

## 本期要目

- |                                     |         |
|-------------------------------------|---------|
| 壹. 學術活動預告(ROCLING, COCOSDA, ISCSLP) | 第三~五頁   |
| 貳. 中文計算語言學期刊 2004~2006 特刊彙編         | 第六~七頁   |
| 參. 專文—自動語言辨認簡介(林奇嶽、王小川)             | 第八~十四頁  |
| 肆. 會員繳費通知單                          | 第十五~十六頁 |

### 博碩士論文獎 7/1 開始申請

#### 名額及獎項：

1. 博士論文優等獎：一名，獎金二萬元，並頒給學生及指導教授獎狀各乙紙。
2. 博士論文佳作獎：一名，獎金一萬元，並頒給學生及指導教授獎狀各乙紙。
3. 碩士論文優等獎：一名，獎金一萬元，並頒給學生及指導教授獎狀各乙紙。
4. 碩士論文佳作獎：三名，獎金各伍仟元，並頒給學生及指導教授獎狀各乙紙。

#### 申請資格：

1. 國內大專院校博碩士班應屆畢業生從事計算語言學相關研究方向者，由其指導教授推薦。
2. 參賽限制：每位指導教授以推薦一篇博士論文及兩篇碩士論文為限。(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。

#### 申請程序：

1. 申請期間：7/1~7/31 日
2. 申請資料：
  - (1) 申請表一份
  - (2) 學位論文三本
  - (3) 指導教授意見書
  - (4) 身分證正、反面影本一份

- (5) 學歷證件影本一份(學生證影本、畢業證書影本或在學證明等)。
3. 申請方式：請至本會網站下載表格並連同上述資料郵寄至本會。

### 九十六年度會費開始繳交

九十五年度「一般個人」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，敬請各位會員如期繳交會費，若您個人資料有所異動，亦請惠予通知本會秘書處。繳費通知單及信用付費請參閱第 15~16 頁。

### 獎助學生出席國際會議

獎助會議：ACM SIGIR-2006  
論文題目：On Hierarchical Web Catalog Integration with Conceptual Relationships in Thesaurus  
獎助學生：陳英祥(元智大學資訊工程學系)  
獎助金額：美金 600 元。

### Mat-2500ExtV 開放商業版

本會已獲國科會授權，開放國語語音資料庫 Mat-2500ExtV 商業版之申請，商業版每套 100,000 元，擬申購之單位請至本會網站參閱詳細說明或聯絡本會秘書處。

## **PACLIC-20 截稿日期為七月一日**

PACLIC-20 今年訂於 11 月 24 日在中國武漢舉行。與會者將可乘地利之便，一覽長江三峽風景。本次會議論文集，洽由國際知名學術出版社出版，即將定案。截稿日期為 7/1，投稿詳情請參看網頁 <http://paclic20.ccnu.edu.cn/>

## **演講公告-語言與性別**

美國語言學會 (LSA) 會員遍佈世界各國，是國際間最重要的語言學學會。該會會長，康乃爾大學的 Sally McConnell-Ginet 教授，應台灣語言學學會與語言學卓越營計畫，將在七月初來台訪問。除了卓越營的演講外，將在七月十二日於輔仁大學舉行的全國語言學論文研討會 (<http://www.ling.fju.edu.tw/NCL2006/>) 中，擔任邀請講席，演講題目為：Language, gender and sexuality: Identities and ideologies (語言，性別與性向：認同與意識型態)。這不但是語言與社會互動的重要議題，也可能是繼情感判別後，下一個語言處理的熱議題。據悉，McConnell-Ginet 教授已應邀就相似議題，在今年的 COCLING/ACL 會中擔任邀請講席。欲聆聽該場演講，除報名 NCL 外，也可能針對該場另外報名，請注意台灣語言學會 (<http://linguist.tw>) 的後續訊息。

## **Fellbaum 與 Miller 榮獲 Zampolli 獎**

日前來台演講有關 WordNet 研究的 Christiane Fellbaum 和 WordNet 的原創人，著名心理學家 George Miller 教授，在 LREC 會中獲頒兩年一度，第二屆的 Zampolli 獎。Zampolli 獎由歐洲語言資源協會 (ELRA) 頒發，目的在表彰計算語言學與人類語言科技領域的重大成就。為紀念 Antonio Zampolli 的貢獻，頒獎的對象著重於語言資源與評估。得獎人除獎章外，還有歐幣一萬元獎金。有關此項獎項，與此次得獎研究的發表，請參看網頁：

[http://www.lrec-conf.org/lrec2006/article.php?id\\_article=45](http://www.lrec-conf.org/lrec2006/article.php?id_article=45)

## **台灣語言學學會-訊息公告**

由輔仁大學外語學院暨語言學研究所及台灣語言學學會共同舉辦的 2006 全國語言學論文研討會 (NCL2006)，將於 95 年 7 月 12-13 日假輔仁大學百鍊展演中心舉行，會議主題為語言學理論與應用，開幕演講將請到美國康乃爾大學語言學系的 Prof. Sally McConnell-Ginet 發表演說，演講題目為 Language and Gender (語言與性別)。活動報名時間自即日起至 6/15(四)截止，報名費用：學生四百元、一般：五百元，研討會報名及相關消息請上輔大網站查詢 ([http://www.ling.fju.edu.tw/NCL2006/index\\_top.htm](http://www.ling.fju.edu.tw/NCL2006/index_top.htm))。活動相關消息公告，以輔大網頁公告為準，若有任何疑問請電輔仁大學語言所系辦 (02)29052553 或 e-mail: G28@mails.fju.edu.tw 詢問。

第三屆語言學卓越營，是由台灣大學語言學研究所及台灣語言學學會所共同籌辦，卓越營研究課程以五場講習課程串連而成，第一場由 Prof. Christiane Fellbaum 主講的詞彙網路-詞彙的意義與結構演講，已於 5/19(五)圓滿結束，接下來尚有四場精采可期的演講將陸續舉行。演講內容、報名方式及相關資訊，請至台灣語言學學會網站查詢 (<http://linguist.tw>)。

卓越營報名台灣語言學學會 (LST) 會員另有優惠。目前台灣語言學學會 (LST) 與中華民國計算語言學學會 (ROCLING) 訂有會員互惠辦法，參加互惠活動之會員，得享有對方學會會員權益 (包括學術活動會員優惠，購買出版品與語料庫會員優惠，獲寄學會通訊與其他相關訊息，及其他雙方學會同意之權益。但不包括學會期刊)。機會難得，請中華民國計算語言學學會 (ROCLING) 踴躍參與會員互惠辦法；若有任何疑問，請電洽中華民國計算語言學學會黃琪小姐 (02)2788-3799 ext. 1502 或 E-mail: [jessie@hp.iis.sinica.edu.tw](mailto:jessie@hp.iis.sinica.edu.tw)。

