

## 第四章

### 實驗分析討論

本章將依據第三章 SAW 系統進行實驗和分析。4.1 節呈現實驗的測試資料，分為實驗語料庫、實驗比對系統、實驗案例三個部分；4.2 節將呈現實驗的評估方法；4.3 節將呈現實驗指標；4.4 節為試題案例評估的結果；4.5 節為問卷評估的結果；4.6 節為學生作文範例，指實際觀察學生的寫作範例，模擬學生使用系統的成效；4.7 節為小結。

#### 4.1 實驗測試資料

此小節將描述實驗使用的語料庫，並論及實驗比對的系統，最後說明實驗測試的案例，詳述如下：

##### 4.1.1 實驗語料庫

在測試的語料庫中，本研究採取 BNC 語料庫和科學人雜誌語料庫。其中 BNC 語料庫共選取約 350 萬筆英文例句，文章年代從 1974-1994，內容分類為自然科學、社會科學、應用科學、哲學、藝術、科幻、金融、休閒、世界九大類，語料庫有內定的 POS 分類，共計 62 個(詳見附錄一)。科學人雜誌(以下簡稱 SA)共選取約 11 萬筆英文例句，文章年代從 2002-2006，內容分為數學、物理、化學、天文、生命科學、醫學、地球科學、生態學、資訊科學、電機、工程、國防軍事、心理學、語言學、人類學、經濟學、科學史、科學哲學、人文科學、其他二十類。實驗依據 MX-POST[11]對 SA 作 POS 標注，MX-POST 的 POS 共計 36 個(詳見

附錄二)。而 SAW 系統僅實做 6 個 POS，即對於 BNC、SA 的 POS 有統籌上的對應，詳盡的統籌方式詳見附錄三，SAW 系統使用的語料庫比較表格如下：

表 4.1 實驗語料庫

	BNC	SA
例句數目	350 萬筆	11 萬筆
文章內容	自然科學、社會科學、休閒、世界等九大類主題	數學、物理、人文科學、其他等二十類主題
年代	1974-1994	2002-2006
POS	62 種	36 種

#### 4.1.2 實驗比對系統

目前市面上有部分利用語料庫協助英語學習的系統，如前述所提及的 VIEW 系統[6][19]。另外，Google 系統有強大的搜尋功能和龐大的網頁資料，亦可當成英語例句的檢索系統，實驗將採 VIEW 系統和 Google 系統與 SAW 系統比較。而 VIEW 使用的語料庫為 BNC 語料庫，例句格式為 concordance，輸入的查詢格式為準確字詞(exact)、萬用字元(wildcard)、單字開頭/結尾(prefix/suffix)和詞性標籤(POS)，然其選取的方針僅只有完整比對，亦沒有排序上的觀念。而 Google 的資料庫為線上網頁，例句格式為 snippet 和 cache，其中 snippet 為網頁的部分內容，cache 為網頁的庫存內容。Google 可輸入的查詢格式為準確字詞(exact)、萬用字元(wildcard)，選取的方針為完整比對和部分比對，排序上以網頁內容主題之相關性為排序機制。此兩系統並非針對英語寫作的協助而設計，故其例句格式均不是針對英語寫作的參照需求。然而 SAW、VIEW、Google 共通點為須依據

查詢條件從語料庫(或線上網頁)選取部分的例句(或網頁)，本研究將比較三者選取特定片語的數量，並依據選取例句的字數予以客觀的句型成本評估。以下為三個系統的比較表格：

表 4.2 實驗比對系統

	SAW	VIEW	Google
corpus(or homepage)	BNC、SA	BNC	線上網頁
例句格式	sentences	concordance	snippet/cache
輸入機制	exact prefix/suffix wildcard POS subsequence	exact prefix/suffix wildcard POS	exact wildcard
選取	完整/部分	完整	完整/部分
排序	MSA	無	以網頁內容主題 之相關性為主

### 4.1.3 實驗案例

本研究為了評估 SAW 系統對英語寫作協助的功效，將從試題中選取重要的英語用法，模擬學習於寫作上可能遭遇的困難。試題來源包括大學聯招和大學學測 83~94 年的考題，共計 16 期，題型分為詞彙、克漏字（文法、文意）、閱讀測驗、翻譯、作文五大類，實驗將從中選取詞彙、翻譯題型來模擬寫作之語言知識需求。本研究同時採取 TOFEL 89~91 的考題，共計 11 期，題型分為詞彙、文法結構兩大類。同理，實驗將從此兩題型模擬學習者於寫作需要協助的情況，以下為兩種試題的比較表格：

表 4.3 實驗案例來源

	大學聯考/學測	TOFEL
期數	83-94(16 期)	89-91(11 期)
實驗題型	詞彙、翻譯	詞彙、文法結構

## 4.2 實驗評估方法

為了客觀評估系統對於學習者的學習成效，實驗評估的方法分試題評估(研究客觀評估)和問卷評估(學習者主觀評估)，試題評估為從試題中選取該試題的重要關鍵字(如單字、片語、子句等)，模擬學習者在寫作過程中可能因關鍵字的不熟悉而產生相對應的困惑，藉此評量系統能否給予學習者相對應的協助。

問卷評估主要的目的為實際瞭解學習者於英語寫作上的認知，並評估本研究系統能否給予學習者適當的協助。在評估系統成效方面，本研究從實驗案例中選取部分試題由受測者模擬測試，並分為兩種施測方式分析系統的成效，第一種為評估受測者於使用系統協助寫作和未使用系統協助寫作上的差異，另一種為評估受測者於使用系統前和使用系統後的差異，藉此兩種施測方式能瞭解 SAW 系統的協助成效。而試題評估和問卷評估選取的試題描述如下：

在試題評估方面，本研究從實驗案例(TOFEL)中選取重要的觀念，其選取的方式除個人主觀的選擇外，並參照線上 TOFEL 的分析網[9]選擇，總共選取 45 個單字片語、12 個不限定子句(含有 subsequence 表達元素的句型)當試題樣本，詳見附錄五。實驗並從三個角度分析成效，第一個角度為量上的評估：比對 45 個單字片語於 SAW、VIEW 和 Google 搜尋例句(或網頁)上的數量差別。第二個角度為單字片語的評估：針對 18 個單字片語，以 SAW 系統不同的表達元素(exact、prefix、POS、wildcard)評估不同認知程度的學習者其於寫作協助上的成

效。第三個角度為子句和認知錯誤的評估：針對 12 個不限定子句分析 SAW 系統予以寫作上的協助，並模擬學習者在認知程度有部分錯誤的情況下，SAW 系統能給予學習者協助上的成效。

而問卷的試題由大學聯招和大學學測以及 TOFEL 選取，題型分為單選題 16 題，翻譯題 3 題，詳盡的問卷內容詳見附錄十、十一、十二、十三。然考量實際英語寫作協助中，學習者給予的查詢條件不見得完全相同，本研究針對問卷的試題，除翻譯題型外，均採取兩種不同的查詢條件，評估在不同查詢條件下，系統是否能有效率給予學習者協助，該部分將置於使用系統協助寫作和未使用系統協助寫作的區別問卷中探討。除了上述試題分數可予以判斷系統的成效指標，本研究斟酌推薦系統的評估方式[36]，將系統評估指標分為以下 5 項：

### 4.3 實驗指標

為客觀並且確實的評估系統的成效，本研究擬定 5 項分析指標，詳述如下：

(1)試題分數：本研究藉由試題模擬學習者於寫作上的協助，試題本有正確的選項，可藉由學習者的分數評估系統的成效，然試題分數僅是單純的描述學習者對於答題上的理解，以推薦系統而言，較重要的指標為學習者的實用度(Usefulness)和滿意度(User Satisfaction)。

(2)推薦度：推薦度是學習者對於系統提供的資訊，其所認為真正實用的程度(Recommend for User-Usefulness: The degree of actual usefulness to a learner)。本研究對於推薦度的評估方式是採取 4 種高低程度的指標，即高推薦度、中推薦度、低推薦度、無推薦度。推薦度是由學習者主觀評估的，本研究對於推薦度的指標將從問卷方式取得。

(3)滿意度：滿意度是學習者對於系統提供的資訊，其整體滿意的程度

(User-Satisfaction: The degree of satisfaction to a learner)。本研究對於滿意度的評估方式是採取 5 個高低程度的指標，即非常滿意、滿意、普通、不滿意、非常不滿意。滿意度亦是由學習者主觀評估的，本研究對於滿意度的指標將從問卷方式取得。

(4)符合度：符合度是依據學習者輸入的查詢條件，比對選取出的結果和學習者預期答案上的相似度(Match: The degree of fit between the actual answers and the retrieved answers based on the different query)。本研究對於符合度的評估方式為在一定選取數量的例句中，真正符合學習者預期答案的比值(Number of relevant retrieve / Number of retrieve)。符合度的精神在於以協助系統而言，我們無法預期學習者能輸入合適的查詢條件，但各種查詢條件和學習者本身預期的結果有其一定的關連度，高符合度代表學習者給定的查詢條件和其所預期的結果逼近，預期給予學習者的推薦度會較高；相反的，低符合度代表學習者給定的查詢條件和其所預期結果有所差距，預期給予學習者的推薦度會較低。而會造成低符合度的原因主要在於兩個因素：(a)模糊條件(vague query)：學習者給定的查詢條件不能確實針對預期結果。例如搜尋 a proportion of 片語，學習者若給定的查詢條件為 a #(wildcard) of，預期結果除了 a proportion of 外，還有 a little of、a bit of 等推薦例句，造成系統的符合度下降；(b)錯誤條件(false query)：學習者給定的查詢條件有一定認知程度上的錯誤。例如搜尋 a proportion of 片語，學習者若給定的查詢條件為 a proportion on，預期結果會顯少 a proportion of 的例句。由此觀之，低符合度本身具有額外的推薦含意，藉由符合低學習者可以回顧其所輸入的查詢條件是否過鬆或者有錯誤，間接得到協助上的成效。

舉個例子來表達上述推薦度、滿意度、符合度上的意義，假設學習者想要搜尋有關 make up 相關的片語例句，依據某種查詢條件查詢下有以下 4 個推薦例句：

S<sub>1</sub>. I would **make** a really nice dinner, and dress **up**.

S<sub>2</sub>. I had **made up** my mind.

S<sub>3</sub>. My friend had wanted specifically to talk to a specialist, so she could **make up** her own mind about surgery.

S<sub>4</sub>. She has to **make up** her own mind.

該學習者因 S<sub>1</sub> 不含有 make up 相關的句型，故判定 S<sub>1</sub> 推薦度和滿意度均是低的。而 S<sub>2</sub> 雖不含有 make up 的直接句型，但學習者可以根據 make 的過去分詞為 made 而自行將 S<sub>2</sub> 轉換成學習者預期的結果(S<sub>2</sub>'. I make up my mind.)，因而學習者認為 S<sub>2</sub> 推薦度和滿意度均是高。至於 S<sub>3</sub>，雖然學習者可直接得到 make up 相關的例句，但因整體的句型長度過長，故學習者認為 S<sub>3</sub> 推薦度普通但滿意度是低的。關於 S<sub>4</sub>，不但可直接得到 make up 相關的例句，例句的句型又很簡短完整，故學習者認為 S<sub>4</sub> 推薦度和滿意度均是高的。而符合度為學習者原先預期的答案(make up)於 S<sub>1</sub>、S<sub>2</sub>、S<sub>3</sub>、S<sub>4</sub> 出現的比率，即  $2/4=50\%(S_3S_4/ S_1S_2S_3S_4)$ 。符合度、推薦度、滿意度的關連圖示如下：

Input: V up	Expected answer: make up	符合度	推薦度	滿意度
S <sub>1</sub> . I would <b>make</b> a really nice dinner, and dress <b>up</b> .		No	Low	Low
S <sub>2</sub> . I had <b>made up</b> my mind.		No	High	High
S <sub>3</sub> . My friend had wanted specifically to talk to a specialist, so she could <b>make up</b> her own mind about surgery.		Yes	Medium	Low
S <sub>4</sub> . She has to <b>make up</b> her own mind.		Yes	High	High

圖 4.1 符合度、推薦度、滿意度的關連圖示

額外強調的是，上述推薦度和滿意度的結果是本研究為了陳述推薦度、滿意度的意義而假設的，實際上推薦度和滿意度是由學習者主觀評估的，真實結果和本研究的假設或許有一定程度上的差異。

本研究以試題案例模擬寫作協助，測試的試題均有明確的關鍵字(即學習者預期的答案是可以得知的)，故本研究可自行推論符合度。而如同前述所提及，符合度和推薦度因有一定的關連性，但因推薦度是由學習者主觀評估，其變數是多的，例如在高符合度時學習者可能因為不甚瞭解例句給予的資訊而認為推薦度是低的；同理，低符合度時因學習者有可能間接得到協助，故推薦度反而有可能高。但整體來說，符合度和推薦度有正相關的線性關係，亦即當符合度高時，推薦度相對高；當符合度低時，推薦度相對低。而由問卷調查結果顯示，上述假設是成立的(詳見 4.5 章節)，符合度和推薦度的相關係數 0.63，代表兩者有正相關的線性關係。而因符合度可以由系統自行推論，本研究的系統評估方式較為彈性，對於推薦系統中重要的推薦度評估指標可以約略由符合度來表示。

(5)句型成本：有鑑於例句格式若過於冗長則無法有效率地於寫作上協助，實驗擬定句型成本的觀念。而所謂句型成本為學習者對於閱讀完一定數量的結果所需花費的成本，實驗以所需閱讀的英語單字字數當成本考量，簡而言之，句型成本為在一定數量結果中學習者所閱覽的英語單字總次數(Cost: the number of vocabularies)。舉例而言，上述例句 S<sub>1</sub> 的英語字數為 10，而例句 S<sub>2</sub>、S<sub>3</sub>、S<sub>4</sub> 分別為 6、20、8，換而言之，在上述結果中，學習者所花費的句型成本為 10+6+20+8=44。而由問卷調查的回饋得知，學習者偏好簡潔有力的句型當寫作參照上的例句。

