

本期要目

壹. ROCLING XVI CFP	第二頁
貳. Student Workshop II CFA	第三頁
參. 「中文計算語言學期刊」徵稿啟事	第四頁
肆. WordNet：結合辭典和本體論的文字的百科全書-張俊盛教授	第五 六頁
伍. 雙語語料庫的應用與電腦輔助語言教學-張俊盛教授	第七 十二頁

第四屆碩士論文獎開始申請

名額：四名

1. 優等獎：一名，獎金二萬元。
2. 佳作獎：三名，獎金各一萬元。

申請資格：

1. 國內大專院校碩士班應屆畢業生及其指導教授從事計算語言學相關研究方向者。
2. 參賽限制：每位指導教授以推薦兩篇論文為限。(含個人指導與共同指導)
3. 申請人應注意其所申請之論文絕無抄襲、剽竊情事，若經發覺除追回其所獲得之獎金、獎狀外，一切後果自行負責。

申請時間：六月一日起至七月十五日止

申請方式請參閱本會網站：

http://rocling.iis.sinica.edu.tw/ROCLING/scholarship/index_cf.htm

學會新聘專任工作人員

「中文計算語言學期刊」自 2005 年起將由一年兩期增刊為一年四期，期刊編輯助理杜易玲小姐自五月起由兼任轉為專任。有關期刊之投稿及其他相關事宜可與之聯絡，徵稿啟事請參閱第四頁。

第二屆學生語言學研討會

「第二屆學生計算語言學研討會」謹訂於九十三年九月三日假福華翡翠灣舉行，本屆研討會依循上屆模式，將與 ROCLING-2004 合併舉行，為了鼓勵學生參與，Student Workshop 免入場費，歡迎從事語言學相關領域之同學踴躍投稿，論文被接受之學生，本會提供一名免費名額參加 ROLCING-2004，徵稿啟事請參閱第三頁。

四月份學生出席國際會議補助

補助會議：ICASSP-2004

論文題目：Lightly Supervised And Data-Driven Approaches To Mandarin Broadcast News Transcription

補助學生：郭人瑋（台灣師範大學資工所碩士生）

補助金額：美金 1,000 元

九十四年度會費繳費通知

本年度個人及學生會員有效期即將於六月三十日到期，為保障各位會員之權益，敬請於七月一日前繳交會費，同時如果您的個人資料有所異動，亦請惠予通知，繳費通知單如信封所附。(本年度新加入之會員，勿需再繳費)

ROCLING XVI:

Conference on Computational Linguistics and Speech Processing

第十六屆自然語言與語音處理研討會

September 2-3, 2004, Howard Pacific Green Bay, Taipei, Taiwan, ROC

<http://www.aclclp.org.tw/rocling2004.html>

CALL FOR PAPERS

Conference Chairs:

Der-Tsai Lee

Academia Sinica

Chin-Chuan Cheng

Academia Sinica

Program Committee:

Lee-Feng Chien, Co-Chair

Academia Sinica

Hsin-Min Wang, Co-Chair

Academia Sinica

Chao-Huang Chang

CCL/TRI

Claire H. H. Chang

National Chengchi University

Jason S. Chang

National Tsing Hua University

Jing-Shin Chang

National Chi Nan University.

Hsin-Hsi Chen

National Taiwan University

Keh-Jiann Chen

Academia Sinica

Kuang-Hua Chen

National Taiwan University

Sin-Horng Chen

National Chiao Tung University

Jen-Tzung Chien

National Cheng Kung University

Zhao-Ming Gao

National Taiwan University

Wen-Lian Hsu

Academia Sinica

Chu-Ren Huang

Academia Sinica

Bor-Shenn Jeng

Chunghwa Telecom Labs

Sur-Jin Ker

Soochow University

Lin-Shan Lee

National Taiwan University

Tyne Liang

National Chiao Tung University

Hsien-Chin Liou

National Tsing Hua University

Ren-Yuan Lyu

Chang Gung University

Chiu-yu Tseng

Academia Sinica

Shu-Chuan Tseng

Academia Sinica

Yuen-Hsien Tseng

Fu Jen Catholic University

Hsiao-Chuan Wang

National Tsing Hua University

H. Samuel Wang

National Tsing Hua University

Jhing-Fa Wang

National Cheng Kung University

Yih-Ru Wang

National Chiao Tung University

Chung-Hsien Wu

National Cheng Kung University

Ming-Shing Yu

National Chung Hsing University

The 16th ROCLING Conference will be held September 2-3, 2004 at Howard Pacific Green Bay in Taipei. Sponsored by Association for Computational Linguistics and Chinese Language Processing (ACLCLP), ROCLING is the most historied and comprehensive conference focused on computational linguistics, speech processing, and related areas in Taiwan. ROCLING XVI will be hosted by Institute of Information Science, Academia Sinica. The conference will feature invited lectures, tutorials, panel discussions, and lecture and poster sessions and two workshops: Workshop on Intelligent Web Technologies and Student Workshop.

