

以英語寫作輔助為目的之語料庫語句檢索方法

劉吉軒、洪培鈞
國立政治大學資訊科學系
jsliu@cs.nccu.edu.tw

李金瑛
國立台灣師範大學英語系

摘要

鑑於現有語言資源工具對於 ESL/EFL 學習者於英語寫作上提供的協助功能有限，本研究致力於提出一種語言資訊檢索方法，能在不同的語言認知程度條件下，從語料庫中找出對使用者之寫作表達需求有參照作用之例句。此方法提供準確字詞、單字的開頭/結尾、英語詞性、一個單字的萬用字元、不限定的子句等表達元素，並使用完整比對或部分比對的兩種比對方式選取例句，而後對於選取的例句使用多重序列排列技術進行相關性評估，最後推薦最符合參照需求的例句。

Abstract

Current language resource tools provide only limited help for ESL/EFL writing. This research proposes a language information retrieval approach to acquire referential sentences from corpus under various levels of users' language cognition. The approach includes a set of expression elements, such as exact words, prefix, suffix, part-of-speech, wildcard, and subsequence. Sentence retrieval involves both exact match and partial match with user query. Finally, the set of retrieved sentences are evaluated and ranked by multiple sequence alignment for relevance to user expression needs.

關鍵詞：資訊檢索，語料庫，寫作輔助

Keywords: Information Retrieval, Corpus, Writing Assistance.

一、緒論

對於大部分的寫作目的而言，寫作是一種深度的表達過程，需要詳盡而深入的描述能力，精準而嚴謹的語意呈現，及適當的詞語選擇安排。對於英語為非母語的學習者(English as Second Language - ESL/English as Foreign Language - EFL)來說，英語寫作尤其是一個困難的過程，常常會因為單字、搭配字(collocation)、詞性組合、常用片語、句型結構等方面的語言知識不足，或是受到母語認知與習慣的牽制影響，而造成用詞、語法、甚至語意上的錯誤[1]。同時，英語寫作也是一個費時的過程，ESL/EFL 的作者常常要花費許多時間，在自身不足的語言知識中搜尋適當的表達方式，或是要借助於字典、辭典等各種語言資源工具，取得初階的詞語資訊。但是，大量時間的投入往往僅能獲得有限的寫作品質的提升，而造成許多 ESL/EFL 作者的挫折與障礙。總結而言，若把 ESL/EFL 作者的英語寫作視為一個資訊處理的過程，這項工作的困難癥結在於資訊量的不足、資訊取得成本的過高、及資訊使用的效益有限。

近年來，語料庫的發展帶來語言研究上的新面向。語料庫是大量語言使用情形的紀錄，不論是以文字或是以口語方式的語言表達內容，都可以被大量蒐集彙整、描述標註，而

形成了自然而真實的語言使用現象呈現[2]。因此，語料庫代表著豐富的語言資源，不僅可以對語言學的研究，提供許多統計分析上的資訊與理解詮釋上的依據，也可以做為語言學習上各種語言使用方式的參照資料[3][4]。同樣的，以英語寫作為目的的觀點而言，語料庫的存在提供了資訊處理工作上相當大的利用空間，若能發展一個有效的語言資訊使用工具，在 ESL/EFL 的英語寫作過程中，針對作者的語言表達需求，提供足夠的、適當的、參照的語言使用資訊，必能有效輔助 ESL/EFL 的英語寫作過程，降低寫作障礙，提升寫作品質。

目前使用語料庫的語言資訊使用工具以 **concordance** 技術為主，提供特定字詞的檢索，將特定字詞在語料庫中出現的上下文，以特定字詞為基準排列呈現，讓使用者觀察特定字詞可能的使用方式及相關字詞的搭配情形。這種工具一般叫做 **concordancer**，不僅可以做為語言學者的研究輔助，也可以做為英語學習者的諮詢參考[5][6][7]。但是 **concordancer** 並非針對非母語人士的英語寫作需求，對 ESL/EFL 的寫作困難所提供的協助仍然不足[8]，其主要限制包括：(1)較有限的查詢機制，只提供詞彙與詞性的條件檢索，無法充分對應 ESL/EFL 作者在語言知識上感到不足或不確定的情形下，所需要的查詢方式；(2)較無彈性的檢索機制，只提供完全符合條件的檢索，在使用者認知錯誤而使用不適當的查詢條件時，將無法提供任何查詢結果；(3)檢索結果的呈現以上下文為主，而非以完整的例句呈現，同時，在結果的提供上，並未考量使用者在寫作上所需資訊的優先次序。

以資訊處理技術的觀點而言，**concordancer** 是一種相當初階的資訊檢索功能，只能針對有限形式的查詢條件，進行簡單的比對檢索，對檢索結果也沒有評估與排序的概念。因此，在語料庫的使用上，**concordancer** 只能針對明確的檢索目標提供少量的有用資訊。這樣的功能特性是無法滿足不同程度的 ESL/EFL 作者，在寫作上所需的多樣化資訊指引與參照協助。寫作是一個語言知識使用與輸出的過程，作者必須將其描述的意圖，明確的用正確的及適當的語彙組合排列，產生具體的文字內容與結構。而 ESL/EFL 作者常因語言知識的不足，在文字內容的生產過程中，常常必須針對疑惑進行探索或尋求答案，甚至在可能的認知錯誤下，也希望得到適當的引導而獲取正確的語言使用資訊。因此，一個以寫作協助為目標的語料庫使用工具，必須提供更有彈性的查詢與檢索機制，呈現更能針對作者語言資訊需求的檢索結果，才能有效利用語料庫，協助作者提升寫作品質。

本研究提出一種語言資訊檢索方法，針對 ESL/EFL 作者在寫作過程上的語言知識不足或不確定，從語料庫中尋找出可用的語言使用範例，進而提供針對性的範例協助。我們的語言資訊檢索方法包含三個模組，第一個模組是一個多元的表達元素模型，能針對寫作過程中的語言資訊需求，包括單字、搭配字、常用片語、句型結構等，提供作者以更多樣的部份資訊的方式來表示其資訊需求。第二個模組是一個彈性的檢索機制，具備精確比對與部分比對兩種功能，能依照作者的資訊掌握程度而調整，從語料庫中比對尋找符合作者語意表達需求的範例。第三個模組是一個評估排序的機制，針對找出的範例，評估其符合作者語意表達需求的程度，並將之排序呈現，以提升範例參照的使用效率。我們以上述方法為基礎，發展了一個以寫作協助為目的的語料庫使用工具 - **SAW (Sentence Assistance for Writing)**系統，並以客觀的指標量測和主觀的問卷調查兩種評估方式來評量 SAW 系統的成效。這兩種評估方式都驗證 SAW 系統能針對 ESL/EFL 作者在其寫作過程中的語言使用資訊需求，給予一定程度的滿足而達到諮詢指引的效果，證明本研究的語言資訊檢索方法確實能達成寫作協助的既定目標。

