

電腦輔助英文字彙出題系統之研究

Toward Computer Assisted Item Generation for English Vocabulary Tests

王俊弘 劉昭麟
政治大學 資訊科學系
{g1753024, chaolin}@nccu.edu.tw

高照明
台灣大學 外國語文學系
zmgao@ntu.edu.tw

摘要

電腦輔助教學的研究熱潮正方興未艾，由電腦自動化產生試題的相關研究也朝各種方向發展。本篇論文主要利用自然語言處理中統計並分析 selectional preference 的技術，與機器學習從訓練資料歸納出法則的精神，發展了一套能協助教師製作英文字彙題的系統。系統所產生的試題主要針對高中生—尤其是高中三年級生而設計，在難度上與大學入學考試英文科的字彙題相仿。除了能依使用者所選定的字彙產生試題，也能在不指定字彙的情況下，產生模擬試題，以提供學生更多自我評量的機會。由於產生試題的過程中並未深入到語意層面的分析，無法確保所有產生的試題語意完整、能夠評量受試者的學習成果，因此試題尚需經由人為篩選，始可用於電腦輔助英文教學。

關鍵字：電腦輔助英文教學，自動化產生試題，大學入學考試，selectional preference

1. 緒論

英文是非常普遍的語言，它已成為不同母語背景的族群間，在溝通與資訊傳遞上的重要載具。為了加快國際化的腳步，近年來政府的教育政策，有明顯加強在學學生英文程度的趨勢，包括國小高年級排入英文課程、公佈國中小學英文基本一千字字彙、建議各大學加強大一英文教學等等。

為了幫助學生在課餘時間自我學習，各類型的輔助教學工具也不斷推陳出新，其中，電腦輔助教學系統（Computer Assisted Tutoring System）可協助教師繁冗的教材準備與出題工作，並且能批改學生的作答情形、提供成績列表與回授訊息給教師等諸多便利性，因而受到廣泛的重視。自動化產生試題（Automatic Item Generation）是電腦輔助教學系統中相當重要的一個環節，它透過既定的演算法，有系統地產生新的試題。自動化產生試題的優點在於能節省人力，提高試題的保密性和多樣性，同時能支援適性化教學系統（Adaptive Tutoring System），階段性地提供不同難度層次的試題，幫助受試者漸進式地成長 [1]。然而，自動化產生試題不容易確保試題的合

理性，機器產生的試題可能有語意不完整，難以作答的問題。因此，在兼顧試題的合理性、節省人力與試題難度的靈活性等考量下，利用電腦輔助的機制以產生試題確實有其必要性。

1. All the flights to and from Kaohsiung were _____ because of the heavy thunderstorm.
(A)advised (B)disclosed (C)cancelled (D)benefited

圖 1. 英文字彙題範例

我們發展的系統選定英文字彙題作為出題目標。圖 1 是九十二學年度學科能力測驗英文科的一道字彙題。字彙題是以單一選擇題的型式呈現，四個選項中，只有一個選項是正確答案，稱之為「待測字彙」，其餘三個選項則稱為「誘答選項」。

系統的試題有兩個主要的來源。其一是直接取材自出版刊物內的文章，經斷句等程序後，產生可做為試題的句子；其二是在既有的文法或規則下產生句子並藉由自然語言處理的技術檢查篩選可用的句子。前者的優點在於句子取得簡單，省略繁冗的前處理，但也造成句子的變化性受限於來源刊物；後者的句子變化性雖高，但需加入許多諸如語法和語意的檢查，包括主詞動詞的一致性（subject-verb agreement）、selectional preference（you **drive** a car, but **fly** an airplane）與定冠詞和不定冠詞的使用（a student...later, the student） [2]。由於我們的試題是針對高中生而設計，在句子的變化性上要求不高，因此採用取材自出版刊物內的句子。而在產生誘答選項時，仍需仰賴語法上的檢查，對誘答選項作消極的限制，以避免產生有多重答案或無答案可選的字彙題。高 [4]所發展的 AWETS (Automatic Web-based English Testing System) 可半自動化產生英文字彙題，句子的來源同樣是取材自出版刊物，但並未對誘答選項如何產生有詳盡的著墨。因此本研究主要延續這工作，加強產生誘答選項的正確性，並將試題的難度導向大學入學考試，鎖定高中生為受試對象。