Papers are invited on substantial, original, and unpublished researches on all aspects of computational linguistics, including, but not limited to the following topic areas.

- | | |
|---|--|
| (a) cognitive linguistics | (l) parsing/generation |
| (b) discourse modeling | (m) phonetics/phonology |
| (c) document database/large corpora | (n) quantitative/qualitative linguistics |
| (d) electronic dictionaries | (o) speech analysis/synthesis |
| (e) information retrieval | (p) speech recognition/understanding |
| (f) language understanding | (q) spoken dialog systems |
| (g) language processing over Internet | (r) spoken language processing |
| (h) machine translation | (s) syntax/semantics |
| (i) NLP and educational applications | (t) Web information extraction |
| (j) morphology | (u) Web corpora |
| (k) computer assisted language learning | (v) others |

Paper Submission:

Prospective authors are invited to submit full papers of no more than 25 A4-sized pages in pdf or MS Word format. Papers will be accepted only by electronic submission through the conference web site. Prospective authors without web access should contact the Program Committee Co-Chair (whm@iis.sinica.edu.tw) before the submission deadline. The submitted papers should be written in either Chinese or English, and in single column, double-spaced format. The first page of the submitted paper should bear the items of paper title, author name, affiliation and email address. All these items should be properly centered on the top, with a short abstract of the paper following.

Best Paper Award:

The best paper will be selected and announced at ROCLING XVI.

Important Dates:

Preliminary paper submission due:	July 5, 2004
Notification of acceptance:	July 26, 2004
Final paper due:	August 9, 2004

Sponsors:

Association for Computational Linguistics and Chinese Language Processing (ACLCLP)
Institute of Information Science, Academia Sinica
Institute of Linguistics, Academia Sinica

Organizing Committee:

Pu-Jen Cheng, Lee-Feng Chien, Wei-Ho Tsai, Hsin-Min Wang, and Jeng-Haur Wang (Academia Sinica)
Qi Huang (ACLCLP)

**ROCLING XVI:
Student Workshop II
第二屆學生計算語言學研討會**

**September 3, 2004, Howard Pacific Green Bay, Taipei, Taiwan, ROC
<http://www.iis.sinica.edu.tw/Conference/ROCLING04/>
CALL FOR ABSTRACTS**

We are pleased to announce that the *ROCLING XVI: Student Workshop II* is now calling for submission of abstracts. The workshop is aimed at providing a forum for student researchers on Computational Linguistics, Natural Language Processing, and Linguistics. We would like to invite student researchers to submit their work to the workshop. In this workshop, students are provided with opportunities to report their completed work or work in progress to a wide audience in the related fields and to obtain useful feedback.

Papers are invited on original work, both completed and on-going research projects with novel ideas/applications, or late-breaking results that are best communicated in an interactive format on all aspects of computational linguistics, NLP, and linguistics, including, but not limited to the following topic areas.

- | | |
|---------------------------------------|--|
| (a) cognitive linguistics | (k) computer assisted language learning |
| (b) discourse modeling | (l) parsing/generation |
| (c) document database/large corpora | (m) phonetics/phonology |
| (d) electronic dictionaries | (n) quantitative/qualitative linguistics |
| (e) information retrieval | (o) speech analysis/synthesis |
| (f) language understanding | (p) speech recognition/understanding |
| (g) language processing over Internet | (q) spoken dialog systems |
| (h) machine translation | (r) spoken language processing |
| (i) NLP and educational applications | (s) syntax/semantics |
| (j) morphology | (t) others |

Paper Submission:

The workshop contains both sections of paper presentation and interactive poster demos. By the indicated date in the workshop schedule, please send an abstract in the form of softcopy in pdf or Microsoft Word format to d928322@oz.nthu.edu.tw. The submitted abstracts should be written in either Chinese or English and in a format of single column, with a maximum of one A4-sized page, optionally with an additional page for data and references. The title, author's name, affiliation, phone and fax numbers as well as email address and section name (paper presentation or poster) should be given on the first page of the submitted abstract. All these items should be properly centered on the top, with the abstract following.

Upon acceptance, authors are required to submit a full paper by the indicated date below. For the poster session, the maximum length for the full paper is 4 pages. For the oral presentation, the maximum length is 12 pages, including references and data.

