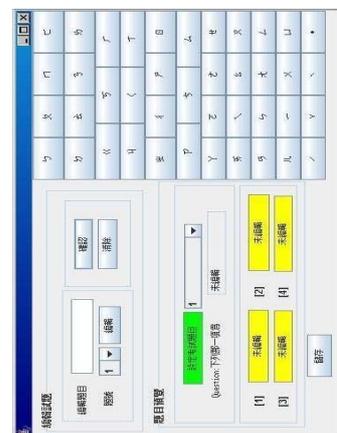


圖 2：編輯四聲辨識題庫畫面

裡的按鈕)右邊的下拉式選單的選項(選項有1、2、3及4,分別代表1、2、3及4的標示在黃色按鈕的內容),選取完之後,再按下**選取設定考試**,即編輯完成了。隨後再按**儲存**按鈕後,儲存到四聲辨識題庫資料庫裡。

### 3.1.2 四聲辨識試卷編輯

圖3為編輯四聲辨識試卷的畫面,這個畫面大概分成三個部份:查詢框架,候選區框架及四個題目框架。**查詢框架**裡的查詢按鈕,用來查詢題庫資料庫裡是否有老師所要考的題目(例如:查詢聲音為“ㄅ”的題目);**候選區框架**是經由查詢所選出來的題目,會先放在候選區裡;一張試卷中,總共會有四題選擇題,所以會有四個題目框架,我們製成試卷時,會把所編輯好的題目,放在題目框架裡,其實最主要的用意在於能夠預覽試卷。



圖 3：四聲辨識試卷編輯畫面

四聲辨識試卷的編輯流程如下所述。首先,我們在圖3按下查詢按鈕後,即出現查詢視窗(圖4),在聲音的文字方塊,輸入查詢題目的條件,就圖4為例,我們想從四聲辨識題庫裡取出聲音是“ㄅ”的題目及答案,結果查詢出來的結果總共有2筆,分別是題號為2和5的內容,內容分別為[題號]2 [題目]ㄅ [1] ㄅ [2]ㄅ [3]ㄅ [4]ㄅ [答案]1 和 [題號]5 [題目] ㄅ [1]ㄅ [2]ㄅ [3]ㄅ [4]ㄅ [答案]2。

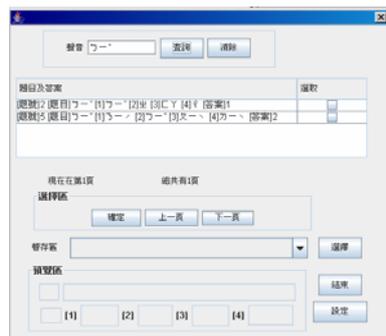


圖 4：四聲辨識試卷編輯畫面

查詢出來的每道題目的後方都有個選取方塊,勾選要題目,按下確認按鈕後,即把所勾選的題目傳送到暫存區。

我們剛才所勾選的題目,都被傳送到暫存區。暫存區是個下拉式的選單,我們從下拉式的選單中,選出一題我們想要編輯的題目,按下選擇按鈕後,在預覽區即顯示出剛所選的題目及在題庫資料庫的題號、答案、及選項內容,按下設定按鈕後就傳送到圖3四聲辨識試卷編輯畫面。

## 3.2 中文克漏詞

圖5為編輯中文克漏詞試題流程,本系統會依據編輯者選擇所要考的詞彙,從基本語料庫搜尋包

含編輯者所指定的詞彙的句子,提供給編輯者做選擇,編輯者選定句子之後,本系統自動把句子中所指定詞彙

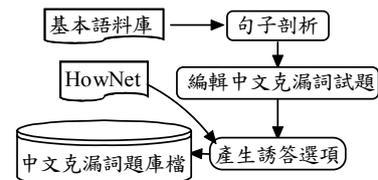


圖 5：編輯中文克漏詞試題流程

代,成為克漏詞試題的題目,編輯者指定的詞彙則為答案,並且加入一些誘答選項後,將題目及誘答選項存入題庫裡,就成一道中文克漏詞試題,我們可以從中文克漏詞題庫裡選取幾道題目後,儲存於中文克漏詞試卷。我們在接著介紹中文克漏詞題庫編輯及中文克漏詞試卷編輯。