#### 4.4 試題案例評估

本章節將陳述上述試題案例關於量上的評估、單字片語的評估、子句和認知錯誤的評估予以實驗結果的呈現，詳述如下：

##### 4.4.1 量上的評估

量上的評估主要考量不同系統選取的例句數量，圖 4.2 將呈現 SAW 系統(使

用 BNC 語料庫)、VIEW 系統、Google 系統於前述所提及的 45 個單字片語上的搜尋數量，其選取的方針均為完整比對。要強調的是，選取數量上的優勢並不代表推薦系統的優勢，推薦系統講究的是質而不是量，換言之，能夠有效率的推薦學習者資訊優於給予學習者大量資訊而讓學習者自行篩選想要的資訊。SAW 系統基於此理念，對於語料庫予以前置的篩選，去除語料庫中含有比較不適宜寫作參照的句型(如 BNC 語料庫中有些文章使用"&"符號代表特殊的字元，如"é "在 BNC 語料庫中以"&eacute;"表示。然特殊字元並不常見於一般正統的英語句型中，SAW 系統予以刪除以提升參照例句的品質)，故相較同樣使用 BNC 語料庫的 VIEW 系統整體，除例句呈現的格式不一樣外，搜尋單字片語的數量略少於 VIEW。

而圖中的數量以 log 呈現，主要呈現前述的觀念，推薦系統強調的是質而不是量，數量上的差距僅在於搜尋數量少時才有顯著性的成果，例如搜尋例句為 1 句和 10 句的重要性，遠大於搜尋例句為 1000 句和 10000 的重要性，log 函數有上述的含意，故以 log 函數呈現量上的評估是恰當的。

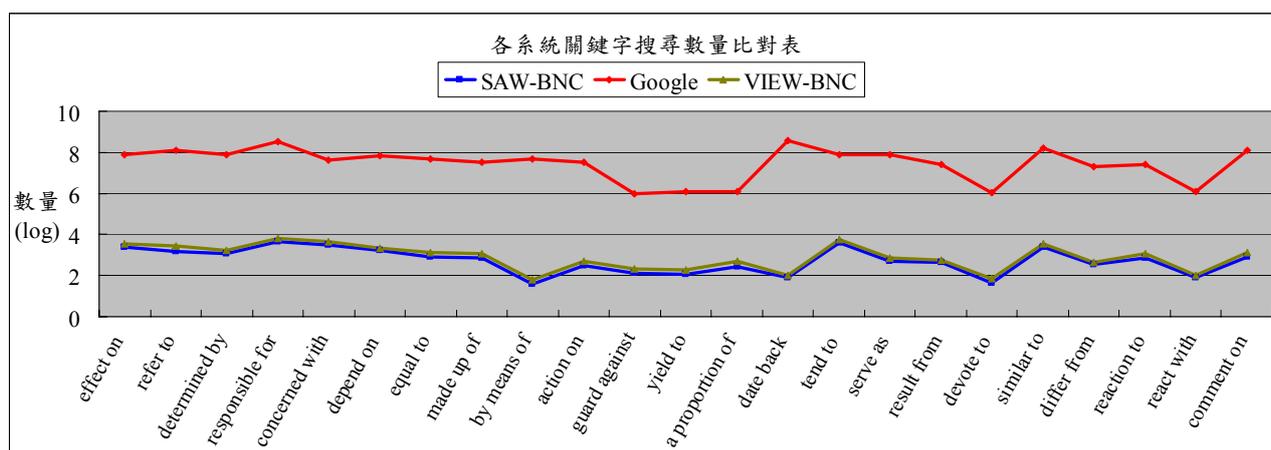


圖 4.2(a) SAW、Google、VIEW 系統於關鍵字的搜尋數量比較

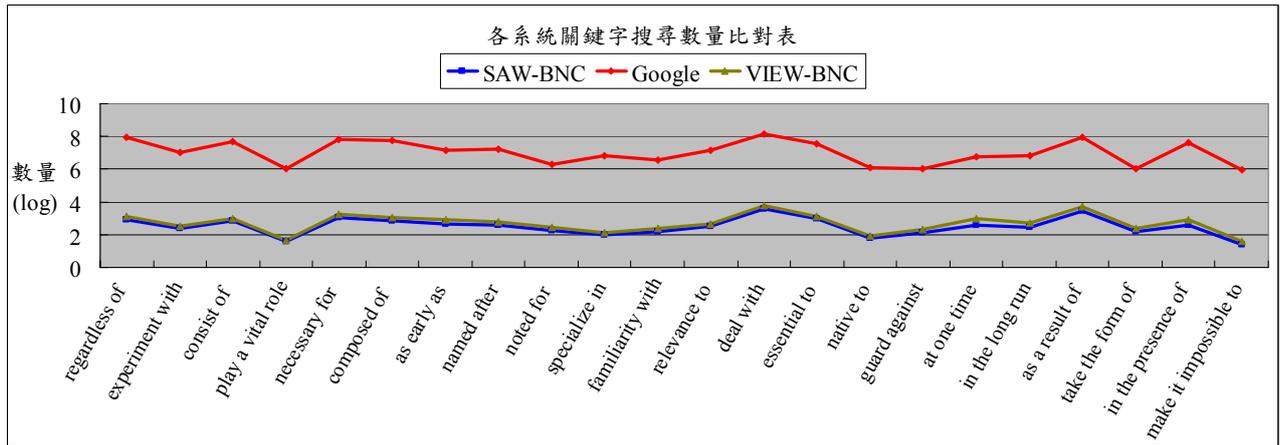


圖 4.2(b) SAW、Google、VIEW 系統於關鍵字的搜尋數量比較

除了 SAW-BNC 語料庫外，本研究亦實驗 SAW-BNC 和 SAW-SA 語料庫搜尋數量上的比較表格，數值同樣以 log 呈現，然 SA 語料庫的句型較少，故會有未搜尋到的結果(以 0 表示)其結果如圖 4.3 所示：

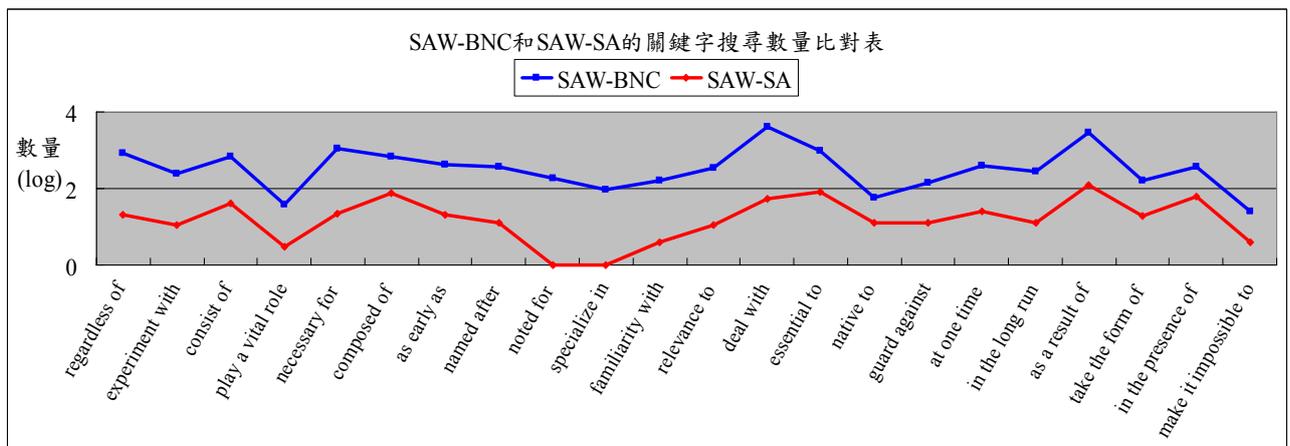
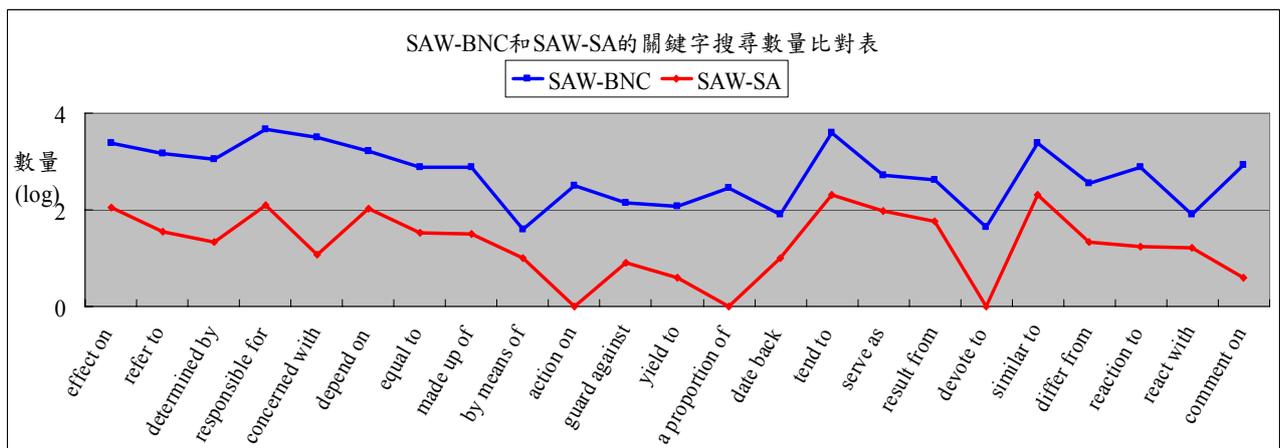


圖 4.3 SAW-BNC、SAW-SA 對於關鍵字的搜尋數量比較

單獨從量上的觀點，Google 搜尋的網頁數量較 SAW-BNC/VIEW-BNC 選取例句數量為多，而 SAW-BNC 選取例句的數量亦多於 SAW-SA。而值得注意的是，觀察圖 4.2 和圖 4.3，從 log 數量上的差距約略為定值，不難發現各單字片語於各個語料庫中出現的頻率約略為定值，亦即若 Google 系統中某單字片語出現的次數佔全體的 1/3，SAW-BNC、SAW-SA、VIEW-BNC 也約略佔全體的為 1/3，換而言之，在此 45 個單字片語中，各個語料數均是平衡的(balance)，亦即不會特別偏重單獨的單字片語。雖 45 個單字片語在整個英語單字片語中佔極少數，然此 45 個單字片語為客觀從 TOFEL 的 11 期試題中每期平均取出 4~5 個當測試樣本，其取捨上並非刻意取捨，如以此為假設，上述各語料庫於所有英語單字片語中，其架構上約為平衡的。

#### 4.4.2 單字片語的評估

單字片語的評估主要藉由試題模擬學習者於寫作上所需的協助，並假設不同符合度的查詢條件，評估對認知程度不同的學習者所能提供的協助成效。實驗將從上述 45 個單字片語挑出 10 個由兩個英語單字組成的片語、以及 8 個由多個英語單字組成的片語當測試樣本。實驗採用 exact、prefix、POS、wildcard 等表達元素當不同的查詢條件，模擬學習者可能得到的不同協助，對於實驗評量的指標以符合度和句型成本表示。

對於兩個英語單字組成的片語，實驗統籌規劃上述 4 種表達元素所產生的不同查詢條件，共計有 4 大類型的查詢模組：(1)exact match：將兩個英語單字完整輸入，如以 specialize in 為例，實驗將輸入 specialize in，並取其符合度、句型成本當實驗評估的指標。(2)prefix：將兩個英語單字其中一個英語單字的第一個英語字母當成 prefix 的引導原則，如以 specialize in 為例，將分為 s% in 和 specialize i% 兩種可能的查詢條件，實驗取此兩種查詢條件的平均符合度、平均句型成本當實驗評量的指標。(3)POS：將兩個英語單字其中一個英語單字以 POS 表達元

素表示查詢，如以 specialize in 為例，將分為 V in 和 specialize P 兩種可能的查詢條件，同理，取兩種不同查詢條件的平均符合度、平均句型成本當實驗評量的指標。(4) wildcard：將兩個英語單字其中一個英語單字以 wildcard 表達元素表示查詢，如以 specialize in，將分為 # in 和 specialize # 的兩種可能的查詢條件，取兩種不同查詢條件的平均符合度、平均句型成本當實驗評量的指標。詳見下列示意圖。