Registration:

Participants are encouraged to join the Association for Computational Linguistics and Chinese Language Processing and register for the ROCLING main conference. The registration fee of the main conference for common students is 1500 NT and for student members 1000 NT. Registered participants can enjoy a banquet and a copy of the proceedings, all the workshops, as well as an excursion. The student workshop is free to all participants. Unregistered participants can buy proceedings on site. Registration fees of the main conference will be waived for student presenters at the student workshop. For co-authored papers, only one of the authors will have this priority.

Jian-Cheng Wu
Co-Chair
Dept. of Computer Science
National Tsing Hua University
d928322@oz.nthu.edu.tw

Yiching Wu
Co-Chair
Graduate Institute of Linguistics
National Tsing Hua University
d898702@oz.nthu.edu.tw

Dian-Sung Wu
Co-Chair
Dept. of Computer & Information Science
National Chiao Tung University
gis92807@cis.nctu.edu.tw

Siaw-Fong Chung
Co-Chair
Graduate Institute of Linguistics
National Taiwan University
claricefong6376@hotmail.com

Important dates for student workshop submission:

Abstract submission due: July 9, 2004
Notification of acceptance: July 16, 2004
Final paper due: August 16, 2004

Call for Papers

Computational Linguistics and Chinese Language Processing

An International Journal of the Association for Computational Linguistics and Chinese Language Processing, Taiwan

Advisory Board

Jyun-Sheng Chang

National Tsing Hua University, Taiwan

Keh-Jiann Chen

Academia Sinica, Taiwan

Ching-Chun Hsieh

Academia Sinica, Taiwan

Lin-Shan Lee

National Taiwan University, Taiwan

Kim-Teng Lua

National University of Singapore,
Singapore

Keh-Yih Su

Behavior Design Corporation, Taiwan

Benjamin T'sou

City University of Hong Kong, China

Hsiao-Chuan Wang

National Tsing Hua University, Taiwan

Jhing-Fa Wang

National Cheng Kung University, Taiwan

Editorial Board

Chung-Hsien Wu (**Editor-in-Chief**)

National Cheng Kung University, Taiwan

Hsin-Hsi Chen

National Taiwan University, Taiwan

Sin-Horng Chen

National Chiao Tung University, Taiwan

Jen-Tzung Chien

National Cheng Kung University, Taiwan

Lee-Feng Chien

Academia Sinica, Taiwan

Chu-Ren Huang

Academia Sinica, Taiwan

Hang Li

Microsoft Research Asia, China

Chin-Yew Lin

University of Southern California, USA

Dekang Lin

University of Alberta, Canada

Helen Meng

Chinese University of Hong Kong, China

Jian-Yun Nie

University of Montreal, Canada

Richard Sproat

University of Illinois at Urbana-Champaign,
USA

Jian Su

Institute for Infocomm Research,
Singapore

Jianhua Tao

Chinese Academy of Sciences, China

Chiu-Yu Tseng

Academia Sinica, Taiwan

Hsin-Ming Wang

Academia Sinica, Taiwan

Kam-Fai Wong

Chinese University of Hong Kong, China

Computational Linguistics and Chinese Language Processing is an

international journal published by the Association for Computational

Linguistics and Chinese Language Processing (ACLCLP), a.k.a. ROCLING.

This journal was founded in August 1996 and is published twice a year

currently. In order to promote research and technical advancement and play a

more active role in this area, this journal is scheduled to publish four issues

per year from 2005. This journal covers all aspects related to computational

linguistics and Chinese speech and language processing. Possible topics for

manuscript submitted to the journal include, but are not limited to:

- Computational linguistics
- Natural language processing
- Machine translation
- Speech analysis/synthesis
- Speech recognition/understanding
- Spoken dialog systems
- Information retrieval and extraction
- Language generation
- Language learning
- Web information extraction/mining

Information for Authors

Computational Linguistics and Chinese Language Processing invites

submission of original research papers in the area of computational linguistics

and Chinese speech and language processing. All papers must be written in

English. Manuscripts submitted must be previously unpublished and cannot

be under consideration elsewhere. Submissions should report significant new

research results in computational linguistics, Chinese speech and language

processing or new system implementation involving significant theoretical

and/or technological innovation. The accepted papers are divided into the

categories of regular papers and short papers. There is no strict length

limitation on the regular papers but it is suggested that manuscripts not exceed

40 double-spaced A4 pages. Short papers are restricted to no more than 20

double-spaced A4 pages. All contributions will be anonymously reviewed by

at least two reviewers.

Submission

Electronic submission: Please e-mail MS-word, PDF, or Postscript files to:

clp@hp.iis.sinica.edu.tw

Hard copy submission: Four copies of the manuscript should be submitted to:

Computational Linguistics and Chinese Language Processing

Ms. I-Ling Tu

Institute of Information Science, Academia Sinica, Nankang, Taipei 115,

Taiwan, R.O.C.