我們搜集網路上的文件以建立語料庫，並利用現有的自然語言處理工具，為語料庫中的字彙標記其詞性、原始型態、與句中其他字彙的依附關係（如某名詞是某動詞的受詞、某副詞用於修飾某動詞）等註解，建立標記化語料庫（Tagged Corpus）。這個標記化語料庫亦是字彙題的題庫來源。此外，由這個自建的標記化語料庫中統計

字彙的頻率（以下簡稱為「詞頻」）與字彙的 selectional preference [7]等數據，這些數據是我們從題庫中篩選試題與產生誘答選項的重要指標。為了使產生的試題能輔助高中生自我評量，我們搜集了近年來日、夜間部大學入學考試與各校模擬考英文科的字彙題，藉由統計試題的詞數、詞頻與誘答選項的組合等資訊，來篩選語料庫中的句子，以期望系統所產生的試題在難度上與大學入學考試近似。

這篇論文的組織架構如下：第二節詳細描述了建立標記化語料庫的步驟；第三節從過去的訓練資料歸納數個大學入學考試英文科字彙題的特徵。第四節是由歸納所得的特徵，從標記化語料庫中產生字彙題的方法。對於所產生的試題在第五節做概要的探討，並於第六節提出總結與未來研究方向。

2. 建立標記化語料庫

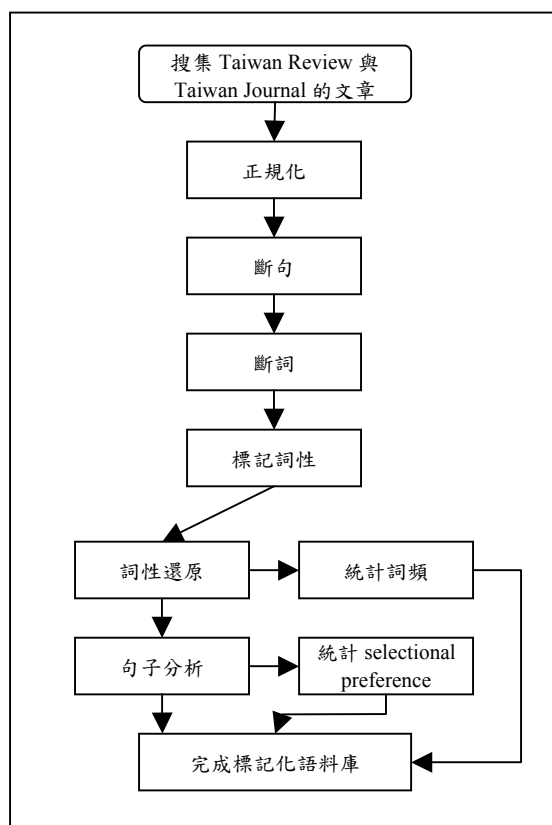


圖 2. 建立標記化語料庫的流程圖

為了能產生難度與大學入學考試相近的英文字彙題，我們從字彙的層面與句型的層面兩方面著手。在字彙的層面上，我們需要得知字彙的詞性、字彙的出現頻率；而在句型的層面上，需要得知句型的結構與 selectional preference 等資訊，因此我們建立標記化語料庫，亦即在語料庫中為每個字彙加入註解來記錄這些資訊，此標記化的語料也同時也是字彙題的來源。語料庫的來源有兩處，行政院新聞局出版的 Taiwan Review

(<http://publish.gio.gov.tw/FCR/fcr.html>)與 Taiwan Journal (<http://taiwanjournal.nat.gov.tw/>) 兩本期刊，從 2002 年一月起，至 2003 年八月迄的部份文章。這些原始的語料經由圖 1 所列的流程建立可用於產生試題的標記化語料庫，以下詳述各步驟的細節。