# ROCLING - XVIII

## Conference on Computational Linguistics and Speech Processing

### 第十八屆自然語言與語音處理研討會

September 7-8, 2006, National Chiao Tung University, Hsinchu, Taiwan, ROC

<http://www.speech.cm.nctu.edu.tw/Rocling2006/index.html>

#### CALL FOR PAPERS

##### Conference Chair:

**Sin-Horng Chen,**  
National Chiao Tung University

##### Program Committee:

**Yih-Ru Wang, Co-Chair,**  
National Chiao Tung University

**Zhao-Ming Gao, Co-Chair,**  
National Taiwan University

**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

**Kuang-Hua Chen**  
National Taiwan University

**Lee-Feng Chien**  
Academia Sinica & Google

**Jen-Tzung Chien**  
National Cheng Kung University

**Hung-Yan Gu**  
National Taiwan University of  
Science and Technology

**Wen-Lian Hsu**  
Academia Sinica

**Bor-Shenn Jeng**  
Chunghwa Telecom Labs

**Sur-Jin Ker**  
Soochow University

**Tyne Liang**  
National Chiao Tung University

**Hsien-Chin Liou**  
National Tsing Hua University

**Wen-Hsiang Lu**  
National Cheng Kung University

**Ren-Yuan Lyu**  
Chang Gung University

**Jeng-Shyang Pan**  
National Kaohsiung University of  
Applied Sciences

**Shu-Chuan Tseng**  
Academia Sinica

**Yuen-Hsien Tseng**  
National Taiwan Normal  
University

**Hsiao-Chuan Wang**  
National Tsing Hua University

**Hsin-Min Wang**  
Academia Sinica

**Chung-Hsien Wu**  
National Cheng Kung University

**Ming-Shing Yu**  
National Chung Hsing University

**Shiaw-Shian Yu**  
CCL/TRI

##### Organization Chair:

**Yuan-Fu Liao**  
National Taipei University of  
Technology

The 18<sup>th</sup> ROCLING Conference will be held at National Chiao Tung University, Hsinchu, on September 7-8, 2006. Sponsored by Association for Computational Linguistics and Chinese Language Processing (ACLCLP), ROCLING is the most historical and major conference in the broad field of computational linguistics, speech processing, and related areas in Taiwan. ROCLING XVIII will be hosted by the Department of Communication Engineering, National Chiao Tung University. The two-day conference will feature invited talks, paper and poster sessions, and Student Workshop

ROCLING XVIII invites the submission of original and unpublished research papers on all areas of computational linguistics, natural language processing, and speech processing, including, but not limited to the following topic areas.

- |   |                                      |
|---|--------------------------------------|
| (a) cognitive/psychological linguistics         | (l) semantic web                     |
| (b) discourse/dialogue modeling                 | (m) semantics/pragmatics             |
| (c) information extraction/text mining          | (n) speech analysis/synthesis        |
| (d) information retrieval                       | (o) speech recognition/understanding |
| (e) language understanding/generation           | (p) spoken dialog systems            |
| (f) lexicon/morphology                          | (q) spoken language processing       |
| (g) machine translation/multilingual processing | (r) syntax/parsing                   |
| (h) named entity recognition                    | (s) text summarization               |
| (i) NLP applications/tools/resources            | (t) web knowledge discovery          |
| (j) phonetics/phonology                         | (u) word segmentation/POS tagging    |
| (k) question answering                          | (v) others                           |

##### Paper Submission:

Prospective authors are invited to submit full papers of no more than 25 A4-sized pages in PDF or Microsoft Word format. Papers will be accepted only by electronic submission through the conference website. Prospective authors without Web access should contact the Program Committee Co-Chair ([yrwang@mail.nctu.edu.tw](mailto:yrwang@mail.nctu.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 XVIII.

##### Important Dates:

Preliminary paper submission due:	July 1, 2006
Notification of acceptance:	August 1, 2006
Final paper due:	August 10, 2006

##### Sponsors:

Association for Computational Linguistics and Chinese Language Processing (ACLCLP)  
Department of Communication Engineering, National Chiao Tung University

# **Oriental COCOSDA 2006**

*International Conference on Speech Databases and Assessment*

**December 9-11, 2006, Malaysia**

**CALL FOR PAPERS**

The oriental chapter of COCOSDA (The International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques) is pleased to announce that the 9th Oriental COCOSDA Workshop will be held on Dec. 9-11, in Penang, Malaysia hosted by the Universiti Sains Malaysia. Oriental COCOSDA is an international workshop held annually by the oriental chapter of COCOSDA. The first preparatory meeting was held in Hong Kong in 1997 and then the past seven workshops were held in Japan, Taiwan, China, Korea, Thailand, Singapore, India and Indonesia.

Oriental COCOSDA workshop in Malaysia will help in boosting the research and development in the field of Speech Technology and will help in enthusing the interest towards Speech Technology in Southeast Asia. Papers are invited on substantial, original and unpublished research on all aspects of speech databases, assessments and speech I/O, including, but not limited to:

## **Topics**

- Speech Databases and Text Corpora
- Assessment of Speech Input and Output Technologies
- Phonetic/Phonological Systems for Oriental Languages
- Romanization of Non-Roman Characters
- Segmentation and Labeling
- Speech Prosody and Labeling
- Speech Processing Models and Systems
- Multilingual Speech Corpora
- Speech Corpus Annotation Tools
- Others

## **Paper Submission**

Papers should describe original work. The paper should emphasize completed work rather than intended work, and should indicate clearly the state of completion of the reported results. Each submission will be reviewed by at least two reviewers. A paper accepted for presentation at Oriental COCOSDA 2006 cannot be presented at any other meeting with publicly available published proceedings. The format and other details of paper submission can be downloaded from the website.

## **Important Dates**

Submission of one page abstract	Aug. 5, 2006
Notification of acceptance of paper	Aug. 26, 2006
Final Manuscript	Sept. 30, 2006
Early Registration for presenters	Oct. 14, 2006
Early Bird Registration on or before	Nov. 3, 2006
Normal Registration on or before	Nov. 23, 2006
Late Registration (with late fee)	Nov 24, 2006 onward
Workshop	Dec. 9-11, 2006

Website: <http://www.usm.my/COCOSDA2006/>

# **ISCSLP 2006**

## **第五屆中文口語語言處理國際會議**

### **December 13-16, 2006, Singapore**

ISCSLP'06 will be held during December 13-16, 2006 in Singapore hosted by the Institute for Infocomm Research (I2R) and the Chinese and Oriental Languages Information Processing Society (COLIPS). This is the 8th year after its inaugural event in Singapore and we welcome ISCSLP back to her birthplace.