### 3.2.1 中文克漏詞題庫編輯

圖6為編輯中文克漏詞試題的畫面。中文克漏詞編輯環境大概分成兩個部份:查詢設定及題目選項編輯。首先輸入詞彙,按下查詢的按鈕,本系統會從語料庫裡搜尋句子,把所有包含此詞彙的句子,傳送到選擇編輯題目的下拉式選單,



圖 6：編輯中文克漏詞試題的畫面



圖 7：題目預覽區的畫面

供編輯者選取。接下去編輯題目選項,假如此題中文克漏詞想要考的詞彙為“學習”,把題目中有“學習”的詞彙以\_\_取而代之,呈現的結果顯示在題目預覽區(圖7),並且選項中的內容則為答案,然後要編輯其他三項的誘答選項(distractor),誘答選項的選擇影響到試題的難易度,是一份很重要的工作,若只考慮以詞頻作為挑選誘選答案是不夠的,王俊弘[4]使用詞義辨析。我們採用兩種方式來編輯誘答選項,第一我們利用唯一的答案跟 HowNet 裡頭的所有的詞的義原做比較,提取與答案有交集義原的詞彙,提供給試題編撰者選擇。如圖8所示,只要輸入“學習”這個詞彙,本系統會提供跟“學習”詞彙有義原關係詞彙的選項,供使用者作為參考及使用。



圖 8：提供參考選項的下拉式選單

第二種方法是編輯者可以自行輸入誘答選項的內容,編輯完成後就儲存到中文克漏詞題庫檔。

### 3.2.2 中文克漏詞試卷編輯

圖9為編輯中文克漏詞試卷的畫面,我們先介紹畫面中各個物件的功能,再介紹編輯的流程。

中文克漏詞試卷編輯畫面分成三個部份:查詢框架,候選區框架及四個題目框架。**查詢框架**裡有

查詢的按  
鈕，用來查  
中文庫句  
漏詞的題  
庫裡；候選  
區是經過  
查詢所  
由選出  
題目，會  
先選  
放在候選  
區裡；一  
張試卷  
裡共有  
四題和  
四題目  
框架  
我們  
製定試  
卷時，  
會把所  
編好的  
題目，  
放在題  
目框架  
裡，供預  
覽試卷  
，一切完  
成之後，  
就按左  
下角儲  
存按鈕  
以儲存  
試卷。



圖 9：中文克漏詞試卷編輯畫面

進行中文  
克漏詞試  
卷的編  
輯時，首  
先在圖  
9中的  
文字方  
塊，輸  
入所要  
編輯的  
詞彙後，  
按下查  
詢按鈕  
後，進  
入查詢  
詞彙畫  
面如圖  
10所示。  
圖10  
為整個  
查詢詞  
彙畫面  
分成三  
個區域，  
分別是  
查詢區、  
查詢結  
果區及  
暫存區。