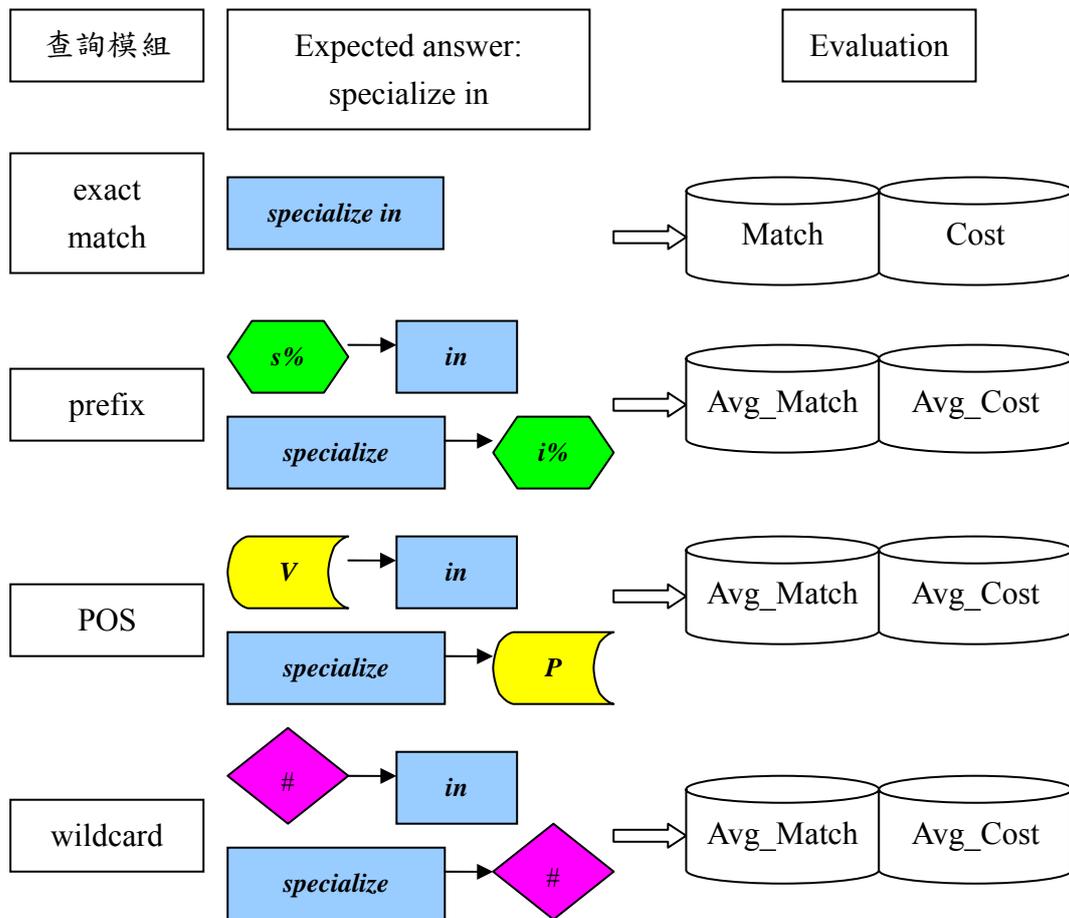


圖 4.4 specialize in 以 exact match、prefix、POS、wildcard 的查詢方式

VIEW 系統雖有上述 exact、prefix、POS、wildcard 表達元素，然其系統並無推薦排序的機制，而 Google 系統則無上述 prefix、POS 表達元素，故 VIEW 和 Google 系統並不列於符合度的比較表當中。VIEW、Google 雖無推薦排序的機制，然 VIEW 系統 concordance 格式中一個句型長度約為 20~25 個單字，Google

的 snippet 格式中一個句型約為 25~30 字個單字。故實驗對於 VIEW、Google 的句型成本以一個例句 22.5 和 25 的平均值來和 SAW 比對。

本研究設定提供 10 個例句，亦即須推薦出前 10 個例句供學習者參照，句型成本是以此 10 個例句的總單字字數當評量依據。圖 4.5 顯示 prefix 查詢模組於此 10 個片語中符合度高於 POS 模組，而 POS 查詢模組符合度高於 wildcard 查詢模組，然其中有一個例外為 native to 片語，POS 查詢模組符合度高於 Prefix 查詢模組符合度 10%(差 1 個例句)，此約略說明 prefix 表達元素為限制較強的查詢條件，而 POS 次之，wildcard 最低。

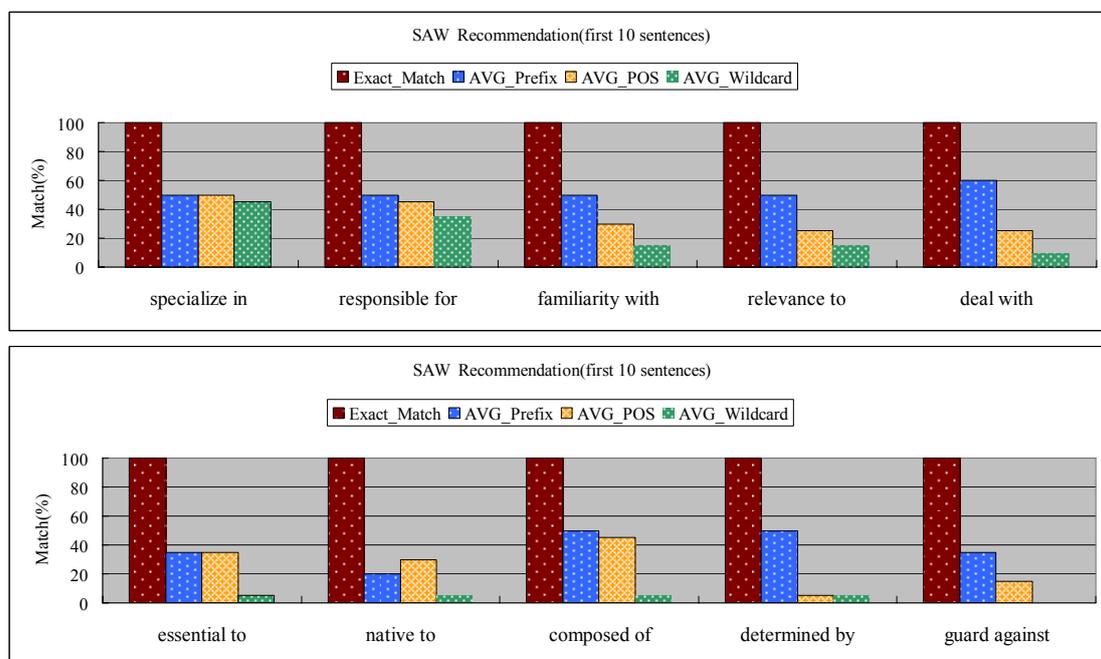


圖 4.5 兩個英語單字符符合度

圖 4.6 顯示，VIEW、Google 句型成本偏高，亦即學習者需花費大量的心力去閱讀結果。而 SAW 系統排序參數中有考量句型長短當推薦機制(詳見附錄五)，故普遍搜尋出的結果句型較簡潔，有利於學習者參照的需求。其中較值得關切的是，在此部分實驗中會造成低符合度的原因為模糊條件(vague query)，亦即學習者若使用較為精準的表達元素(如 prefix)，符合度會高於使用較不精準的表達元

素(如 wildcard)。然而於檢索模組中較不精準的查詢條件其選取的例句數量較多，而 SAW 系統會依據例句長短的機制推薦合宜的例句給學習者，所以當選取例句數量越大時，平均前 10 個例句的字數總和會越少，故在此情形下符合度和句型成本的關係即成正比。

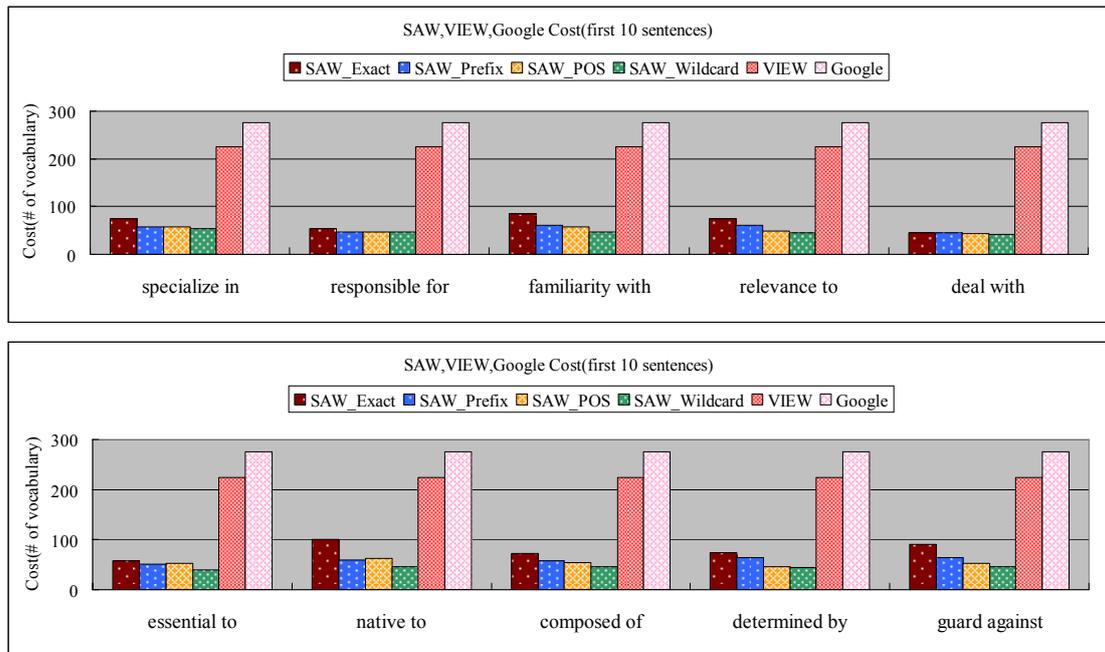


圖 4.6 兩個英語單字 SAW、VIEW、Google 的句型成本

上述僅考量兩個英語單字組成的片語，本研究亦測試三個和四個英語單字組成的片語，評估上述查詢模組的符合度和句型成本。和上述兩個英語單字片語查詢模組唯一不同的是，三個英語單字的 prefix 查詢模組可能的查詢條件有三個，例如以 a proportion of 為例，可能有 a% proportion of、a p% of、a proportion o% 三種不同的查詢條件，此時，實驗是取此三種查詢條件的平均符合度、平均句型成本當實驗評量的指標。同理，以 a proportion of 為例，POS 查詢模組可能有 O proportion of、a N of、a proportion P 三種不同的查詢條件；而 wildcard 查詢模組可能有 # proportion of、a # of、a proportion # 三種不同的查詢條件，此時，實驗是取此三種查詢條件的平均符合度、平均句型成本當實驗評量的指標。另外，關於四個英語單字的 prefix、POS、wildcard 查詢模組可能的查詢條件有四個，此

時，實驗是取此四種查詢條件的平均符合度、平均句型成本當實驗評量的指標。

圖 4.7 顯示，關鍵字越多 SAW 的符合度成效會比只有兩個關鍵字的成效好，原因在於多個英語單字組成的片語，其本身結構性較強，在其中一個關鍵字為模糊的情況下，學習者還是能得到較多的協助。而結果亦顯示，prefix 模組的成效高於 POS 模組，POS 模組高於 wildcard 模組，顯示 prefix 表達元素為限制條件較強的查詢條件，而 wildcard 表達元素限制條件是最為寬鬆的。

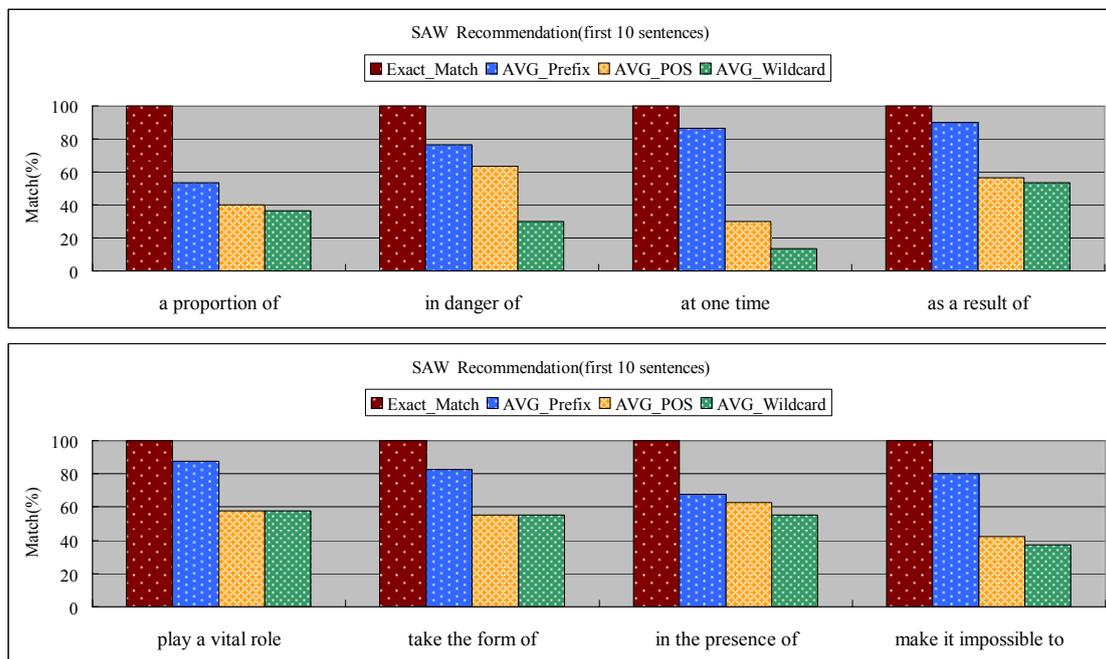


圖 4.7 多個英語單字符符合度

圖 4.8 顯示，關鍵字越多，經 MSA 比對後句型長度會比兩個單字的結果為高，原因在關鍵字越多，符合查詢條件的例句其字數亦較多。和兩個單字的句型成本結果類似，在此情形下符合度和句型成本是成正比的。