1. 正規化 (formalization)：從 Taiwan Review 與 Taiwan Journal 取得的網路版文件包含有標題、子標題、圖片、圖片的說明文與內文五個部份，正規化的工作將前四者刪除，僅保留內文的部份。

2. 斷句 (sentence segmentation)：正規化後的文件是以一行儲存一段文字的形式存在，為了便於統計詞頻與句子分析 (sentence parsing)，必須將段落中的各別的句子分割開來。Reynar [10] 等人利用 maximum entropy 的資訊來偵測一個句子的結尾，實驗於 Brown 與 WSJ 等著名的語料庫有 97.5%到 98.0%的正確率，因此我們直接使用作者提供的工具來完成斷句的工作。

3. 斷詞 (tokenization)：斷詞與斷句同樣是統計詞頻與句子分析的前處理。英文的詞與詞之間大都是以空白作為間隔，斷詞的工作相當容易，例外發生於兩個狀況：第一是大部份的文件格式中，標點符號會緊接在詞之後，無空白相隔，在我們的語料庫也不例外；其次是縮寫詞的情形，如常見的“I”與“am”縮寫成“I’m”，“is”與“not”縮寫成“isn’t”。因此斷詞的工作是將標點符號與詞分離，並將縮寫詞拆開。例如句子“The stock didn't rise \$5.”經斷詞後成為“ The stock did n't rise \$ 5.”。名詞的所有格與阿拉伯數字並不在我們斷詞的考慮範圍內，譬如人名“Kevin”的所有格“Kevin's”與阿拉伯數字“1500”，皆是當成一個詞處理。

4. 標記詞性 (part-of-speech tagging)：將語料標記詞性著眼點很多，當統計詞頻時可依不同詞性分別統計；另一方面，則有助於提高句子分析的正確性；而在產生誘答選項時，也提供一項重要的資訊。我們利用 Ratnaparkhi [8] 所開發的 Maximum Entropy Part-Of-Speech Tagger 作為標記詞性的工具，測試語料在斷句與斷詞無誤的情形下，標記詞性的正確率約可達到 96.5%。

5. 詞形還原 (lemmatization)：詞形還原的目的在於消除大小寫 (Book 與 book)、不同時態動詞 (go 與 went)、不同數量級名詞 (desk 與 desks) 與不同比較級形容詞 (big 與 bigger) 之間的差異。我們利用上一步驟所得到詞的詞性來做詞形還原，當“classified”一詞被標記為動詞時 (中譯為被分類)，需要將其還原成“classify”，若當其被標記為形容詞 (中譯為分類的、類別的) 或名詞 (中譯為分類廣告)，則無需做詞形還原的動作。經由詞性還原，始可正確地統計詞頻。

6. 統計詞頻：由於我們假定某一詞在語料庫中出現頻率的高低，可作為其難度的參考依據之一

一出現頻率越低，難度越高。此外，在產生誘答選項時也將依賴詞頻的資訊。因此我們對語料庫中出現過的詞皆統計其詞頻。分兩個方向作統計，一是不考慮詞的詞性，二是考慮詞的詞性。

7. 句子分析：透過句子分析，可以將一個句子解析成樹狀結構，以方便我們了解句中詞與詞、詞與子句的依附關係，如：某片語所修飾的詞為何、某動詞的主詞與受詞各是何詞等資訊，這是統計 selectional preference 的必要前置工作。我們使用 Lin [6] 撰寫的 Minipar 工具來對語料庫中的句子做分析。圖 2 是句子“A polite person never interrupt others while they are discussing important matters.”經由 Minipar 解析後所產生的輸出。Minipar 可分析出“person”一詞是“interrupt”的主詞、“others”一詞是“interrupt”的受詞、“polite”用於形容名詞“person”與“while”子句是用於修飾“interrupt”一詞等資訊，有利於我們統計 selectional preference。

```
> (
E1 (( Fin C * )
1 (A ~ Det 3 det (gov person))
2 (polite ~ A 3 mod (gov person))
3 (person ~ N 5 s (gov interrupt))
4 (never ~ A 5 anod (gov interrupt))
5 (interrupts interrupt U E1 i (gov fin))
E3 (( person N 5 subj (gov interrupt) (antecedent 3))
6 (others ~ N 5 obj (gov interrupt))
7 (while ~ SentAdjunct 5 mod (gov interrupt))
E8 (( Fin C 7 compl (gov while))
8 (they ~ N 10 s (gov discuss))
9 (are be be 10 be (gov discuss))
10 (discussing discuss U E8 i (gov fin))
E4 (( they N 10 subj (gov discuss) (antecedent 8))
11 (important ~ A 12 mod (gov matter))
12 (matters matter N 10 obj (gov discuss))
)
```