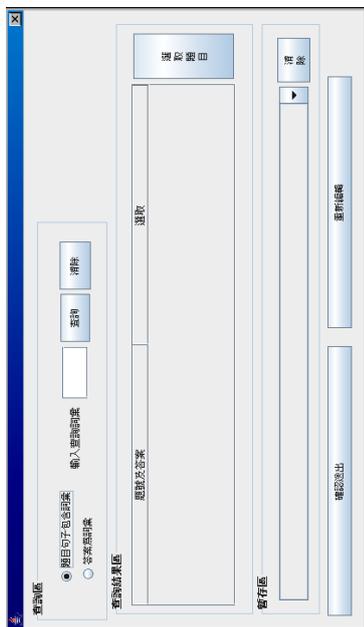


圖 10：查詢詞彙的畫面

接下來介紹編輯的流程，我們分成三個步驟來說明。

◆ 步驟一

首先在查詢區



圖 11：查詢區的畫面

中的文字方塊裡，輸入查詢的詞彙，如圖 11 所示，先點選查詢條件的選取方塊，查詢條件有二種，一種是題目句子包含詞彙，另一種是答案為詞彙，第一種查詢的條件是找尋題目句子中有包含某個詞彙的題目（不表示此詞彙為這道題目的答案），第二種查詢條件是編輯者能夠針對某個詞彙，找尋出題目的答案為該詞彙。選取完畢之後，再輸入所要查詢的詞彙，按下查詢按鈕，本系統從中文克漏詞題庫裡取出所包含所輸入詞彙的題目，顯示在查詢結果區的表格，即進行步驟二。

◆ 步驟二

在結果顯示區中，如圖 12 所示，查詢出來的結果總共有兩筆，分別是題號為 7 和 8 的內容，內

容分別為  
[題號]7 [題  
目]我覺得  
——中文  
是一件很  
愉快的事  
[答案]學習  
[1]討論  
[2]打球  
[3]學習  
[4]打球  
[答案號  
碼]2 和 [題  
號]8 [題  
目]我們要  
——照顧  
自己[答案]  
學習 [1]討  
論 [2]學  
習 [3]認  
真 [4]勤  
奮 [答案  
號碼]3。查  
詢出來  
的每道  
題目後  
方都有  
一個選  
取方塊  
，勾選  
要題目  
，按下  
選取題  
目按鈕  
後，即  
把所勾  
選的題  
目，傳  
送到暫  
存區。

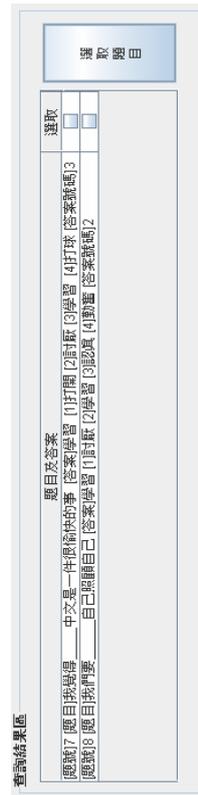


圖 12：查詢結果區

◆ 步驟三

如圖 13 所示，暫存區裡的下拉式選單內容為編輯者選取題目的結果，我們從下拉式的選單中，選出一題我們想要編輯的題目，即按下確認送出按鈕，會把下拉式選單的資料傳送到圖 9 的候選區。

尋找包含“學習”這個詞彙题目的句子，編輯選定某個句子之後，會把此题目的選項及答案，列在候選區暫時瀏覽，在預覽區即顯示出剛所選的題目以及在題庫資料庫的題號、答案及選項內容，編輯者也可以在候選區來修改題目、答案及選項內容，在圖 9 中文克漏詞試卷編輯畫面中間有選擇題號的下拉式選單、編輯及重新編輯按鈕功能鍵。這三個功能鍵主要的目的為：

- ◆ 題號下拉式選單功能為決定候選區裡試題在此試卷中的題號。
- ◆ 編輯按鈕功能為，根據題號下拉式選單所決定的題號，把候選區的試題的题目、選項內容及答案丟到所指定題號的题目框架。
- ◆ 重新編輯按鈕功能並不是放棄之前的編輯動作，而是根據題號下拉式選單所決定的題目框架編號，把該題目框架裡的試題题目、選項內容及答案傳送到預覽區，進行修改題目、答案、及選項內容的動作。

當試卷編輯完成之後，按下左下角儲存按鈕後，即儲存到中文克漏詞試卷檔。

3.3 改錯字題庫編輯

圖 14 為編輯改錯字題庫的畫面，我們先介紹整個畫面中各個物件的功能後，再介紹整個編輯的流程。

改錯字題庫畫面分成四個區域，分別是輸入文字區、編輯區、暫存區及試題預覽區。輸入文字區域裡有輸入一句話的文字方塊及編輯按鈕，使用者先在文字方塊裡輸入題目，接下來按下編輯後，會把所輸入的文字傳送到編輯區的選擇區中下拉式