Please see the CLCLP Web page at

<http://rocling.iis.sinica.edu.tw/CLCLP/CLCLP11.htm>

喬治米勒在聖塔菲

WordNet：結合辭典和本體論的文字的百科全書

張俊盛

清華大學資訊工程學系

自然語言實驗室

在大學作研究有一個好處，每年總會有一兩次機會到最好的旅遊景點去開會。其中，讓我記憶猶新的是 1996 年在美國新墨西哥州聖塔菲(Santa Fe)舉行的國際機器翻譯會議。我從台灣出發，抵美後途中停留舊金山，再搭螺旋槳飛機，飛越加州死谷到聖塔菲去開會。抵達後，才臨時找平價的 Day Inn 汽車旅館投宿。旅館看起來雖不起眼，卻靠近熱鬧的廣場 (The Plaza)，會後可以很方便的參觀週末的西班牙雕刻祭。旅館到大街的路上，兩旁都是一間間兼具西班牙與印地安風味的夢幻粉紅泥磚屋 (adobe)。一位與會的英國學者大嘆：「我的天呀！聖塔菲就像一座主題樂園，連加油站都是泥磚屋！」英國人對美國風 (Americana)，當然不是讚賞而是挖苦。

國際機器翻譯會議的會場，設在沙漠中的聖約翰學院 (St. John's College)內。聖約翰創立於 1665 年，算來有 300 多年的歷史，以直接閱讀經典名著的大學學程 (Great Books Curriculum) 而聞名。沙漠中有道路，但沒有到會場的巴士，也叫不到計程車。第二天起了個大早，健行一小時，經過了歐姬芙美術館 (Georgia O'Keeffe Museum) 才到會場。我先坐下來喘一口氣，也趁開會前的空檔，和鄰座閒聊兩句。我們談到最近很多學者利用普林斯頓大學的 WordNet 做研究 (可以在 www.cogsci.princeton.edu/~wn/ 免費下載) 源自心理學研究的 WordNet 可以說是文字的百科全書，包含了將近二十萬個英文字義及其語意關係，在自然語言處理、搜尋引擎等研究中，常見 WordNet 的身影。我抱怨說：「美中不足的是 WordNet 沒有主題資訊，例如無法列出所有和美術有關的辭彙。」

這時候，主席交代過會議的開場的事務性報告，開始介紹邀請主題講座，喬治米勒 (George Miller) – 心理學大師，有名的 7 加減 2 定理的發現者，也就是 WordNet 計畫的主持人。米勒由我的旁邊起身，走到台前，準備作他的大師開示。有眼不識泰山，鄰座就是 WordNet 之父，而我竟然當他的面挑剔起 WordNet！

大師暖場的笑話沒有奏效，聽眾全板著臉：或許是時差在作祟吧。米勒嘲弄地說：「天啊！你們機器翻譯專家平常都是這麼嚴肅嗎？」這下子才打破僵局，一屋子原本正經八百的專家哄堂大笑。接下來，大家都很捧場，該笑的時候就笑。米勒接著說，心理學家想從小孩身上了解語意。他曾實地的去看老師的作業。老師要小學生用「eat out」造句 (如「酸雨侵蝕了銅質的屋頂」)，學生交回來「My family ate out a lot recently」，大出老師意料之外。又是哄堂大笑。

學生很自然的以日常的脈絡下來解讀「eat out」，造出句子。我們莞爾一笑之餘，可以再深入思考。我想到喬治米勒的另外一大貢獻 – 語意由上下文決定的理論。喬治米勒認為語意和

上下文，是一體的兩面。至今，這個想法仍然很有啟發性 – 最近史丹福大學的辛瑞奇舒茲 (Hinrich Schütze) 的博士論文中，就以實驗證明了英文字「suit」的不同意思可以由 7,467 個「suit」例句，自動分析出來！

再回頭來談 WordNet 有什麼好處。以自動詢答 (question answering) 這個熱門的研究題目做例子：我們可以問電腦「Who invented electric light?」或「Who invented lightbulb?」。電腦由 WordNet 中，知道「electric light」和「lightbulb」在 WordNet 中，同屬一個同義集 (synset)，因此指的是同一個東西。WordNet 還提供廣義、狹義、相似、部分、全體等 18 種詞彙語意關係。有了 WordNet，電腦就能應付變化多端的問題，容易在「大英百科全書」檔案中，找到愛迪生發明電燈泡的一段相關文字，來正確的回答問題：He [Edison] created the world's first industrial-research laboratory, ... invented ... the incandescent **lightbulb**. (大英百科全書)。如果沒有 WordNet，電腦就無法順利的將問題中的 electric light 和大英的 lightbulb 連起來，正確回答問題。

