

行政院國家科學委員會補助專題研究計畫成果報告

※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※

※

※

※

智慧型電腦輔助翻譯與教學--

※

※

網路中英語料庫對應句檢索系統之建構

※

※

※

※

※

※※※※※※※※※※※※※※※※※※※※※※※※※※※※※※

計畫類別：個別型計畫

計畫編號：NSC 89-2411-H-002-074

執行期間：89年8月1日至90年7月31日

計畫主持人：高照明

zmgao@ccms.ntu.edu.tw

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

執行單位：台灣大學外國語文學系

中華民國 90 年 10 月 15 日

行政院國家科學委員會專題研究計畫成果報告

智慧型電腦輔助翻譯與教學--

網路中英語料庫對應句檢索系統之建構

Building an Intelligent Web-Based Bilingual Chinese-English Concordancer for Computer-Assisted Language Learning and Translation

計畫編號：NSC 89-2411-H-002-074

執行期限：89年8月1日至90年7月31日

主持人：高照明 執行單位：台灣大學外國語文學系

計畫參與人員：台灣大學資訊工程研究所研究生 林桂光 梁菁秀

中文摘要 本計劃利用中英平行語料庫建構一個可應用於電腦輔助語言教學與翻譯的網路程式。此程式可從平行語料中自動擷取包含中英文關鍵詞的句子及其對應的翻譯，並可以擷取與輸入的中英文句子最近似的翻譯。

關鍵詞：中英平行語料庫，電腦輔助語言教學，電腦輔助翻譯，中英雙語對應句檢索

Abstract : In this project, a web-based bilingual concordancer is developed which can retrieve example sentences containing a keyword and their corresponding translations from parallel Chinese-English corpora. This useful and practical tool is capable of retrieving and ranking relevant examples and their translations in terms of their similarity to an input sentence and can be fruitfully used in computer-assisted language learning and translation.

Keywords : bilingual concordancer, parallel Chinese-English corpora, computer-assisted language learning, computer-assisted translation

A perennial problem for language learners and translators is that they are often

confounded by which word they should use to convey an idea. Monolingual and bilingual dictionaries are not very helpful to resolving this problem, because many expressions involve word combinations such as collocations and lexical phrases, which are either difficult to find or not listed in the dictionaries. For second language learners, the task of searching for the right expressions in the target language is like looking for a needle in a hay. It is time-consuming, painful, and often turns out to be in vain.

With the advent of the world wide web and the development of corpus linguistics, a new approach to tackling the problem has emerged. This promising new approach draws on a bilingual concordancer which can retrieve bilingual examples from large bilingual corpora. Unlike a traditional monolingual concordancer, it allows users to query an unfamiliar word in both the source and target language. This greatly facilitates language learners in looking for appropriate expressions.

The origin of such a tool may in part be attributed to a technical report by Martin Kay at Xerox in 1980, which was not published until 1997 (Kay (1997)). In that seminal paper, Kay proposed a computer-aided translation system which can access previous translation examples created by professional translators. Similar

idea was reintroduced by Harris (1988), who suggested using "bitexts" (i.e. translation equivalents coupled together unit by unit) as a computer-aided tool for professional translators. It was not until the mid 1990s that Kay's idea was finally realized (cf. Dagan and Church (1997)). Unfortunately, this kind of powerful tool is not widely used in Taiwan. Part of the reason is that it presupposes the correct identification of bilingual correspondences of the source and target language at the sentence level. In computational linguistics, this procedure is called sentence alignment and has been actively researched in recent years. Approaches to sentence alignment can be categorized into statistics-based, dictionary-based, and hybrid method, none of which performs well in Chinese-English bilingual corpora according to the experiments reported in Gao (1997).

A simple method to circumvent the problem of sentence alignment is to look for bilingual texts where there are clearly identifiable markers for text alignment. The internal structure of Bible makes it a good resource for aligning texts at paragraph level, because each paragraph (more precisely passage) in Bible is marked by a numerical indicator. The difficult task of text alignment is thus simplified to matching numerical indicators in the English Bible and the Chinese Bible. Figure 1 shows the user interface of the bilingual concordancer using Chinese and English Bible. Figure 2 is the output of a query.

Another important Chinese-English bilingual resource is Sinorama Magazine. Articles in the Sinorama Magazine are written in Chinese and translated into English, Spanish, and Japanese. The English translations are all checked by native speakers of English. Despite its unauthentic nature, the English translations are important resources for Chinese learners of English who wish to learn English expressions related to Chinese culture or events that takes place in Taiwan.

Sentence alignment in the Sinorama bilingual magazine is complicated by several factors. As with most translations, additions

and omissions of sentences or even paragraphs are quite common in the Sinorama magazine. Another difficulty is caused by the abundance of nonliteral translations in the Sinorama magazine. In the absence of a highly robust algorithm of text alignment, we perform paragraph alignment using indentation as paragraph markers. We assume that paragraph correspondence can be established if the bilingual texts have the same number of paragraph markers. Our approach is of course subject to errors, because it ignores the fact that there might be additions or omissions involved in the translation process. Furthermore, indentation might not be a reliable clue for paragraph markers. In inspecting the results of our program, we do find quite a few errors in paragraph alignment. To remedy this inadequacy, we provide hyperlinks for users to view the source and target texts. To facilitate browsing the bilingual texts, the keyword is highlighted and the paragraph containing the keyword is highlighted in another color. The proposed corresponding paragraphs are also highlighted for ease of reference (Figure 4). As the program is designed for advanced learners of English, it is reasonable to expect that they are capable of deciding whether the aligned paragraphs are correct. If they find errors in the alignment of paragraphs, they can view the whole texts and find the correct alignment around the paragraph proposed by the program. The tool, therefore, is still very useful even though it is error-prone.

As illustrated, a bilingual concordancer can accept a query word in the source or target language and present example sentences (or paragraphs) containing the input word juxtaposed with their (sentence or paragraph) translations. As expressions might take the form of a phrase or a sentence, it is more desirable to design a bilingual concordancer which can take a sentence as input. But given the small size of Chinese-English bilingual corpora available, the chance of finding examples identical to the input sentence is slight. To make the tool useful, it should have the capability of extracting and ranking examples in terms of

their relevance to the input query. With this powerful tool, translators and language learners can use the intelligent bilingual concordancer to find the closest match to the input sentence and its translation.