二、英語寫作及語言資源工具

英語學習與使用的「聽、說、讀、寫」四個面向中，「寫」往往是較為困難的部分，除了需要較為廣泛、深入、及精準的語言知識使用之外，也常會受到母語認知語習慣的牽制影響，而無法使用正確的英語詞彙及結構。例如對於「創作音樂」的表達方式，母語為中文的學生多以 *make*、*create*、*produce* 和 *music* 搭配，然而正確的英語詞彙應該是 *compose music*。研究指出 ESL/EFL 作者所需的前提知識有三個，分別為英語單字、英語搭配和英語文法句型[8]。英語單字指作者對於英語字彙的認知能力；英語搭配指作者對於特定詞彙搭配上的認知能力；英語文法句型指作者對於英語文法句型本身組成結構的認知能力。

相關研究亦指出[9][10]，在英語的使用上，詞語搭配(collocation)是 ESL/EFL 作者較為薄弱的一環，需要較多的語言資訊協助。詞語搭配是指共同出現的情形高於一般機率的特定詞彙組合。例如，在英語的使用習慣中，單字 *problem* 常和 *cause*、*create*、*solve* 一同搭配，而較少用 *make*，原因為詞語搭配為一種特定的結合方式，近似於語言使用上的約定與習慣。Ilson 等學者將複雜的英語詞語搭配分為文法搭配(grammartical collocation)與詞彙搭配(lexical collocation)兩類[11]。文法搭配是指包含一個主要單字(dominant word)及一個介係詞、冠詞或連接詞的片語，例如， *determined by*。而詞彙搭配是指名詞、動詞、形容詞、副詞等特定字詞之間的慣用組合，例如， *compose music*。詞彙搭配沒有主要單字，也是 ESL/EFL 作者感到較為困難及容易發生錯誤的部分。

語料庫使用工具泛指結合語料庫資源對於語言使用上提供資訊的工具。語料庫資源，是學習語言上最有成效的參考資源，語料庫使用工具能針對指定的詞語，提供許多的使用方式與情境範例，讓使用者有較佳的參照與學習成效。以 *Word Sketch* 為例[12]，該系統藉由語料庫的使用，提供英語詞語使用情形描繪(lexical profiling)與搭配詞(collocation)顯著性的資訊。另一個語料庫使用工具的範例為美國 Brigham Young University 研發的 VIEW[13]系統，使用者可輸入完整的或部分的字詞及詞性標籤(POS)等，查詢的結果則以 concordance 的方式呈現。圖一顯示以 *determined by* 的詞語查詢的結果。

The screenshot shows the VIEW concordancer interface. The search string is 'determined by'. The results table is as follows:

DISTRIB	WORD/PHRASE	TOKENS REG1	PER MIL IN REG1 [100,000,000 WORDS]
1	DETERMINED BY	1675	16.75

Below the table, there is a 'KEYWORDS IN CONTEXT' section with 11 results:

- A04 artist's intentions, which, it can be argued, are **determined by** society. Or, the effect of a painting may t
- A04 masters, is more minute; and the composition, completely **determined by** a diagonal, is stricter than th
- A04 achievements. The content of an article will tend to be **determined by** its length; no newspaper article l
- A05 . Both novels see the world in colonial colours --; as **determined by** empires, in the furtherance of which
- A07 by the parents, the mode of religious instruction to be **determined by** them. It opened the following Sep
- A08 , he wrote. My sense of how to go on **determined by** the vividness of my imagination of what it will be ill
- A0B this blended surprisingly well with the company's corporate identity as **determined by** Milner Gray of the
- A0C a restaurant's success depends on cuisine. The rest is **determined by** decor, the welcome you give and t
- A0E its involvement with television mean that it becomes completely dominated and **determined by** financia
- A0K haunt the executive, who saw that control could best be **determined by** the installation of a professiona
- A0K transformations, creating homologies which reaffirm operational practice and which are **determined by** a

圖一：VIEW concordancer 的語言資訊提供範例

由相關研究可知，語言資源工具對於 ESL/EFL 作者的寫作協助是相當重要的，然而不論是字典/辭典或是目前的語料庫使用工具對於寫作需求上的滿足仍然是相當有限的。不同的 ESL/EFL 作者因其語言知識程度的不同及認知的不同，往往對於相同的表達意圖會使用不同的查詢條件，而目前的語言資源工具仍然缺乏足夠彈性的表達方式對應於各種 ESL/EFL 作者可能需要的查詢條件。舉例而言，當 ESL/EFL 作者想要諮詢有關 **by and large** 的使用方式，有的作者能正確的以完整的片語組合查詢。但是更多的作者卻對此片語的認知不足或不確定，甚至有所誤解，而嘗試使用 **large** 或 **by large** 等查詢條件。這些不正確的或是模糊的查詢條件，在目前的語言資源工具中並無法得到直接的及有效的協助。這也正是 ESL/EFL 寫作需求與目前的語言資源工具之間矛盾與吊詭之處，對英語的語言知識愈不足的 ESL/EFL 作者愈需要語言資源工具的協助，而使用語言資源工具的成效卻以語言知識掌握的程度為門檻，造成中等程度以下的 ESL/EFL 作者並無法在寫作的過程中，善用語言資源工具以提升寫作品質，甚至形成了學習進步上的瓶頸。

三、句子檢索與推薦

為了解決在語言知識不足的情形下使用現有語言資源工具的成效不彰的問題，我們提出了一個語言資訊檢索方法，能接納不正確的或是模糊的查詢條件，盡可能的找出可以參照的使用範例，讓使用者探索或確認其真正需要的表達方式。同時，在檢索結果的提供上，我們也對選取出來的使用範例，進行與使用者語言資訊需求的相關性評估，再根據其個別相關程度建立排序，以推薦的諮詢參照先後次序，嘗試讓使用者能以最有效率的方式確認其適合的表達方式。此方法讓使用者以彈性的查詢條件表達其不同程度的語言認知，搜尋檢索的對象是語料庫，查詢結果是以語料庫中的個別句子為單元排序呈現，提供使用者在一個完整的例句中觀察、學習及確認特定的表達方式。

我們將語言資訊檢索方法定義如下：給定使用者的語言資訊需求，句子檢索與推薦機制的目的是回傳一個經過排序的多個例句所組成的例句集合。此方法可分為三個模組：表達元素模組、檢索模組、及排序模組。表達元素模組是讓使用者能將其語言資訊需求及完整或部分認知，轉換成多個表達元素(expression elements)，並藉由彈性的組合搭配，提供寬廣的需求與認知的對應。檢索模組從表達元素的組合中轉換成相對應的查詢條件，從語料庫中比對選取可能符合使用者語言資訊需求的例句。而排序模組則針對檢索模組所選取的例句，評估其個別符合使用者語言資訊需求的程度並建立排序，再回傳此排序過的例句集合給使用者。