在米勒的啟發下，此刻許多研究者都在努力建立各種語言的 WordNets：Spanish WordNet、Chinese WordNet 等。WordNet 儼然已經成為語言研究者的必修課；因為 WordNet 可以透過語意關係，連接到其他相關字，社會大眾也多了一個比一般辭典更方便的工具。在認知科學早就立下不朽之言之米勒，為了研究語意，晚年又大膽的闖入辭典學的領域，創立結合辭典和本體論的 WordNet。大師不受領域限制的貢獻，影響領域之廣，讓人敬佩！

(註：本文經原作者與原出版社之同意，大部分內容轉載自《科學人》2004 年 5 月，頁 27。)

前瞻性數位英文學習中心 CANDLE: 雙語語料庫的應用與電腦輔助語言教學

張俊盛

清華大學資訊工程學系

自然語言實驗室

壹、前言

近一年來，本實驗室已經與六個教育機構中，十位來自各領域的專家—包含電腦科學、電腦輔助語言學習，以及英語語言教學的研究者¹，共同致力於「數位學習國家型科技計畫(Corpora And NLP (Natural Language Processing) for Digital Learning of English, 簡稱 CANDLE)」的研究。本計畫為國科會的整合型計畫，為期三年，以雙語語料庫為研究中心。本研究議題不僅在計算機學界極具前瞻性，也能讓台灣學生在學習系統中，善用母語長處及原有的文化背景學習英文，可謂同時兼顧學習者之學習心理及電腦輔助學習的研究計畫。

本計畫的目的，在於利用自然語言處理工具，建立「因材施教」的英文個別化學習系統。根據學習者的英文程度，提供適當的聽、說、讀、寫等教材與合適的練習題，精練學習者的英語技能。利用本雙語語料庫及統計式的自然語言處理工具，除了達到學習適性化的效果之外，並可採用「歸納式」的英文學習法學習英文，此為以往傳統英文教學所無法達到的新式學習法。

貳、CANDLE 計畫簡介

CANDLE 計畫為期三年，將由四個子計畫共同合作達成，包括（一）自然語言處理及評量工具、（二）智慧型個人化閱讀環境、（三）利用翻譯及寫作練習學習第二外語，與（四）雙語語料庫及文化相關議題教學。

以上第一個子計畫為其他三個子計畫提供必備的自然語言處理工具及追蹤系統，以記錄學習者線上的學習歷程、觀察學習者線上的學習行為，作為評估整體教學活動的依據，並進一步提出最合適的系統架構。第二個子計畫針對學習者的學習行為及學習歷程，設計出一個符合直覺性思考操作，並且具備完整學習功能的個人自學閱讀環境。第三個子計畫是設計一個電腦化的環境，提供學習者寫作及翻譯練習來增進英文學習。第四個子計畫則是探討學習者如何利用既有文化背景及母語優勢來學習第二外語。

這四個子計畫不管是在自然語言處理，或是英語學習領域方面都極具創新。我們預期在三年內能夠建造出一個內容完整、豐富的雙語語料庫，並提供具親和力、直覺性思考的智慧型介面，讓學習者能透過自然語言工具所研發的新興語言學習工具，利用電腦化的環境以最有效益的方式來學習第二外語。

¹這六所教育機構包括清華大學、台灣師範大學、台灣大學、東海大學、萬能科技大學，以及台中技術學院。領導研究學者為清大外語系劉顯親(Liou, H.C.)、清大資工系張俊盛(Chang, J.S.)等教授。合作研究者包括清大外語系葉由俐(Yeh, Y.L.)、師大英語系陳浩然(Chen, Howard)、師大英語系林至誠(Lin, C.C.)、東海外語系廖美玲(Liaw, M.L.)、台中技術學院資管系游耿能(You, G.N.)、萬能科技大學資管系莊暢(Chuang, C.)、台大外文系高照明(Gao, Z.M.)，以及清大資工系張智星(Jang, J.S.)等教授。

本計畫三年之進度為：第一年處理語料、研發自然語言處理工具、數位學習管理系統、線上學習教材以及英文學習單元；第二年進行學生上機測試之「形成性」評量，修改各單元，並完成系統雛型；第三年進行「總結性」評量，讓更多學生上機測試，並且修改系統、完成結案報告。