The central question of building such a program is how to compute the similarity between a sentence and an input sentence. The following procedure is proposed for an intelligent Chinese-English bilingual concordancer.

1. **Language Identification:** A program is written to check if the input is Chinese or English.
2. **Tokenization:** As there is no delimiter between words in Chinese, a word segmentation program is required to process the Chinese input as well as the Chinese corpus.
3. **Lemmatization:** Through lemmatization, English words with inflection (whether regular or irregular) will be converted into their lemmas (i.e. basic forms).
4. **Indexing:** All the words in the Chinese and English texts are indexed so that sentences containing a particular word can be retrieved in no time.
5. **Term Weighting:** Every Chinese and English word is assigned a weighting depending on its frequency. Content words are given a weight based on Inverse Document Frequency (IDF), which assumes that the importance of a word is in inverse proportion to the number of documents in which it occurs. In other words, the more documents in which a word occurs, the lower IDF value it has. IDF captures the generalization that words of low indexing value such as function words occur in most documents.
6. **Scoring Function:** All sentences in the bilingual texts which contain a matched word with the input sentence are retrieved. The similarity of the input sentence and the retrieved sentences is computed based on the sum of the IDF of the matched terms. The sentence with the highest score is considered most similar to the input query.

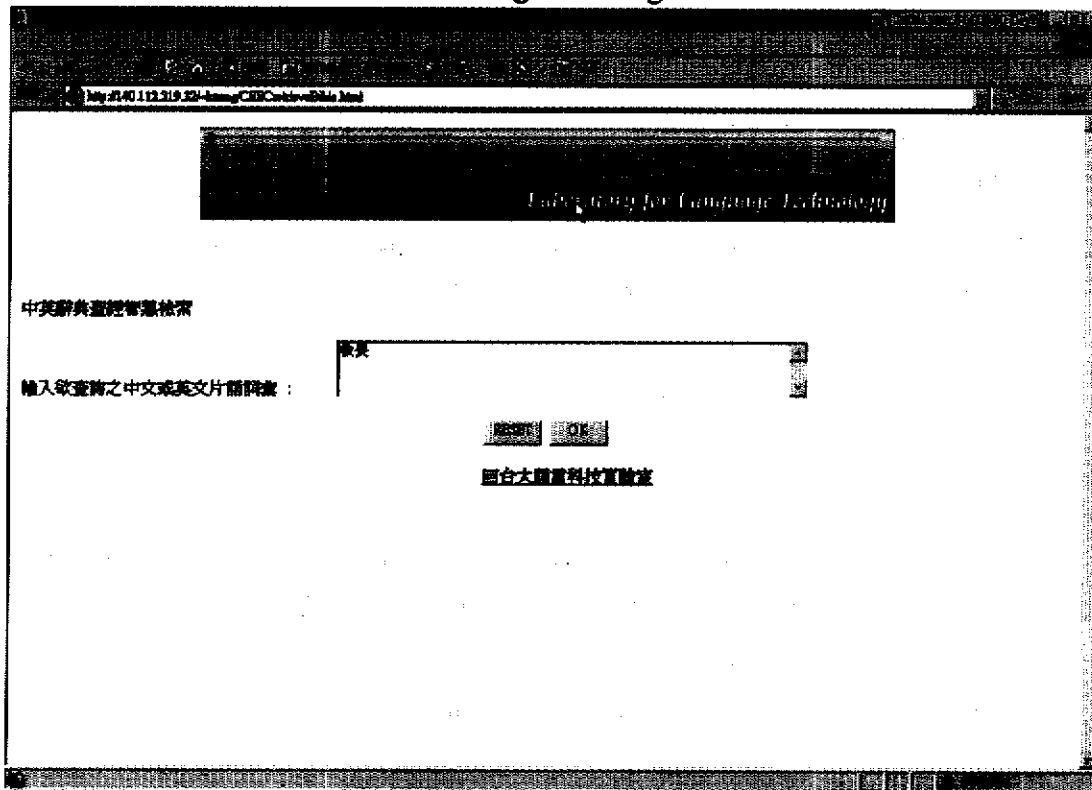
The algorithm above proposes a simple measure to calculate semantic similarity, which is used in building an intelligent web-based Chinese and English bilingual concordancer. To test its performance, we create a bilingual database by extracting bilingual sentences from a bilingual dictionary. The interface and the output of the program are shown in Figure 5 – 7. As can be seen from Figure 6 and 7, the program works reasonably well and is a valuable and convenient tool to search for appropriate English expressions. To enhance its functions, more Chinese-English bilingual corpora will be collected. Meanwhile, more robust and sophisticated methods of performing sentence alignment and calculating semantic similarity will be experimented in our next NSC project.