Singapore, popularly known as "The Garden City", is situated at the southern tip of the Malaysian Peninsula in South-East Asia and has a rich and interesting history tracking back to 1819 when the British started a trading post which later developed into an important commercial and military imperial base. Singapore is a small but prosperous cosmopolitan state diversified with 4 main ethnic groups, namely, Chinese, Malays, Indians and Eurasians. With museums exhibiting rich collections of historical information and relics, fun theme parks, zoos and night safaris, bustling shopping and dining heavens, Singapore provides an interesting stop to unwind from the daunting stress of today's society.

We invite your participation in this premier conference, where the language from ancient civilizations embraces modern computing technology. The ISCSLP'06 will feature world-renowned plenary speakers, tutorials, exhibits, and a number of lecture and poster sessions on the following topics:

- Speech Production and Perception
- Phonetics and Phonology
- Speech Analysis
- Speech Coding
- Speech Enhancement
- Speech Recognition
- Speech Synthesis
- Language Modeling and Spoken Language Understanding
- Spoken Dialog Systems
- Spoken Language Translation
- Speaker and Language Recognition
- Indexing, Retrieval and Authoring of Speech Signals
- Multi-Modal Interface including Spoken Language Processing
- Spoken Language Resources and Technology Evaluation
- Applications of Spoken Language Processing Technology
- Others

#### **Schedule**

- Full paper submission by Jun. 15, 2006
- Notification of acceptance by Jul. 25, 2006
- Camera ready papers by Aug. 15, 2006
- Early registration by Nov. 1, 2006

Website: <http://www.iscslp2006.org/index.htm>

## 中文計算語言學期刊 2004~2006 特刊彙編

### **New Trends of Speech and Language Processing**

**Vol. 9 No. 2 August 2004**

1. Toward Constructing A multilingual Speech Corpus for Taiwanese (Min-nan), Hakka, and Mandarin Chinese  
*Ren-Yuan Lyu, Min-Siong Liang, and Yuang-Chin Chiang*
2. Multiple-Translation Spotting for Mandarin-Taiwanese Speech-to-Speech Translation  
*Jhing-Fa Wang, Shun-Chieh Lin, Hsueh-Wei Yang, and Fan-Min L)*
3. Latent Semantic Modeling and Smoothing of Chinese Language  
*Jen-Tzung Chien, Meng-Sung Wu, and Hua-Jui Peng*
4. Multi-Modal Emotion Recognition from Speech and Text  
*Ze-Jing Chuang, and Chung-Hsien Wu*
5. Multiband Approach to Robust Text-Independent Speaker Identification  
*Wan-Chen Chen, Ching-Tang Hsieh, and Eugene Lai*
6. An Innovative Distributed Speech Recognition Platform for Portable, Personalized and Humanized Wireless Devices  
*Yin-Pin Yang*

**Guest Editors: Jhing-Fa Wang and Chung-Hsien Wu**

### **Annotated Speech Corpora**

**Vol. 10 No. 2 June 2005**

1. Automatic Segmentation and Labeling for Mandarin Chinese Speech Corpora for Concatenation-Based TTS  
*Cheng-Yuan Lin, Jyh-Shing Roger Jang, and Kuan-Ting Chen*
2. The Formosan language Archive: Linguistic Analysis and Language Processing  
*Elizabeth Zeitoun, and Ching-Hua Yu*
3. Mandarin Topic-Oriented Conversations  
*Shu-Chuan Tseng*
4. MATBN: A Mandarin Chinese Broadcast News Corpus  
*Hsin-Min Wang, Berlin Chen, Jen-Wei Kuo, and Shih-Sian Cheng*
5. TAICAR: The Collection and Annotation of an In-Car Speech Database Created in Taiwan  
*Hsieh-Chang Wang, Chung-Hsien Yang, Jhing-Fa Wang, Chung-Hsien Wu, and Jen-Tzung Chien*

**Guest Editors: Hsin-Min Wang and Shu-Chuan Tseng**

### **Selected Papers from ROCLING XVI**

**Vol. 10 No. 3 September 2005**

1. Using Lexical Constraints to Enhance the Quality of Computer-Generated Multiple-Choice Cloze Items  
*Chao-Lin Liu, Chun-Hung Wang, and Zhao-Ming Gao*
2. Collocational Translation Memory Extraction Based on Statistical and Linguistic Information  
*Thomas C. Chuang, Jia-Yan Jian, Yu-Chia Chang, and Jason S. Chang*
3. Detecting Emotions in Mandarin Speech  
*Tsang-Long Pao, Yu-Te Chen, Jun-Heng Yeh, and Wen-Yuan Liao*
4. Modeling Pronunciation Variation for Bi-Lingual Mandarin/Taiwanese Speech Recognition  
*Dau-Cheng Lyu, Ren-Yuan Lyu, Yuang-Chin Chiang, and Chun-Nan Hsu*

**Guest Editors: Hsin-Min Wang, Wei-Ho Tsai, and Lee-Feng Chien**

### **Selected Papers from CLSW-5**

**Vol. 10 No. 4 December 2005**

1. The Sinica Sense Management System: Design and Implementation  
*Chu-Ren Huang, Chun-Ling Chen, Cui-Xia Weng, Hsiang-Ping Lee, Yong-Xiang Chen and Keh-Jiann Chen*
2. From Frame to Subframe: Collocational Asymmetry in Mandarin Verbs of Conversation  
*Mei-Chun Liu and Chun Edison Chang*

3. Feature Representations and Logical Compatibility between Temporal Adverbs and Aspects  
*Shih-Min Li, Su-Chu Lin and Keh-Jiann Chen*
4. 信息處理用現代漢語虛詞義類詞典研究和工作單設計  
(Study on the Machine Tractable Thesaurus Dictionary of Contemporary Chinese Functional Words for Information Processing and Design Information Terms for Dictionary Entries)  
*陳群秀(Qunxiu Chen)*
5. An Unsupervised Approach to Chinese Word Sense Disambiguation Based on Hownet  
*Hao Chen, Tingting He, Donghong Ji and Changqin Quan*
6. 以句式爲本的多義詞詞義辨識  
(Word Sense Disambiguation Based on Syntactic Construction)  
*蔡美智(Mei-Chih Tsai)*
7. 從構式語法理論看漢語詞義研究  
(A Construction-Based Approach to Chinese Lexical Semantics)  
*王惠(Hui Wang)*
8. 現代漢語中的形式動詞  
(Dummy Verbs in Contemporary Chinese)  
*俞士汶、朱學鋒、段慧明(Shiwen Yu, Xuefeng Zhu and Huiming Duan)*
9. A Synchronous Corpus-Based Study on the Usage and Perception of Judgement Terms in the Pan-Chinese Context  
*Oi Yee Kwong and Benjamin K. Tsou*
10. 《人民日報》語料庫命名實體分類的研究  
(The Chinese Named Entity Categorization Based on the People's Daily Corpus)  
*夏迎炬、于浩、西野文人(Yingju Xia, Hao Yu and Fumihito Nishino)*
11. 双向考察和驗證：并列成分中心語的語義關係和 CCD 的名詞語義分類体系  
(Bidirectional Investigation: The Semantic Relations between the Conjuncts and the Noun Taxonomy in CCD)  
*吳云芳、李素建、李芸、俞士汶(Yunfang Wu, Sujian Li, Yun Li and Shiwen Yu)*
12. Source Domains as Concept Domains in Metaphorical Expressions  
*Siaw-Fong Chung, Kathleen Ahrens and Chu-Ren Huang*
13. 隱喻性成語的語義映射(Semantic Mapping in Chinese Metaphorical Idioms)  
*李芸、李素建、王治敏、吳云芳(Yun Li, Sujian Li, Zhimin Wang and Yunfang Wu)*
14. 基于現代漢語語法信息詞典的詞語情感評價研究  
(Research on Lexical Emotional Evaluation Based on the Grammatical Knowledge-Base of Contemporary Chinese)  
*王治敏、朱學鋒、俞士汶(Zhimin Wang, Xuefeng Zhu and Shiwen Yu)*