圖 13：暫存區

選單做為編輯字的依據。編輯區裡包含選擇區及提供錯別字選項區；選擇區裡有下拉式選單及選取



圖 14：改錯字試題編輯畫面

按鈕，下拉式選單內容為輸入文字區的文字方塊內容，使用者在編輯區的下拉式選單內容，選取中文字來編輯錯別字，按下選取按鈕後，本系統根據使用者所選擇的內容（選擇區裡下拉式選單內容的國字），找出發音相同的國字與字形相似的國字。

提供錯別字選項區裡有同音字下拉式選單及相似字下拉式選單。同音字下拉式選單內容為本系統所找出來的同音字，而相似字下拉式選單為本系統所找出來的相似字，分別在兩個下拉式選單右方都有確定按鈕，同音字下拉式選單右方的確定按鈕的功能是將兩部份（選擇區所選下拉式選單內容以及同音字下拉式選單所選的內容）內容分別傳送到暫存區的正確字文字方塊及錯字文字方塊。

按下相似字下拉式選單右方的確定按鈕，則將傳送選擇區所選下拉式選單內容以及相似字下拉式選單所選的內容。暫存區的主要目的是暫時儲存編輯者目前所編輯出來結果，把目前的結果儲存在正確字文字方塊及錯字文字方塊。在兩個文字方塊的右方，各有清除的按鈕分別清空目前所編輯的正確字及錯別字。試題預覽區的功能是來預覽試題最後的結果，而試題預覽區有 3 個文字方塊，而這 3 個文字方塊的目的如下。

- ◇ 題目文字方塊：預覽改錯字的題目。
- ◇ 正確字文字方塊：預覽此題的正確解答。
- ◇ 錯別字文字方塊：預覽編輯者所決定那些錯別字來測驗學生。

試題預覽區右方有功能區，功能區裡有兩個按鈕，分別為儲存及清除按鈕，按下儲存按鈕，即將題目的內容、正確字及錯別字內容儲存到改錯字題庫檔裡；按下清除按鈕，則清空試題預覽區裡的所有文字方塊內容。

接下我們分成三個步驟來介紹編輯的流程。

#### ◇ 步驟一

在輸入文字區中的文字方塊裡，輸入一句文字，如圖 15 所示，然後按下編輯按鈕，會把所輸入句子中的每個國字傳送到編輯區的下拉式選單，即進行步驟二。

#### ◇ 步驟二

在編輯區中，如圖 16 所

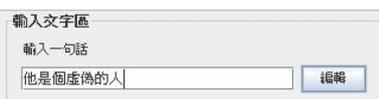


圖 15：輸入文字區的畫面



圖 16：編輯區的畫面

示，我們針對某個國字進行編輯（例如：我們在選擇區選擇“偽”這個字），按下選取

按鈕，本系統自動產生同音字（例如：提供與“偽”同音的字），分別提供儲放在同音字下拉式選單及相似字下拉式選單，供編輯者作參考。假如，編輯者決定同音字為錯別字，就按下確定按鈕（同音字下拉式選單右方），同樣地，編輯者決定相似字為錯別字，則按下確定按鈕（同音字下拉式選單右方）。

#### ◇ 步驟 3

暫存區顯示目前編輯改錯字题目的情況，如果要繼續編輯時，就重覆步驟一到步驟二的動作，假使編輯完成之後，即按下編輯完成按鈕，會將暫存區的資料傳送到試題預覽區，如圖 17 所示，編輯者可以在試題預覽區裡預覽最後的編輯情況，假使不再做任何修改，在試題預覽區右方有個儲存按鈕，按下儲存按鈕即把試題預覽區的資料儲存到改錯字題庫裡。