References

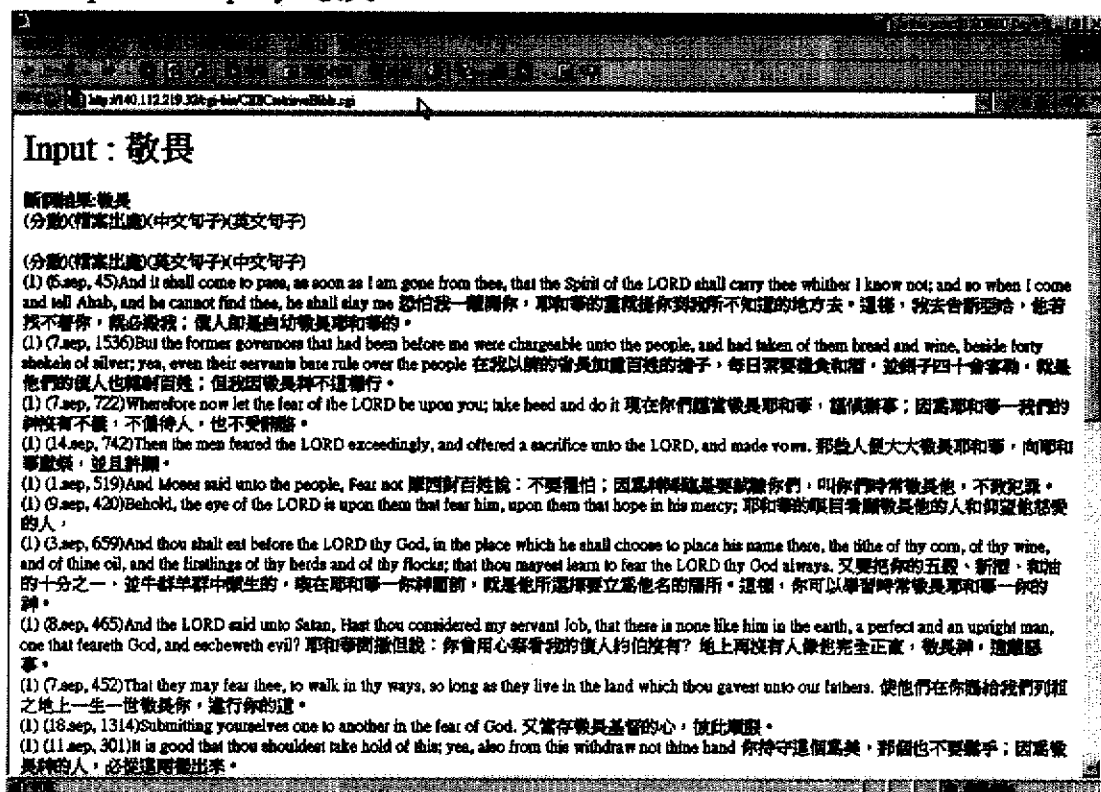
- Dagan, I., Church, W, and Gale, W. (1993) "Robust Bilingual Word Alignment for Machine Aided Translation." In Proceedings of the Workshop on Very Large Corpora: Academic and Industrial Perspectives, pp. 1-8, Ohio.
- Dagan, I. (1996) "Bilingual Word Alignment and Lexicon Construction." Tutorial paper given at the International Conference on Computational Linguistics, Copenhagen.
- Fung, P. and Church, K. (1994) "K-vec: A New Approach for Aligning Parallel Texts." Proceedings of the International Conference of Computational Linguistics, pp.1096-1102, Kyoto.
- Fung, P. and McKeown, K. (1994) "Aligning Noisy Parallel Corpora Across Language Groups: Word Pair Feature Matching by Dynamic Time Warping." Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation, pp. 81-88.
- Fung, P. and McKeown, K. (1997) "A Technical Word- and Term-Translation Aid Using Noisy Parallel Corpora Across Language Groups." Machine Translation, Vol. 12, Nos. 1-2., pp. 53-87.
- Gao, Zhao-Ming. (1997) Automatic

- Extraction of Translation Equivalents from A Parallel Chinese-English Corpora. Ph.D. Thesis. Department of Language Engineering. University of Manchester Institute of Science and Technology.
- Hutchins, J. (1998) The Origins of the Translator's Workstation. *Machine Translation*, Vol 13, No. 4, pp. 287-307.
- Haruno, M. and Yamazaki, T. (1996) "High-Precision Bilingual Text Alignment Using Statistical and Dictionary Information." *Proceedings of Annual Conference of the Association for Computational Linguistics*, pp. 131 -138.
- Jones, D. and Somers, H. (1995) "Bilingual Vocabulary Estimation from Noisy Parallel Corpora Using Variable Bag Estimation." In *JADT III Giornate Internazionali di Analsi Statistica dei Dati Testuali*, pp. 255-262, Rome.
- Kay, M. and Röscheisen, M. (1993) "Text-Translation Alignment." *Computational Linguistics*, Vol. 19, No 1, pp 121-142.
- Kay, M. (1997) "The Proper Place of Men and Machines in Language Translation", *Machine Translation*, Vol. 12, No. (1/2), pp.3-23, 1997
- Ker, S.-J. and Chang, J.-S. (1997) "A Class-based Approach to Word Alignment." *Computational Linguistics*, Vol. 23., No. 2, pp. 313-343.
- Kumano, A. and Hirakawa, H. (1994) "Building an MT Dictionary from Parallel Texts Based on Linguistic and Statistic Information." in *Proceedings of International Conference on Computational Linguistics*, pp. 76-81, Kyoto.
- Macklovitch, M. and Hannan, M. (1998) "Line 'Em Up: Advances in Alignment Technology and their Impact on Translation Support Tools", *Machine Translation*, Vol. 13, No. 1, pp. 41-57.
- Melamed, D. (1995) "Automatic Construction of Clean Broad-Coverage Translation Lexicons." In *Proceedings of 2nd Conference of the Association for Machine Translation in the Americas*, Montreal.
- Somers, H. and Ward, A. (1996) "Some More Experiments in Bilingual Text Alignments." In Oflazer, K. and Somers, H.(eds) *Proceedings of the Second International Conference on New Methods in Language Methods in Language Processing*, pp. 66-78, Ankara.
- Utsuro, T. et al. (1994) "Bilingual Text Matching Using Bilingual Dictionary and Statistics." in *Proceedings of International Conference on Computational Linguistics*, pp. 1076-1082, Kyoto.
- Wu, D. and Xia, X. (1995) "Large-Scale Automatic Extraction of an English-Chinese Translation Lexicon." *Machine Translation*, Vol. 9, pp. 285-313.

