行政院國家科學委員會專題研究計畫　期中進度報告

運用自然語言處理於建置網路英語學習環境(2/3)

計畫類別：整合型計畫
計畫編號：NSC94-2524-S-032-006-
執行期間：94 年 05 月 01 日至 95 年 07 月 31 日
執行單位：淡江大學資訊工程學系

計畫主持人：郭經華

計畫參與人員：王尹伶、呂明峰、柳佳儀

報告類型：精簡報告
報告附件：出席國際會議研究心得報告及發表論文
處理方式：本計畫可公開查詢

中　華　民　國 95 年 5 月 19 日

行政院國家科學委員會補助專題研究計畫 □ 成 果 報 告 ■期中進度報告

分散式認知, 電腦運算與隨處的數位語言學習-

運用自然語言處理於建置網路英語學習環境(2/3)

計畫類別：□ 個別型計畫　　■ 整合型計畫
計畫編號：NSC94-2524-S-032-006-
執行期間：94 年 05 月 01 日至 95 年 07 月 31 日

計畫主持人：郭經華
共同主持人：
計畫參與人員：王尹伶、呂明峰、柳佳儀

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告
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

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：淡江大學

中　華　民　國　九十五年　五月 十五 日

行政院國家科學委員會專題研究計畫成果報告

計畫編號：NSC 94-2524-S-032-006
執行期限：94 年 05 月 01 日至 95 年 07 月 31 日
主持人：郭經華
計畫參與人員：王尹伶、呂明峰、柳佳儀

一、　中文摘要

本計畫旨在探討如何將自然語言處理工具及技術，運用在英語教學上。在計畫的第一年我們陸續完成了幾個自然語言處理工具，包含詞性註解工具 (Part-of-speech tagger)，字義辨識工具(Word Sense Identifier)以及語言難度過濾工具(Language Difficulty Filter)。在計畫的第二年，除了持續強化以上工具以及新研發搭配字探索工具(Collocator)。此外為了實現無所不在的學習輔助工具，我們研發了UWiLL 線上工具列 (UWiLL toolbar)，整合所研發的自然語言處理工具，並且使之可以用於在任何網頁上蒐尋出所需的語言學習資訊。實現本計畫所述求的無所不在的學習情境。

關鍵詞：自然語言處理，詞性註解，字義辨識，語言難度過濾

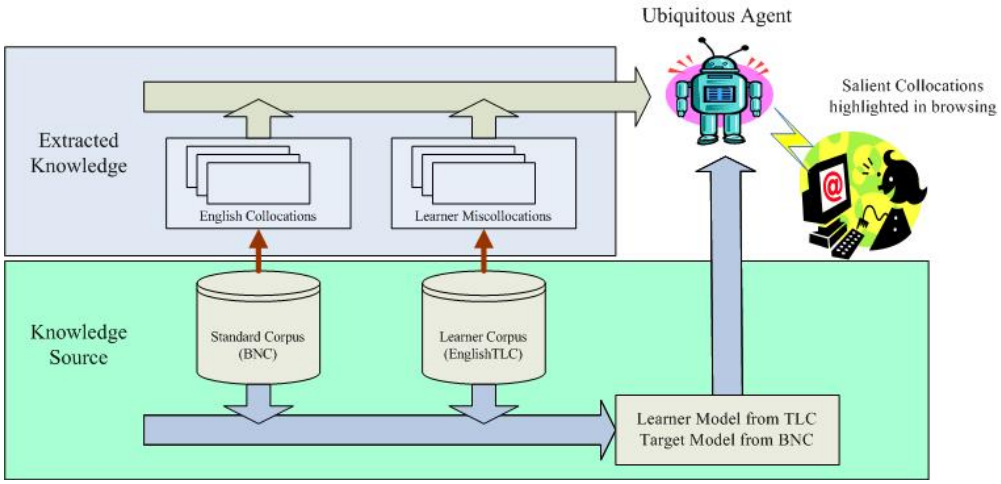
ABSTRACT

The purpose of this project is how to use Natural Language Processing(NLP) in Computer Assisted Language Learning(CALL). In the first year, we developed some prototyped tools, including part-of-speech tagger(POS tagger), word sense identifier, and language difficulty filter. In the second year, we worked on the development of collocation explorer. Furthermore, we developed the UWiLL toolbar, which allows users surfing the web to retrieve information desired for English learing. The approach that we were taken is to realize the ubiquitous learning environment as we proposed in this project.