### 3.4 量詞

量詞是表示數量單位的詞，具有修辭的作用，目前常用的量詞大概有五百多種[13]。

有了那麼多量詞，需要使用量詞時，不再只是使用“個”這樣的量詞。本系統所考的量詞題

目，編輯者先自行輸入一段句子，本系統從輸入的句子中擷取量詞部份，並且把量詞部份挖空作為填充題，提供給學生測試。

圖 18 是編輯量詞試題的畫面，整個畫面分成三個區域，分別是輸入句子區、選擇量詞區以及題目預覽區。

再來介紹編輯的流程，分成三個步驟來說明。

#### ◇ 步驟一

##### 一

先，在輸入句子區中的文字方塊裡，輸入一句文字（例如：我們在文字方塊裡輸入“那邊有一張桌子和一顆蘋果”的句子），如圖 19 所示。輸入完畢之後，按下確定按鈕，會把所輸入句子傳送到中文斷詞系統，本系統分析所收到的斷詞結果中所附帶的詞類訊息，判斷是否有量詞存在，則會把量詞傳送到選擇量詞區。

在選擇量詞區中，出現了兩個量詞，我們可以對一個或一個以上的量詞進行編輯，只需按下量詞的按鈕，按鈕的背景會變成紅色（背景為紅色的按鈕則是要考的量詞），假使要取消某個背景為紅色的量詞按鈕，則按下背景為紅色的量詞按鈕，此量詞按鈕的背景就不會出現紅色，選擇所要考的量詞



圖 17：試題預覽區的畫面

圖 18：量詞試題編輯畫面

圖 19：輸入句子區的畫面



圖 18：量詞試題編輯畫面

圖 19：輸入句子區的畫面

圖 19：輸入句子區的畫面

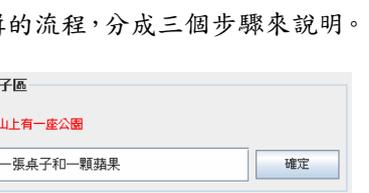


圖 19：輸入句子區的畫面

完畢之後，預覽題目區就可以預覽編輯的題目，例如：我們按下“張”及“類”二個量詞，在題目預覽區則立即顯示題目為“那邊有一( )桌子和一( )蘋果”。

#### ◆ 步驟三

預覽題目區顯示目前編輯量詞題目的情況，假如編輯完成之後，即按下儲存按鈕，會將預覽題目區的資料傳送到量詞題庫檔裡。

## 4 相關文字處理技術

### 4.1 中文克漏詞

1. 另一方面\_\_\_又對名牌所標榜的精緻及品質愛不釋手
  - (1) 卻 → 正確答案
  - (2) 卻病 → 誘答選項
  - (3) 卻步 → 誘答選項
  - (4) 卻而不受 → 誘答選項

圖 20：中文克漏詞範例試題

圖 20 是一題克漏詞試題的範例，挖掉了一個詞的句子稱為題幹 (stem)，被挖掉的詞即是該試題唯一的答案 (key)，而其它三個選項稱為誘答選項。

接下來我們要進行編輯誘答選項，本系統提供兩種方式產生誘答選項的詞彙，提供給編輯者作為參考。

第一種方式，在 HowNet 中，每個詞彙都有 8 個欄位，尤其最後一個欄位名稱為 DEF，描述著這個詞彙的義原關係，HowNet 大概使用了一千七百個多個義原關係 [8]，來定義中英雙語知識辭典中的每個詞彙，並且建有描述各個義原之間的關係的分類樹，例如：“讀書”一詞由“從事”、“學”與“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， $s = \{\text{從事, 學, 教育}\}$ ，我們去尋找 HowNet 所有的詞彙，並且把詞彙的義原與集合  $s$  做比對，與  $s$  有交集的詞彙都找出來後，接下來做排序 (由多到少)，把排序後的詞彙提供給編輯者做為誘選答案。