1. The User Interface of the Chinese-English Bilingual Concordancer



2. Output of the query “敬畏”



3. Output of the query “風潮”using Sinorama Bilingual Magazine

http://43.113.229.128/cgi-bin/CIS/Content/Text_List

Input : 風潮

新聞結果查詢

(分檔)(檔案出處)(中文母子)(英文母子)

(1) (8901098.asp, 68) 風潮風潮, 如今似已告一段落。真箇未來, 景氣的谷底一轉還不易度過。沈潛靜賞, 重慶正確心態, 才是台灣經濟力求自保的不二法門吧。

[看中文全文](#) [看英文全文](#)

(1) (8901073.asp, 8) 陳水扁會中樞下馬, 確實出乎許多人的意料, 以市府副議長高票度高達七成的市政選票成就連署的陳水扁, 得票率較上屆增加了一、三三個百分點。不過由於阿扁市長權力被電玩、拼球遊戲、博奕、賭公娼等娛樂作風, 以及與國會爭奪不斷, 不但引起八大行黨以及阿扁陣營團體人士的反感風潮, 也使得許多選民對於阿扁個人的領導作風持保留態度。再加上國新圖、新兩黨黨源合流的壓力, 阿扁最終仍以百分之四十五、九一的高得票率落選。 The highly charged emotions did not abate with the end of the election. After Chen Shui-bian's defeat was confirmed, many inconvertible supporters gathered outside his election headquarters to wish him well, and have since attended a number of moving post-election rallies which have been held across the island to thank people for their support. At the same time, Chen Shui-bian has received much encouragement and support to stand for president at the next election due in the year 2000. Faced with the concerned voice of his supporters, Chen has said that the very fact that many regarded the recent poll as a preamble to the presidential election did not help his campaign. Nevertheless he thanked everyone for their support and is now resting, carefully considering the next step in his political career.

[看中文全文](#) [看英文全文](#)

(1) (8907002.asp, 14) 問題是, 每個人都看到這帶令人驚嘆的成就, 可是它不是一種簡單的行條公式可以隨意複製成功, 事實上, 從商品向米進軍的作, 這個商業價值開始運作了, 好萊塢的年度大片、中繼片和小成本製作, 本來就有所區隔, 以年度大片爭相題材、籌備、佈局、戰術的佈局, 漸漸向內流轉, 一出中就帶動風潮, 真是已到了地球村世界同步的境界。 As ordinary people in modern society, fads-whether they be from Hollywood or Japan (as we relate in this month's stories on Taiwan's Japanophiles)-affect our daily lives and thoughts. But a deeper look at them often leads to thoughts of cultural conflict and cultural convergence. In China, these are the very themes that trace back to the May Fourth Movement 80 years ago, on which we offer a retrospective in this issue. These issues are list from entitled today.

[看中文全文](#) [看英文全文](#)

(1) (8907006.asp, 20) 在香港電影界有二十多年資歷的導演吳宇森, 一九八六年以《英雄本色》名震江湖, 掀起了男性情懷英雄片風潮。之後, 吳宇森又拍了《喋血街頭》、《喋血街頭》等多部英雄片在港院線中掙利, 但隨著香港電影進入改革期, 久久無法平復的阿爾登電影。 One day in 1989, John Woo's agent Terence Chang received an international call from a script reader at TriStar Studio, inquiring if Woo would be interested in doing some work in the US. But the Hong Kong film industry was riding high at the time, and Woo's response was: "Who needs Hollywood?"

[看中文全文](#) [看英文全文](#)

4. Users can view the whole texts by clicking the hyperlinks.

http://43.113.229.128/cgi-bin/CIS/Content/Text_List

Input : 風潮

新聞結果查詢

(分檔)(檔案出處)(中文母子)(英文母子)

(1) (8901098.asp, 68) 風潮風潮, 如今似已告一段落。真箇未來, 景氣的谷底一轉還不易度過。沈潛靜賞, 重慶正確心態, 才是台灣經濟力求自保的不二法門吧。

[看中文全文](#) [看英文全文](#)

(1) (8901073.asp, 8) 陳水扁會中樞下馬, 確實出乎許多人的意料, 以市府副議長高票度高達七成的市政選票成就連署的陳水扁, 得票率較上屆增加了一、三三個百分點。不過由於阿扁市長權力被電玩、拼球遊戲、博奕、賭公娼等娛樂作風, 以及與國會爭奪不斷, 不但引起八大行黨以及阿扁陣營團體人士的反感風潮, 也使得許多選民對於阿扁個人的領導作風持保留態度。再加上國新圖、新兩黨黨源合流的壓力, 阿扁最終仍以百分之四十五、九一的高得票率落選。 The highly charged emotions did not abate with the end of the election. After Chen Shui-bian's defeat was confirmed, many inconvertible supporters gathered outside his election headquarters to wish him well, and have since attended a number of moving post-election rallies which have been held across the island to thank people for their support. At the same time, Chen Shui-bian has received much encouragement and support to stand for president at the next election due in the year 2000. Faced with the concerned voice of his supporters, Chen has said that the very fact that many regarded the recent poll as a preamble to the presidential election did not help his campaign. Nevertheless he thanked everyone for their support and is now resting, carefully considering the next step in his political career.

[看中文全文](#) [看英文全文](#)

(1) (8907002.asp, 14) 問題是, 每個人都看到這帶令人驚嘆的成就, 可是它不是一種簡單的行條公式可以隨意複製成功, 事實上, 從商品向米進軍的作, 這個商業價值開始運作了, 好萊塢的年度大片、中繼片和小成本製作, 本來就有所區隔, 以年度大片爭相題材、籌備、佈局、戰術的佈局, 漸漸向內流轉, 一出中就帶動風潮, 真是已到了地球村世界同步的境界。 As ordinary people in modern society, fads-whether they be from Hollywood or Japan (as we relate in this month's stories on Taiwan's Japanophiles)-affect our daily lives and thoughts. But a deeper look at them often leads to thoughts of cultural conflict and cultural convergence. In China, these are the very themes that trace back to the May Fourth Movement 80 years ago, on which we offer a retrospective in this issue. These issues are list from entitled today.