Keywords: NLP, CALL, Part-of-speech tagging, Word sense identification, Language difficulty filter, collocation explorer.

二、緣由與目的

本計畫旨在探討如何將自然語言處理工具及技術，運用在英語教學上。經由總計畫的協調與整合，透過所提的三個子計畫，運用數位工具來掌握無所不在的學習情境。在計畫的第一年我們陸續完成了幾個自然語言



圖一、The schema represents the components of Collocator, the browser-based agent.

處理工具，包含詞性註解工具(Part-of-speech tagger)，字義辨識工具(Word Sense Identifier)以及語言難度過濾工具(Language Difficulty Filter)。在計畫的第二年，除了持續強化以上工具以及新研發搭配字探索工具(Collocatior)。此外為了實現無所不在的學習輔助工具，我們研發了 UWiLL 線上工具列(UWiLL toolbar)，整合所研發的自然語言處理工具，並且使之可以用於在任何網頁上蒐尋出所需的語言學習資訊。下面的內容中將詳細介紹這些工具，並說明這些工具如何與開發的網路代理器(Web agent)整合，及如何運用於英語教學。

三、討論與結果

圖一所示為所設計的UWiLL工具列之運作概念[1]，利用此內嵌於瀏覽器的代理者，可提供使用者在閱讀網頁時適時的語言學習協助。我們已逐步的將所設計語言學習工具整合在工具列上，其中包含了在第一年計畫中所實現的字義辨識工具(Word Sense Identifier)以及第二年的 Collocator 和已有初步成果的 Lexical Chunks Detector[3]。

字義辨識工具：字義辨識工具主要提供系統了解字義的能力，藉此提供學習者語言學習上的輔助。我們在提出一內嵌式瀏覽器語言學習工具，稱為 Word Spider，此工具是利用網頁文章中單字的相關性來提升學習者的閱讀能力。例如圖一的例子，當使用者瀏覽一網頁，看到一生字，antibiotics，在一般的認知下，學習者可能會利用查字典來得知此字的字義；然我們認為，一以英語為母語的人就算遇到未知生字，她們也不會查字典，而會利用此文章中的其他線索來得知此生字的大概含義。舉個中文的例子，當我們看到”貝里斯”時，我們也不知它是代表人名還是其他含意，但在閱讀文章中其他部分，如”貝里斯首府為”的時候，我們就可猜它是一個地名，甚至可猜到是一國家。運用相同原理，英文也應該可提供相關訊息。

圖二、UWiLL toolbar 上的 Collocation 功能。

語言難度過濾工具：語言難度過濾工具主要是希望提供適合學習者程度的閱讀教材而設計的。此一工具於第一年中已有成果，本年度著重於分類技術的研發[4]，希望可依使用者的興趣推薦合適難度的文章。

搭配字選擇工具： 圖二所示為本計畫所研發的搭配字選擇工具(Collocator)，此一工具的重要特點是將語言學習脫離固定網站學習模式，來到無所不在的學習情境，這才是一個真實的現況。譬如，當學習者來到了 CNN 網站某網頁，運用所提供的 UWiLL Collocator，其可掌握網頁所有的搭配字，如圖二。之後，使用者可以以 highlight 的方式找出所在的位置，了解使用的句字，或是回到 IWiLL 網站中，取得更多的範例。在技術面上，找出搭配字的做法，係經 IWiLL 團隊多年研究的成果，於此不在贅述。