圖 3. Minipar 輸出範例

8. 統計 selectional preference：英文的動詞對於其主詞與受詞的選擇通常有所偏好，例如“assassinate”的主詞和受詞通常是人，而“eat”的主詞傾向於動物、受詞傾向是食物。不止是動詞，形容詞對其所形容的名詞與副詞對其所修飾的動詞的皆有所偏好，這就是所謂的 selectional preference [7]。藉由統計 selectional preference 的資訊，在產生誘答選項時，可以消極地排出不合語意的字彙。我們以圖 3 中的句子說明統計的方法。由 Minipar 分析句子後得知“person”一詞是“interrupts”的主詞，而在第 5 步驟中已知“person”的原型仍是“person”，於是查詢 HowNet [3] 對“person”這個詞所定義的類別—human，並記錄之，也就是以詞對類別（word to class）的方式統計 selectional preference。形容詞與副詞的 selectional preference 也是採取相同的方法統計。表 1 列出三個動詞“eat”、“see”與“find”對其受詞的 selection preference 所屬的類別的比例，受詞的類別僅列出較具代表性的 human（人）、food（食品）與 amount（多少），其餘的類別歸入 others（其他）。

表 1. Selection preference 統計資料（部份）

	human	food	amount	others
eat	1.2%	54.6%	0%	44.2%
see	26.9%	2.5%	55.3%	15.3%
find	18.1%	1.9%	39.3%	40.7%

目前我們的標記化語料庫中共有 107244 個句子，包含 2573993 個詞次（word token）與 35090 個詞型（word type）。

3. 歸納大學入學考試英文科字彙題特徵

為了產生難度上和大學入學考試英文科字彙題相似的題目，我們將近十二年日、夜間部大學入學考試與各校模擬考英文科的字彙題做為訓練資料，歸納其特徵，以做為系統出題時參考依據或限制條件。

訓練資料共有兩百六十五道字彙題，包括八十一學年度至八十九學年度的大學聯考英文科的字彙題共九十題，九十學年度至九十二學年度的學科能力測驗英文科的字彙題共三十題，八十一學年度至八十五學年度的夜間部大學聯考試英文科的字彙題共五十題，與各校模擬考題共九十五題 [13,14]。經過簡單的統計與分析，歸納出下列幾項共通性。

1. 試題的詞數少且變化不大。將字彙題經斷詞並統計詞數後可發現，一道字彙題所包含的詞數介於 7 個詞到 28 個詞之間，平均 14.01 個詞，標準差僅有 3.77 個詞。這項特徵被我們用來限制產生的試題在詞數上的範圍。
2. 試題中的詞除了待測字彙外多屬常用詞。將試題中的待測字彙與人名、地名等專有名詞排除後，經由統計可得的詞頻分佈表如表 2。我們定義詞頻排名第一的詞是語料庫中出現頻率最高的詞。由此數據可知有高達 89.6% 的詞在我們所搜集的語料庫中屬於出現頻率最高的兩千詞範圍內。在產生試題時，系統將避免試題中含有過多的低頻詞。

表 2. 訓練資料的詞頻分佈表

頻率範圍	詞數	比例	頻率範圍	詞數	比例
1-2K	3094	89.6%	6K-8K	40	1.16%
2K-4K	165	4.78%	8K+	69	2.00%
4K-6K	85	2.46%			