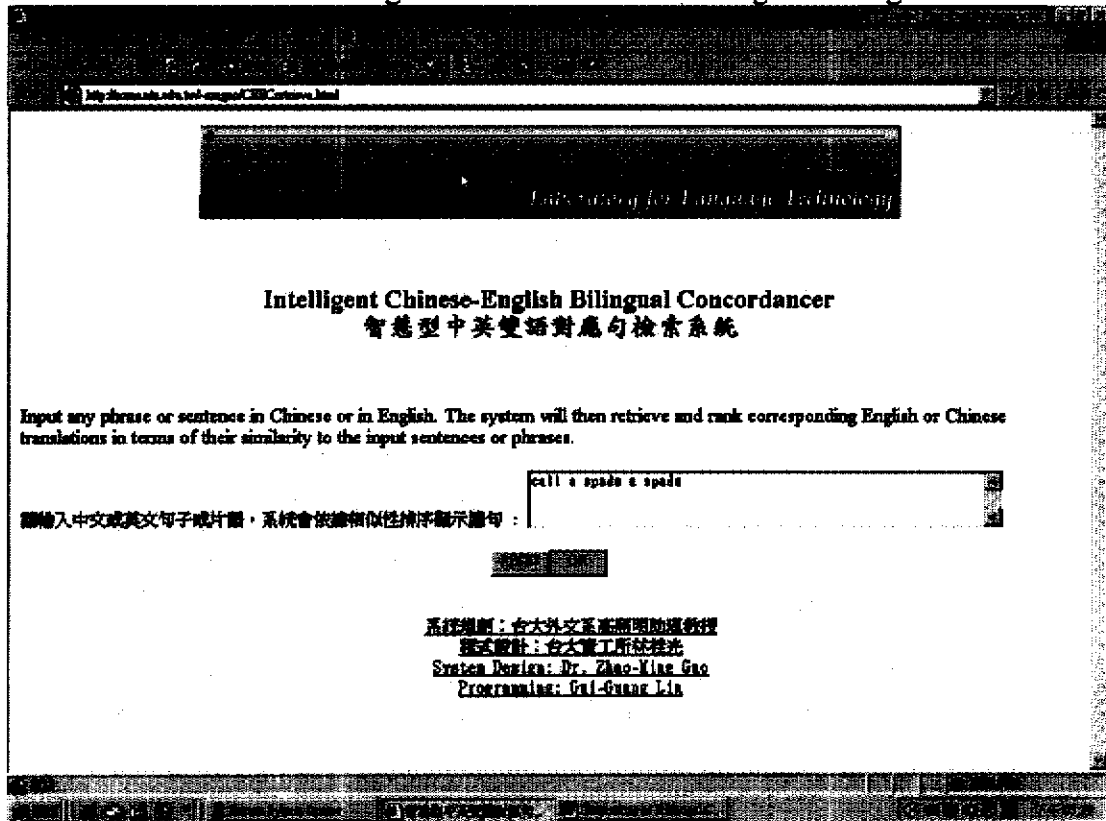
[看中文全文](#) [看英文全文](#)

(1) (8907006.asp, 20) 在香港電影界有二十多年資歷的導演吳宇森, 一九八六年以《英雄本色》名震江湖, 掀起了男性情懷英雄片風潮。之後, 吳宇森又拍了《喋血街頭》、《喋血街頭》等多部英雄片在港院線中掙利, 但隨著香港電影進入改革期, 久久無法平復的阿爾登電影。 One day in 1989, John Woo's agent Terence Chang received an international call from a script reader at TriStar Studio, inquiring if Woo would be interested in doing some work in the US. But the Hong Kong film industry was riding high at the time, and Woo's response was: "Who needs Hollywood?"

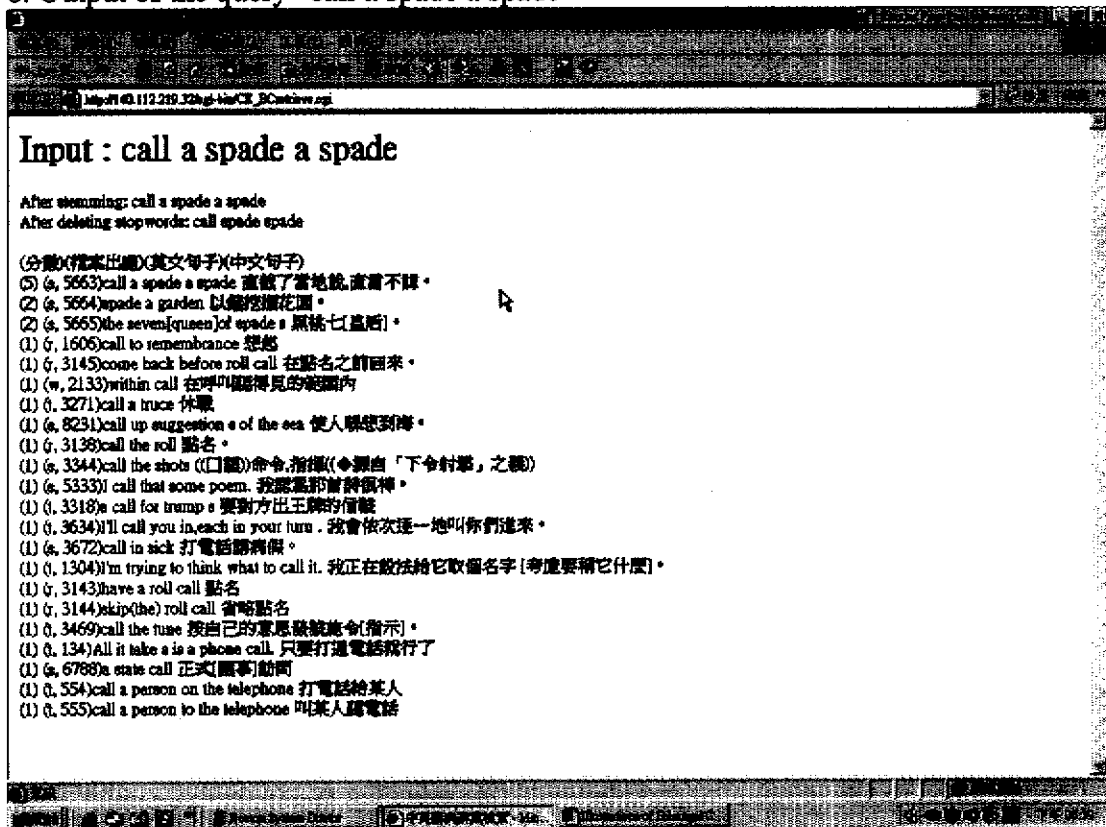
[看中文全文](#) [看英文全文](#)