**Guest Editors: Donghong Ji and Chu-Ren Huang**

**Human Computer Speech Processing**

**Vol. 11 No. 1 March 2006**

1. Using Duration Information in Cantonese Connected-Digit Recognition  
*Yu Zhu and Tan Lee*
2. Modeling Cantonese Pronunciation Variations for Large-Vocabulary Continuous Speech Recognition  
*Tan Lee, Patgi Kam, and Frank K. Soong*
3. A Maximum Entropy Approach for Semantic Language Modeling  
*Chuang-Hua Chueh, Hsin-Min Wang, and Jen-Tzung Chien*
4. Robust Target Speaker Tracking in Broadcast TV Streams  
*Junmei Bai, Hongchen Jiang, Shilei Zhang, Shuwu Zhang and Bo Xu*

**Guest Editors: Jen-Tzung Chien, Tan Lee, and Bo Xu**

# 自動語言辨認簡介

林奇嶽、王小川

國立清華大學電機工程研究所

d913920@oz.nthu.edu.tw hcwang@ee.nthu.edu.tw

## 摘要

語音信號之自動語言辨認研究，大約始於 1970 年代，在這近三十年中，各地的研究先進們無不絞盡腦汁，尋求有效的解決方法。拜前人的研究經驗以及電腦高速化所賜，這些努力終於在 1990 年代中期綻放光彩，許多重要的方法被提出並加以實現，奠定了近十年在此領域研究的基礎。而國內對於此主題較為陌生，且少有涉獵，因此本篇文章將以深入淺出的方式，向諸位介紹語音信號之自動語言辨認。

## 1. 背景介紹

自動語言辨認 (Automatic Language Identification)，文獻上常以縮寫 ALI 或 LID 來表示，可做為多國語言語音資訊檢索系統或是人工接線生的前處理。多國語言語音資訊檢索系統是指可以用不同語言語音作資訊檢索，此系統以平行處理方式所架構而成，當一段未知所屬語言的語音輸入系統時，可將輸入語音以平行的方式導入各個語言的語音檢索系統，再從中挑選出一組較為合理的檢索資訊。以上這種嘗試錯誤 (Try and Error) 的處理方法略嫌笨拙且較無效率，若能把自動語言辨認的技術加在前端，預先將輸入語音的所屬語言辨認出來，然後只要把輸入語音導入所屬語言的語音資訊檢索系統即可，此舉可大幅降低系統的運算時間，迅速地將資訊回應給使用者。若是做為人工接線生之前處理，則是替代人工及時的進行語言辨認。有一個發生在 AT&T Language Line 的小故事，Language Line 是提供多國語言間及時口譯的服務，宣稱可處理多達 150 種語言。這個故事是一位研究語言辨認的研究生，操著南印度及斯里蘭卡特有的坦米爾語 (Tamil) 打電話到 Language Line，結果接線生花了三分鐘才把通話引導到正確的坦米爾語通

譯，能引導成功，並非接線生用嘗試錯誤的方法找出正確結果，而是該研究生大發慈悲地用英文說出 "Tamil" 這個單字所致。若是發生緊急事件，三分鐘的延誤勢必造成嚴重的後果，適當地利用語言辨認技術，將可大幅降低因人工判斷錯誤而造成的延遲。

而第一篇關於自動語言辨認的文章，一般咸信是始於 1970 年代中期到 1980 年之間，由德州儀器公司 (Texas Instruments) 和美國空軍所合作的一系列四篇論文。文章內容在國內不易取得，但透過 Y.K. Muthusamy 的博士論文 [1]，可以大致瞭解當時是針對每個語言的聲音，取出具代表性的音段當做參考樣型 (reference template)，透過樣型比對和樣型出現的頻率來決定測試語料的所屬語言。這四篇文章粗略的介紹，如實驗設定、語言數目、辨認率等資訊，可以參考 [1] 文獻回顧的部份。1977 年，A.S. House 和 E.P. Neuberg [2] 在 JASA 發表了一篇極具指標性的文章，雖然他們並未使用真實的語音資料，但他們提出的兩個先知灼見：(1) 使用廣義音標分類 (Broad Phonetic Category)；以及 (2) 採用馬可夫過程 (Markov Process) 來建構廣義音標間的轉移機率，這個見解大大地影響了後人在此課題的研究，而從這個觀點所延伸出去的方法，已被證實是最有效的方法之一。

而 1980 年代，在 IEEE 所舉辦有關語音處理的國際會議中，並不常見相關之研究，原因不外乎是一方面受限於電腦運算速度的限制，另一方面還在摸索探尋合適的方法，此時期實為萌芽階段。進入 1990 年代，拜電腦高速化所賜，諸多堪稱經典的方法陸續被提出並加以實現，不論質跟量皆有突破性的發展，而方法之適用性也都經過大量語料的佐證，這十年可謂自動語言辨認的黃金年代。即便時光推



移，來到 2000 年，在短短幾年內雖有為數不少的文章發表，但大部份的系統架構依舊有著 1990 年代方法的影子，只是系統表現得更為快速、更為準確。表一 整理了在 IEEE Xplore 以 "language identification" 為關鍵字，所尋找出有關自動語言辨認的論文數目，我們可以約略一窺自動語言辨認的演進。