本研究使用的語料庫，包括光華雜誌（Sinorama）語料庫、美國之音（Voice of America）語料庫、香港立法局資料（Record of HK Legislative Court），與《大家說英語》（Let's Talk in English）、《空中英語教室》（Studio Classroom）等資料。光華雜誌語料庫由遠流出版社授權，每月由政府官方資訊處發行，歷時近三十年（1976-2002），內含近四千萬字關於台灣的中英雙語資料。在目前既有的數個官方發行雜誌中，《光華雜誌》最受好評，因其涉入的主題層面包羅萬象，包括表演藝術、文學、繪畫或書法等等，藉此可讓學生在學習系統中善用母語長處及原有之本國背景知識來學習第二外語。

本計畫內容將由六個參與計畫的教育機構的大一學生實際使用，並希望將來能繼續推廣，讓高中生也能使用。我們在研究過程中同時收集「學習者語料」（Learner corpora），未來將公開供其他研究者利用。目前部分成果，已可見於計畫網址(<http://candle.cs.nthu.edu.tw/>)中。

參、CANDLE 第一年成果介紹

過去一年來，在研發語料庫處理及自然語言工具、智慧型英語閱讀自學環境，以及探究雙語語料庫及英文文化教學等方面，皆已有初步成果。我們已開發的相關自然語言處理工具與技術包括：英語詞塊切分系統（English Chunker）²、雙語辭彙索引典（TOTAL recall）、雙語搭配詞查詢系統（TANGO），此外，搭配詞之寫作輔助系統（Collocation Checker）、英語語音辨識引擎（Automatic Speech Recognition，簡稱 ASR）、線上教材、電腦輔助學習教材、依據語境不同提供不同的中文註解，以及附註錯誤標記之學習者語料庫也都正在研發中。以下將分項說明。

一、語料庫處理及自然語言工具之研發

（一）TOTAL recall 雙語辭彙索引典

1. 本引典為利用中、英文雙語語料庫建構的檢索系統。以下簡介 TOTAL recall 介面：

²英語詞塊切分系統(English Chunker)，利用 CoNLL (Conference on Computational Natural Language Learning) 2000 年會議所提供的訓練資料，建立詞塊模組，可將英語的語句(sentence)，切分成若干個詞塊(chunk)，每個詞塊自成一個單元(unit)；其單位相當於一個詞組(phrase)，例如：名詞組(NP)、動詞組(VP)、介詞組(PP)等。



- A. Text Collection: 選擇欲查詢之資料範圍，目前提供《光華雜誌》(Sinorama) 1990-2000 年的語料，以及香港立法局 (Record of HK Legislative Court) 的雙語語料庫。
- B. Search Time: 顯示系統查詢資料所花的時間。
- C. Query: 輸入欲查詢字串。搭配 E 中的單語 (mono mode) 或雙語查詢模式 (bilingual mode)，在單語查詢模式中，可同時輸入中文及英文字串查詢；在雙語查詢模式中，可輸入中文或英文字串查詢相對的英文或中文。
- D. Items/page: 設定查詢結果畫面一頁顯示的搜尋筆數。
- E. 提供單語查詢 (mono mode) 及雙語查詢 (bilingual mode) 模式。在雙語查詢模式中，結果會將翻譯為中文時相同的英文詞彙予以群組並以其中一筆資料顯示。
- F. Order by: 設定查詢結果的排序。查詢結果可以按照關鍵詞出現次數、關鍵詞所在的中文或英文句子長度，由短至長順序排列。
- G. 顯示共有幾頁查詢結果。
- H. Citations: 此項目出現於雙語查詢模式，這是由於系統將翻譯為中文時相同的英文詞彙予以歸類成一筆資料顯示，使用者可點選此部份查看所有其他資料。
- I. Text: 可查看範例來源的整篇文章，而 BiText 則是查看範例來源文章之段落。

以上查詢介面的說明，顯示 *TOTAL recall* 雙語辭彙索引典含有以下特色：

2. 含中、英雙語語料，便利使用者參照。學習者能利用母語優勢，藉由電腦化的語言學習介面學習第二外語。
3. 介面設計簡單明瞭並具彈性，學習者可依據個人需求與喜好，選擇語料庫、設定結果顯示格式，以及排序方式。
4. 標明詞語相對應的翻譯，並且提供全文、段落、句子以及短語結構(sub-sentential)³的顯示方式，避免讓學習者在紛雜的大量資料中找尋顯示結果，而能節省學習者寶貴的時間。

(二) TANGO 雙語搭配詞查詢系統

TANGO 的介面及功能如下：

³短語結構，為比句子還短的結構單位，例：如「根據這項『可能構成妨害』驗證」(Under the "likely to prejudice" test).