<p>三合一選舉地區政治版圖 ——北京幾天, 國民黨立與民主進步 黨、陳水扁</p>	<p>令人驚嘆的成就, 真是已到了 month's stories cultural convergence s issue. These are:</p>	<p>Reviewing the Political Map Mayors Change in Taipei and Kaohsiung, KMT Takes Control of the Legislative Yuan</p>	<p>複製成功, 事實上, 從商品向米 年度大片爭相題材、籌備、佈局、 in modern society, fads-whether they thoughts. But a deeper look at them ck to the May Fourth Movement 80</p>
<p>包括台北、高雄兩市第二屆市員及市 議員與第三屆立院的三合一選舉, 在 民國八十七年十二月五日開幕, 北高 兩市與民選, 市議員與市議員不 大, 台北市議會以維持三合一選中, 選出立院部分, 國民黨與民主進步黨 競選。</p>	<p>香港的導演吳宇 森) 等多部英雄 片 Terence Chang But the Hong K</p>	<p>December 5th 1998 marked an im-portant step in the democratization of Taiwan, with the second election for the mayors and city councils in Taipei and Kaohsiung and the fourth election for the Legislative Yuan. The mayoral elections in both Taipei and Kaohsiung saw incumbents</p>	<p>掀起了男性情懷英雄片風潮。之 情緒入改革期, 久久無法平復的 at TriStar Studio, inquiring if Woo Woo's response was: "Who needs</p>

5. User Interface of the Intelligent Web-Based Chinese-English Bilingual Concordancer



6. Output of the query "call a spade a spade"



7. Output of the query “考試不及格”

http://10.112.219.32/gb-hk/CE_E/Carbivw.cgi

Input : 考試不及格

考試-5.91 不及-7.20 格-5.42
(分數)(標準出處)(中文句子)(英文句子)
(18.53)(f,140) 失敗的考試 考試不及格。 a failed examination
(18.53)(f,2067) 他那一科考試不及格。 He flunk ed the course.
(18.53)(f,118) 瑪麗考試不及格。 Mary fail ed in her exams.
(12.62)(e,1893) 考試及格[不及格] pass [fail] an examination
(12.62)(e,3667) 他對於考試不及格覺得很沮喪。 He was very sick at failing in the examination.
(12.62)(e,828) 他為了考試而努力讀書,結果只落得不及格。 He studied hard for the exam only to fail.
(12.62)(f,2493) 她預感自己會考試不及格。 She forebode d her failure in the exam.
(12.62)(e,1334) 他努力讀書[工作],否則他當時一定會不及格[失敗] He worked hard; otherwise he would have failed.
(12.62)(f,1) 他的歷史成績得了劣等[不及格]。 He got an F in history.
(11.33)(f,1592) 我考試及格 I got through the examination.
(11.33)(p,533) 我們老師說我們大部分的人考試都及格。 Our teacher pass ed most of us.
(11.33)(a,1382) 她考試及格了,所以現在興高采烈。 walking on air .
(11.33)(e,1085) 他僅以及格分數地通過(英文)考試。 He barely scrape d through(in English).
(7.20)(e,3506) 完全不及物[及物] 動詞。 a complete intransitive [transitive] verb
(7.20)(w,2239) 他迫不及待地接受該提議,我對此並不感到神奇。 I don't wonder at his jumping at the offer.
(7.20)(b,788) 不及物動詞的過去分詞構成完成式[已... 5 [be+
(7.20)(v,313) 不及物[及物] 動詞 an intransitive [a transitive] verb
(7.20)(h,126) 他不及她一半用功。 He is not half as hard working as she.
(7.20)(e,1926) 不及物動詞 an intransitive verb = a verb -
(7.20)(a,333) 它遠不及以前漂亮 It is not nearly so pretty as it was before.
(7.20)(u,1076) 這篇小說不及他的最佳作品 This novel isn't up to his best.
(7.20)(o,684) 他迫不及待地抓住了那個機會。 He leap ed at the chance.
(7.20)(e,2687) 我們將不得不及時趕到那裏 We shall have to hurry to get there in time.
(5.91)(w,693) 他還遠遠考試還沒得過 He is still a long way from passing his exam.
(5.91)(o,3155) 他在這次考試中,申論題部分答錯了。 He tripped up on the essay part of the exam.
(5.91)(e,1392) 考試得優等成績。 get a good result in the exam
(5.91)(e,1892) 英文考試 an examination in English = an English -

8. Output of the query “我今天心情很低落”

http://10.112.219.32/gb-hk/CE_E/Carbivw.cgi

Input : 我今天心情很低落

我-1.61 今天-5.33 心情-6.48 個-3.26 低落-8.36
(分數)(標準出處)(中文句子)(英文句子)
(15.3)(f,808) ((美))我今天情緒低落[萎靡] I feel down [up] today.
(14.84)(b,2404) 心情鬱悶,煩躁情緒低落 get the hump
(11.35)(f,1485) 我覺得很好(指心情或健康) I feel first-class .
(10.2)(f,1496) 今天早上我覺得精神很好。 I feel first-rate this morning.
(10.2)(f,2159) 今天我很忙 I today I am very busy.
(9.74)(e,1934) 他心情很愉快。 He was in excellent spirits.
(9.74)(m,1506) 覺得心情很好。 feel like a million (dollars)
(8.59)(w,661) 今天海上的浪很大。 The sea is very wavy today.
(8.59)(f,1914) 今天很熱。 It is flipping hot today.
(8.59)(h,123) 今天天氣很好。 It's not half fine today.
(8.59)(e,3238) 今天過得真快[令人覺得很短暫]。 Today was a short day.
(8.59)(h,2129) 今天很熱。 It is hot today.
(8.59)(e,1942) 除了這裏[南方] 以外,今天各地的天氣都很好 except here [in the South].
(8.59)(e,2809) 今天很冷 It is cold today.
(8.59)(n,699) 今天很溫和。 It is nice and warm today.
(8.59)(p,1125) 今天很冷。 It's perishing today.
(8.59)(d,2337) 今天的天氣真的很高和。 It is distinctly warm today.
(8.36)(m,2219) 士氣高[低沉,低落]。 M morale is high [low,falling].
(8.36)(e,1550) 那消息使士兵們士氣低落。 The news chill ed the morale of the soldiers.
(8.36)(d,3350) 連幾天他似乎情緒十分低落。 His spirits seem to be droop ing these days.
(8.36)(e,838) 還有誰的,不等的,(還有誰的)品質低劣的。 I
(8.09)(e,505) 這首曲子使我心動。 That tune get's me.
(8.09)(f,2879) 我沒有氣力[心情]做這件事。 I can't rise to it.
(8.09)(m,2156) 我有心情[沒心情]工作。 I was in the [no] mood for work.
(8.09)(a,386) 我了解你的心情 I understand how you feel.
(8.94)(f,806) 我今天早上發不出聲音。 I have no voice this morning.
(8.94)(w,1221) 你造成了今天的我。 You have made me what I am today.