第二種方式是採用中研院“現代漢語語料庫”一詞泛讀的學習工具 [12]，本系統會連到一詞泛讀的網頁，假如答案為 A 詞彙，會自動連到中研院一詞泛讀的網頁，並且傳送參數 A 詞彙，網頁會回傳有關 A 詞彙的近義詞，本系統收集網頁所回傳的結果後，提供 A 詞彙的近義詞給編輯者。

### 4.2 改錯字

在錯別字試題編輯方面，本系統會針對編輯者所選的字，提供同音字及相似字二種功能。

在同音字方面，本系統使用詞庫檔為 tsi.src [17]，也是新酷輸入法 [9] 所採用的詞庫檔。本系統會根據編輯者所選字，從詞庫檔找出與編輯者所選字具有相同發音的國字，全部列出來，供編輯者作為選擇，例如：別出“心”裁，常會被誤用為別出“新”裁。

在相似字方面，本系統會根據編輯者所選字，列出這個字的倉頡碼。我們採用倉頡碼的原因，因為倉頡碼是根據字的形狀來做分類 [5]，倉頡字母的分類，總共分成二十五類，每個字根有相對應的字根符號。

接下來介紹，如何把中文字區分為“分體字”及“連體字”。

一個中文字可以分為兩個或兩個以上的部份，則稱為“分體字”，而“分體字”才有“字首”與“字身”。第一個切割出來的部份，則稱為“字首”，其餘的部份稱為“字身”。“分體字”的分割方式如下。

◆ 字首字身上下重疊：例如：音。

◆ 字首字身左右並排：例如：仁。

◆ 字首而外而內包含字身：例如：國。

我們會記錄這個中文字的字首與字身，所以分體字會有兩組倉頡碼，並且把字首及字身的中文字都轉換成倉頡碼 (字首及字身的倉頡碼個數都不超過五個)，例如：“借”這個字為是由字首“人”與字身“昔”所組合而成的，並且把“人”轉成倉頡碼為“人”，而“昔”轉成倉頡碼為“廿日”。我們在處理一些中文字，有些中文字滿足之前所描述的“分體字”分割方式其中一項，但是字首或者是字身卻沒有這個中文字時，於是我們仍舊是以倉頡碼取碼方式來取碼，並且強迫把此中文字歸類為“分體字”，將來要編輯這一類型的中文字，就可以尋找此中文字字首及字身的倉頡碼 (例如：“聖”因為合乎了“字首字身上下重疊”的規則，所以此字為分體字，但此字的字首卻沒有中文字，所以我們仍舊採用倉頡碼取碼規則，“聖”的字首倉頡碼為“尸十口”，而字身為“壬”，所以倉頡碼“竹土”)。

我們建立一個構字式檔，假如此中文字為分體字時，將會把此中文字字首及字身的倉頡碼儲存在構字式檔，所以分體字在構字式檔裡會有兩組倉頡碼，第一組為字首的倉頡碼，第二組為字身的倉頡碼。

如果中文字無法滿足之前所描述的“分體字”分割方式其中一項，則稱為“連體字”。我們會把此中文字的倉頡碼儲存在構字式檔，所以連體字在構字式檔只有一組倉頡碼。

接下來要說明如何去尋找相似字的步驟，當編輯者選取某個中文字時，本系統會去尋找構字式檔，如果有兩組倉頡碼，此中文字為分體字，如果只有一組倉頡碼 (此中文字的倉頡碼)，則為連體字；我們依照構字式檔來判別此中文字為分體字或連體字，分體字或連體字尋找相似字的條件以及方法也會有所不同。假設  $\beta$  為尋找相似字的條件 ( $\beta$  有  $n$  個倉頡碼)，如何決定  $\beta$  的尋找條件如下。

分體字與連體字尋找相似字的條件 ( $\beta$ ) 有所不同，我們分成兩點來介紹。

1. 假設此中文字的構字式檔裡只有兩組倉頡碼 (此中文字為分體字)， $\beta$  為此中文字“字身”的倉頡碼，我們不用中文字字首的倉頡碼，因為字首一般都為中文字的部首。下列四點則為編輯者所選的字的字身倉頡碼，來取得  $\beta$  的方法。