(一) 表達元素模組

表達元素模組必須彈性的允許不同語言程度的使用者表示其完整或部分認知，我們的方法目前涵蓋下列表達元素(expression elements)：

1. 準確字詞(exact words)：準確字詞讓使用者提供拼字無誤的英文單字或單字的組合表達其語言資訊需求的認知。
2. 單字的開頭/結尾(prefix/suffix)：單字的開頭/結尾讓使用者以單字的部份資訊來表達其不完整的認知。在 ESL/EFL 寫作過程中，使用者常因單字的完整拼字方法遺忘而無法有效率的進行諮詢。我們以"%"符號來代表單字開頭或結尾的表達方式。
3. 一個單字的萬用字(wildcard)：一個單字的萬用字讓使用者表示不確定或不特定的一個單字，當成欲諮詢的表達方式的一部分。我們以"# "符號來代表一個單字的萬用字元的表達方式。

4. 英語詞性(POS)：當語料庫含有英語詞性標籤時，使用者可以運用英語的詞性標籤當成諮詢條件。英語的詞性標籤種類繁多，本研究將語料庫的詞性標籤大致分為六類，分別為介係詞(PREP)、形容詞(ADJ)、名詞(N)、副詞(ADV)、動詞(V)以及其他詞性(Other)。
5. 不限定英語子句(subsequence)：不限定英語子句代表零個至多個不限定的英文單字序列，其考量為英語句型中常有特定片語結構結合不限定子句的情形，如 *either ... or* 或 *rather ... than*。我們以"*"來代表不限定單字序列的英語子句。

上述這些表達元素部分是以 *regular expression* 之概念為基礎，但提供了更多元的語言表達需求之空間，可以讓使用者依照其語言認知情形與表達需求，選用適當的表達元素，代表部份確認而部分設定範圍的查詢條件。例如，使用者要表達中文之「濃茶」之意思，但不確定英語中表示「濃」之形容詞為何，就可以用 *ADJ tea* 來表達其認知與需求。另外，這些表達元素也可以被彈性的組合，形成一個表達元素序列，例如，*a pro% P* 或 *would rather V than V*。這代表使用者在較多的語言認知下，所提供的較充分的表達意圖，而構成更強的查詢條件限制。

(二)檢索模組

檢索模組的目的為從語料庫中選取符合使用者表達需求的例句。本研究同時採用完整比對和部分比對的選取法則，完整比對是要求語料庫的例句選取和使用者限定的查詢條件必需完整的吻合。部分比對則容許料庫的例句選取和使用者限定的查詢條件有部分的差異。我們預期 ESL/EFL 作者經常會有語意認知不完整或不正確的情形，若只依照使用者的限定條件進行完整比對，將導致於諮詢成效取決於使用者認知程度的矛盾。部分比對的選取方式提供了較大的彈性空間，容許使用者在不完整或不正確的認知下，也能獲得有用的參照資訊。

舉例而言，一個程度不高的 ESL/EFL 作者對於 *not only ... but also* 的片語結構可能只約略有 *only* 及 *also* 的認知。使用者可以提出 *only also* 的需求表達方式，經由部分比對的選取法則，也可以檢索出使用 *not only ... but also* 的例句。而在使用者的認知是錯誤的情形下，部分比對也可以選取出可能相關的例句，提供使用者判斷其認知是否正確的機會。例如，使用者以 *an university* 為查詢條件，其中"*an*"是錯誤的用詞，正確的用詞應為"*a*"，若以完整比對的方式，是無法搜尋出任何例句的，而部分比對則能找出含有 *an* 開頭而 *university* 在後的例句。使用者可以從例句的檢索結果評斷其原先認知的 *an university*，可能含有詞語用法搭配上的錯誤，以致於選取出的結果並無 *an university* 緊密相連的例句，而間接得到諮詢的協助。

(三)排序模組

排序模組的目的為針對選取出來的例句，評估其符合使用者的語意表達需求程度，並將之排序，以提升使用者參照例句的成效。本研究採用多重序列排列(Multiple Sequence Alignment – MSA)的技術[14]，進行查詢條件與選取例句之間的相關程度評估與排序。給予多個長度不同的序列(sequence) S_1, S_2, \dots, S_n ， $n \geq 3$ ，一個多重序列排列為相同長度的序列 A_1, A_2, \dots, A_n ，其中 A_1 對應 S_1 、 A_2 對應 S_2 ，依此類推到 n 。而 A_1, A_2, \dots, A_n 序列中允許元素之間出現「間隔」，以"-"表示。

假設有兩個序列 $S_1 = \text{CCAATA}$ 、 $S_2 = \text{CCAT}$ ，序列元素為 $\Sigma = \{A, C, T\}$ ，則排列的結果可能為 $S_1 = \text{----CCAATA}$ ， $S_2 = \text{CCAT-----}$ ，或 $S_1 = \text{CCAATA}$ ， $S_2 = \text{CCA-}$

T-。兩種結果均為 S_1 、 S_2 的 MSA，兩種排列方式之間的優劣或合宜程度，可經由計算所有位於同一相對位置元素的比對分數總和來評估。通常我們以一個置換矩陣 (substitution matrix) 來代表上述所有元素間的比對分數。而兩個序列之間的最優排列結果 (optimal MSA)，可以透過 Needleman-Wunsch algorithm [14] 求得。然而排序模組的工作需要對兩個以上的選取例句進行排列與評估，在運算時間的考量下，我們採用 the center star algorithm，以一個參考序列當基準用以比對其他的序列，尋求較佳解而不是最佳解。對於長度均為 k 的 n 個序列，the center star algorithm 的效能為 $O(k^2n^2)$ [15]。本研究以學習者輸入的 query 為參考序列，可省略尋找中心序列的步驟。

本研究將 MSA 的參數矩陣依照表達元素的模組分為 11 個比對單位，分別為英語詞性標籤 6 個：P、J、N、D、V、O (介係詞、形容詞、名詞、副詞、動詞、其他詞性)；以及準確字詞 (exact)、單字開頭/結尾 (prefix/suffix)、萬用單字 (wildcard) 和用以對應於不限定字串的間隔 (gap)。比對參數上的取捨以準確字詞最高 (100)、單字開頭片尾其次 (50)、詞性標籤次之 (25)、萬用單字最低 (5)，其考量為查詢條件的精準程度區分。若以 = 代表準確字詞，% 代表單字開頭/結尾，# 代表萬用單字，- 代表不限定字串的間隔，X 代表無關之單字。當學習者輸入的 query 為 S_0 ：a pro% P， S_0 的 MSA 代表元素為 "= % P"，選取的例句為 S_1 、 S_2 、 S_3 ：

- S_1 . This(X) posed(X) a(=) particular(X) **problem(%)** for(P) an(X) agent(X).
- S_2 . Listening(X) to(X) all(X) these(X) personal(X) accounts(X) has(X) had(X) a(=) **profound(%)** effect(X) on(P) us(X).
- S_3 . Increasingly(X) acid(X) rain(X) is(X) a(=) **problem(%)** in(P) Europe(X) too(X).