The screenshot shows the TANGO search interface. At the top, there is a search bar with the word 'effort' entered. To the right, there is a dropdown menu for 'Text corpus' set to 'Sinorama 1990-2000'. Below the search bar, there are radio buttons for 'Verb', 'Noun', and 'Adjective', with 'Noun' selected. There is also a section for 'Collocation type' with buttons for 'VN' and 'AV'. The search results are displayed in a table with columns for the word, its frequency, and example sentences with Chinese translations. The results are: 'make effort(41)', 'make effort at(2)', 'make effort regarding(1)', 'make effort to(1)', 'put effort(21)', 'put effort into(12)', 'put in effort(5)', 'put effort for(2)', and 'put out effort(1)'. Each result includes an example sentence and its Chinese translation.

A. Text corpus: 選擇欲查詢之資料範圍，目前提供《光華雜誌》(Sinorama) 1990-2000 年之雙語語料庫，未來並會提供美國之音(Voice of America)與英國國家語料庫(British National Corpus)的語料。

B. 設定查詢之單詞之詞性，目前提供動詞、名詞以及形容詞之使用。

C. Collocation type: 設定查詢之單詞與搭配詞之排列型態，目前提供「動詞+名詞」以及「形容詞+名詞」之搭配。

D. 和搭配詞之應用在資料庫文章裡頭出現的頻率。

片語往往是學習英文最困難的部份之一，而 TANGO 系統利用英語詞塊切分技術，學習者只要輸入一個單詞或是相關搭配詞後，系統就會將查詢項目傳回雙語資料庫，找尋相關搭配詞，並且歸納出相關的片語與句式，便利學習者了解各種情境使用模式、掌握片語使用時機。因此，TANGO 系統特色包括：

1. 利用的英語詞塊切分技術，讓學生能夠在擁有大量資訊的雙語語料庫中，取得與單詞相關的應用片語。有助於使學生確實了解單詞與其他搭配詞的關聯性和使用訣竅。
2. 依據搭配詞中的搭配字進而歸類所搜尋到的資料，查詢結果按照頻率排序，學生可知哪些片語較為常用，可優先學習常用片語。

(三) Collocation Checker 錯誤搭配詞之偵測及改正

此為正在研發的系統之一。期能在學習者於線上寫作區寫作後，系統便可自動偵測出錯誤的搭配詞，並針對錯誤提供正確解答。本系統預計將有以下功能：

1. 利用英語詞塊切分技術，判定文章中的動詞名詞結構。
2. 將所張貼文章中的動詞、名詞與大量語料庫中所抽取的動詞、名詞作比對。
3. 當文章中的動詞、名詞組合，與語料庫之動詞、名詞組合不符時，動詞部份將會有所標示。
4. 移動游標至所標示之動詞上，即可顯示系統從語料庫中搜尋到符合使用情境下之正確動詞。

(四) ASR 英語語音辨識引擎的研發

這個部份目前亦正處於研發階段。為協助建立此英語自動語音辨識系統，目前做法為：

1. 使用 TIMIT 語音資料庫(TIMIT Acoustic-Phonetic Continuous Speech Corpus)⁴與美國之音

⁴ TIMIT 語音資料庫，內含 630 位以美式英語為母語人士的語音資料。其語音資料，由德州儀器(Texas Instruments, 簡稱 TI)錄製、麻省理工學院(Massachusetts Institute of Technology, 簡稱 MIT)轉寫(transcribe)。

- (Voice of America)中的資料為訓練材料，建立英語自動語音辨識引擎；
2. 收集五十位清華大學英文系學生的英語語音資料，建立 CANDLE 英語語音語料庫，以協助建立英語自動語音辨識引擎；
 3. 以軟體切分英語語音音節，比對 TIMIT 和 CANDLE 語音語料庫，找出清華學生最常犯的語音錯誤；
 4. 建立一個發音評分模式，分析以英語為第二語言的學習者(L2 learner)與母語者(native speaker)發音上的不同。

(五) 附註錯誤標記之學習者語料庫

在未來的線上英文寫作系統中，除了可以標記學習者寫作文章中的錯誤，亦可將這些錯誤的型態作歸類。例如，系統可以標記句法上的錯誤，並將錯誤型態歸類；自動判斷錯誤是否來自同位語使用錯誤，或是片語使用錯誤等等，使學習者能利用錯誤庫(Error library)做更有效率的寫作練習。目前 CANDLE 正針對一般學習者可能發生的寫作錯誤作分析與歸類，以便日後能達成線上寫作系統自動勘誤之目標。

除了上述這些自然語言處理工具及線上學習英語工具之外，正在發展中的尚包括：依據語境提供不同的中文註解、自動產生閱讀測驗的問題與答案、建構擬人化的模擬對話式交談練習與評量以及將文本的文字資料，自動轉成語音檔(Text-to-speech)的技術等。