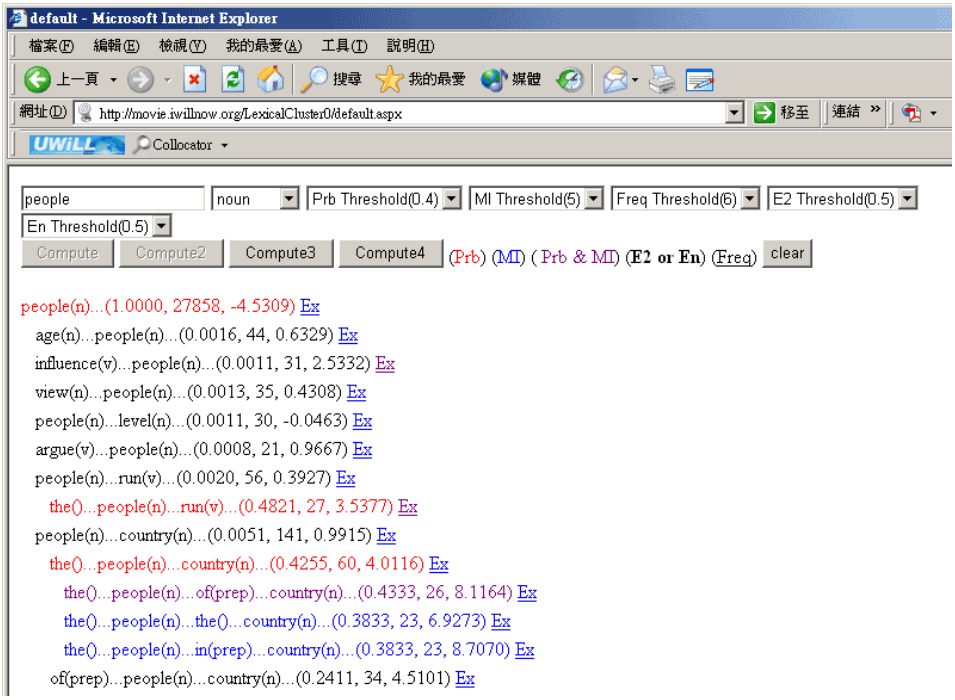
以上所述的各項有助於數位語言的工具部分已經整合在 UWiLL 的工具列中，未來可進一步在評量其成效與設計相關的應用情境，融入教學或是學習之中。

Lexical Chunks Detector：

圖三所示為 Lexical Chunks Detector 工具，此工具的重點是以單一個字為出發點，在 BNC 語料庫中找出所有相關的字句，藉由設定的 probability 與 mutual information 去標註出符合的組合，讓使用者可以快速的看出有用的 Lexical Chunk，點入字句後方例子連結，就可以看到完整的句子應用。

四、計畫成果自評

計畫執行至今，在團隊的合作下，許多困難透過定期的團隊研討，亦已一



圖三、Lexical Chunks Detector

一克服，上述的工具才得完成。接下來的執行重點，將放在實際的現場實驗上，我們已與高中教師建立默契，擬將這些工具實地的使用，進一步了解使用狀況，以期得知工具的效能並進以改善；再做必要的修正。換言之，我們一貫的信念是著立於研發對學習有助益的數位語言工具，這是將學習情境與自然語言處理工具和學習認知結合的工具，而非僅是自然語言處理工具。這些功能經過評量與修正後，將與 IWiLL 平台再做結合與實際應用[2]，提供完善的數位英語學習環境。我們已將部份成果投稿至國際期刊與研討會中發表[1][2][3][4]。其他相關的成果也陸續在整理中，將於近期內投稿。

五、參考文獻

[1] David Wible, Chin-Hwa Kuo, Meng-Chang Chen, Nai-Lung Tsao, Tsung-Fu Hung, “A Ubiquitous Agent for Unrestricted Vocabulary Learning in Noisy Digital Environments,” Lecture Notes in Computer Science, 2006. **SCI extended**

[2] Chin-Hwa Kuo, David Wible, Meng Chang Chen, Hsin-Yi Huang, and Sharon Kuo, “On the Application of ICT in Learning English as a Second Language: IWiLL,” The Internet Society Wit2006, June 12-14, 2006.

[3] David Wible, Chin-Hwa Kuo, Meng Chang Chen, Nai-Lung Tsao, and Tsung-Fu Hung, “A Computational Approach to the Discovery and Representation of Lexical Chunks,” TALN 2006, April 2006.

[4] Chin-Hwa Kuo, Meng Chang Chen, and Tzu-Chuan Chou, “On the Approach of Automatic Adjustments for Gaussian-Mixture Clustering,” *The Tamkang Journal of Science and Engineering*, June, 2006. **EI**