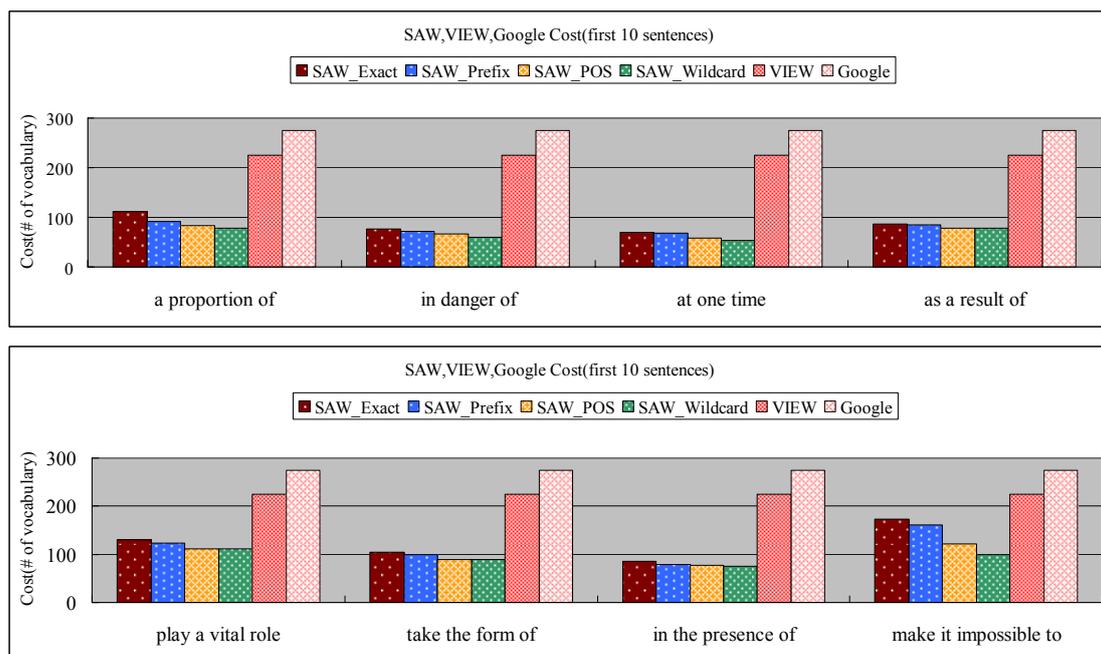


圖 4.8 多個英語單字 SAW、VIEW、Google 的句型成本

#### 4.4.3 子句和認知錯誤的評估

本研究除上述表達元素外，另外有 subsequence 的表達元素和部分比對的選取機制，上述機制能有效率的表達英語不限定子句上的概念，在學習者部分認知錯誤或不確定的情形下提供較大範圍的參照協助。VIEW 系統和 Google 系統並無所謂 subsequence 的表達元素，故不列入符合度和句型成本比較。本研究從選取的 12 個子句中評估 subsequence 表達元素的查詢成效(enable \* to; derive \* from; expose \* to; not only \* but also; would rather \* than; distinguish \* from; divide \* from; expand \* into; provide \* with; the same \* as; either \* or; so \* that)，並刻意對於 12 個子句中提供部分錯誤的查詢條件輸入。舉例而言，對於 enable ... to 的片語，刻意以 enable \* and 和 enable \* but 的部分錯誤查詢條件模擬學習者認知上的錯誤，並以平均符合度和平均句型成本來描述系統的協助成效，示意圖如下：

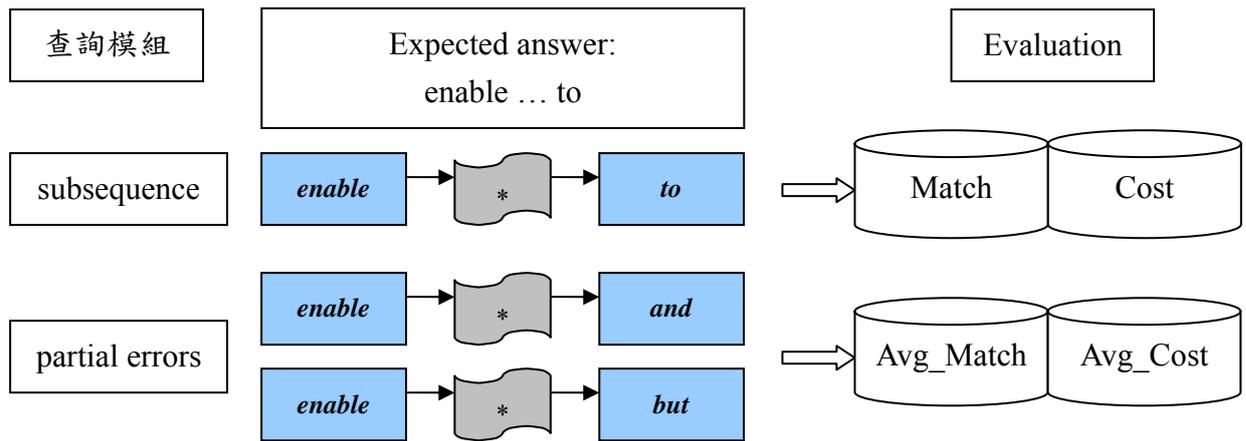


圖 4.9 enable ... to 以 subsequence 和 partial errors 的查詢方式

而此 12 個子句中，均有一個重要的關鍵字，依序為 to, from, to, but, than, from, from, with, into, as, or, so。本實驗將部分錯誤的認知均以 and 和 but 來取代上述的關鍵字，唯一例外的是 not only \* but also 中的 but，因 but 是正確用法，實驗將 but 取代為 or 當部分錯誤認知的查詢條件。

實驗的結果顯示，本研究的子句查詢模組均能對學習者提供相當程度的協助（符合度 100%）。在學習者部分錯誤的認知方面，查詢結果會因子句關鍵字彼此間的強度而有不同的成效，例如 enable \* to 因本身關鍵字較強，故即使將 to 輸入成 and 或 but，亦能藉由系統的部分比對得到一定的協助。相反的，諸如 either \* or 和 so \* that 子句中的關鍵字彼此的關連強度不強，縱使經過部分比對的選取方式，給予學習者的符合度亦有限，然對於低符合度的情形，本研究會建議學習者檢查關鍵字是否有錯誤(false query)或者不夠嚴謹(vague query)。然由問卷的結果顯示，事實上學習者亦能於低符合度的情況下，觀察推薦例句結果而評估本身是否有錯誤(false query)或者不夠嚴謹(vague query)的查詢條件發生，可以說間接讓學習者得到一定的寫作協助。

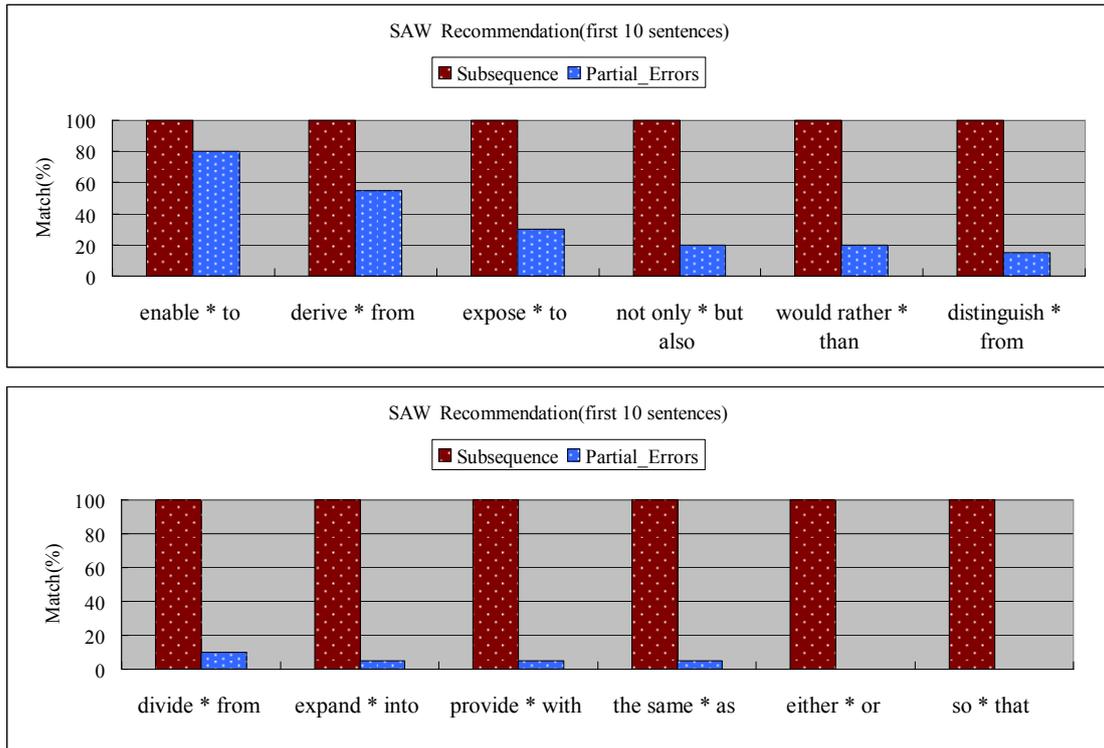


圖 4.10 subsequence 和 partial\_errors 查詢方式符合度

值得注意的是，在此實驗中低符合度代表錯誤查詢(false query)，含有錯誤的查詢條件相對於正確的查詢條件搜尋出的例句篇數是少的，而經系統前 10 篇的句型成本評估，其句型成本是較高的，在此情形下，符合度和句型成本成反比。

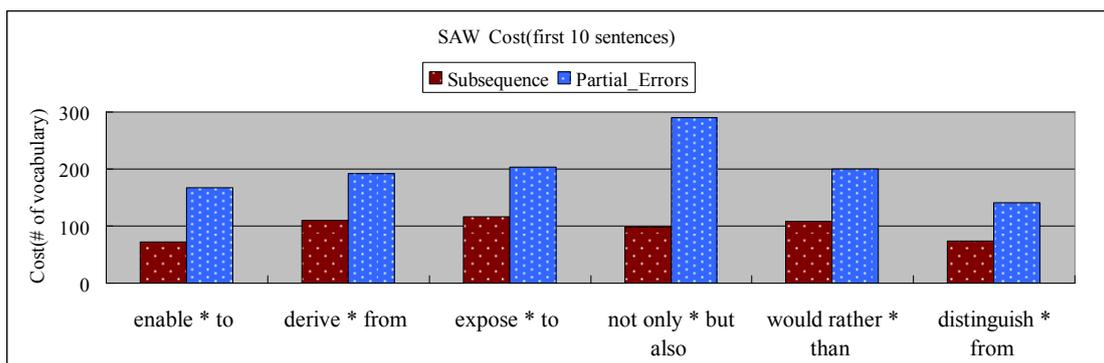


圖 4.11(a) subsequence 和 partial\_errors 查詢方式句型成本

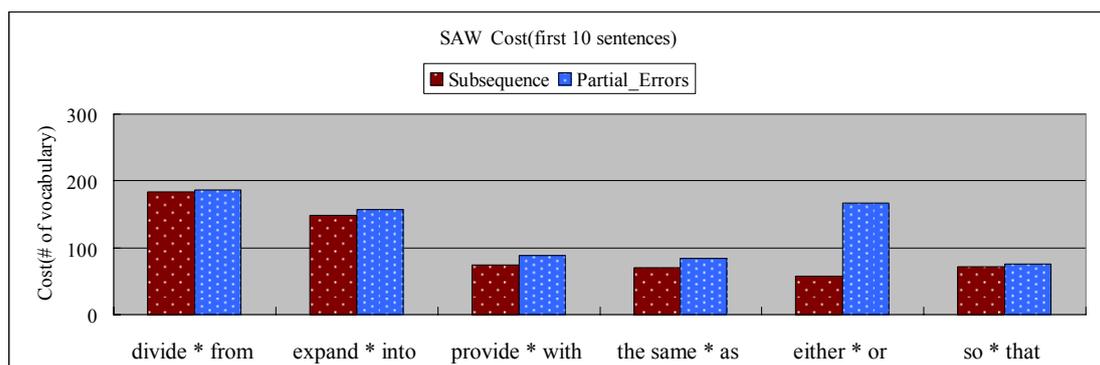


圖 4.11(b) subsequence 和 partial\_errors 查詢方式句型成本

#### 4.4.4 試題案例小結

在試題案例評估中，研究從 TOFEL 試題選取重要的單字片語和句型模擬學習者於寫作上的查詢需求，而從輸入的機制而言，SAW 系統相較於 VIEW 系統和 Google 系統擁有規劃較完善的表達元素，表示 SAW 系統能較彈性地表達學習者於英語寫作時的認知需求。而 SAW 系統呈現的例句亦比 VIEW 系統的 concordance 和 Google 的 snippet 格式上不但完整有利於學習者於寫作上的參照，且整體閱讀所需花費的句型成本也較為少，證明 SAW 相較於 VIEW 和 Google 系統有較佳的英語寫作協助成效。除此之外，SAW 系統能充分利用不同的輸入機制呈現不同學習者的認知需求，而經由符合度的驗證，證實 SAW 系統對於上述認知程度不同的學習者均能給予一定程度的寫作參照，充分證明 SAW 系統於試題評估客觀面上，擁有不錯的協助成效。

#### 4.5 問卷案例評估

問卷評估主要的目的為實際瞭解學習者於英語寫作上的認知，並評估 SAW 系統能否給予學習者適當地協助。問卷內容分為認知調查(詳見附錄十)和試題分析兩部分(詳見附錄十一、十二、十三)。認知調查瞭解受測者使用既有寫作協助工具的頻率，並瞭解受測者對於寫作背景的三个前提知識(單字、詞語搭配、文

法句型)的困擾程度。試題分析有 19 道試題(16 題單選 3 題翻譯)，針對 19 道試題，SAW 系統提供相對應的 BNC 參照例句，而 16 題單選試題有兩種不同查詢條件的參照例句(符合度不同)，3 題翻譯由 SAW 系統提供單一版本的參照例句。藉由試題分數上的差異，以及受測者提供的推薦度、滿意度，可以瞭解 SAW 系統能否給予受測者一定程度的寫作參照協助。

問卷施測時先填寫認知調查，而後試題分析的部分將分為以下兩種不同的施測方式：(1)有無使用系統的區別：了解有無使用系統協助是否會影響學習者作答，施測時將對象分為有使用和沒使用系統協助的兩組，而考量到實際寫作時學習者對於 SAW 系統可能使用不同的查詢條件，施測時將有使用系統協助的組別再細分為使用高符合度和低符合度參照例句的組別，如此可以瞭解在不同認知程度的學習者所輸入的不同查詢條件下，SAW 系統的使用與否是否均能給予一定程度的寫作協助。

簡而言之受測者將分為三組：第一組填寫無參照例句的 19 道試題；第二組填寫高符合度參照例句的 19 道試題；第三組填寫低符合度參照例句的 19 道試題。此部分問卷將以下列四個角度分析：(a)試題分數：分析在有無使用系統協助下，受測者的試題分數是否有所區別；(b)推薦度：瞭解在有使用系統協助下，受測者對於提供的參照例句推薦度；(c)推薦度和符合度的關係：此部分將分析符合度和推薦度的關聯，分析的方法為從試題的符合度和施測者給予的對應推薦度，延伸推得推薦度和符合度兩個評量指標的關連；(d)句型長度和滿意度：瞭解在有使用系統協助下，受測者對於 SAW 系統提供的參照例句長度是否感到恰當，和瞭解受測者對於 SAW 系統的滿意度。