表一: IEEE Xplore 以 "language identification" 為關鍵字所搜尋到的文章數目

年代	1980 之前	1980 至 1989	1990 至 1999	2000 至今
文章數目	No	5	50+	40+

接下來的幾個章節當中，我將介紹幾個常用的方法，以及自動語言辨認的未來展望。

## 2. 自動語言辨認系統

自動語言辨認需要在各個語言當中，尋找具有鑑別力的樣型，而常用的鑑別樣型則大致可區分為四大類型 [3]：

1. 音韻資訊 Phonology
  - 各語言使用的音素集合 (phoneme set) 具有差異性
  - 各語言某些音素出現的頻率具有差異性
  - 各語言的音序 (phonotactic) 具有差異性
2. 韻律資訊 Prosody
  - 各語言的音節長短、音節複雜度、基週曲線和重音位置具有差異性
3. 形態資訊 Morphology
  - 各語言的字根 (word root) 和語彙 (lexicon) 具有差異性
4. 句法資訊 Syntax
  - 各語言的句法具有差異性

在以上四個類型當中，前兩者較廣泛被應用在系統上，原因不外乎以現階段的語音技術發展，我們可以萃取出較為可靠的音韻資訊和韻律資訊，而且所建構出來的系統也不至於太過複雜。後兩者所用的語言形態和句法屬於高階

層次資訊，一般說來高階層次資訊可以提供最佳的鑑別能力，但之前相關技術未臻成熟，無法有效地應用在系統當中。近年來高階層次資訊萃取技術已有蓬勃發展之趨勢，在可預期的未來，這類的資訊將被應用在自動語言辨認系統上。

接著談到系統實作，回顧近 30 年的發展，依據其所使用的方法和資訊，可以約略地歸納成以下五大類，

1. 基於頻譜相似度的方法
2. 基於韻律的方法
3. 基於音序辨認的方法
4. 基於多國音素單元的方法
5. 基於大字彙連續語音辨認的方法

在往後的章節當中，將會對每一種方法做一個概括性的介紹，有興趣者亦可參考三篇關於自動語言辨認回顧的文章 [3, 4, 5]。

### 2.1 基於頻譜相似度的方法

這是最早期應用在語言辨認系統的方法，從頻譜中萃取出來的傳統特徵參數皆可拿來比對，D. Cimarusti [6] 用 15 個口腔截面積係數、15 個自相關係數、5 個共振峰頻寬、15 個倒頻譜係數、15 個濾波器係數、5 個共振頻頻率、15 個對數口腔截面積比和 15 個口腔反射係數，組成一個 100 維的特徵向量。J.T. Foil [7] 和 F.J. Goodman [8] 也利用從共振峰萃取出來的資訊，來作為他們的語言辨認依據，F.J. Goodman 更指出，從共振峰取出來的參數比 LPCC 來得有用。M.A. Zissman [9] 利用高斯混合模型 (Gaussian mixture model) 來建立倒頻譜和差分倒頻譜的統計特性。不過到了 1990 年代中期之後，利用頻譜相似度的方法已較少為人所使用。

### 2.2 基於韻律的方法

韻律特性包含基頻變化、振幅大小、音節長度、重音位置等，其中以基頻變化對語言辨認較為有用，在 M. Savic [10] 的研究指出，用基頻資訊可區分西班牙語和中文；S. Itahashi

[11] 則是利用多邊直線來近似基頻曲線，並萃取出 21 個相關的特徵參數，在少量的實驗語料下得到不錯的效果，除此之外，他並認為在雜訊干擾的環境下，取出的基頻資訊較具有強健性。近期 J. Rouas [12] 利用子音母音的長度及其複雜度，外加基頻的統計特性來進行語言辨認，而他之後的研究 [13] 更利用傳統的 N 連文法，來描述長時間和短時間內韻律參數的變化。C.Y. Lin 和 H.C. Wang [14, 15] 捨棄傳統的多邊直線近似法，改採 Legendre 多項式來描述基頻曲線，在此法中，基頻曲線先投影到由低階 Legendre 多項式所構成的子空間中，接著再使用高斯混合模型或各態經歷馬可夫模型來描述投影量的統計特性，這種基頻的描述法已在實驗中證實其有效性。

人類感知的研究方面，K.M. Muthusamy [16] 在受測者完成實驗後，詢問受測者是依何種知識來辨認語言，大部份的受測者回應能以聲調的變化來辨認出中文和越南話；而也有受測者反應，西班牙語能藉由它快速的發音速度來區分。F. Ramus [17] 則利用語音合成的方式，將語音中豐富的音素資訊去除，只保留韻律的部份，結果發現受測者可以從韻律訊息區辨某幾類的語言。利用韻律方法的缺點在於，當語言數增加時，其辨認能力會大幅下降，因此韻律資訊較為適合區分語言的廣義分類，或做為其他方法的輔助資訊。

### 2.3 基於音序辨識的方法

基於音序辨認的諸多方法之中，較常採用的是稱之為"音素辨識器後接語言模型" (PRLM, Phone Recognizer followed by Language Models) 的系統架構 [18, 19]，PRLM 也是目前表現相當成功的方法，而且系統複雜度不高，一個基本的系統很容易實現出來，基本的系統架構圖可參考 圖一。架構圖分成四大部份：預處理、音素辨識器、語言模型、後端處理。首先要先取得語言 X 的語料來訓練出音素辨識器，音素模型只需採用內文無關 (Context-Independent) 聲學模型即可，舉例來說，若音素辨識器的語言選擇中文，我們只需訓練 22 個聲母 (initial) 和 40 個韻母 (final)

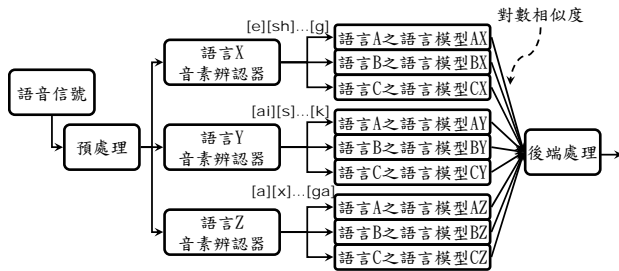
的聲學模型。而在訓練 PRLM 的後接語言模型時，先把各個語言的訓練語料經過預處理轉換成一連串的特徵向量，再通過語言 X 的音素辨識器來產生一連串的音素序列，這裡要注意的是在進行音素解碼時不使用任何的語言模型，意即採用自由內文文法 (Context-Free Grammar)，接著把每個語言所得到的音素序列，訓練出對應的語言模型。至於在評估階段，未知語言的測試語音一樣先經過預處理，通過語言 X 的音素辨識器得到音素序列，這個音素序列再由後端數個語言模型個別予以評分，產生最高相似度的語言模型即判定為測試語音的所屬語言。PRLM 比較有趣的地方在於，前端音素辨識器的使用語言可以與欲辨認的語言無關，就如圖一所示，X、A、B、C 可以代表四個不同的語言，但只有 A、B、C 才是我們想要判別的語言。