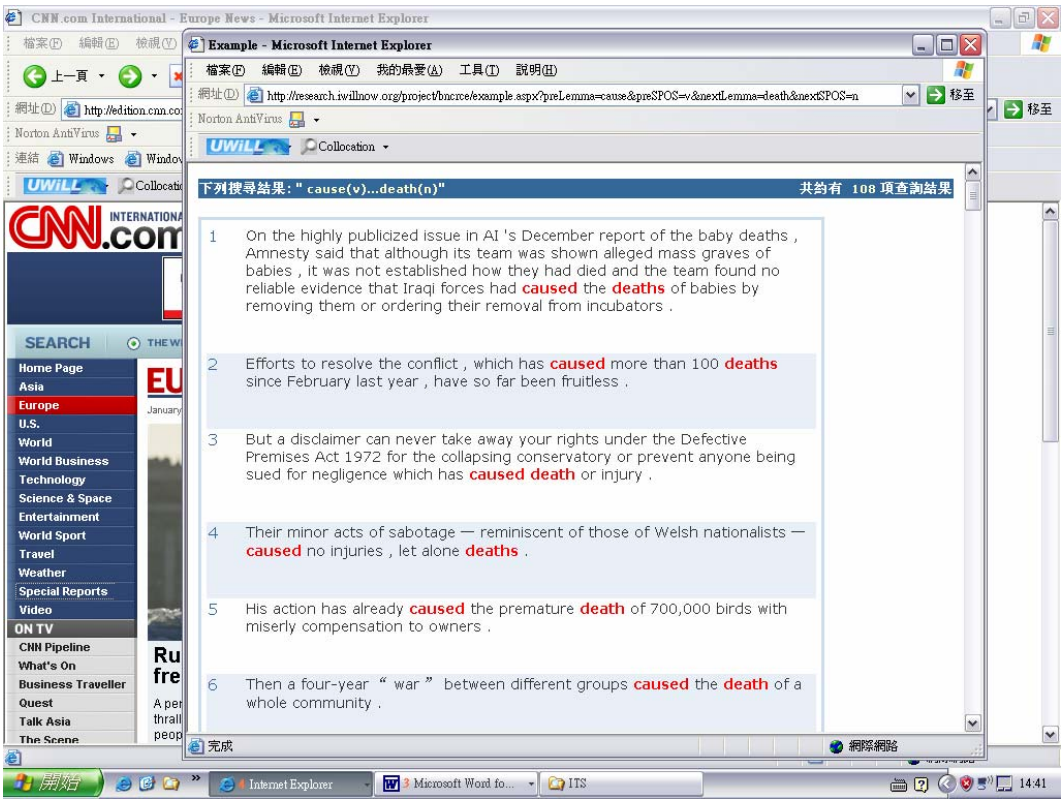
行政院國家科學委員會 數位學習國家型科技計畫

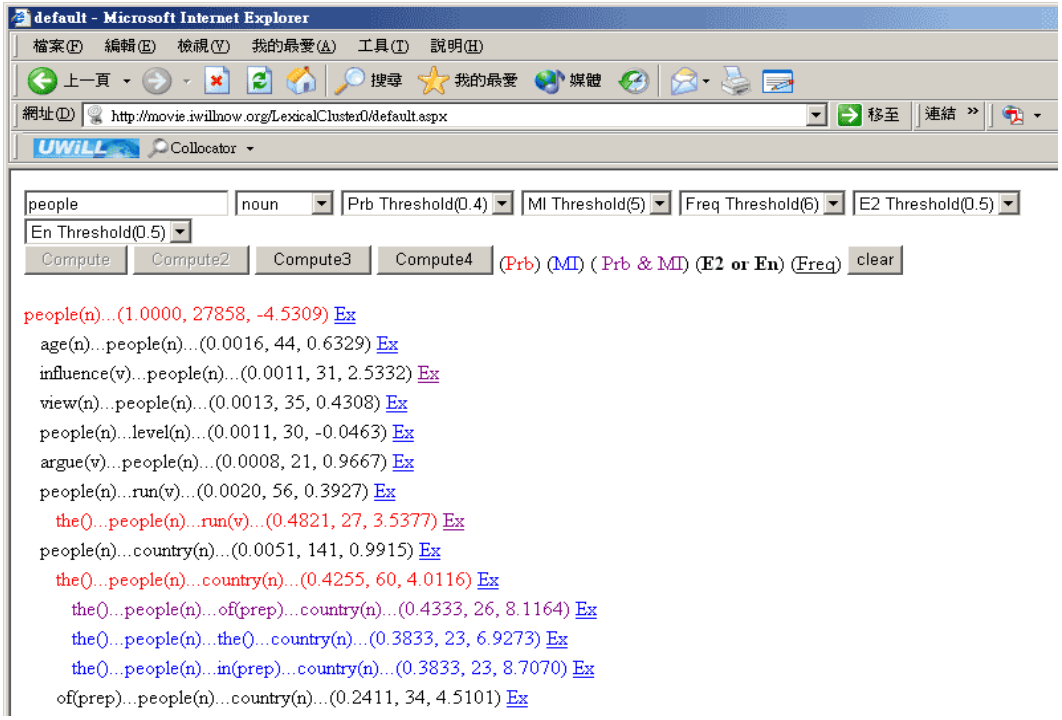
九十四年度研究成果資料表

國科會補助計畫	計畫名稱：分散式認知, 電腦運算與隨處的數位語言學習-運用自然語言處理於建置網路英語學習環境(2/3) 計畫主持人：郭經華 研究人員：王尹伶、呂明峰、柳佳儀 計畫編號：NSC 94-2524-S-032-006
研究摘要	中文： 在計畫的第一年我們陸續完成了幾個自然語言處理工具，包含詞性註解工具(Part-of-speech tagger)，字義辨識工具(Word Sense Identifier)以及語言難度過濾工具(Language Difficulty Filter)。在計畫的第二年，除了持續強化以上工具以及新研發搭配字探索工具(Collocator)。此外為了實現無所不在的學習輔助工具，我們研發了UWiLL 線上工具列(UWiLL toolbar)，整合所研發的自然語言處理工具，並且使之可以用於在任何網頁上蒐尋出所需的語言學習資訊。實現本計畫所述求的無所不在的學習情境。
	英文： The purpose of this project is how to use Natural Language Processing(NLP) in Computer Assisted Language Learning(CALL). In the first year, we developed some prototyped tools, including part-of-speech tagger(POS tagger), word sense identifier, and language difficulty filter. In the second year, we worked on the development of collocation explorer. Furthermore, we developed the UWiLL toolbar, which allows users surfing the web to retrieve information desired for English learing. The approach that we were taken is to realize the ubiquitous learning environment as we proposed in this project.
研究特色	本計畫旨在探討如何將自然語言處理工具及技術，運用在英語教學上，依總計畫所述求的理念，將這些工具實作在Web-based browser agent 中，以實現無所不在的語言學習。利用已完成的自然語言處理工具，包含字義辨識工具(Word

	Sense Identifier)、語言難度過濾工具(Language Difficulty Filter)、Lexical Chunks Detector 以及搭配字選擇工具(Collocator)，讓使用者在閱讀瀏覽網頁時，可得到語言學習工具協助。
研究成果可應用範圍	本計畫所實作出的語言學習輔助工具，除了可以讓一般使用者在瀏覽網頁時使用，配合目前實作出的 Web-based Learning Environment，更能發揮其功效。因為一個固定的學習環境可提供更多個人化的資訊，有利自然語言處理的成效。另外也可以運用在學習者語料庫（Learner Corpus）上，分析學習者寫作內容，取得錯誤資訊，進而幫助學習者了解其錯誤原因。