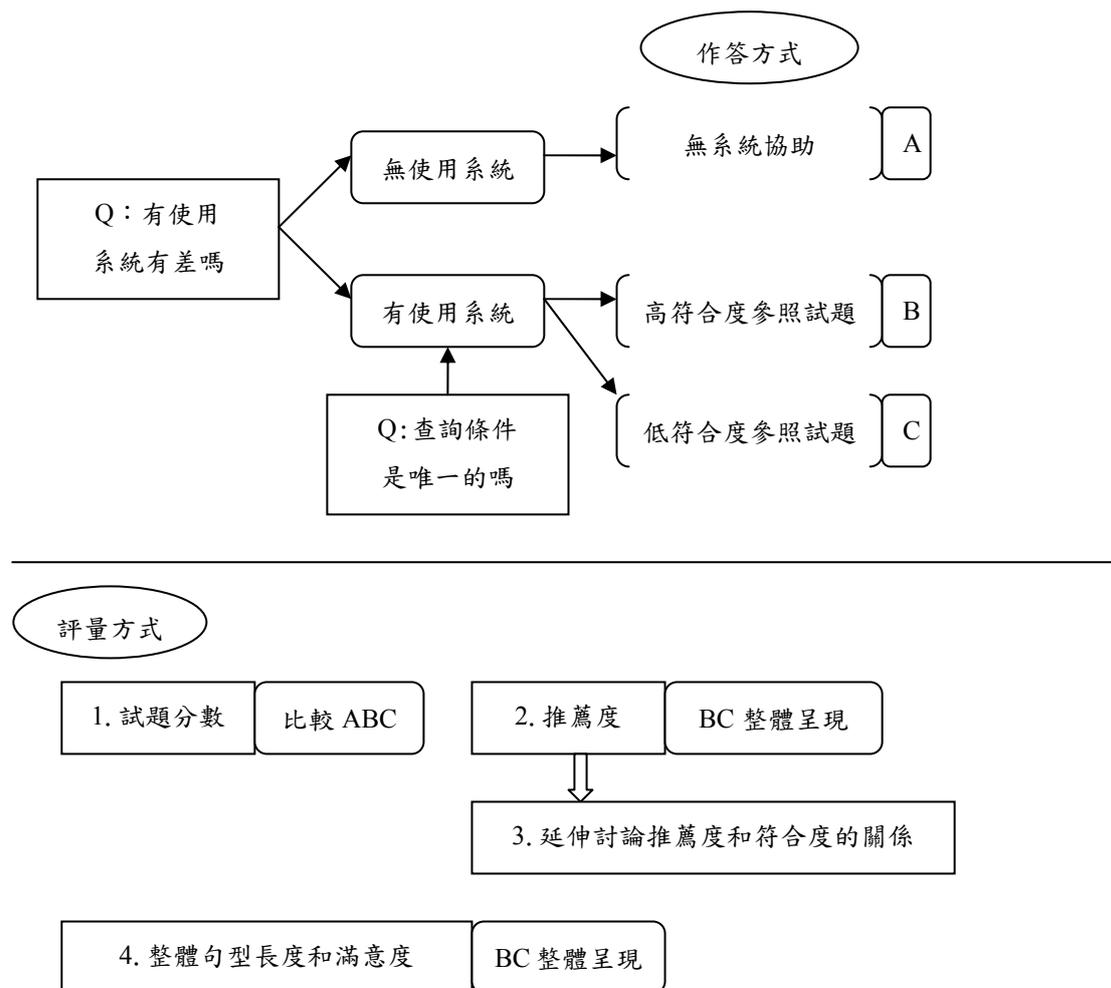


圖 4.12 有無使用系統的區別問卷施測方法

(2)使用系統前後的區別：若有無使用系統的區別結果為強調系統的整體協助成效，本部分則是強調個人的協助成效，而藉由受測者使用系統前和使用系統後的分析，我們可以更加確信學習者在實際需要協助時，SAW 系統能否給予協助。施測時先填寫無參照例句的 19 道試題，收回試題後提供高符合度參照例句的 19 道試題，由施測者再次填寫試題。如此可藉由使用系統前後上的區別瞭解當學習者實際需要協助時(使用系統前有疑問的地方)，SAW 系統能否給予協助(使用系統後後使否還有疑問)。此部分問卷將以下述角度分析：(a)試題分數：分析在使用系統前後，受測者的試題分數是否有所區別；(b)推薦度：瞭解在使用系統協助後，受測者對於提供的參照例句推薦度；(c)句型長度和滿意度：瞭解在使用系統協助後，受測者對於 SAW 系統提供的參照例句長度是否感到恰當，

和瞭解受測者對於 SAW 系統的滿意度。

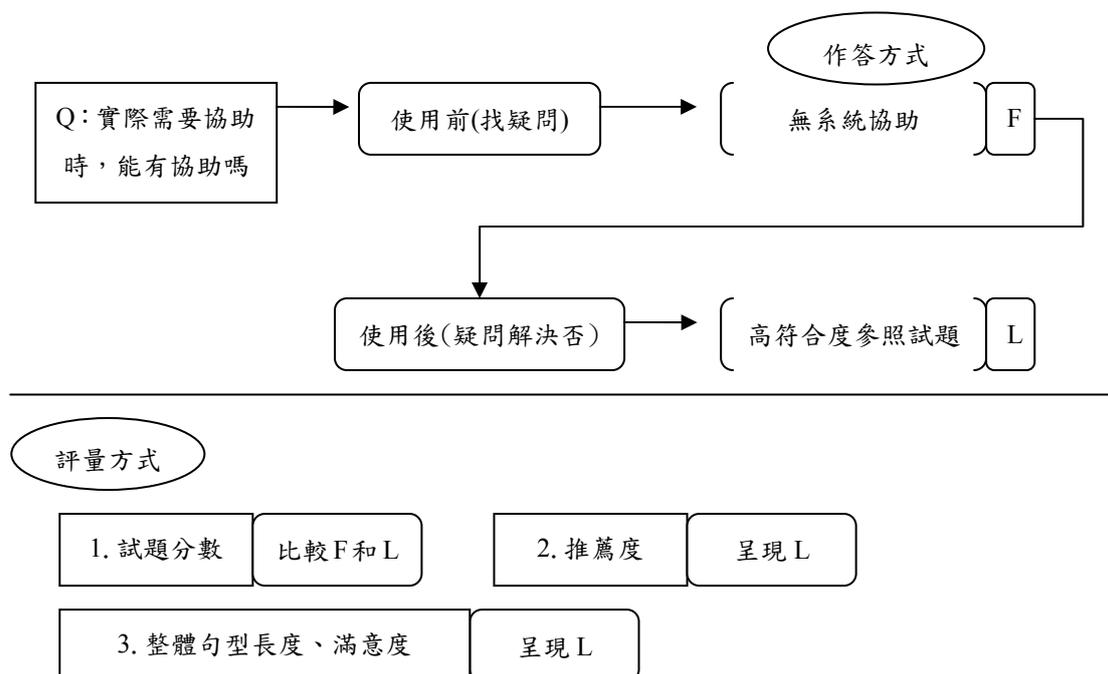


圖 4.13 使用系統前後的區別問卷施測方法

問卷施測的對象為 125 位的五專生。而依據英語字彙程度測試問卷(詳見附錄九) [10]，可瞭解學習者對於英語字彙的認知度，此次五專生對於英語常用 1000 單字(附錄九中 TEST 1,Level 1)約有 80.44%的認知度；對於英語常用 2000 單字(附錄九中 TEST 1,Level 2)有 46.10%的認知度。而根據 Nation[31]的研究，英語常用 1000 單字出現於英語文章的頻率為 74%，英語常用 2000 單字頻率為 81%。換而言之，問卷對象平均不超過 2000 單字的字彙程度，亦即平均每五個單字中受測者約有一個單字不熟識。

問卷回收的份數為 125 份，其中無效的問卷有 12 份(如繳交白卷、猜題作答(如選項只有 1 和 2，受測者選 3 或 4；抑或受測者對於選項全作答 1、2、3 或 4))，總計有效問卷為 113 份，其中有無使用系統上的區分問卷共 86 份，細分為無使用系統參照的 44 份、使用高符合度參照例句的 20 份、使用低符合度參照例句的

22 份問卷；而使用系統前和使用系統後的問卷共 27 份。問卷調查的結果如下：

#### 4.5.1 受測者對於寫作協助的認知調查

關於認知調查的部分有效問卷為 113 份，認知調查的目的為瞭解受測者使用既有寫作協助工具的頻率和對於寫作背景的三个前提知識(單字、詞語搭配、文法句型)的困擾程度。以 1~5 的分數代表學習者的使用頻率或困擾程度，分數越高代表學習者的使用頻率越高或困擾程度越大，結果以 113 份問卷的平均分數呈現。

而由圖 4.14 發現，此次受測者普遍常使用協助工具輔助寫作(3.55)，相形之下，並約略覺得協助工具的功能有限(3.33)而提供的例句過少(3.10)。而本研究能確實克服所謂功能有限的障礙、並可使用完整的語料庫而予以較多的參照例句，實際滿足上述受測者的建議需求。

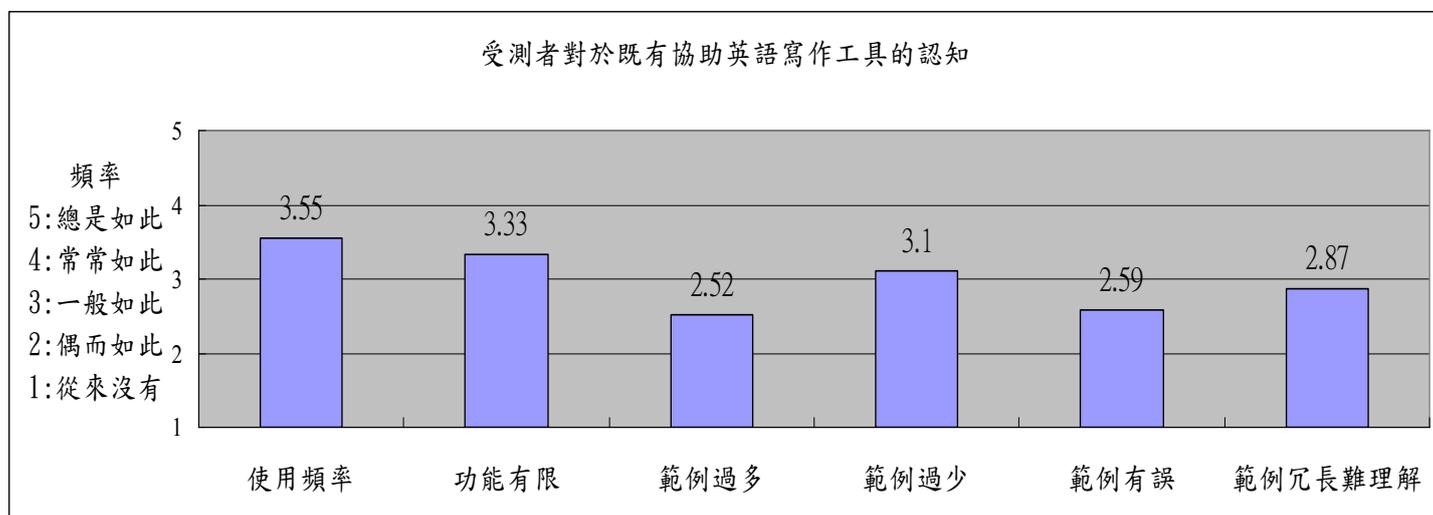


圖 4.14 受測者對於既有協助英語寫作工具的認知

至於寫作的大三前提知識受測者的困惑度，結果顯示受測者對於英語寫作有相當的困擾(單字困惑為 3.61、詞語搭配困惑為 3.65、文法句型為 3.47)，結果如

圖 4.15 顯示。

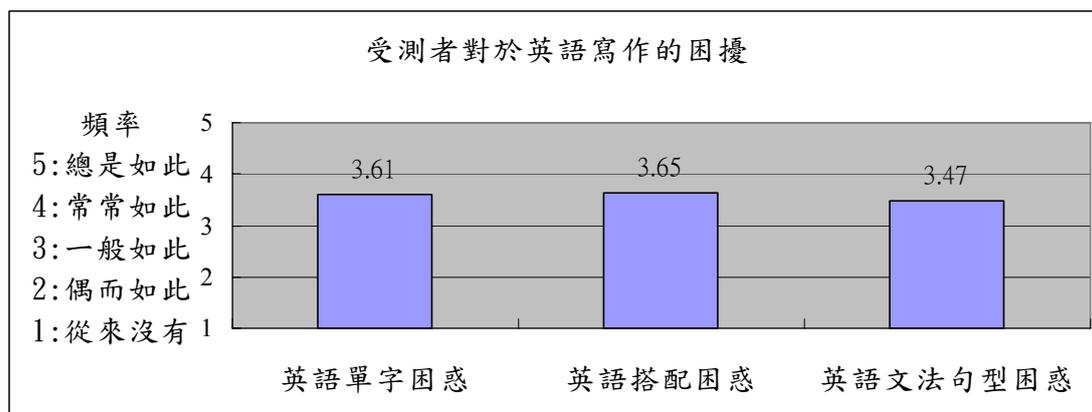


圖 4.15 受測者對於英語寫作的困擾

而由圖 4.14 和圖 4.15 發現，此次受測者普遍對於選項的選擇較為保守(最高的選項分數和最低的選項分數差距不到 1)，然單字、搭配、文法句型的困擾分數相對性地高，此更加證實受測者對於寫作的三大前提知識相當不足，無法撰寫完善地英語寫作。