其中 S_1 、 S_2 、 S_3 粗體的單字依序為選取的 "a" (exact)、"pro%" (prefix) 和 "P" (介係詞)，括弧內的符號為 S_1 、 S_2 、 S_3 的 MSA 代表元素。以代表學習者的語意表達需求之查詢條件 (S_0) 為中心點，依據 the center star algorithm 排序，可得等長的序列 A_0 、 A_1 、 A_2 、 A_3 ，並以 sum-of-pair score 來計算排序後的分數，假設 A_1 相對於 A_0 的分數為 C_1 、 A_2 相對於 A_0 的分數為 C_2 、 A_3 相對於 A_0 的分數為 C_3 ，其中 $S(x,y)$ 代表 x 元素比對 y 元素的分數，我們可得以下的結果：

A_0	-	-	-	-	-	-	-	-	=	%	P	-	-	-
A_1	-	-	-	-	-	-	X	X	=	X	%	P	X	X
A_2	X	X	X	X	X	X	X	X	=	%	X	P	X	-
A_3	-	-	-	-	X	X	X	X	=	%	P	X	X	-

$$C_1 = S(-,-) + \dots + S(=,=) + S(%,X) + S(P, %) + S(-,P) + \dots + S(-,X) = 93$$

$$C_2 = S(-,X) + \dots + S(=,=) + S(%,%) + S(P, X) + S(-,P) + \dots + S(-,-) = 139$$

$$C_3 = S(-,-) + \dots + S(=,=) + S(%,%) + S(P, P) + S(-,X) + \dots + S(-,-) = 169$$

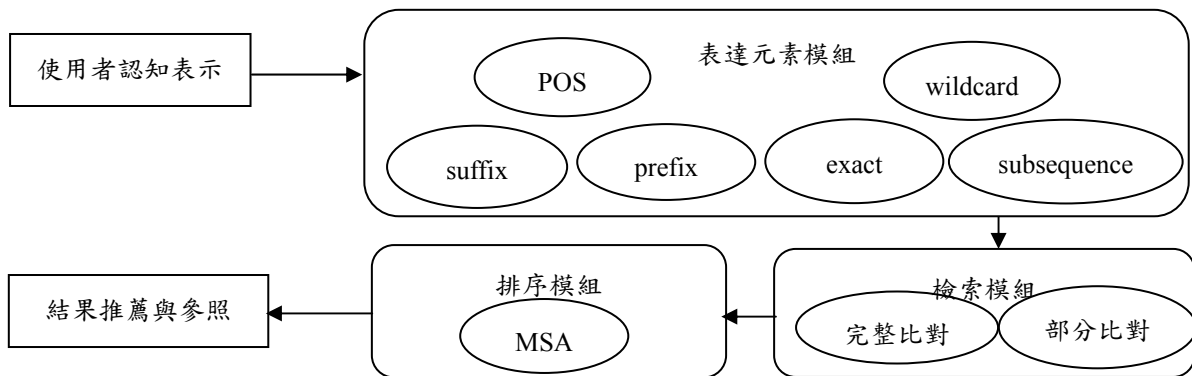
根據 sum-of-pair score 的分數，SAW 系統會推薦給學習者的例句依序為 S_3 、 S_2 、 S_1 ：

- S_3 . Increasingly acid rain is **a problem in** Europe too.
- S_2 . Listening to all these personal accounts has had **a profound effect on** us.
- S_1 . This posed **a particular problem for** an agent.

(四) 實作系統與處理流程範例

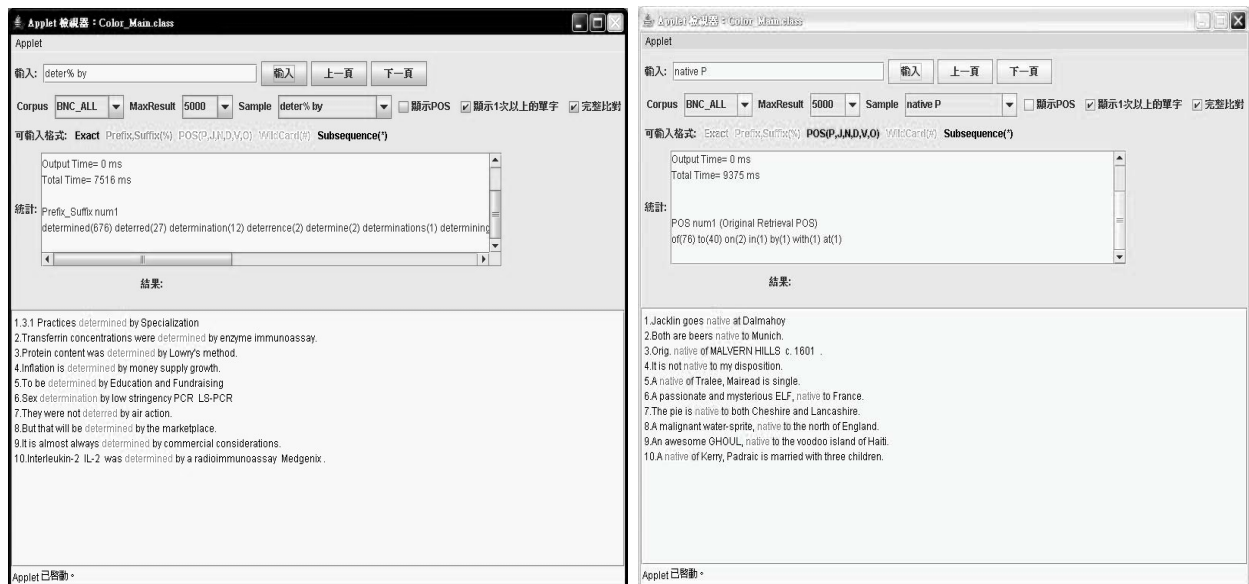
我們以上述的語言資訊檢索方法實際開發完成一個稱為 SAW (Sentence Assistance for Writing) 的雛型系統，其表達元素模組的部分包含了六種認知表示的基本方式；檢索模組則可分別進行完整比對與部分比對，以選取符合查詢條件的例句；最後，排序模組則