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

90 年 6 月 20

附件三

報告人姓名	高照明	服務機構 及職稱	國立台灣大學外國語文學系 助理教授
會議 時間 地點	6月1日—2日 香港理工大學	本會核定 補助文號	NSC 89-2411-H-002-074 90 臺會綜二字 0021987 號
會議 名稱	(中文) 2001 年國際資訊科技與多媒體英語學習會議 (英文) ITMELT2001		
發表 論文 題目	(中文) 語料庫句型檢索系統 (英文) A Pattern Retrieval Program for Exploring Corpus-Based Phraseology		

報告內容應包括下列各項：

一、參加會議經過

此次會議的主辦單位是香港理工大學 Hong Kong Polytechnic University, 共 29 篇論文發表, 參與者將約 3 百人, 會議的規模相當龐大。除了來自香港各大學與國外的學者外大都是中小學教師。因此有幾個場次是 penal session 或 tutorial。

二、與會心得

有 2 篇論文值得注意。Meena Singhal 博士所發表的 CALL for Reading in English: an Interactive Web Program for College-Level ESL Students 以網路互動式的教學方法訓練以英語為第二語言的學生閱讀的各種策略, 進而增進閱讀與寫作的的能力。Birgid Winkler 博士所發表的 Working with an English Learners' Dictionary on CD-ROM 以實際觀察與訪談紀錄探索學生使用電子辭典的方式, 從而瞭解一般英語學習者的需求。

三、考察參觀活動(無是項活動者省略)

四、建議

1. 從 Meena Singhal 與 Birgid Winkler 及其他學者的論文中, 我們可以發現電腦輔助語言教學不只是運用現有資訊科技, 包括人工智慧, 第二語言習得理論, 測試理論等其它相關理論與實證研究是電腦輔助語言教學成功的重要基礎。
2. 鑑於國內對於尖端電腦輔助語言教學理論與實際仍然遠落後歐美甚至香港, 國科會與教育部有必要推動此一領域之研究。

五、攜回資料名稱及內容

Bruce Morrison, David Gardner, Ken Keobke, and Mary Spratt (2001) ELT Perspectives on IT & Multimedia. Selected Papers from the ITMELT 2001 Conference 1st & 2nd, 2001. Published by the English Language Centre, The Hong Kong Polytechnic University.

六、其他

國內雖有類似會議但水準有待提升。

A Pattern Retrieval Program for Exploring Corpus-Based Phrasology

GAO Zhao-Ming
Department of Foreign Languages and Literatures
National Taiwan University
zmgao@ccms.ntu.edu.tw
<http://ccms.ntu.edu.tw/~zmgao/pattern.htm>

Data-Driven Language Learning

- Tim Johns (1991) "language-learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data."
- Vance Stevens(1993) "With DDL, the learner's own discovery of grammar based on evidence from authentic language use becomes central to the learning process."

The Importance of Concordancers in Data-Driven Language Learning

- Gregory Hadley (forthcoming) "Data-driven learning is very much a "research-then-theory" method of studying grammar. Language learners start with a question, and then come to their conclusions after analyzing the corpora with a concordancer program."

Vance Stevens(1993)

- "... concordancers are the type of software that most closely approaches fulfilling the potential of computers in language learning. In a sense, they are working approximations of expert systems. They bring cognitive and analytic skills in students to bear on the manipulation of comprehensive databases for the purpose of solving real-language problems."

Key Function of Concordancers

- A concordancer allows a user to explore authentic language use by inspecting the context of an input keyword in the corpora.

Major Inadequacy of Traditional Concordancers

- They often display hundreds or even thousands of lines of unorganized data. Looking for a particular usage is just like looking for a needle in a hay.

Some of the Features of an Advanced Concordancer

- 1. It should support queries involving syntactic patterns.
- 2. It should facilitate the retrieval of collocations.

What is a pattern?

- Hunston and Francis (2000", p. 37)
- "The patterns of a word can be defined as all the words and structures which are regularly associated with the word and which contribute to its meaning.

The Role of Patterns in Lexicography, Language Learning, and Linguistics

- 1. Subcategorization: Chomsky (1957, 1965)
- Verb Patterns:
- 2. Hornby (1975) *Guide to Patterns and Usage in English*
- *Oxford Advanced Learners' Dictionary*
- 3. *Collins Cobuild Grammar Patterns*
- Hunston and Francis (2000) *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*

The Role of Patterns in Lexicography, Language Learning, and Linguistics

- 4. Collocations: Benson et al. (1997) *The BBI Dictionary of English Word Combinations*
- 5. Levin (1993) *English Verb Classes and Alternations*
- 6. Construction Grammar: Fillmore, Kay, and Goldberg

What can be found using our pattern retrieval program?

- 1. Verb Patterns
- 2. Idioms
- 3. Collocations

The screenshot shows a web browser window with the title "English Pattern Retrieval Program". Below the title, there is a text input field containing the pattern "ADV det N". To the right of the input field is a search button. Below the input field, there is a section titled "Or Select from the following patterns." with a dropdown menu showing "VNPB".

English Pattern Retrieval Program

Input a pattern consisting of keywords and the following syntactic categories.
[N V ADV ADV O NP VP PP INF S *] e.g. Input *V responsibility* to find a verb followed
the word *responsibility*.