3. 相較於第 2 點，待測字彙的詞頻並沒有集中於前兩千個高頻詞的傾向，但我們發現這兩百六十五個待測字彙中，僅有三個屬於教育部所公佈的中小學基本千字表 [15]。套用於從我們自建的語料中所統計的詞頻排名表，也同樣除了這三個字彙外，其餘的詞的詞頻排名都在一千以外。換句話說，待測字彙的詞頻雖然分佈很廣，但普遍不落在前一千個高頻詞的範圍內。

4. 待測字彙的詞性固定。我們發現待測字彙的詞性不外乎四種：動詞（包括過去分詞、現在分詞等時態）、名詞（包括動名詞）、形容詞與副詞。

5. 誘答選項的詞性與待測字彙一致。兩百六十五題中有兩百六十二題的誘答選項與待測字彙相同。因此我們在產生待測字彙是，將限定其詞性與待測字彙相同。

6. 誘答選項與待測字彙的詞頻排名差距不大。針對兩百六十二題誘答選項與待測字彙有相同詞性的題目，計算誘答選項與待測字彙的詞頻排名差距的平均值與標準差可得表 3。注意不同詞性的詞是分開計算詞頻排名的。當誘答選項與待測字彙是名詞時，誘答選項與待測字彙的平均詞頻排名差距會比副詞時來得高，這是由於我們的語料庫中有一萬三千多種不同的名詞，而副詞則只有一千三百餘種。

表 3. 誘答選項與待測字彙的平均詞頻差距

誘答選項與待測字彙的詞性	動詞	名詞	形容詞	副詞
題數	93	79	53	37
詞頻排名差距平均值	679.6	2228.3	1659.1	206.5
詞頻排名差距標準差	397.2	1320.1	755.6	134.3

7. 待測字彙為何有規則可循。由第 3 與第 4 點可知，歷屆大學入學考試英文科字彙題的待測字彙在詞性、詞頻上皆有一定的限制。此外，為了在試題裡提供足夠的資訊使得四個選項中僅有一個為正確答案的條件成立，句子中的其他字彙或子句傾向與待測字彙有相關性。以 85 年大學聯考英文科的一題字彙題為例(圖 4)，“that he does not accept others' opinions”子句提供了充份的資訊說明 John 是一個不採納他人建言的人，可得知待測字彙是選項 B 的 “intimate” 一詞。

4. John is so _____ that he does not accept others' opinions.
 (A)delicate (B)intimate (C)obstinate (D)considerate

圖 4. 英文字彙題

我們試探性地使用不同的法則判斷句子中那一個詞是待測字彙，發現以下的法則可達到不錯的正確率。共分成兩個階段：第一階段從詞的層面上考量。由詞的詞性與詞頻排除句中極不可能是待測字彙的詞，這些詞包括：(1)詞性不屬於動詞、名詞、形容詞與副詞四者之一、(2)專有名詞、(3)屬於教育部所公佈的中小學基本千字表、(4)屬於我們搜集的語料庫中最高頻的一千詞。剩餘的詞稱為「可能待測字彙」。

第二階段則著眼於分析句子的結構。當句中存在一詞或一子句為某可能待測字彙的修飾詞或修飾子句時，該可能待測字彙有較高的機率是待測字彙，因此我們對可能待測字彙作排序，順位第一者就是我們認為最有可能是待測字彙的詞，排序的方法如下。

- 7.1. 若句中存在一詞為某可能待測字彙的修飾字或與之有依附關係，則該可能待測字彙的分數加 1。
- 7.2. 若存在一子句為某可能待測字彙的修飾子句，則該可能待測字彙的分數加 2。
- 7.3. 順位的決定即依據可能待測字彙的分數，若分數相同，則以排在句中較前的可能待測字彙的順位較高。

以 “He was very shy, so his smile was barely noticeable when he met his teacher.” 一試題為例，經第一階段的篩選後，只剩下 “barely” 與 “noticeable” 兩字為可能待測字彙，而 “noticeable” 的順位比 “barely” 高，這是由於句中存在 “when he met his teacher” 子句為 “noticeable” 的修飾子句，又存在 “barely” 一字修飾 “noticeable”，因此 “noticeable” 得到 3 分。而 “barely” 只得到一分，因為句中僅存在 “noticeable” 一字與之有依附關係。