#### 4.5.2 有無使用系統的區別

有無使用系統的區別將分為試題分數、推薦度、推薦度和符合度的關係、整體句型長度和滿意度四個小節分析：

(1)試題分數：在有無使用系統上的區別有效問卷 86 份中，單選試題有 44 份問卷為受測者沒有使用系統的協助而填寫試題、20 份為使用系統較高符合度的參照例句填寫試題(平均符合度為 62.48%)、22 份為使用系統較低符合度的參照例句填寫試題(平均符合度為 34.69%)。其中單選試題共有 16 題，每題為 0.625 分，滿分為 10 分。而由結果發現有使用系統參照的分數均高於未使用系統協助的分數，其中無使用系統參照的分數為 4.42 分，標準差為 1.95；而使用高符合度的參照例句其分數為 6.69 分，標準差為 2.23，比無使用系統參照的分數高 2.27

分；而使用低符合度的參照例句其分數為 5.97 分，標準差為 1.85，比無使用系統參照的分數高 1.55 分。

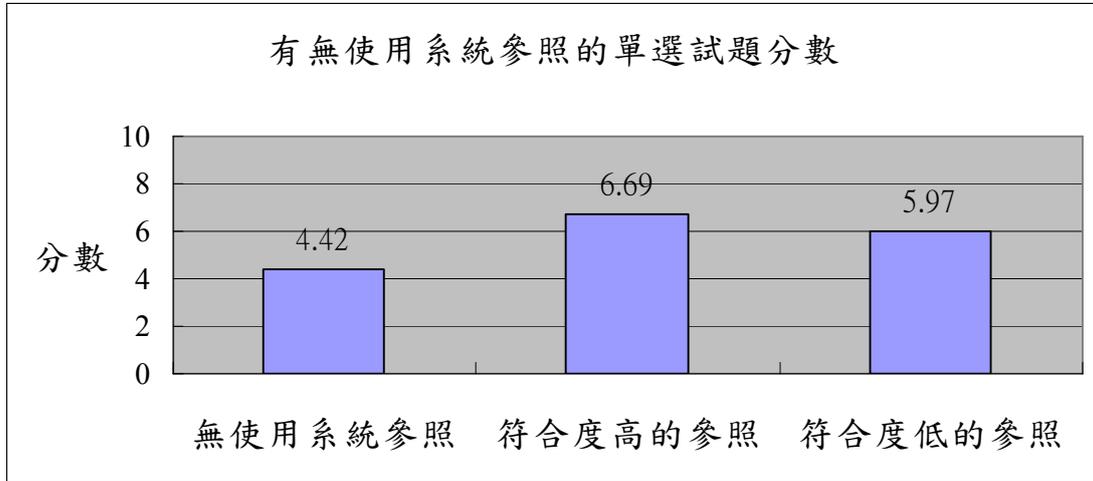


圖 4.16 有無使用系統參照的單選試題分數

而為了驗證上述結果並非偶然發生，研究以統計學的顯著水準(significance level)說明 SAW 系統確實有一定的協助顯著性。假設受測者的分數分佈為常態分佈，並以無使用系統參照的試題分數當所有受測者的母體。假設 Null Hypothesis 為  $H_0$ :SAW 系統沒有成效(平均分數為 4.42), Alternative Hypothesis 為  $H_1$ :SAW 系統有成效(平均分數大於 4.42)，在顯著水準 5%下，可得拒絕  $H_0$  的臨界分數為 X：

$$X = \mu + \frac{Z_{(0.05)} * \sigma}{\sqrt{N}}$$

其中  $\mu$  為母體的平均值(4.42)， $Z_{(0.05)}$  為在 5%顯著水準下的標準統計量 (Z-score, 查表可得  $Z_{(0.05)}=1.65$ )， $\sigma$  為母體的標準差(1.95)，N 為樣本的數量。如以高符合度參照例句的 20 份樣本為例(N=20)，拒絕  $H_0$  的臨界分數為  $4.42+(1.65*1.95/\sqrt{20})=5.14$ ，因實驗所得的平均分數為 6.69 大於臨界分數 5.14，可以否定  $H_0$ ，亦即在 5%顯著水準下 SAW 系統對於受測者是有協助的。

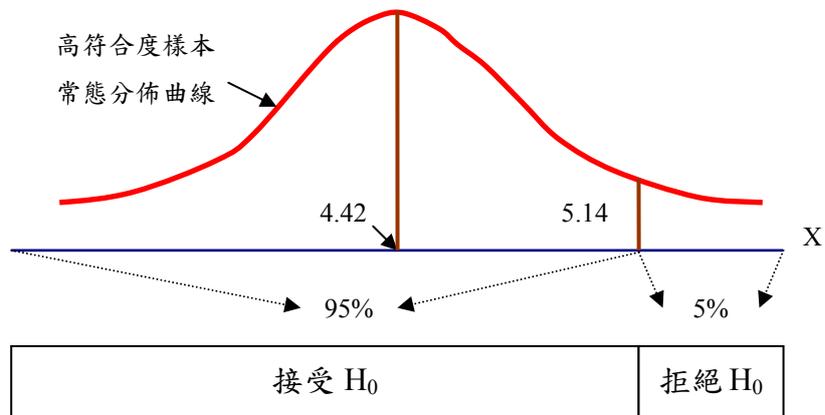


圖 4.17 高符合度參照例句樣本於顯著水準 5% 下拒絕平均分數為 4.42 的示意圖

依據同樣的統計分析方法，對於低符合度參照例句的 22 個樣本，其臨界分數為  $4.42 + (1.65 * 1.95 / \sqrt{22}) = 5.10$ ，而 SAW 系統得到的分數為 5.97，因 5.97 大於臨界分數 5.10，證明在 5% 顯著水準下 SAW 系統對於受測者仍然是有協助的。

根據高符合度和低符合度參照例句的分數結果，我們發現縱使對於不同認知程度的學習者，雖學習者給予的查詢條件可能有所不同，SAW 系統亦能針對不同的學習者給予其一定顯著性的協助。

除單選試題外，問卷有三題翻譯試題，翻譯試題能更確實地模擬學習者於寫作上所遇到的困難，其中在上述 86 份問卷中，有 44 份問卷為受測者沒有使用系統的協助而填寫試題，42 份為系統給予統一參照例句而填寫試題(即上述單選試題高低符合度中的 20 和 22 份問卷均予以統一的參照例句)。

其中翻譯試題本研究的評分標準為句型結構 50%、用詞 25%、其他 25%(包含語意、文法等)，三題試題以滿分 10 分為標準。而無使用系統參照的平均分數為 4.20，標準差為 2.56；而有使用系統參照的平均分數為 5.83，標準差為 2.46，分數高於無使用系統的問卷 1.63 分。同樣依據統計的顯著性分析，假設以無使

用系統參照的對象為母體，而使用系統參照的對象為樣本，在顯著水準 5% 下臨界分數為 4.97，因 5.83 大於 4.97，即在 5% 顯著水準下 SAW 系統對於受測者翻譯試題是有協助的。

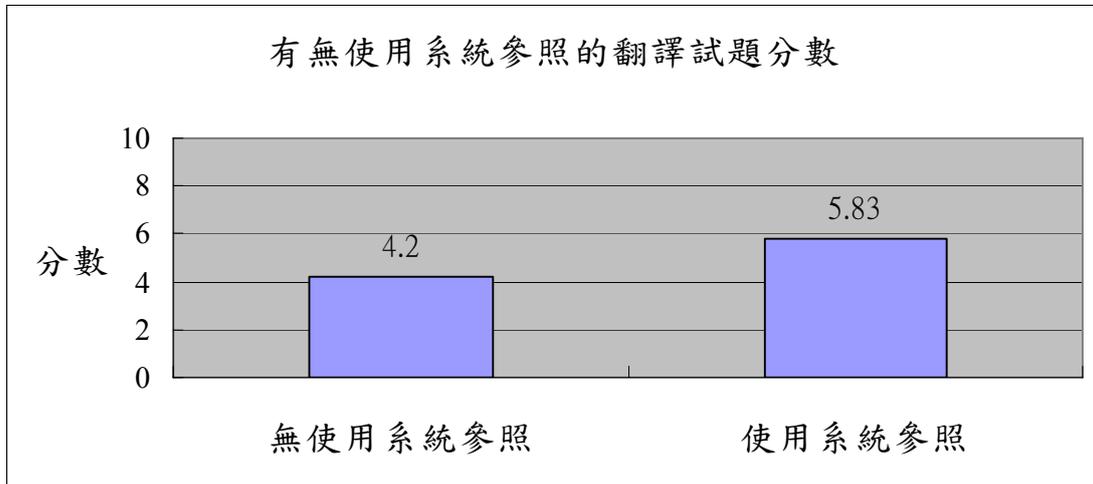


圖 4.18 有無使用系統參照的翻譯試題分數

(2)推薦度：推薦度是由學習者評估系統是否有實用性(Usefulness)，問卷以分數 1~4 分為標準，分數越高代表受測者越覺得有實用性。而結果顯示在使用系統參照的 44 份問卷中，其中單選試題使用高符合度參照例句的問卷平均為 2.78、使用低符合度參照例句的問卷平均為 2.70，而翻譯試題平均為 2.80。整體而言，有使用系統的平均推薦度介於 2.7~2.8 之間，表示受測者認為 SAW 系統是有一定的協助實用性。然此數值不算太高的原因是受限於受測者平均字彙程度不即 2000 字，以 BNC 的例句深度會造成受測者發生不甚瞭解例句的狀況，以致於受測者認定 SAW 系統的推薦度不高，而上述的假設亦經由受測者的問卷回饋證實此情況是發生的。

(3)推薦度和符合度的關係：本研究評估的方式有兩個很重要的指標，一個是由學習者評估的推薦度，一個是由系統自行評估的符合度，本小節將證實符合度和推薦度是有一定關連的，而問卷調查的結果證明了此論述。

在有使用系統區別的問卷單選試題部分有高符合度參照和低符合度參照的問卷，亦即 16 題單選中有 32 個符合度數值，同時有 32 個受測者給予的平均推薦度數值，本研究將上述所有 32 個關連點以 X-Y 分佈圖表示之，其中 X 座標為試題符合度，Y 座標為平均推薦度，並採用統計學的相關係數 (correlation coefficient) 來了解符合度和推薦度的關係，而相關係數的定義如下：

設有兩組樣本  $x_1, x_2, \dots, x_n$  及  $y_1, y_2, \dots, y_n$ ，其樣本平均數分別為  $\bar{x}$ ,  $\bar{y}$ 。樣本標準差分別為  $S_x, S_y$ ，且兩組樣本之共同變異數 (covariance)  $S_{xy}$  定義為

$$S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

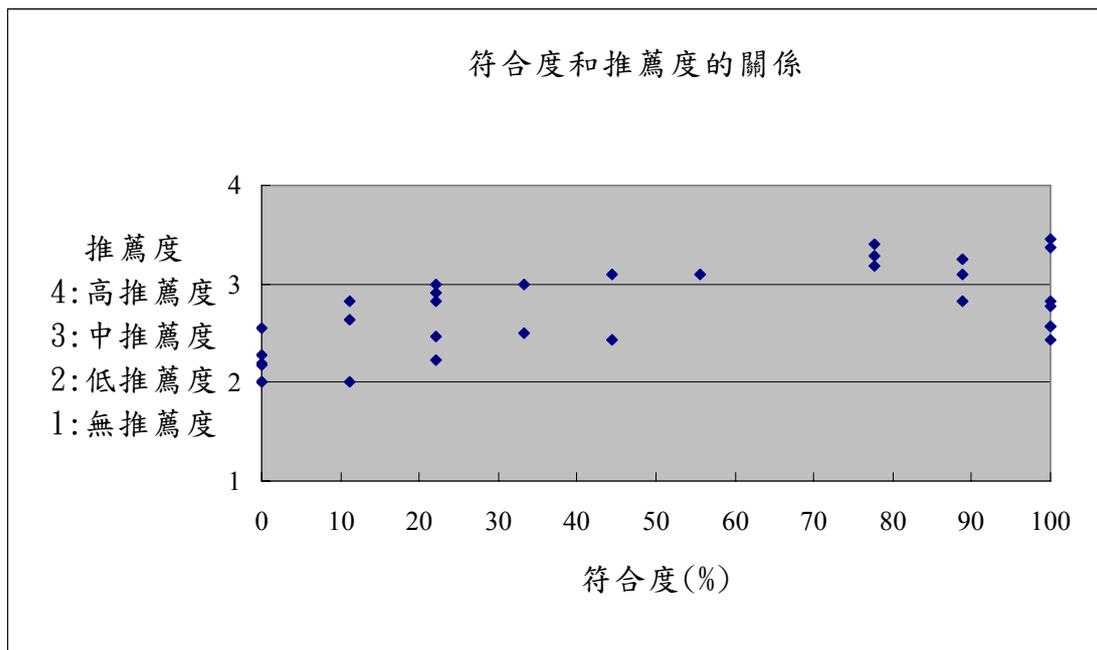
則相關係數  $r$  定義為

$$r = \frac{S_{xy}}{S_x S_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

相關係數的值介於 -1 與 1 之間，而越接近 -1 或 1 時表示兩樣本越相關 (兩樣本越呈線性關係)，當接近 -1 時為負相關，而當接近 1 時為正相關。而由結果發現符合度和推薦度的相關係數為 0.63，亦即兩種指標有一定程度的正相關線性關係。而觀察符合度低的點，受測者的推薦度也約略有 2 以上 (低推薦度以上)，此證實學習者亦能從符合度低的情況下間接得到協助資訊。

表 4.4 符合度對應推薦度的數值(符合度,推薦度)

以下為 32 個符合度和推薦度的 X-Y 分佈圖座落點(符合度(%),推薦度(1~4))			
(0.0 , 2.00)	(0 , 2.00)	(0.0 , 2.18)	(0.0 , 2.19)
(0.0 , 2.28)	(0 , 2.55)	(11.1 , 2.00)	(11.1 , 2.64)
(11.1 , 2.82)	(22.2 , 2.22)	(22.2 , 2.46)	(22.2 , 2.82)
(22.2 , 2.91)	(22.2 , 3.00)	(33.3 , 2.50)	(33.3 , 3.00)
(44.4 , 2.44)	(44.4 , 3.10)	(55.5 , 3.10)	(77.7 , 3.19)
(77.7 , 3.19)	(77.7 , 3.40)	(88.8 , 2.82)	(88.8 , 3.10)
(88.8 , 3.25)	(99.9 , 2.44)	(99.9 , 2.56)	(99.9 , 2.56)
(99.9 , 2.78)	(99.9 , 2.82)	(99.9 , 3.37)	(99.9 , 3.46)