圖一: PRLM 架構

M.A. Zissman [18] 把數個 PRLM 系統平行接在一起，形成"平行音素辨識器後接語言模型" (PPRLM, Parallel PRLM) 的架構。如 Figure 2 所示，前端使用三個語言 X、Y、Z 的音素辨識器，每個音素辨識器後面分別接著三個不同語言的語言模型，而欲辨認的語言為 A、B、C 三種語言。普遍來說，PPRLM 的表現比 PRLM 來得更佳，只是在系統設計時，未必能夠取得多個語言的語料來訓練音素辨識器。為了降低前端音素辨識器的設計門檻，P.A. Torres-Carrasquillo 改用"高斯混合模型表徵器" (GMM Tokenizer) 來取代音素辨識器的地位 [20]，高斯混合模型表徵器是一個以音框為處理單元的系統，與前述以音素為處理單元有所不同，因此其優點是不需要任何標註語料即可建構出系統，而缺點則是以音框為單位過於精細，使得音框跟音框之間的表徵值變化過於頻繁，為了解決過於精細的問題，原本使用的特徵參數可以用"位移差分化倒頻譜特徵" (shifted delta cepstral features) [21] 來取代，其原理是沿

著時間軸，以一定間隔取出數個特徵向量予以重新組合而成，但合適的間隔該為多久還是一個值得深究的課題。



圖二: PPRLM 架構

## 2.4 基於多國音素單元的方法

一般來說，不同語言有各自的音素集合，若比較這些音素集合，可以發現其中有些音素的聲學特性很類似，有些卻非常的與眾不同，這些獨特的音素就可以做為辨別語言的工具。在文獻上 [22]，聲學特性類似的音素以 poly phoneme 稱之，即意味著這個音素存在於諸多語言之中，舉例來說，[u] [ū] [U] 三個音素在聲學上相當接近，可以用一個音素 [u] 來代表這三個音素 [23]；另一方面，聲學特性不同的音素則稱為 mono phoneme，而這些大相庭逕的音素則是用來辨認不同語言的利器之一。poly phoneme 和 mono phoneme 可藉由人工 [24] 或是資料驅動 (data-driven) 的方式篩選出來，但前者需要具備通曉跨語言音素間聲學知識的專家，大部份情況下並不容易找到合適的人選，也因此人力有限的情況下，資料驅動的方法較為可行。在 O. Anderson 和 P. Dalsgaard [22, 25, 26] 的作法中，先將數個語言的音素集成一個\*\*字？\*\*集合，再利用音素辨識的結果，建立每個音素和其他音素之間的混淆矩陣 (confusion matrix)，再從中把幾個相似度高的音素合併，組成 poly phoneme 集合；而相似度較低的音素則分別歸類在各語言的 mono phoneme 集合。就如 [25] 中所提到表示法，四國語言的音素可以歸類成

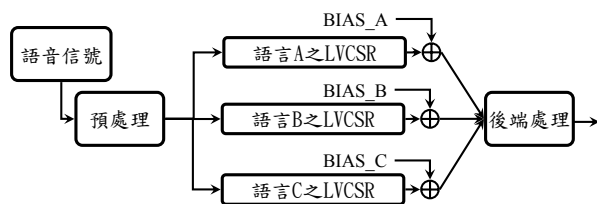
$$\Psi = \{ \Psi^p \cup \Psi_{UK}^m \cup \Psi_D^m \cup \Psi_G^m \cup \Psi_I^m \}$$

$$= \{ \psi_0, \psi_1, \psi_2, \dots, \psi_K \}$$

$\Psi$  是歸類後的音素\*\*字？\*\*集合，共含有  $K+1$  個音素，其中  $\Psi^p$  代表四國語言共享的 poly phonemes 所構成的集合；而  $\Psi_{UK}^m$ 、 $\Psi_D^m$ 、 $\Psi_G^m$ 、 $\Psi_I^m$  則分別代表英荷德義四個語言獨自的 mono phonemes。辨認時是計算各語言 mono phonemes 出現的頻率，看哪個語言的 mono phonemes 出現最多次，就認定輸入語音屬於該語言。P. Dalsgaard [26] 甚至以實驗來檢視該取多少個 mono phonemes 和 poly phonemes，在英德西三國語言的實驗中，結果顯示總共用 20 個 mono phonemes 即可得到不錯的辨認結果，即使增加 mono phonemes 的數目，辨認率也不會有太大的改變。相反地，增加 poly phonemes 的數目反而會對辨認結果造成反效果，實驗中 poly phoneme 的數目分別設定為 10、20、30、40、50，結果只用 10 個時的辨認率最高，若增加數目則辨認率會開始下降。

## 2.5 基於大字彙連續語音辨識的方法

這個方法看似與基於音序辨識的方法雷同，但兩者之間還是有些許差異性存在，在這個方法當中，是把數個語言的大字彙連續語音辨識器 (Large Vocabulary Continuous Speech Recognizer) 如圖三所示地以平行的方式架構起來，在辨認時，一段未知語言的語音信號分別通過這些大字彙連續語音辨識器，看哪一個語言的大字彙連續語音辨識器輸出較高的分數，就判別輸入語音屬於該語言。S. Mendoza [27] 建立英文、日文、西班牙文三個語言的語音辨識器，並在這三個語言間進行兩國語言辨認以及三國語言辨認，最後的辨認結果相當優異。在實驗之中，他們也發現利用語音辨識器來進行語言辨認時，聲學相似度會有偏差值 (bias) 的現象產生，所以要先扣掉偏差值再進行辨認才能得到一個較好的結果。T. Schultz [28] 和 J.L. Hieronymus [29] 也秉持相同的精神，利用四國或五國的語音辨識器，並結合聲學、音序、字彙、文法等訊息來進行辨認。這個方法與前述的「音素辨識器後接語言模型」相較起來，需要更多的人工標註資料和語料庫，也因此進入門檻相對地增高許多。



圖三:基於大字彙連續語音辨識的方法

### 3. 語言辨識語料庫

早期的語言辨識研究並無公認的語料庫，因此較難比較各家系統孰優孰劣，情況在 1992 年有所改變，Y.K. Muthusamy 收集整理出第一個專門用來評估語言辨識的語料庫：OGI-TS 語料庫 [30]，其全名為 Oregon Graduate Institute Multi-Language Telephone Speech Corpus。語料庫中共包含 10 種不同的語言：英文、法文、德文、日文、中文、韓文、西班牙語、越南話、波斯語 (Farsi)、坦米爾語。每種語言共收集 90 通的電話錄音，並以 50、20、20 的數目區分成訓練集、發展集和測試集。錄音的內容大致可區分成三大類：

1. 固定或常用的語句
  - 包含錄音者的母語、常用的語言、一星期七天的唸法、數字 0 到 10。
2. 領域限定的語句
  - 描述家鄉的景色，家鄉的氣候，房間的裝潢，最近一餐吃些什麼。
3. 未限定的語句