將 265 筆訓練資料經由上述兩個階段作測試，可得圖 5 的結果。其中有兩筆訓練資料在第一階段已將待測字彙排除。若不計這兩筆，有 141 筆訓練資料的待測字彙被排在第一順位，也就是 53.2% 的正確率，平均而言，待測字彙的排名是第 1.70 名。因此，在產生字彙題時，我們只認定順位最高的前兩個字彙可做為待測字彙。

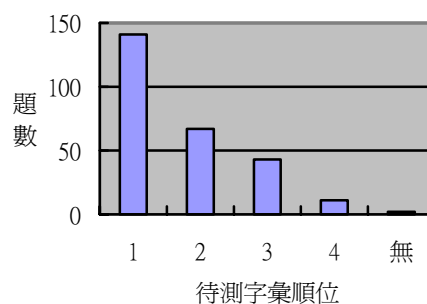


圖 5. 待測字彙順位分佈表

上述七項特徵，是我們系統目前產生英文字彙題的過程中重要的參考依據。

4. 產生英文字彙題

系統可依兩種不同的輸入條件產生英文字彙題。第一種是由使用者指定待測字彙、待測字彙的詞性與出題的數量，可經由以下四個步驟完成出題：

1. 從標記化語料庫中搜尋含有待測字彙的句子。句子的長度必需介於 7 到 28 個詞之間（這符合我們從訓練資料所統計的結果—試題的詞數介於 7 到 28 個詞之間）。除了待測字彙外，句子中不能有任何一個詞的詞頻排名在語料庫中排在一萬五千名之後，也不能同時存在兩個詞的詞頻排名在一萬名之後（這亦符合訓練資料中除待測字彙外多屬常用詞的特性）。標記化語料庫中 107244 個句子，經過此步驟篩選後，適合作為字彙題的僅有 37617 個句子，為了減少產生字彙題時從標記化語料庫中搜尋句子的動作，將這 37617 個句子另外儲存，與原標記化語料庫分隔開來，做為加快此步驟的快取之用。
2. 利用詞性標記、字彙頻率的統計資料，隨機挑選與待測字彙同詞性且詞頻排名相近的候選誘答選項—這些候選誘答選項需經由下一步驟的檢驗始可成為誘答選項。若待測字彙是動詞，候選

誘答選項挑選的範圍是與待測字彙的詞頻排名相差正負七百名之內，若待測字彙是名詞，則範圍限制在正負兩千名之內，形容詞與副詞分別是正負一千七百名與正負兩百名之間。這些數值的都是歸納自歷屆大學入學考試英文科字彙題的特徵而決定，與表 3 的統計結果相仿。

3. 檢驗候選誘答選項的 selectional preference，是否與待測字彙有相似的傾向，若無，則表示該候選誘答選項取代待測字彙後會使得原試題不合理，可將其列為誘答選項。這裡定義候選誘答選項的 selectional preference 的表現是否和待測字彙相似的條件略為嚴格，以 “If you want to become a good tennis player, you have to **sharpen** your skill.” 一試題為例，其中待測字彙是 “sharpen”，由 Minipar 分析後得知 “sharpen” 的受詞是 “skill” 一字，而 “skill” 在 HowNet 中的分類是 knowledge (知識)，於是我們依第 2 步的方法隨機產生一候選誘答選項 “enhance”，從 selectional preference 統計資料中發現 “enhance” 的受詞的定義中亦有 knowledge (不論其比例高低)，因此我們排除 enhance 成為誘答選項。重複 2 與 3 兩個步驟，產生三個誘答選項。