採用 MSA 演算法進行檢索出來的例句與查詢條件之間的排列方式並評估其相關程度，再依照相關性依序推薦呈現給使用者。SAW 系統的整體架構如圖二所示。



圖二：SAW 雛型系統架構圖

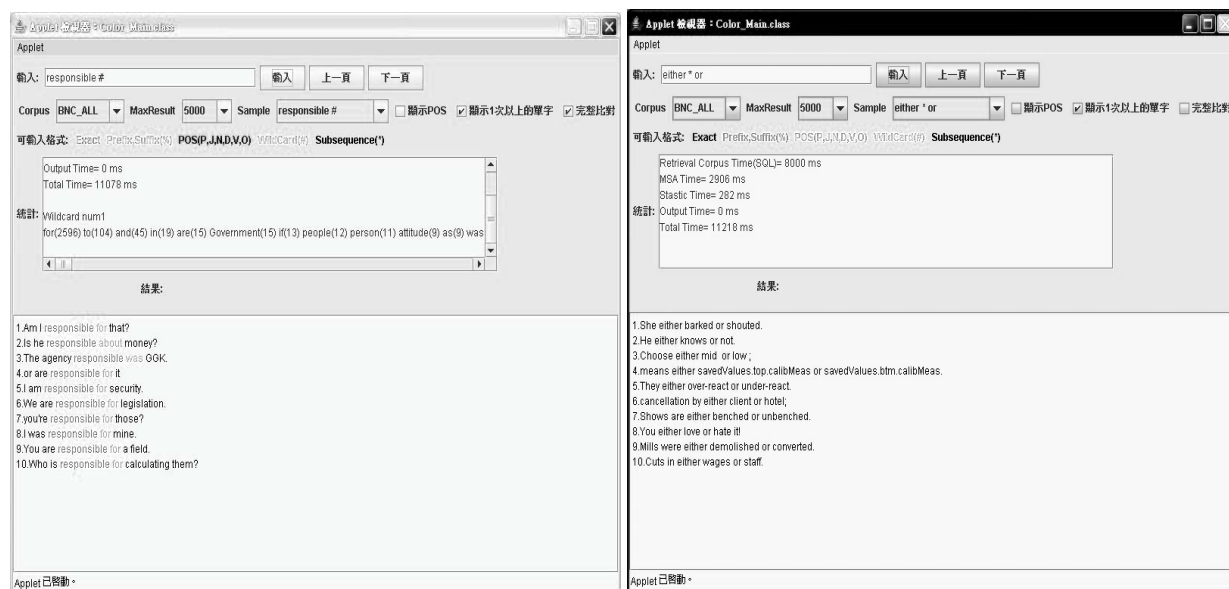
假設使用者想要搜尋 not only ... but also 的例句，而其認知的表示方式為 not only but also，四個緊密相連的英語單字，而語料庫中有三個相關例句，分別為 S₁：We must also make sure that future generations not only read, but also have a real enthusiasm for visiting bookshops and libraries. S₂：This was not only humiliating but also very awkward for Baldwin. S₃：This is not only easier, but also more fun. 首先，使用者藉由準確字詞表達其所認知的表示方式；檢索模組中的完整比對因語料庫中並無 not only 和 but also 緊密相連的例句，故無法搜尋出任何例句；而部分比對因容許不同單字間可以有間隔，而可選取出 S₁、S₂、S₃ 例句。而排序模組再根據整體例句的結構，例如以句子的長短為考量因素，將 S₁、S₂、S₃ 的例句排序為 S₃、S₂、S₁ 的順序以供使用者參照。



圖三：SAW 系統分別以 deter% by (左)及 native P(右)的查詢結果

最後，我們展示 SAW 系統的介面及運作結果，查詢的條件分別為 deter% by、native P、responsible #、either * or，亦即以 prefix、POS、wildcard、subsequence 等表達元素所組成的查詢條件當測試樣本。其中 deter% by、native P、responsible # 將採取完整比對的機制，either * or 則採取部分比對的機制，其結果畫面分別由圖三與圖四呈現。使用者可以選擇特定語料庫進行檢索，可以選擇完整比對或部分比對。檢索結果是精簡而完整的

例句，而非如圖一所示以上下文為主的 concordance，系統提供每頁 10 個例句，使用者可視需要檢視下頁的例句。另外，SAW 系統也提供了檢索結果的統計資訊。例如，deter% by 的檢索結果例句中，使用 determined by 的數量有 676 個，使用 deterred by 的數量有 27 個，使用 determination by 的數量有 12 個，使用 deterrence by 的數量有 2 個，使用 determine by 的數量有 2 個，使用 determining by 的數量有 1 個，等等。這些統計資訊也將提供給使用者更進一步判斷或探索的指引。



圖四：SAW 系統分別以 responsible # (左)及 either * or (右)的查詢結果

四、實驗評估

本實驗採取 BNC(British National Corpus)語料庫，該語料庫收錄的文章類別包括自然科學、社會科學、應用科學、哲學、藝術、科幻、金融、休閒、世界九大類，年代範圍為 1974-1994，共有約 350 萬個例句。語料庫中有內定的 POS 細目分類，共計 62 個。本實驗將 BNC 的 62 個細目 POS 分類對應為 6 個 POS 的大類。

(一)評估方法

為了客觀評估 SAW 系統對於 ESL/EFL 作者的參照成效，我們進行了兩種評估方法，分別為模擬測試(實驗客觀評估)和問卷調查(使用者主觀評估)。在模擬測試方面，我們從英語能力測驗的試題中，選取重要的英語表達方式(如單字、片語、句型結構等)，模擬 ESL/EFL 作者於寫作上可能遭遇的困難，再由實驗者提供數種不同的查詢方式，測試 SAW 系統的檢索結果，並觀察評估其效能。試題來源包括大學聯招和大學學測的考題，共計 16 期，及 TOFEL 的考題，共計 11 期。總共選取 45 個單字片語、12 個句型結構 (含有 subsequence 表達元素的句型)當試題樣本。本實驗從二個角度分析成效，第一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)，模擬不同認知程度的使用者，以評估對檢索結果所能得到的參照諮詢成效。第二個角度為句型結構和認知錯誤的評估：針對 12 個句型結構，模擬使用者在認知有部分錯誤的情況下，SAW 系統能給予使用者參照諮詢上的成效。

問卷調查的主要目的為以使用者的觀點評估 SAW 系統所提供的參照諮詢成效。我們選

取部分英語能力測驗試題，包括單選題 16 題及翻譯題 3 題，由受測者模擬英語能力測驗，再以分組測驗的結果，比較受測者於使用 SAW 系統協助和未使用 SAW 系統協助之間測驗成績上的差異。另一部份的比較為同一組的受測者，在未使用 SAW 系統前和使用 SAW 系統後的測驗成績差異。

為客觀並且確實的評估 SAW 系統的成效，本研究參考推薦系統的評估方式[16]，擬定 5 項觀察指標，詳述如下：

1. 試題測驗分數：本研究藉由試題模擬學習者於寫作上的困難，而試題具備了錯誤選項及正確的答案，因此，可藉由學習者的測驗分數來評估系統的參照諮詢成效。
2. 推薦度：對於系統提供的語言使用資訊，推薦度是使用者所認為真正實用的程度。本研究對於推薦度採取 4 種高低程度的量度，即高推薦度、中推薦度、低推薦度、無推薦度。推薦度是由使用者主觀評估的，本研究對於推薦度的衡量將以使用者的回應方式取得。
3. 滿意度：對於系統提供的語言使用資訊，滿意度是使用者對其整體滿意的程度。本研究對於滿意度採取 5 個高低程度的量度，即非常滿意、滿意、普通、不滿意、非常不滿意。滿意度亦是由使用者主觀評估的，本研究對於滿意度的衡量將以使用者的回應方式取得。
4. 符合度：符合度是依據使用者提供的查詢條件，系統比對選取出的結果和使用者預期答案上的相似度。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 retrieved / number of retrieved)，也就是一般資訊檢索中的準確度(precision)之定義。
5. 例句成本：例句成本為使用者閱讀完一定數量的單字序列所需花費的成本，實驗上以所需閱讀的英語單字字數為成本量。例句成本的評估為反應簡潔有力的例句，對於參照諮詢上的效果可能較好。