而語句的長度也分成三種，3 秒、10 秒和 45 秒。除此之外，各個語言尚有七種廣義音標分類 (broad phonetic category) 的標註資訊，如母音、摩擦音、塞音、靜音或未帶聲，以及三種不同位置的響音，關於這個語料庫的詳細資訊可以參見 Y.K. Muthusamy 的博士論文。

但在 OGI-TS 語料中，每個語言的語料總時數大約為 180 分鐘，訓練語料大約才佔 100 分鐘，對於模型訓練來說，並不是很足夠。有鑑於 OGI-TS 語料庫的不足，在 1990 年代中期之後，大多加上 CallFriend 和 CallHome 兩個較為大型的語料庫，CallFriend 含有 12 種

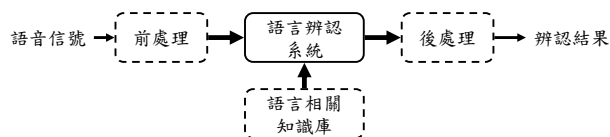
語言，除了上述的十種語言之外，還加入阿拉伯話和印度話，每種語言的訓練語料約有 10 個小時，其中英文、中文和西班牙語更高達 20 個小時；而 CallHome 含有 6 種語言：英文、阿拉伯話、德文、日文、中文、西班牙語，訓練語料則超過 10 個小時。在 2005 年 OGI 更推出一套新的語言辨識語料庫 CSLU : 22 Languages Corpus，扣除方言共含括 21 種語言，比上述的語料庫包含更多不同類型的語言，例如多了廣東話、捷克語、匈牙利話、馬來語 (Malay)、義大利語、波蘭語、葡萄牙語、俄語、瑞典語、斯華西里語 (Swahili)。這些語料庫皆由 Linguistic Data Consortium [31] 所發行，可透過他們的網站購買取得。

### 4. 語言辨識的未來展望

在經過近十年的努力之後，利用不同語言之間音序差異的方法，已被諸多實驗證實相當地管用，話雖如此，還是有些地方需要去加以改進。就以目前表現最為突出的"平行音素辨識器後接語言模型"來說，運算時間花費較多是它的缺點之一，在文獻 [32] 中提到，前端掛載六國語言音素辨識器的 PPRLM，單位時間語音信號的處理速度是 14x，也就是說要花十四秒鐘來處理一秒鐘的輸入語音；而 PPRLM 後端每種語言的語言模型，也需要大量的訓練語料來提供其可靠性。這兩種尚待克服的問題，也讓語言辨識系統離實際應用還有一段路要走。

圖四是一般語言辨識系統的基本架構，以"前處理"來說，要關注的問題在於是否能找到與語言無關的切割方法，如此一來可取代原本來自某些語言的音素辨識器，避免掉需要音素標註語料的問題，目前針對這個問題所嘗試的解決方案只有高斯混合模型表徵器，當然，利用表徵器的方法也有其缺點，因此語言辨識系統的前處理是尚待深究的部份。而"語言相關知識庫"則冀望引入更多來自各語言的知識，利用每種語言獨特的結構，如是否有聲調的變化，音節有哪些特殊的組成方式等，用這些相對少量，但極俱彰顯各語言獨特性的知識來減少訓練語料的需求，尤其當手邊沒有足夠多的訓練語料時，擁有語言相關的知識能以簡馭繁。另

外, "後處理"來自於當系統是採平行架構時, 該如何同時處理從各個子系統所傳遞出來的決策或訊息, 傳統的作法是直接把子系統產生的分數累加起來, 或是利用統計或迴歸的方式替這些分數建立一個分類器, 最後再加以判別。在這裡是希望能找到更具有理論架構的結合方法, 套用在最後的決策或是直接內嵌在各個子系統當中, 以期達到更好的語言分類結果。



圖四：語言辨認系統的架構圖

## 5. 結語

在這篇文章當中, 簡介了語言辨認發展的歷史, 數種適用於語言辨認的方法, 以及專門評估語言辨認系統效能的語料庫。雖然文中提到的每種方法都有其優缺點, 但以目前的發展來看, 基於音序的辨認方法為首要之選, 即使沒有足夠的語料建構 PPRLM, 利用單一語言音素辨識器的 PRLM 亦能提供不錯的效果。除此之外, 並提到語言辨認系統一些尚待解決的問題, 包括前處理、後處理, 以及語言相關知識等部份, 若能克服所述的難題, 語言辨認系統離實際應用將會更進一步。

## 參考文獻

- [1] Y.K. Muthusamy, "A Segmental Approach to Automatic Language Identification," in Phd. dissertation of Oregon Graduate Institute of Science & Technology, 1993.
- [2] A.S. House and E.P. Neuberg, "Toward automatic identification of the language of an utterance. I. Preliminary methodological consideration," in Journal of the Acoustical Society of America, 62(3):708-713, 1977.
- [3] M.A. Zissman, and K.M. Berkling, "Automatic language identification," in Speech Communication, Vol. 35, pp. 115-124, 2001.
- [4] Y.K. Muthusamy, E. Barnard, and R.A. Cole, "Reviewing automatic language identification," in IEEE Signal Processing Mag., Vol. 11, no. 4, pp.33-41, October 1994.
- [5] M.A. Zissman, "Comparison of four approaches to automatic language identification of telephone speech," in IEEE Trans. Speech and Audio Processing, Vol. 4, no. 1, pp. 31-44, January, 1996.
- [6] D. Cimarusti, and R.B. Ives, "Development of an automatic identification system of spoken languages-Phase I," in IEEE ICASSP, pp. 1661-1664, 1982.
- [7] J.T. Foil, "Language identification using noisy speech," in IEEE ICASSP, pp. 861-864, Tokyo, 1986.
- [8] F.J. Goodman, A.F. Martin, and R.E. Wohlford, "Improved automatic language identification in noisy speech," in IEEE ICASSP pp. 528-531, 1989.
- [9] M.A. Zissman, "Automatic language identification using Gaussian mixture and hidden Markov models," in IEEE ICASSP 1993, Vol. 2, pp. 399-402, 1993.
- [10] M. Savic, E. Acosta, and S.K. Gupta, "An automatic language identification system," in IEEE ICASSP, pp. 817-820, 1991.
- [11] S. Itahashi, J.X. Zhou, and K. Tanaka, "Spoken language discrimination using speech fundamental frequency," in Proc. of ICSLPD94, Yokohama, Japan, 1994, pp.1899-1902.
- [12] J. Rouas, J. Farinas, F. Pellegrino, and R. Andre-Obrecht, "Modeling prosody for language identification on read and spontaneous speech," in Proc. of ICASSPD2003, Hong Kong, China, April 2003, Vol. I, pp. 40-43.
- [13] J. Rouas, "Modeling long and short-term prosody for language identification," in Proc. of EUROSPEECH, pp. 2257-2260, 2005.
- [14] C.Y. Lin, and H.C. Wang, "Language identification using pitch information," in Proc. ICASSP 2005, Philadelphia, USA, Vol. 1, 2005, pp.601-604.