ADV det N

Or Select from the following patterns.

VNPB

http://112.102.200.134/eng/104.htm Log

Meanwhile, [details] [details] of the personal lives of political, business and media personalities fill the newspapers - for example, extramarital affairs, privacy claims and more - while on the popular TV matchmaking show True Love Match, the hopeful participants pour bitter stories of failed past relationships.

scandal's [detail] [detail]?

on the storyline to the discourse and clothing of the actors, these mini-series, known in Japan as "read dramas," only reflect the [style] [style] of everyday life in the 1990s.

to see these until nearly eight o'clock, trying vainly to put together the [details] [details] of Lois Royman's revelation, wondering whether the perplexities that vexed her mind were real or fanciful, but warmed to the heart by affection that the older woman seemed instinctively to feel for her.

http://www.english.cn/eng/pattern.htm

English Pattern Retrieval Program

Input a pattern consisting of keywords and the following syntactic categories.
[N V ADV ADJ O NP VP PP DNF E *] e.g. Input V responsibility to find a verb followed by the word responsibility.

For syntax:

Or Select from the following patterns.

http://112.102.200.134/eng/105.htm Log

as Huhua have long been searching for the [true] [spirit] of Huhua culture and their place in larger society.

or even then many would be given a way to neighbors to add to the [divine] [work].

Many police officers receive the [lightness] [spirit] [Duan Gong].

Some without a theme, reflect some [particular] [spirit], can no longer attract an audience, so naturally you have the emergence of the professional exhibition curator.

Some scholars' advocacy of the [scientific] [spirit] and academic freedom is still well remembered today.

It are examples of wisdom and express the [inner] [spirit] of the subjects.

He, almost overwhelmed with the weight of knowledge that had been poured in upon her [inspiring] [spirit], and only knowing whether what she had heard should make her glad or sorry, leaned back against the door - post, and calmly listened to the voice, as it came nearer and nearer.

And this [contending] [spirit] began to be developed at a very early age in children.

She had a [proper] [spirit], but now her heart was cold with a new fear, and she felt, with Whitell, that her father can be obeyed and his temper kept within bounds, until God provident than a way of escape.

http://www.english.cn/eng/pattern.htm

English Pattern Retrieval Program

Input a pattern consisting of keywords and the following syntactic categories.
[N V ADV ADJ O NP VP PP DNF E *] e.g. Input V responsibility to find a verb followed by the word responsibility.

Or Select from the following patterns.

System Design By Gao Zhao-Ming
E-mail: gaozhao@ccit.com.cn

http://112.102.200.134/eng/106.htm Log

"How do you [put] [them] [all in] there?"
snc10.ppt

"Every time we catch him in an impossibility, he [tells] [us] [the explanation] [it is in the city]."
snc10.ppt

"The doctor gladly [gave] [them] [the medicine], and in a few days they began to show remarkable signs of recovery."
snc10.ppt

Such wild, unusual, and anti-social behaviors did not go unnoticed by the other natives, who quickly extracted the crowd outside from the tribal camp, [calling] [them] [benjamin] of the curtain system.
snc10.ppt

No matter how early he walked there, his Christmas cards arrived on his and his Valentine's arrived on December 24 revealing confusion and uncertainty among the people and [giving] [the Problem Solved] [an answer] to those the fables against him.
snc10.ppt

"Only the artist's extreme commitment to [complete] [truth] [to] [him] to admit that a fair trace of imitation was [needed, but] only for the briefest of moments, across the lovely Asian's brow."
snc10.ppt

http://www.english.cn/eng/pattern.htm

Input a pattern consisting of keywords and the following syntactic categories.
[N V ADV ADJ O NP VP PP DNF E *] e.g. Input V responsibility to find a verb followed by the word responsibility.

Or Select from the following patterns.

System Design By Gao Zhao-Ming
E-mail: gaozhao@ccit.com.cn
Programmer: Ling Jing-Jin

Laboratory for Language Technology
Department of Foreign Languages and Literatures
National Tsing Hua University

The Ministry of Education (MOE) (1998) is required to submit a report to the President of the Republic of China (Taiwan) (1998) and the National Education Commission (NEC) (1998) on the implementation of the plan. Li Yang, a professor of history at National Taiwan University, has reservations about the effectiveness of this plan in spite of the fact that he himself is a scholar who was involved in getting the Ministry of Education to adopt this plan.

The intention of this proposal was so obvious that the Republic was (declared) the status and the (plan) was immediately instituted.

One day he thought of a plan.

Finally, he (chose) to start broadcasting in September because that is when advertising agencies (plan) their next year's advertising plan.

This wave of enthusiasm for (protecting) old master engineers got to start to the TRA's (plan) to preserve railway history.

He was (arrived) to (plan) a super-skycraper for a site within Taipei.

* The executive (took) place according to Watson's (plan), and at four o'clock she sped back to her night work and preparation for supper, leaving Patsy with a great bunch of early wildflowers for Troy's mother.

I do not (have) what you'd do without me to (plan) for you!

- ## How is the program implemented?
- 1. Corpus Collection
 - 2. Identify Sentence Boundary
 - 3. Preprocess the corpus with Brill's Part-of-Speech Tagger
 - 4. Use Abney's Partial Parser
 - 5. Index words and syntactic structures

The output of Brill's tagger

- Such/JJ an/DT act/NN is/VBZ
 abhorrent/JJ to/TO my/PRP\$ sense/NN
 of/IN right/NN ./.

Output of Abney's Parser

```

  • in
  • [
  •   [such Such]
  •   [to to]
  •   [my my]
  •   [is is]
  •   [abhorrent]
  • ]
  • [
  •   [
  •     [my my]
  •     [sense sense]
  •     [of of]
  •     [right right]
  •   ]
  • ]
  • [
  • ]
  
```

- ## Future Research
- 1. Increase the corpus size.
 - 2. Enhance the accuracy of structure match using other taggers and parsers.
 - 3. Allow more flexible pattern search.
 - 4. Optimize the system.
 - 5. Investigate the relationship of meaning and patterns.