二、智慧型英語閱讀自學環境研發

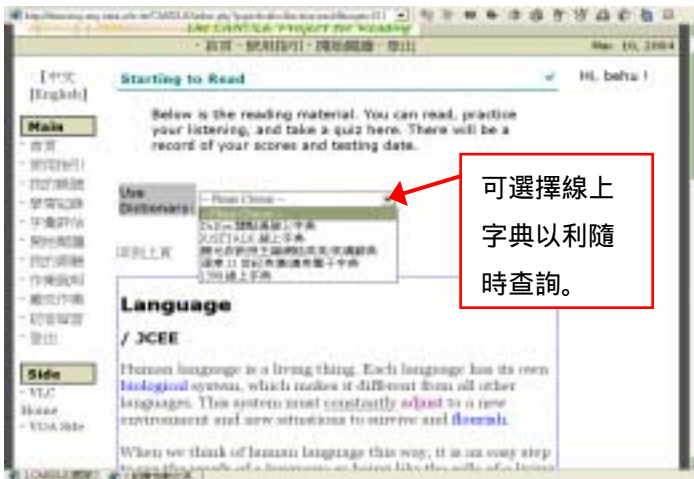
(一) 建置數位學習管理系統

研發 CANDLE 英語學習網站之電腦輔助管理系統，包含：

1. 學習者會員之登錄管理：將學習者基本資料建檔，以利後續建立學習者線上學習活動歷程、掌握學習者學習狀態，並對系統研發之良莠有一客觀的評估，作為往後網站發展方向之重要指標；
2. 建立學習者線上學習活動歷程：將學習者在網站中作的各項學習活動、學習時間、測驗分數予以紀錄，讓學習者能檢視自我學習成效，並提供教學者了解學習者學習效果之參考；
3. 提供立即回饋：如即時答案判定、寫作錯誤之標記與正確英語用法參考等；
4. 分數報表。

(二) 提供適性化的學習輔助工具及介面

希望能建構智慧型個人閱讀環境，能依據學習者個人需求，提供適性之教材。例如，學習者可以光華雜誌語料庫選擇有興趣之文章閱讀，甚且可自行在網路上找尋有興趣的文章，放入系統內觀看，而系統能夠利用單字困難度分析工具分析文章難度，讓學習者在學習過程中充分了解自我能力。介面位置在 <http://elearning.eng.ntnu.edu.tw/CANDLE/>：



在閱讀過程中，學習者可以利用線上字典或是依據情境提供之中文註解，快速達到學習的效果。此外，目前正陸續研發其他相關之語言學習工具，例如自動語言程度評分系統、自動學習活動及內容產生系統、自動測驗產生系統、閱讀策略訓練及速讀訓練等等。

三、雙語語料庫及英文文化教學之探究

除了上述研究，在系統研發的過程中，也必須對學習成效、工具、教材等進行評估。目前，有七篇碩士論文正在進行相關的研究，包括：

1. 藉助英文詞塊切分系統自動抽取動詞與名詞的搭配詞；
2. 錯誤搭配詞的偵測及改正；
3. 常用片語及單字的學習；
4. 測試商業研發之自動作文評分系統 My access 及雙語語料庫 TOTAL recall 對大學生作文及錯誤改正之影響；
5. 測試非英語系的大學生，使用 TOTAL recall 雙語語料庫，學習文化課程的效果；
6. 電腦輔助教學對於大學生動詞、名詞搭配詞學習之影響；
7. 線上廣泛英文閱讀(incidental vocabulary learning)及設定特定單字量，對於單字學習的影響。

其它協同主持人主持的文化教學部份，由台灣師範大學及東海大學各建一文化教學平台，測試融合 TOTAL recall 雙語語料庫至大學部非英語本科系學生，文化學習的課堂上之成效；另外針對聽說能力教學也將建置 *Candle talk* 學習平台，利用對話來讓學生學習社會文化策略。

伍、結語

這十個月來，我們已經設計出許多前瞻性的英語學習工具，系統也漸具雛型。未來二年內，預期將能順利完成其他工具的研發。也希望在這二年內，可以將 CANDLE 系統提供給更多的學習者使用，實際了解其學習成效以作為 CANDLE 完善的指標。

陸、參考資料

國立清華大學（2004，2月）。前瞻性數位英文學習中心 CANDLE 之研發：應用(雙語)語料庫及電腦化學習支援精簡成果報告。發表於行政院國家科學委員會主辦的成果報告會議。

國立清華大學（2003，12月）。前瞻性數位英文學習中心 CANDLE 之研發：應用(雙語)語料庫及電腦化學習支援。發表於行政院國家科學委員會主辦的成果報告會議。