(二)實驗結果

1. 模擬測試之單字片語評估

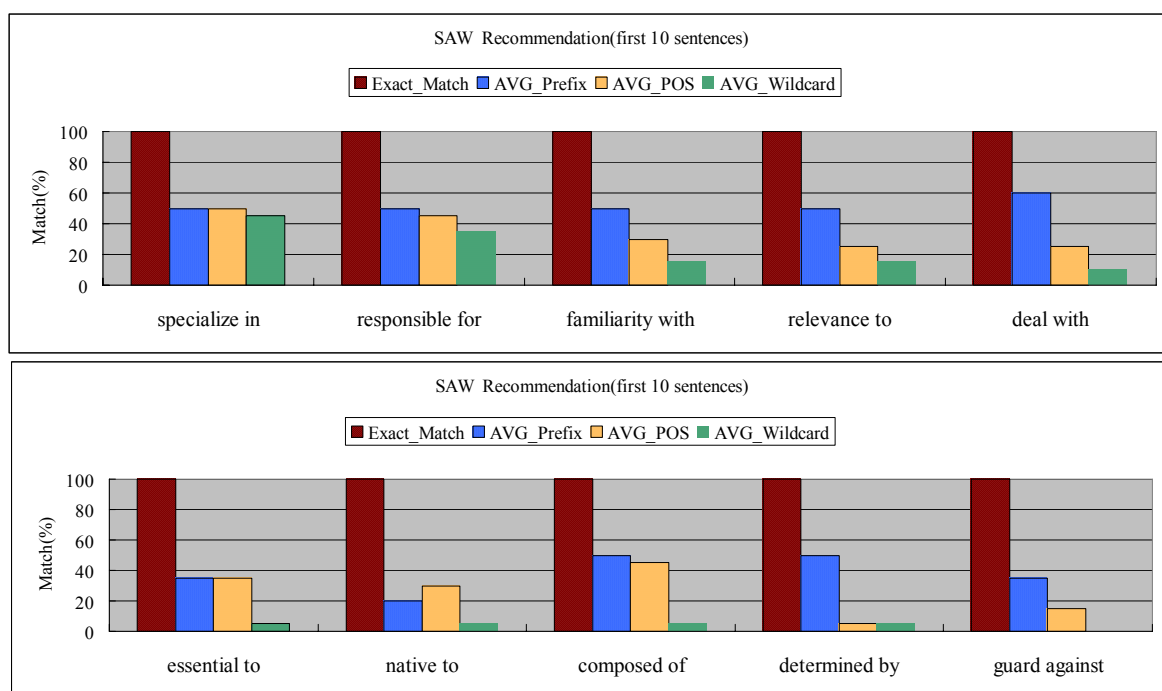
單字片語的評估主要藉由試題模擬學習者於寫作上所需的協助，並假設不同符合度的查詢條件，評估對認知程度不同的學習者所能提供的協助成效。實驗以 10 個由兩個英語單字組成的片語及 8 個由多個英語單字組成的片語當測試樣本，並採用 exact、prefix、POS、wildcard 等表達元素形成不同的查詢條件，模擬學習者可能的不同認知及所得到的不同參照結果，再以結果的符合度進行評量。

本實驗採用 4 大類型的查詢方式：(1)exact：完整提供兩個英語單字，如 specialize in。(2)prefix：將兩個英語單字其中一個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，再取此兩種查詢條件所得檢索結果的平均符合度。(3)POS：將兩個英語單字其中一個以 POS 表達元素查詢，如以 specialize in 為例，將分為 V in 和 specialize P 兩種可能的查詢條件，同樣的，取兩種不同查詢條件所得檢索結果的平均符合度。(4)wildcard：將兩個英語單字其中一個以 wildcard 表達元素查詢，如以 specialize in，將分為 # in 和 specialize # 的兩種可能的查詢條件，再取兩種不同查詢條件所得檢索結果的平均符合度。

實驗設定 SAW 系統每次提供 10 個例句，亦即須推薦出前 10 個例句供使用者參照。圖

五顯示 10 個由兩個英語單字組成的片語的資訊需求，對於 SAW 系統所提供的參照資訊之符合度評估。一如預期的，exact 查詢條件的結果符合度最高，為 100%。另外，prefix 查詢條件於此 10 個片語中符合度高於 POS 條件；而 POS 查詢條件符合度高於 wildcard 查詢條件，其中有一個例外為 native to 片語。這些結果約略說明 prefix 表達元素為限制較強的查詢條件，而 POS 次之，wildcard 最低。

除了兩個英語單字組成的片語之外，本研究亦測試了三個和四個英語單字組成的片語，評估上述不同查詢條件的檢索結果之符合度。和上述兩個英語單字片語查詢方式唯一不同的是，三個英語單字的 prefix 查詢方式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所得檢索結果的平均符合度。同樣的，以 a proportion of 為例，POS 查詢方式可能有 O proportion of、a N of、a proportion P 三種不同的查詢條件；而 wildcard 查詢方式可能有 # 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所得檢索結果的平均符合度。另外，關於四個英語單字的 prefix、POS、wildcard 查詢方式可能的查詢條件有四個，此時，實驗是取此四種查詢條件所得檢索結果的平均符合度。