◆ 字身倉頡碼剛好取一碼：例如：字身為人， $\beta$  為“人”。

◆ 字身倉頡碼剛好取二碼：例如：字身為二， $\beta$  為“一一”。

◆ 字身倉頡碼剛好取三碼：例如：字身為寺， $\beta$  為“土木戈”。

◆ 字身倉頡碼超過三碼，仍舊取三碼 (頭、二及尾碼共三碼)，取三碼的目的是因為倉頡碼檔裡的每一個國字，字身倉頡碼最多三碼。例如：字身為馬， $\beta$  為“尸手火”。

2. 假設中文字的構字式檔裡只有一組倉頡碼 (此中文字為連體字)，這一組倉頡碼超過兩碼，則  $\beta$  取此中文字構字式檔的倉頡碼，總共三碼 (頭、二及尾碼共三碼)；若這一組倉頡碼未超過兩碼，則  $\beta$  為此中文字的倉頡碼。

我們收集了一萬多個中文字的倉頡碼 [1]，儲存在倉頡碼檔，接下來介紹如何利用  $\beta$  從倉頡碼檔裡尋找相似字。

依據編輯者所選中文字為分體字或者是連體字，我們依照上述原則選定判斷相似字所需的  $\beta$ ，

其尋找相似字的方法也有所差異，以下來分別說明。

◇ 假設編輯者所選的中文字為分體字，尋找相似字的方式是利用  $\beta$  去比對倉頡碼檔每一個中文字倉頡碼的後幾碼（因為倉頡碼檔裡，每一個國字的字身倉頡碼最多三碼，而且又位於倉頡碼的後幾碼），例如： $\beta$  總共有三碼，只需搜尋其他中文字的最後三個倉頡碼，只要比對相同，就成了相似字，像“河”的字身為“可”，字身的倉頡碼為“一弓口”，而“何”的字身為“可”，字身的倉頡碼為“一弓口”，二字的字身倉頡碼都同為“一弓口”，所以“何”為“河”的相似字。

◇ 假設編輯者所選的中文字為連體字，尋找相似字的方式又分為兩種。

(1) 此中文字構字式倉頡碼總共四碼或四碼以上時， $\beta$  只需去比對其他中文字倉頡碼的後三碼，只要比對相同，就成了相似字。

(2) 此中文字構字式倉頡碼少於四碼時，本系統自動尋找其他中文字倉頡碼中包含著  $\beta$  的倉頡碼，此字就成為了相似字。

#### 4.3 量詞

判斷那些詞為量詞的步驟如下。

步驟 1：本系統會把句子送到中央研究院中文詞庫小組的中文斷詞系統[3]。

步驟 2：本系統會接收中文剖析服務回傳詞性的結果。

步驟 3：量詞詞性的標記為“M”，來判定那些詞為量詞。

編輯者如果輸入一句話為“今天天氣很好”，本系統會把句子送到中央研究院中文詞庫小組中文剖析服務，回傳結果為 今天(N) 天氣(N) 很(ADV) 好(Vi)，系統判斷“今天天氣很好”這個句子完全沒有任何的量詞。

與上一個例子比較，假如編輯者輸入一句話為“山上有一座公園”，則回傳結果為 山(N) 上(N) 有(Vt) 一(DET) 座(M) 公園(N)，本系統會判斷回傳的結果的詞性，所以“山上有一座公園”這個句子有量詞。

#### 5 結論

使用者使用本系統只需開啟網頁，經由網頁連接到 JNLP (Java Network Launching Protocol) 來取得 JAR 檔 (JAR 檔裡包含著本系統的 class 檔、聲音檔及一些文件檔)，把 JAR 檔下載到使用者端的電腦來執行。