4. 重複 1 到 3 三個步驟，直到產生的題目數量與輸入條件相符合。

圖 6 與圖 7 是系統輸出的兩個範例。

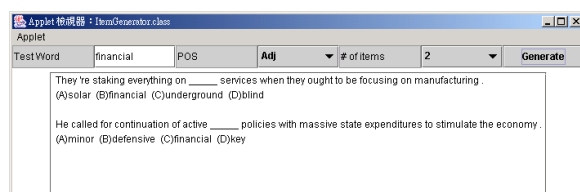


圖 6. 待測字彙 financial，系統的輸出情形

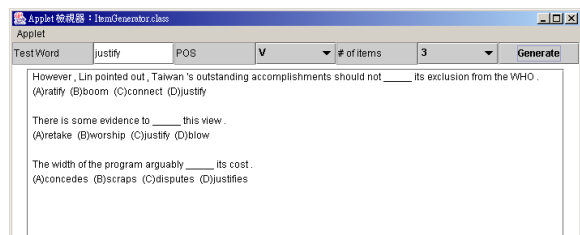


圖 7. 待測字彙 justify，系統的輸出情形

第二種輸入條件，是當使用者沒有指定待測字彙，僅決定出題數量，則系統會隨機決定待測字彙的詞性（動詞、名詞、形容詞與副詞四種詞性的機率均等），並依第三節的第七點，搜尋標記化語料庫的句子中，含有可能是待測字彙的詞（即順位第一與第二位）其詞性符合之前隨機決定的詞性，將該詞設為待測字彙。有了待測字彙與待測字彙的詞性，經由上述的步驟 2 到步驟 4 即可產生字彙題。

圖 8 是未指定待測字彙的情形下，系統輸出的範例。

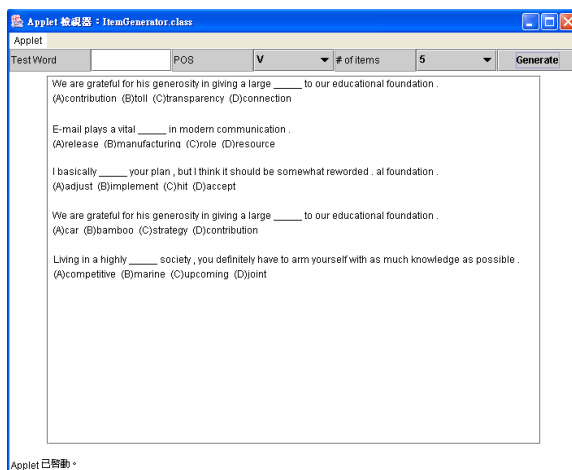


圖 8. 未指定待測字彙時，系統輸出情形

5. 出題結果探討

系統的出題目標除了希望試題合理之外，在難度上也希望能與大學入學考試近似。由於目前尚未由受試者或專家學者們實際測試難度上與大學入學考試的近似程度，因此以下僅就試題的合理性做概要的探討。

在建立標記化語料庫的過程中，斷句、斷詞、標記詞性與詞形還原等環環相扣的步驟皆無法達到百分之百的正確率，其後的統計詞頻與統計 selectional preference 也連帶受到影響，因此系統所產生的試題，仍有少數會出現以下兩種類型的錯誤。一是因斷句不正確，造成試題的句子不完整；雖然這是一個問題，不過作為一個出題輔助系統，這樣的訛誤並不會直接衝擊到試題的品質。另一個問題是統計 selectional preference 時的錯誤，可能會造成有多重答案相容的情形。除此之外，selectional preference 在確保單一選擇題僅有一個正確答案的效果上，是另人滿意的。

然而，只使用 selectional preference 的資訊來產生誘答選項，缺點是誘答選項的產生便受制於 selectional preference。以 “The couple had **financial** problems in the first year of their marriage. They even couldn't afford a meal at a fancy restaurant.” 為例，在待測字彙是 “financial” 的情形下，若是能考慮整段文字的語意，會發現句中這對夫婦面臨的是經濟上的難題，因此 “security”、“emotional” 等字都適合作為誘答選項。倘若只依賴 selectional preference 產生選項，則完全依 “problems” 一詞來限制能形容它的形容詞，上述的字彙都不能用，排除許多可能性。

6. 結論與未來展望

在這篇論文中，我們呈現了一個半自動化產生英文字彙題的系統，可在使用者指定待測字彙或僅指定出題數量下，產生難度與大學入學考試英文科字彙題相仿的試題。誘答選項的產生是仰賴 selectional preference 的統計資訊，以避免造成