圖五：單字片語的符合度評估

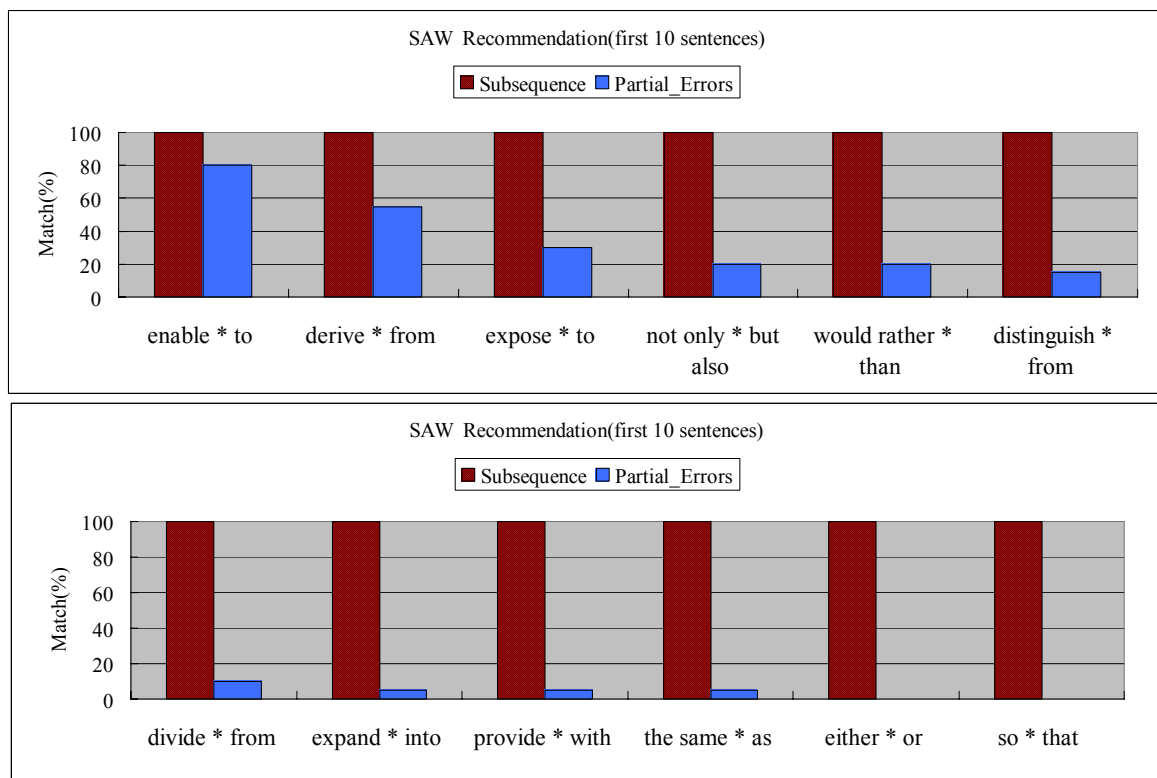
實驗結果顯示，查詢條件中包含的字數越多，SAW 系統提供的參照資訊的符合度越高，原因在於多個英語單字組成的片語，其本身結構性較強，在其中一個字為模糊的情況下，學習者還是能得到較針對其表達需求的參照協助。而結果亦顯示，prefix 查詢方式的成效高於 POS 查詢方式，POS 查詢方式成效高於 wildcard 查詢方式，顯示 prefix 表達元素為限制條件較強的查詢條件，而 wildcard 表達元素的限制條件最為寬鬆。

2. 句型結構和認知錯誤的評估

本實驗選取 12 個句型結構，包括：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 ，以評估 subsequence 表達元素的查詢方式之參照成效，並刻意對於 12 個句型結構中提供部分錯誤的查詢條件。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，並以平均符合度來描述系統的參照協助成效。此 12 個子句中，均有一個重要的關鍵介係詞，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。

實驗結果如圖六顯示，在句型結構查詢條件下 SAW 系統均能對學習者提供針對性的參照協助(符合度 100%)。在學習者部分錯誤的認知方面，查詢結果會因句型結構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參照協助。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，提供給學習者的參照符合度亦有限。但是，由問卷調查的結果顯示，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參照協助。



圖六：subsequence 和 partial errors 查詢條件結果之符合度評估

3. 問卷調查評估

試題回應有 19 道試題(16 題單選 3 題翻譯)， SAW 系統分別提供與個別試題相對應的 BNC 語料庫中的相關例句，做為受測者作答時的參照。藉由試題測驗分數上的差異，以及受測者提供的參照推薦度與滿意度，評估 SAW 系統給予受測者寫作參照協助的程度。實驗將受測者分為四組：第一組填答 19 道試題時，並無任何參照例句；第二組填答同樣的 19 道試題，而且有高符合度的參照例句；第三組填答同樣的 19 道試題，但提

供的為低符合度的參照例句。第四組施測時先在無參照例句的情形下填答同樣的 19 道試題，收回試題後，再於提供高符合度參照例句的情形下，填答同樣的 19 道試題。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[17]，可瞭解受測者對於英語字彙的認知度，亦約略代表其英文程度。受測的五專生對於英語常用 1000 單字約有 80.44%的認知度；對於英語常用 2000 單字有 46.10%的認知度。而根據 Nation 的研究[18]，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，問卷調查測試對象平均不超過 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，包括繳交白卷和猜題作答(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。

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答試題、20 份為使用系統較高符合度的參照例句填答試題(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答試題(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，證實 SAW 系統能給予學習者一定程度的寫作上的參照協助。此部分的相關問卷分析結果彙整於表一中。

表一：有無使用系統之問卷結果分析

	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)	---	2.78	2.70
翻譯試題推薦度(1~4)	---	2.80	
句型長度(1~3)	---	2.09	
滿意度(1~5)	---	2.76	

為了驗證上述結果並非偶然發生，我們以統計學的顯著水準(significance level)說明 SAW 系統確實有一定的協助成效顯著性。假設母體的受測分數為常態分佈，並以無使用系統參照的試題分數當母體的受測分數。假設 Null Hypothesis (H_0) 為 SAW 系統沒有成效(平均分數為 4.42)，Alternative Hypothesis (H_1)為 SAW 系統有成效(平均分數大於 4.42)，在顯著水準 5%下，對於高符合度參照例句的 20 份樣本，可得拒絕 H_0 的臨界分數為 $4.42+(1.65*1.95/(20)^{1/2}) = 5.14$ ，因實驗所得的平均分數為 6.69，大於臨界分數 5.14，所以可以否定 H_0 ，亦即在 5%顯著水準下 SAW 系統對於受測者是有協助成效的。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 $4.42+(1.65*1.95/(22)^{1/2})=5.10$ ，而使用 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成效的。

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。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答試題，42 份為系統給予同樣的參照例句而填答試題。翻譯試題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。結果顯示，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成效的。

推薦度是由使用者評估系統是否有實用性(utility)，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到 2000 字，BNC 語料庫的例句的程度較深，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 系統的參照結果推薦度不高，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。

表二：使用系統前後之間卷結果分析

	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)	---	2.59
翻譯試題推薦度(1~4)	---	2.53
句型長度(1~3)	---	2.16
滿意度(1~5)	---	3.16

在使用系統上前後的區別有效問卷 27 份中，單選試題共有 16 題，每題為 0.625 分，滿分為 10 分。相關問卷分析結果彙整於表二中。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。同樣的，若以使用系統前的結果為母體，使用系統後的結果為樣本，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成效的。

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；使用系統後的平均分數為 5.42，標準差為 2.60，分數增加為 0.86 分。同樣的，以使用系統前的結果當母體，而使用系統後的結果當樣本，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成效的。

(三)結果討論

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，學習者亦認為能從 SAW 系統中得到一定的參照協助，證實 SAW 系統對於寫作有其一