- [15] C.Y. Lin, and H.C. Wang, "*Language identification using pitch information in the ergodic Markov model*," to appear in Proc. ICASSP 2006.
- [16] Y.K. Muthusamy, N. Jain, and R.A. Cole, "*Perceptual benchmarks for automatic language identification*," in IEEE ICASSP, Vol. 1, pp. 333-336, 1994.
- [17] F. Ramus, and J. Mehler, "*Language identification with suprasegmental cues, A study based on speech resynthesis*," in Journal of Acoustic Society of America, 105(1), pp.512-521, 1999.
- [18] M.A. Zissman, and E. Singer, "*Automatic language identification of telephone speech messages using phoneme recognition and N-gram modeling*," in IEEE ICASSP 1994, Vol. 1, pp. 305-308, 1994.
- [19] T.J. Hazen, and V.W. Zue, "*Automatic language identification using a segment-based approach*," in Proceeding of Eurospeech 93, Vol. 2, pp. 1303-1306, September 1993.
- [20] P.A. Torres-Carrasquillo, D.A. Reynolds, and J.R. Deller Jr., "*Language identification using Gaussian mixture model tokenization*," in IEEE ICASSP 2002, Vol. 1, pp. 757-760, 2002.
- [21] P.A. Torres-Carrasquillo, D.A. Reynolds, E. Singer, M.A. Kohler, R.J. Greene, and J.R. Deller, Jr., "*Approaches to language identification using Gaussian mixture models and shifted delta cepstral features*," in ICSLP, pp. 77-81, Denver, Sep. 2002.
- [22] O. Anderson, P. Dalsgaard, and W. Barry, "*On the use of data-driven clustering technique for identification of poly- and mono-phonemes for four European languages*," in IEEE ICASSP, Vol. 1, pp. 121-124, 1994.
- [23] D. Zhu, M. Adda-Decker, and F. Antoine, "*Different Size Multilingual Phone Inventories and Context Dependent Acoustic Models for LID*," in EUROSPEECH, pp. 2833-2836, Lisbon, Sep., 2005
- [24] K.M. Berkling, T. Arai, and E. Barnard, "*Analysis of phoneme-based features for language identification*," in IEEE ICASSP, Vol. 1, pp. 289-292, 1994.
- [25] P. Dalsgaard, and O. Anderson, "*Application of inter-language phoneme similarities for language identification*," in ICSLP, pp. 1903-1906, Yokohama, 1994.
- [26] P. Dalsgaard, O. Anderson, H. Hesselager, and B. Petek, "*Language identification using language-dependent phonemes and language-independent speech units*," in ICSLP, pp. 1808-1811, 1996.
- [27] S. Mendoza, L. Gillick, Y. Ito, S. Lowe, and M. Newinan, "*Automatic language identification using large vocabulary continuous speech recognition*," in IEEE ICASSP, pp. 785-788, 1995.
- [28] T. Schultz, I. Rogina, and A. Waibel, "*LVCSR-based language identification*," in IEEE ICASSP, pp. 781-784, 1995.
- [29] J.L. Hieronymus, and S. Kadambe, "*Robust spoken language identification using large vocabulary speech recognition*," in IEEE ICASSP, pp. 1111-1114, 1997.
- [30] Y.K. Muthusamy, R.A. Cole, and B.T. Oshika, "*The OGI Multilanguage telephone speech corpus*," in Proc. of ICSLPD92, Banf, Alberta, Canada, October 1992, Vol. 2, pp.895-898.
- [31] "<http://www ldc.upenn.edu/>"
- [32] E. Singer, P.A. Torres-Carrasquillo, T.P. Gleason, W.M. Campbell, and D.A. Reynolds, "*Acoustic, phonetic, and discriminative approaches to automatic language identification*," in Proc. EUROSPEECH, pp. 1345-1348, 2003.

# 中華民國計算語言學學會

## 九十六年度會員資格更新暨會員資料異動通知

各位會員您們好！

九十五年度「一般個人」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，敬請各位會員如期繳交會費，同時如果您的個人資料有所異動，亦請惠予通知。

中華民國計算語言學學會會員十大獨享權益如下：

1. 免費寄發《中文計算語言學期刊》(Computational Linguistics & Chinese Language Processing) (一年四期)
2. 免費寄發「計算語言學通訊」(一年六期)
3. 享有學會出版品優惠價(相關出版品請參閱學會網頁)
4. 享有中研院資訊所詞庫、語料庫等資料庫授權之優惠價
5. 享有學會主辦會議之優惠註冊費
6. 學生會員可申請參加國際計算語言學相關會議之獎助
7. 學生會員可申請計算語言學研討會之最佳論文獎並獲獎學金
8. 透過電子郵件遞即時公布計算語言相關之研究活動
9. 享有學會網頁提供之豐富研究資源
10. 會員網路討論區(email discussion list)

繳交會費       會員資料異動單(請勾選您想要的選項)

### 一、會員資料異動(煩請詳細填寫)

姓名：\_\_\_\_\_ 會員別：終身      個人      學生

學歷：\_\_\_\_\_

現職：\_\_\_\_\_

地址：\_\_\_\_\_

電話：\_\_\_\_\_

傳真：\_\_\_\_\_

E-mail：\_\_\_\_\_ (請務必提供)

### 二、會費：

- |                      |                |
|----------------------|----------------|
| 1. 會費：終身會費：\$ 10,000 | 團體年費：\$ 20,000 |
| 個人年費：\$ 1,000        | 學生年費：\$ 500    |

2. 繳費方式：

- 郵政劃撥（帳號：19166251 號 戶名：中華民國計算語言學學會）  
 支票（支票抬頭請加註「中華民國計算語言學學會」）  
 信用卡（請加填信用卡付款單）

3. 繳費期限：九十六年度會費請於 95 年 7 月 1 日前繳交

註：終身會員、今年度已繳交者及今年新加入之會員，勿需繳納會費。

三、96 年度會費有效期間：95 年 7 月 1 日至 96 年 6 月 30 日。

四、信用卡付款單

姓名：\_\_\_\_\_（請以正楷書寫）

卡別： VISA CARD       MASTER CARD       JCB CARD

卡號：——— 有效期：\_\_\_\_\_

卡片後三碼：\_\_\_\_\_ 發卡銀行：\_\_\_\_\_

金額：\_\_\_\_\_ 持卡人簽名：\_\_\_\_\_

五、上述資料填妥後，請傳真至本會。

六、通訊資料：

地址：台北市 115 南港區研究院路 2 段 128 號 中研院資訊所（轉）

電話：(02)2788-3799 轉 1502 傳真：(02)2788-1638

E-mail：aclclp@hp.iis.sinica.edu.tw

網頁：http://www.aclclp.org.tw/