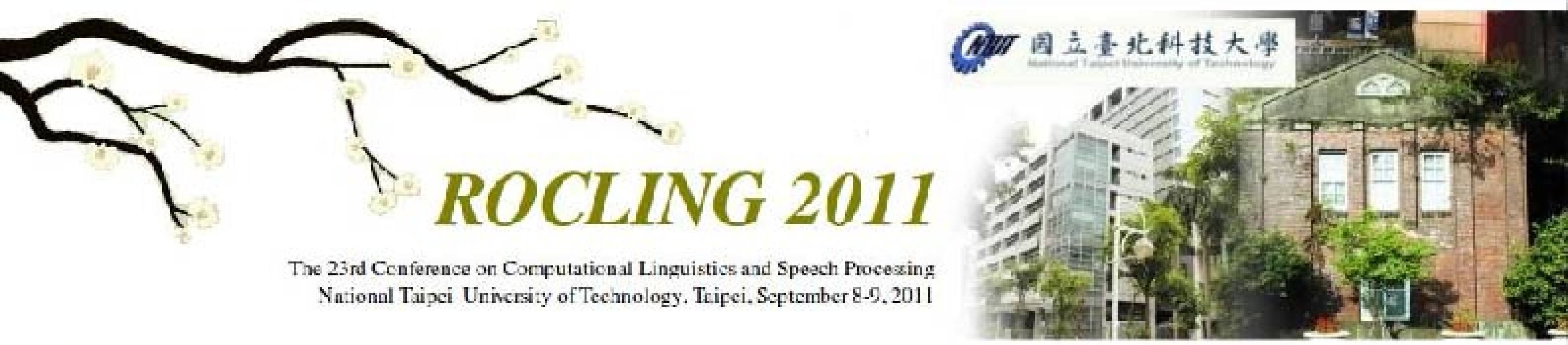
23rd Conference on
Computational Linguistics and
Speech Processing

第二十三屆
自然語言與語音處理研討會



ROCLING 2011

The 23rd Conference on Computational Linguistics and Speech Processing
National Taipei University of Technology, Taipei, September 8-9, 2011

時間：2011.9.8 ~ 2011.9.9
地點：國立台北科技大學科技大樓國際會議廳
網址：http://sites.google.com/site/rocling2011/

主辦單位： 國立台北科技大學電子工程系
中華民國計算語言學學會

協辦單位： 國科會工程科技推展中心
中央研究院資訊科學研究所
中華電信研究所
資訊工業策進會
工研院資通所
元智大學資訊管理學系

贊助單位： 無敵科技(股)公司
賽微科技(股)公司
致遠科技(股)公司

序 言

本年度的ROCLING共收到投稿數為 30 篇，每篇論文都至少經 2 位該領域的專家學者審查，最後議程委員會共接受 12 篇oral presentation論文和 13 篇poster presentation論文，包含了語音辨認與合成、機器翻譯、語音學與音韻學之分析及應用、自然語言處理之應用、工具與資源、及語音識別和理解等領域，此審查結果維持了ROCLING歷屆以來一貫的論文品質，並兼顧多層面研究人員的參與，在此非常感謝論文審查委員的把關。

今年的議程安排，除了最新的學術論文的發表外，也邀請三位語音及自然語言處理領域專家給予專題演講，包括新加坡Institute for Infoc omm Research李海洲博士、Google, Taiwan 總經理簡立峰博士與中國東北大學計算機軟件研究所朱靖波博士。此外，王駿發教授也熱心幫忙組織一個Panel Discussion Session ，關心語音及自然語言科技的未來應用與發展。非常感謝他們的幫忙。

我們同時要感謝國科會工程科技推展中心、中央研究院資訊科學研究所、中華電信研究所、資訊工業策進會、工研院資通所、無敵科技、賽微科技與致遠科技的協辦與贊助。另外，也謝謝台北科技大學電子系語音訊號處理實驗室，多媒體訊號處理實驗室，與元智大學資管系自然語言處理與文字探勘實驗室的同學們在各項事務上的協助。

最後，感激各位與會先進的積極參與和支持，使本次研討會得以順利舉行。

大會主席 廖元甫
議程主席 蔡偉和、禹良治 謹識
2011 年 9 月 8 日

Organization

Conference Chair

Yuan-Fu Liao	廖元甫	National Taipei university of Technology

Program Committee Co-Chairs

Wei-Ho Tsai	蔡偉和	National Taipei university of Technology
Liang-Chih Yu	禹良治	Yuan Ze University