定的協助成效。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。

我們另外也以學生實際的寫作錯誤進行測試，如 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. 我們使用 SAW 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都可以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的七個例句得知正確的動詞形式，其符合度為 7/10，證實以上述兩種查詢方法，SAW 系統均能給予學習者相關的參照協助。

五、相關研究

近年來，運用資訊技術協助英語學習的研究引起極多的興趣，我們僅討論與本研究較為直接相關者。首先是 CMU 的 REAP 系統[19]，其主要目的為提供學習者英語閱讀的協助，研究方法包括依照學習者認知程度，以測驗的方式評估學習者的詞彙程度，在使用 REAP 系統前後，是否有顯著性的差異，並以問卷調查和統計數值整合分析 REAP 系統的成效。本研究參考了 REAP 系統的評估方式評估 SAW 系統的成效。另外，Gsearch 系統主要目的為在語料庫中建立句法準則(syntactic criteria)，而學習者可以依據建立的句法準則有效率地從語料庫中選取合宜的例句[20]。Gsearch 系統提供的句法準則較為彈性，對於不同的學習者能依據多樣的句法準則有效率地查詢。本研究延續其理念，但特別關注於認知不足情形下之表達方式查詢，而這也是 ESL/EFL 作者最為關鍵的困難。最後，TANGO 系統[21]提供中英雙語的詞語搭配關係抽取與查詢，尤其著重在動詞與名詞的相互搭配資訊。本研究提供較多元的需求表達與查詢方式，並對檢索結果進行評估與排序，以提升使用者參照的效率。同時，本研究以更嚴謹的方式評量系統的成效。

六、結論

本研究鑑於現有語言資源工具對於 ESL/EFL 學習者於英語寫作上提供的協助功能有限，致力於提供一種語言資訊檢索方法，能有效率地根據不同學習者的認知程度予以一定的參照諮詢協助，而此方法分為表達元素模組、檢索模組、排序模組三個部份。為了證實此方法的成效，我們實作完成 SAW 系統。SAW 系統可使用準確字詞(exact)、單字的開頭/結尾(prefix/suffix)、英語詞性(POS)、一個單字的萬用字元(wildcard)、不限定的子句(subsequence)等做為表達元素(expression elements)，並可使用完整比對或部分比對的兩種比對方式，於檢索模組中選取例句，而後對於選取的例句使用 MSA 技術為排序模組的評估方法，最後提供推薦的相關例句供學習者參照。

為了評估 SAW 系統的成效，實驗分為模擬測試(實驗客觀評估)和問卷調查(使用者主觀評估)兩種方式，藉由試題分數、推薦度、滿意度、符合度、句型成本五項指標，驗證在上述兩種評估方法下，SAW 系統均能得到一定程度的協助成效。除此之外，我們並依據學生的作文範例，將範例模擬輸入於 SAW 系統中，發現 SAW 系統亦能給予學習者一定的參照協助，更加驗證 SAW 系統的協助成效。綜觀以上幾點，本研究的語言資訊檢索方法確實能針對 ESL/EFL 學習者於英語寫作時給予適度的協助。

參考文獻

- [1] A. Aghbar, Fixed Expressions in Written Texts: Implications for Assessing Sophistication, East Lansing, MI: National Center for Research on Teacher Learning, ERIC Document Reproduction Service No. ED 352808, 1990.
- [2] T. McEnery and A. Wilson, *Corpus Linguistics*, Edinburgh University Press, Edinburgh, 1996.
- [3] S. Conrad, The Importance of Corpus-based Research for Language Teachers, *System* 27: 1-18, 1999.
- [4] A. B. M. Tsui, ESL Teachers' Questions and Corpus Evidence, *International Journal of Corpus Linguistics* 10(3): 335-356, 2005.
- [5] I. de O'Sullivan and A. Chambers, Learners' Writing Skills in French: Corpus Consultation and Learner Evaluation, *Journal of Second Language Writing* 15(1):49-68, 2006.
- [6] Jean-Jacques Weber, A Concordance and Genre-informed Approach to ESP Essay Writing, *ELT Journal* 55(1): 14-20, 2001.
- [7] Y. C. Sun, Learning Process, Strategies and Web-based Concordancers: a Case Study, *British Journal of Educational Technology* 34(5): 601-613, 2003.
- [8] H. Yoon, An Investigation of Students' Experiences with Corpus Technology in Second Language Academic Writing, Ph.D. Dissertation, The Ohio State University, USA, 2005.
- [9] C. C. Shei and H. Pain, An ESL Writer's Collocational Aid, *Computer Assisted Language Learning* 13(2): 167-182, 2000.
- [10] B. Altenberg and S. Granger, The Grammatical and Lexical Pattern of MAKE in Native and Non-native Student Writing, *Applied Linguistics* 22(2):173-195, 2001.
- [11] R. Ilson, B. Morton, and B. Evelyn, *The BBI Dictionary of English Word Combinations*, Amsterdam: John Benjamin Publishing Company, 1997.
- [12] A. Kilgarriff and D. Tugwell, WORD SKETCH: Extraction and Display of Significant Collocations for Lexicography, *Proceedings of the Collocations Workshop, ACL 2001*, pp. 32-38.
- [13] M. Davies, The Advantage of Using Relational Databases for Large Corpora, *International Journal of Corpus Linguistics* 10(3): 307-334, 2005.
- [14] 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* 48: 443-453, 1970.
- [15] Y. L. Francis, N. L. Ho, T. W. Lam, W. H. Wong, and M. Y. Chan, Efficient Constrained Multiple Sequence Alignment with Performance Guarantee, *IEEE Computational Systems Bioinformatics* 2: 337-346, 2003.
- [16] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, Analysis of Recommendation Algorithms for E-Commerce, *Proceedings of ACM Conference on Electronic Commerce*, Minneapolis, Minnesota, U.S.A., 2000.
- [17] English Word Tests, [Online]. Available: <http://www.lex tutor.ca/> [Accessed: May. 15, 2007].
- [18] Paul Nation, Measuring Readiness for Simplified Material: A Test of the First 1,000 Words of English, *In Simplification: Theory and Application RELC* 31:193-203, 1993.
- [19] M. Heilman, K. Collins-Thompson, J. Callan, and M. Eskenazi, Classroom Success of an Intelligent Tutoring System for Lexical Practice and Reading Comprehension, *Proceedings of the Ninth International Conference on Spoken Language Processing*, Pittsburgh, USA, 2006.
- [20] S. Corley, M. Corley, F. Keller, M. Crocker, and S. Trewin, Finding Syntactic Structure in Unparsed Corpora: The Gsearch Corpus Query System, *Computers and the Humanities* 35(2): 81-94, 2001.
- [21] T. C. Chuang, J. Y. Jian, Y. C. Chang, and J. S. Chang, Collocational Translation Memory Extraction based on Statistical and Linguistic Information, *Computational Linguistics and Chinese Language Processing*, Vol. 10, No. 3, September 2005, pp. 329-346.