#### 4.19 符合度和推薦度的關係

(4)整體句型長度和滿意度：整體句型長度為瞭解受測者對於 SAW 提供的參考例句，於長度上是否感覺恰當，問卷以 1、2、3 的分數表示長度過短、適中使、過長。而滿意度為受測者對於 SAW 系統的滿意度，問卷以 1~5 的分數表示，分數越高代表受測者越滿意。此部分以有使用系統的 44 份問卷以平均分數呈現。

受測者對於例句長度的感覺平均為 2.09(適中)。因滿意度受限於受測者的字彙程度不及 2000 字，而 BNC 的例句普遍會造成受測者閱讀上的困難，導致受測者的平均滿意度為 2.76(約為中偏下)，這證明在受測者認知程度不同的情況下，須有不同程度的語料庫當推薦方法的語料庫(或資料庫)，才能有效率地給予所有學習者一定程度的滿意度。

以下為有無使用系統區別的試題分數、推薦度、整體句型長度、滿意度的問卷數值：

表 4.5 有無使用系統區別的問卷數值總表

數值越大，代表分數越高、推薦度越高、覺得句型越長或滿意度越高(最低值~最大值)(X:無意義)			
	無使用系統	符合度高的參照	符合度低的參照
單選試題分數(0~10)	4.42	6.69	5.97
翻譯試題分數(0~10)	4.2	5.83	
單選試題推薦度(1~4)	X	2.78	2.70
翻譯試題推薦度(1~4)	X	2.80	
句型長度(1~3)	X	2.09	
滿意度(1~5)	X	2.76	

### 4.5.3 使用系統前後的區別

使用系統前後的區別將分為試題分數、推薦度、整體句型長度和滿意度三個小節分析：

(1)試題分數前後的差異：在使用系統上前後的區別有效問卷 27 份中。其中單選試題共有 16 題，每題為 0.625 分，滿分為 10 分。而由結果發現使用系統前受測者的平均分數為 4.99，標準差為 2.49；使用系統後的平均分數為 6.53，標準差為 2.00，分數增加 1.54 分。同理，若以使用系統前的分數為母體，使用系統

後的分數為樣本，在 5%顯著水準下的臨界分數為 5.71。因 6.53 大於 5.71，證實 SAW 系統對於單選試題是有協助性的。

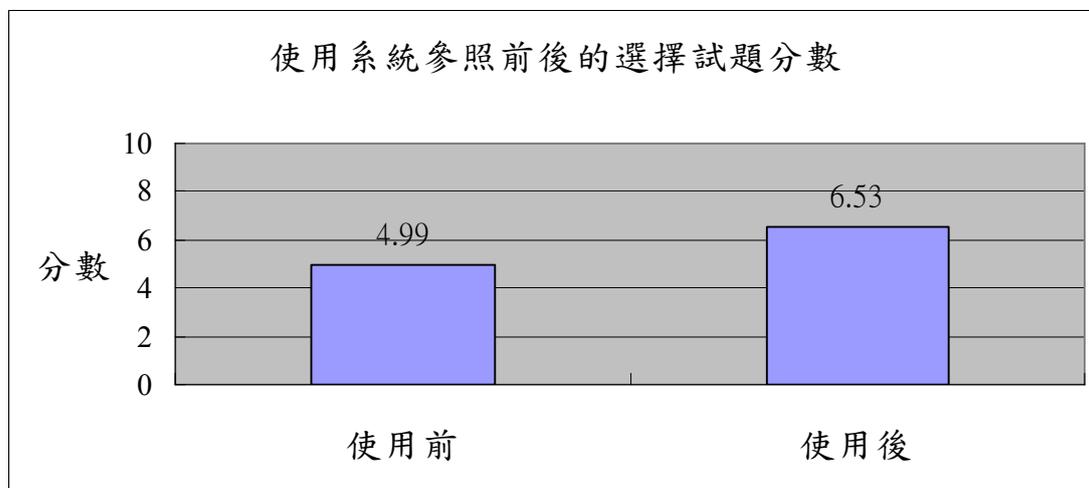


圖 4.20 使用系統參照前後的單選試題分數

在使用系統上前後的區別中翻譯試題同樣以句型結構 50%、用詞 25%、其他 25%(包含語意、文法等)為評分標準，滿分為 10 分。而使用系統前的平均分數為 4.56，標準差為 2.59；使用系統後平均分數為 5.42，標準差為 2.60，分數增加為 0.86 分。同理，以使用系統前的分數當母體，而使用系統後的分數當樣本，在 5%顯著水準下的臨界分數為 5.38。因 5.42 大於 5.38，證實 SAW 系統對於翻譯試題是有協助性的。

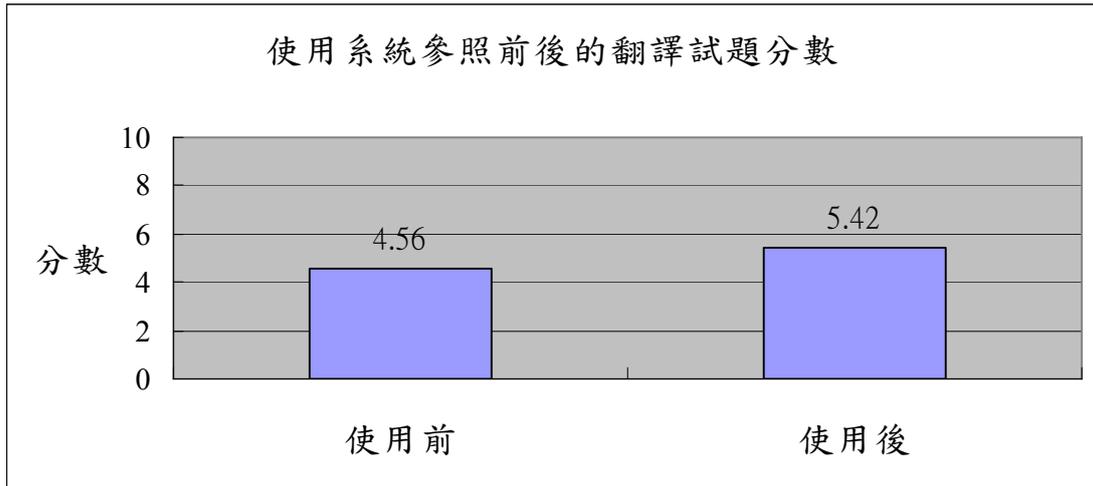


圖 4.21 使用系統參照前後的翻譯試題分數

(2)推薦度：推薦度是由學習者評估系統是否有實用性(Usefulness)，分數的取捨以 1~4 分為標準，分數越高對於系統越覺得有實用性。而結果顯示在使用系統前後上的區別有效問卷 27 份中，受測者對於單選試題的推薦度平均為 2.59；而翻譯試題推薦度平均為 2.53，同理，會造成推薦度無法相當高的原因還是在於受測者詞彙上的不足，導致發生無法理解例句的狀況而給予較低的推薦度。

(3)整體句型長度和滿意度：整體句型長度為瞭解受測者對於 SAW 提供的參考例句，於長度上是否感覺恰當，問卷以 1、2、3 的分數表示長度過短、適中使、和過長。而滿意度為受測者對於 SAW 系統的滿意度，問卷以 1~5 的分數表示，分數越高代表越滿意 SAW 系統。

整體而言，在使用系統前後上的區別有效問卷 27 份中，受測者對於句型長度的感覺平均為 2.16(適中)，而滿意度平均為 3.16(普通)。

以下為使用系統前後區別的試題分數、推薦度、整體句型長度、滿意度的問卷數值：

表 4.6 使用系統前後區別的問卷數值總表

數值越大，代表分數越高、推薦度越高、覺得句型越長或滿意度越高(最低值~最大值)(X:無意義)		
	使用系統前	使用系統後
單選試題分數(0~10)	4.99	6.53
翻譯試題分數(0~10)	4.56	5.42
單選試題推薦度(1~4)	X	2.59
翻譯試題推薦度(1~4)	X	2.53
句型長度(1~3)	X	2.16
滿意度(1~5)	X	3.16

#### 4.5.4 問卷小結

在問卷調查中，研究證實無論從有無使用系統的角度或者使用系統前後的角度，SAW 系統如從試題分數的觀點切入，能給予學習者分數上的協助；而從推薦度的觀點切入，受測者亦認為能從 SAW 系統中得到一定的協助。而研究更進一步的驗證符合度和推薦度的關係，當符合度越高，推薦度也相對高，當符合度越低，推薦度也相對低，兩者呈現近似正相關的線性關係，故本研究使用符合度評估系統的成效是合理的。然因本問卷施測的對象普遍字彙程度不超過 2000 字，而 BNC 的例句對於受測者普遍偏難，容易造成受測者的推薦度和滿意度下滑，而此部分可藉由採取較適宜的語料庫讓結果更完善。

#### 4.6 學生作文範例

本研究亦從上述問卷施測對象得到相關的學生作文範例，學生作文範例為學生實際英語寫作的作業，並已經過教師批閱，觀察學生作文範例能更明確地瞭解學習者於寫作過程中常犯的錯誤，以下將以三個案例來表達學生若妥善使用 SAW 系統，能確實給予其協助：

### 學生 A. I need a nest of glasses.

此學生所犯的錯誤為不瞭解 glasses 所需搭配的冠詞，如該學生輸入 a # of glasses 於 SAW 系統中查詢，由前十句例句應不難發現合宜的冠詞應該為 tray、set、couple 或 pair(glasses 於當中的解釋不同)，而非 nest，符合度為 10/10，換而言之，此學生若於寫作時參照 SAW 系統，應能避免調上述的錯誤，詳盡的結果如下所示：

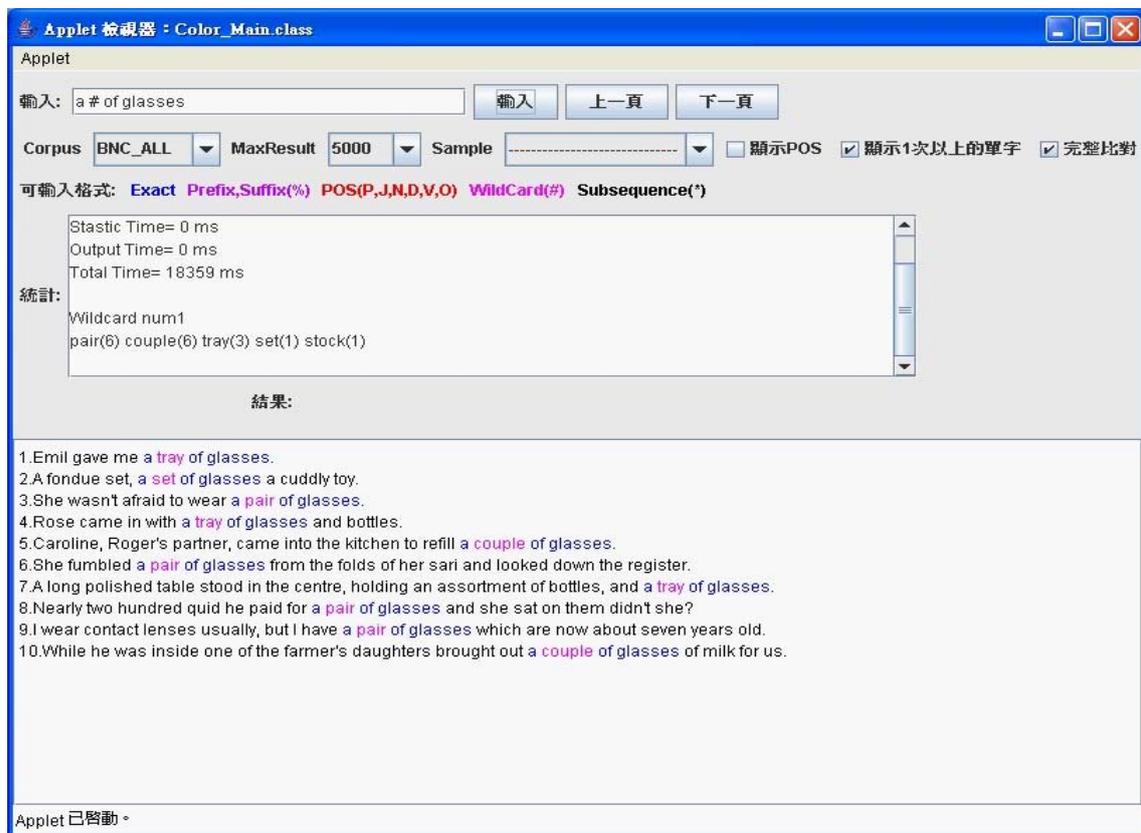


圖 4.22 a # of glasses 的查詢結果

### 學生 B. I would rather going shopping than staying home.

此學生所犯的錯誤為對於 would rather ... than 的句型不太熟悉，其中 would rather 後若接動詞，應以原形表達，亦即正確的句型應為 I would rather go shopping than stay home. °

同理，系統以 would rather V than V 和 would rather \* than \* 模擬學習者可能的查詢條件，而由結果發現，使用 would rather V than V 的查詢條件可以於前五句推薦例句中的第一、第二、第三、第四、第五句例句中得知正確的動詞形式，符合度為 5/5；而使用 would rather \* than \* 的查詢條件可以於前十句推薦例句中除第二、第七、第九例句得知正確的動詞形式，其符合度為 7/10。證實對於上述兩種查詢方法，SAW 系統均能給予學習者相關的協助。