本系統能夠輔助教師編輯題目時，常見題目類型有基本聽力測驗、中文克漏詞、改錯字及量詞等；編輯紙本考卷就不需要手工剪貼的，畫面不但美觀而且也比較方便，在基本聽力測驗中，會提供語音的功能，讓學生邊聽邊作答，是一般紙本題目所不能相比擬的。在編輯題目時，希望系統能夠即時提供出題資訊，讓教師做編輯題目的參考，這樣教師就不需要為了要編什麼樣的題目來傷腦筋了。

E-Learning 現在非常的盛行，希望教材能夠在不同的平台上執行，SCORM 收集各方的優點 [14]，已成為廣泛的標準，並且可依照使用者的需求，產生符合 SCORM 標準的教材，引用別人的教材，編輯成自己的教材，讓本系統更有延展性。

#### 致謝

本研究承蒙教育部頂尖大學計畫與國科會研究計

畫 NSC-94-2213-E-004-013-MY2 之部分補助，謹此致謝。

#### 參考文獻

- [1] 中文全字庫，  
<http://www.cns11643.gov.tw/web/index.jsp>。
- [2] 中央研究院-中文剖析樹檢索系統。
- [3] 中央研究院中文詞庫小組中文斷詞系統。
- [4] 王俊弘，利用自然語言處理技術自動產生英文克漏詞試題之研究，碩士論文，國立政治大學，資訊科學研究所，台北，台灣，2003。
- [5] 多讀多寫網上培訓計劃，  
[http://mcu.edu.tw/~ccwu/med\\_course/chenjen.doc](http://mcu.edu.tw/~ccwu/med_course/chenjen.doc)。
- [6] 注音符號第二式（維基百科，自由的百科全書），  
<http://zh.wikipedia.org/>
- [7] 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，2005。
- [8] 曾慧馨，中文動詞自動分類研究，碩士論文，國立政治大學，語言學研究所，台北，台灣，2001。
- [9] 智慧型注音輸入法發展總部，  
[http://chewing.csie.net/chewing\\_dict\\_edit.html](http://chewing.csie.net/chewing_dict_edit.html)。
- [10] 楊媛茜、楊捷扉、張嘉銘及張俊盛，電腦輔助閱讀測驗自動出題，第十七屆自然語言與語音處理研討會論文集，2005。
- [11] 鄭守益及梁婷，中文句子相似度之計算與應用，第十八屆中華民國計算語言學學會論文集，2006。
- [12] 鄭錦全、黃居仁、羅鳳珠、蔡美智、黃郁純、陳錦宇、呂奇蓉、韓雅琪、李嘉真，《全球華語文數位教與學資源中心》，  
<http://elearning.ling.sinica.edu.tw/index.html>，中央研究院語言學研究所，2005。
- [13] 賴慶雄，實用量詞手冊，螢火蟲出版社，2005。
- [14] 蘇俊銘、曾憲雄、蘇東興及蔡文能，網路教材標準化之研究，網路教學系統平台與內容標準化研討會論文集，83-90，2003。
- [15] G. Conole and J. Bull, Pebbles in the Pond: Evaluation of the CAA Central, *Proceedings of the Sixth Computer-Assisted Assessment Conference*, 63-73, 2002.
- [16] HowNet, <http://www.keenage.com>.
- [17] libtabe - the way to deal with Chinese,  
<http://libtabe.sourceforge.net>。
- [18] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*, The MIT Press, 1999.
- [19] C.-L. Liu, C.-H. Wang and Z.-M. Gao, Using Lexical Constraints for Enhancing Computer-Generated Multiple-Choice Cloze Items, *International Journal of Computational Linguistics and Chinese Language Processing*, 10(3), 303-328, 2005.
- [20] M.-S. Lu, J.-H. Lin, Y.-C. Wang, Z.-M. Gao, C.-L. Liu, and C.-Y. Chang, Prototypes of Using NLP Techniques for Assisting Translation and Authoring of Test Items, submitted for publication, 2007.
- [21] 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*, 17-22, 2003.