研究成果預期效益／商機	本計畫所倡導的隨處學習，除具吸引力外，也非常符合現代人生活的習慣與模式。尤其所設計的數位語言學習工具部分是力基於學習者語料庫來設計，如此很自然地融入了文化背景的考量。是以，對以英語為第二語言的習得方面，具有貢獻。再經由實際使用實驗的回饋，可以改善教學策略與學習方法，並發展學習模式。在實際將受測資料分析後，應能提供一健全的使用模型供其他開發者參考，希望此模型能激勵更多創意實現在 Web browser agent 及其他隨處學習元件上。
研究成果圖片	如附件所示


UWiLL
Collocation
cause death
cause people
cut gas
face world
get hotel
have plan
make home
open window
pay service
plunge water
Collocator
Highlight
Show examples
CNN
SEARCH
Home Page
Asia
Europe
CNN.com
Updated 0253 GMT (1053 HKT)


Example - Microsoft Internet Explorer
Collocation
下列搜尋結果: " cause(v)...death(n)"
共約有 108 項查詢結果
1 On the highly publicized issue in AI 's December report of the baby deaths , Amnesty said that although its team was shown alleged mass graves of babies , it was not established how they had died and the team found no reliable evidence that Iraqi forces had caused the deaths of babies by removing them or ordering their removal from incubators .
2 Efforts to resolve the conflict , which has caused more than 100 deaths since February last year , have so far been fruitless .
3 But a disclaimer can never take away your rights under the Defective Premises Act 1972 for the collapsing conservatory or prevent anyone being sued for negligence which has caused death or injury .
4 Their minor acts of sabotage — reminiscent of those of Welsh nationalists — caused no injuries , let alone deaths .
5 His action has already caused the premature death of 700,000 birds with miserly compensation to owners .
6 Then a four-year " war " between different groups caused the death of a whole community .



本計畫所實作的 UWiLL 搭配字選擇(Collocator)與 Lexical Chunks Detector 工具。