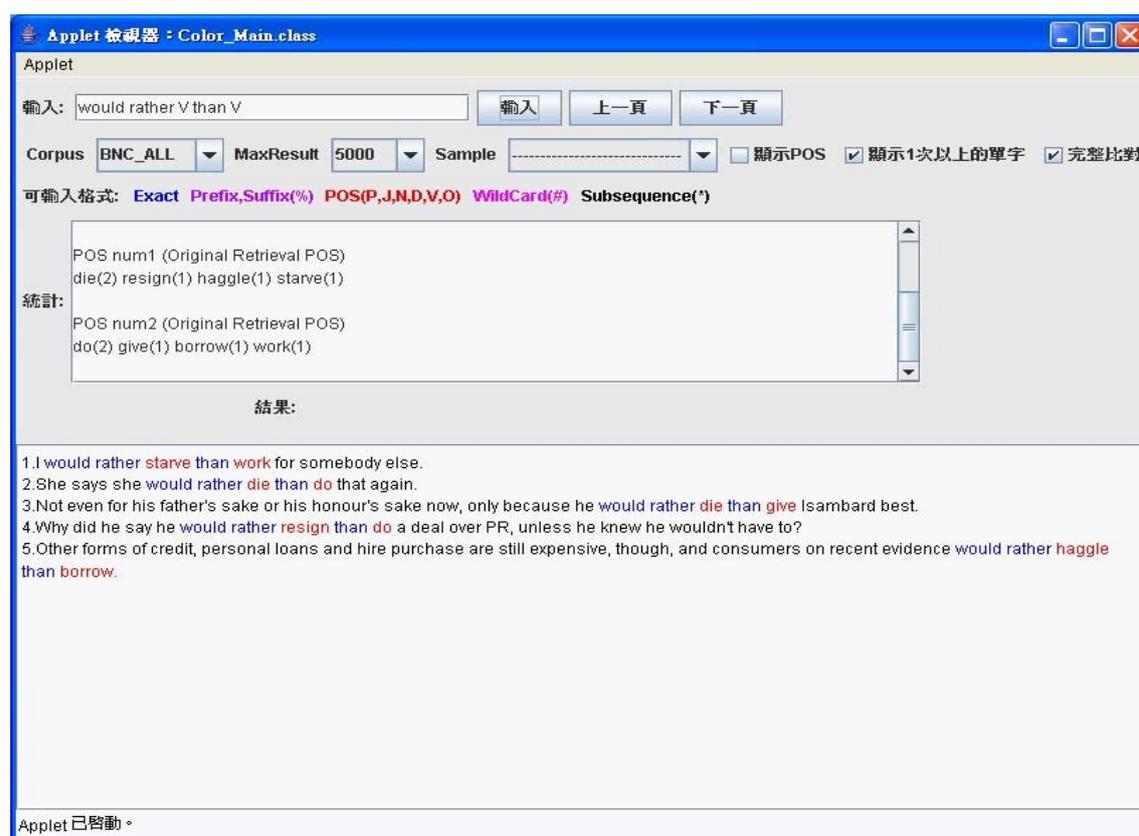


圖 4.23(a) would rather V than V 和 would rather \* than \* 的查詢結果

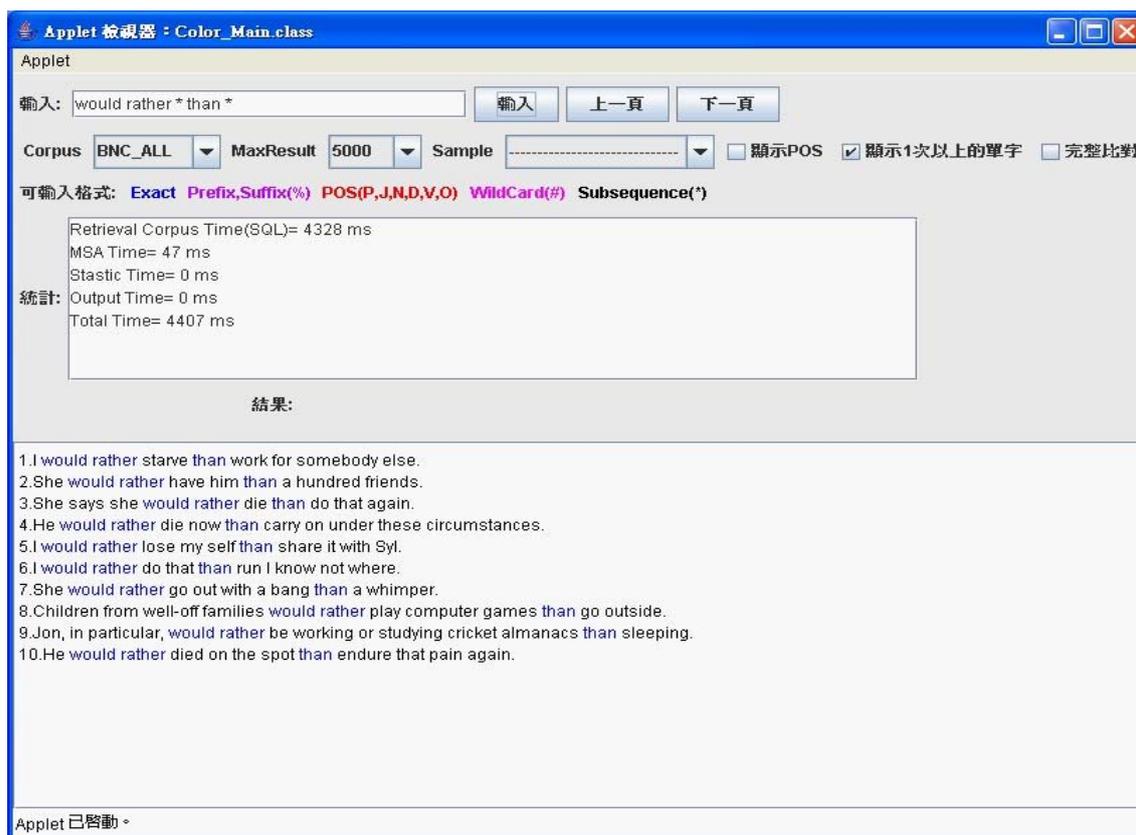


圖 4.23(b) would rather V than V 和 would rather \* than \* 的查詢結果

學生 C. My **mon** always get sick so my father tells me you have to \_\_ care of her.

此學生所犯的錯誤為對於 mother(mom)的單字不太熟悉，並且對於 to take care of 的句型不甚瞭解，實驗將上述的關鍵字分為兩部分，分別為 mo% \* father 和 to care of，其中 to take care of 故意以部分錯誤的認知方法輸入(少了 take)，模擬學習者於部分錯誤認知上的查詢。而由結果發現，使用 mo% \* father 的查詢條件可以於前十句推薦例句中除第一、第六、第十例句中得知正確的 mother 拼法，符合度為 7/10；而使用 to care of 的查詢條件可以於前十句推薦例句中除第六、第九句中得知正確的用法應為 to take care of，符合度為 8/10。

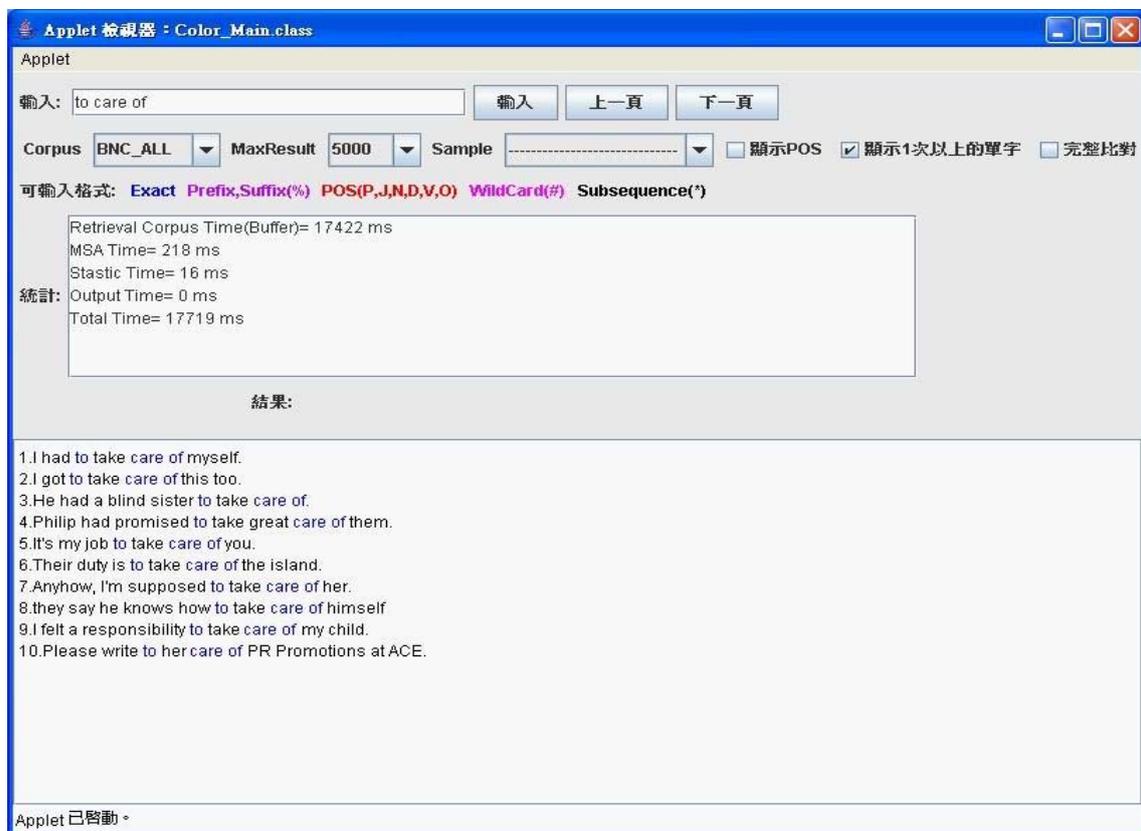
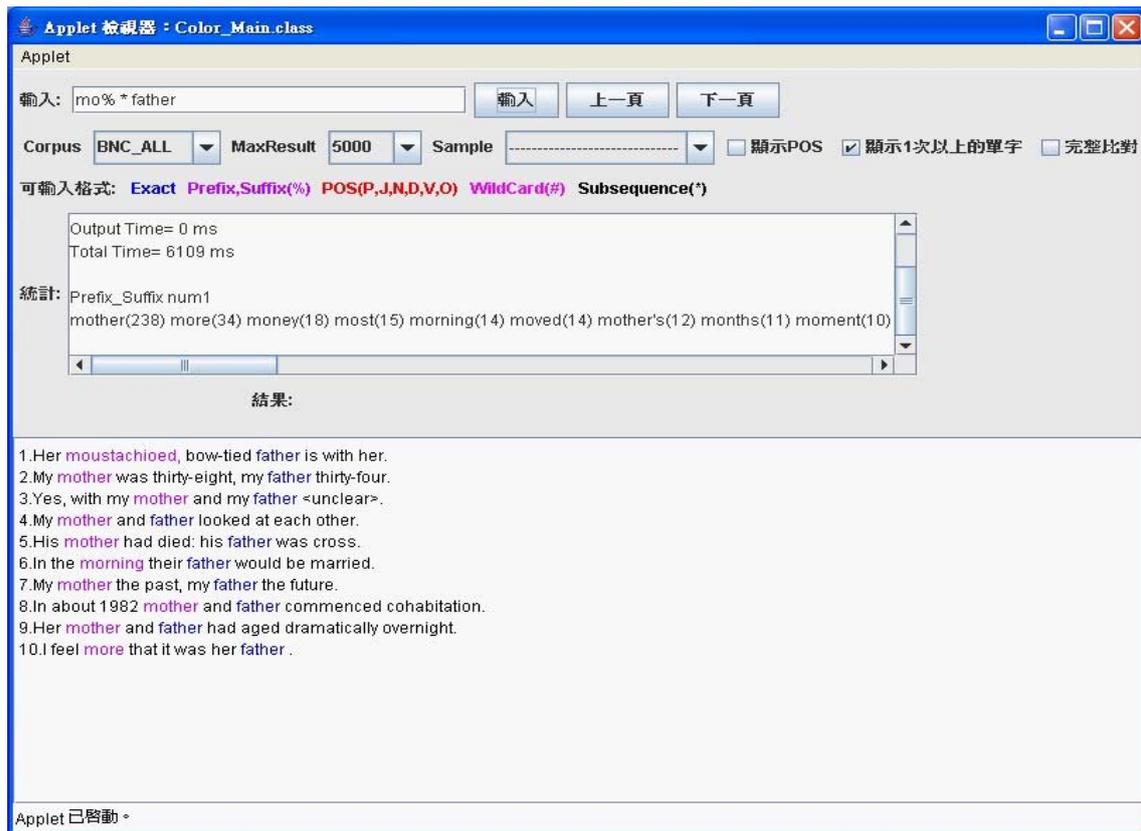


圖 4.24 mo% \* father 和 to care of 的查詢結果

由以上的三個學生作文範例，可以理解學習者於真實寫作時，可以靈活使用本研究的 exact、wildcard、POS、subsequence、prefix 的表達元素當不同的查詢條件，並於推薦的例句中得到一定的參照協助。

#### 4.7 實驗分析小結

前述六個小節分別描述論文的實驗語料庫(BNC、SA)，以及論文比對的系統(VIEW、Google)，並論及本研究的案例來源(TOFEL、大學聯招和學測)，而後定義論文擬定的五項指標(試題分數、推薦度、滿意度、符合度、句型成本)。而經由試題和問卷的整合評估，符合度和推薦度約略成正相關的線性關係，且根據低符合度所代表的兩層含意(vague query、false query)，學習者亦能從中間接得到一定程度上的協助。整體而言，SAW 系統能給予不同認知程度的學習者一定程度的寫作參照協助。除此之外，本研究亦將據學生的作文範例，輸入於 SAW 系統中測試，並發現 SAW 系統能確實給予一定的協助。