單一選擇題卻有多重答案的情形。在難度上是依據歷屆考題歸納試題中的詞與句型等特徵，對所產生的試題作限制。由於系統產生的試題尚未有完整的語義檢查機制，因此尚需經由人為篩選。

尋找英文語料中 concordance 或是 collocation [7] 相關資訊的技術雖然可以用於包含產生字彙題目等英文字彙的教學活動，但是單純的 concordance 和 collocation 訊息尚且沒有提供足夠的語意層次的訊息，沒有足夠的語意訊息，字彙题目的產生就不容易作的盡善盡美。我們目前朝向以語意為基礎的輔助出題系統，企圖藉由語意訊息的輔助來改進系統所產生的题目的適性度和多樣性。

未來系統改進的工作將朝四個方向努力。首先，句子的分析將深入語意的層面，檢驗試題的合理性，減少人為的篩選工作；其次，考慮英文字彙中一詞多義的情形，使用者不僅可指定待測字彙，亦可針對該字彙的特定詞義，由系統產生試題；再者，研究其他題型的自動出題的可能性，包括片語（包括 concordances 和 collocations）、會話與克漏字測驗；最後，我們計劃結合學生塑模（student modeling），由學生塑模接收受試者的答題狀況，評估學習進度，並提供補足性的教材加強其尚未完全習得的部份，幫助受試者自我成長，以達到線上適性化教學系統的目的。

致 謝

本研究承蒙國科會研究計畫 NSC-92-2213-E-004-004 之部份補助。

參考文獻

1. I.I. Bejar and P. Yocom. 1991. A Generative Approach to the Modeling of Isomorphic Hidden Figure Items, *Applied Psychological Measurement*, 15(2), pp. 129-137.
2. P. Deane. 2003. Automatic Item Generation via Frame Semantics: Natural Language Generation of Math Word Problems. <http://www.ets.org/research/dload/ncme03-deane.pdf>
3. Z. Dong and Q. Dong. 2000. HowNet - Chinese Message Structure Base. <http://www.keenage.com/>
4. Z.-M. Gao. 2002. An Automatic Web-Based Computer-Adaptive Vocabulary Testing System, *Proceedings of the Conference and Workshop on TEFL & Applied Linguistics*.
5. Z.-M. Gao and C.-L. Liu. 2003. A Web-Based Assessment and Profiling System for College English, *Proceedings of the 11th International Conference on Computer Assisted Instruction*.
6. D. Lin. 1998. MINIPAR - A Principle-Based Broad Coverage Parser. <http://www.cs.ualberta.ca/~lindek/minipar.htm>
7. C. D. Manning and H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*, The MIT Press.
8. A. Ratnaparkhi. 1996. A Maximum Entropy Part-Of-Speech Tagger. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing Conference*, University of Pennsylvania, pp. 133-142.
9. P. Resnik. 1997. Selectional Preference and Sense Disambiguation. In *Proceedings of the Applied Natural Language Processing Workshop on Tagging Text with Lexical Semantics: Why, What and How?*, pp. 52-57.
10. J. C. Reynar and A. Ratnaparkhi. 1997. A Maximum Entropy Approach to Identifying Sentence Boundaries, In *Proceedings of the 5th Conference on Applied Natural Language Processing*, pp. 16-19.
11. I. I. Bejar, R. R. Lawless, M. E. Morley, M. E. Wagner, R. E. Bennett and J. Revuelta. 2002. A Feasibility Study of On-the-Fly Item Generation in Adaptive Testing, *GRE Board Professional Report No. 98-12P*.
12. B. Martin and A. Mitrovic. 2002. Automatic Problem Generation in Constraint-Based Tutors. In *Proceedings of the 6th Conference on Intelligent Tutoring System*, pp. 388-398.
13. 蔡方。2002。近十年大學聯考試題英文試題暨詳解。建興文化事業有限公司出版。
14. 大學入學考試中心網頁 <http://www.ceec.edu.tw/>
15. 教育部國教司國民教育行政資訊傳播網 <http://www.eje.ntnu.edu.tw/>