Program Committee Members

Guo-Wei Bian	邊國維	Huafan University
Chia-Hui Chang	張嘉惠	National Central University
Jason S. Chang	張俊盛	National Tsing Hua University
Jing-Shin Chang	張景新	National Chi Nan University
Yi-Hsiang Chao	趙怡翔	Ching Yun University
Berlin Chen	陳柏琳	National Taiwan Normal University
Chia-Ping Chen	陳嘉平	National Sun Yat Sen University
Chien-Chin Chen	陳建錦	National Taiwan University
Hsin-Hsi Chen	陳信希	National Taiwan University
Keh-Jiann Chen	陳克健	Academia Sinica
Kuang-Hua Chen	陳光華	National Taiwan University
Sin-Horng Chen	陳信宏	National Chiao Tung University
Tai-Shih Chi	冀泰石	National Chiao Tung University
Jen-Tzung Chien	簡仁宗	National Cheng Kung University
Chih-Yi Chiu	邱志義	National ChiaYi University
Hung-Yan Gu	古鴻炎	National Taiwan University of Science and Technology
Wei-Tyng Hong	洪維廷	Yuan Ze University
Wen-Lian Hsu	許聞廉	Academia Sinica
Jeih-weih Hung	洪志偉	National Chi Nan University
Jyh-Shing Jang	張智星	National Tsing Hua University
Chih-Chung Kuo	郭志忠	Industrial Technology Research Institute

June-Jei Kuo	郭俊桔	National Chung Hsing University
Wen-Hising Lai	賴玟杏	National Kaohsiung First University of Science and Technology
Chao-Lin Liu	劉昭麟	National Chengchi University
Jyi-Shane Liu	劉吉軒	National Chengchi University
Chuan-Jie Lin	林川傑	National Taiwan Ocean University
Shou-De Lin	林守德	National Taiwan University
Richard Tzong-Han Tsai	蔡宗翰	Yuan Ze University
Yuen-Hsien Tseng	曾元顯	National Taiwan Normal University
Hsiao-Chuan Wang	王小川	National Tsing Hua University
Hsin-Min Wang	王新民	Academia Sinica
Yih-Ru Wang	王逸如	National Chiao Tung University
Chin-Sheng Yang	楊錦生	Yuan Ze University
Cheng-Zen Yang	楊正仁	Yuan Ze University
Ming-Shing Yu	余明興	National Chung Hsing University
Chung-Hsien Wu	吳宗憲	National Chen Kung University
Gin-Der Wu	吳俊德	National Chi Nan University
Jui-Feng Yeh	葉瑞峰	National ChiaYi University
Shih-Hung Wu	吳世弘	Chaoyang University of Technology

Program Overview

September 8, 2011 (Thursday) 9:10 ~ 20:00		
09:10-10:00	Registration	
10:00:10:10	Opening Ceremony	Prof. Leehter Yao Chair: Prof. Yuan-Fu Liao
10:10-11:10	Invited Talk: Machine Transliteration – Translating the Untranslatables	Speaker: Prof. Haizhou Li, Institute for Infocomm Research, Singapore Chair: Prof. Hsiao-Chun Wang
11:10-11:40	Coffee Break	
11:40-12:40	Oral Session 1: Speech Recognition and Synthesis	Chair: Prof. Chia-Ping Chen
12:40-13:30	Lunch	
13:30-14:30	ACLCLP meeting for future directions/Poster Session 1:NSC Project reports	
14:30-15:30	Invited Talk: Opportunities and Technology Challenges for Search Engines in the mobile internet	Speaker: Dr Lee-Feng Chien, General Manager, Google Chair: Prof. Hsin-Hsi Chen
15:30-16:00	Coffee Break/IJCLCLP editors meeting(資工系系辦公室會議室科技大樓 3 樓)	
16:00-17:00	Panel Discussion: Frontier of speech science and technology for real life	Panelists: 吳宗憲教授，簡立峰博士 郭志忠博士，沈家麟博士 Chair: Prof. Jhing-Fa Wang
17:00~18:00	Walking to banguet place (美麗信飯店)	
18:00-20:00	Banquet (美麗信飯店 buffet)	

September 9, 2011 (Friday) 9:30 ~ 16:20		
9:30-10:30	Invited Talk: Some Issues on Statistical Machine Translation Using Source and Target (or) Syntax	Speaker: Prof. Jingbo Zhu, Northeastern University, ShenYang, China Chair: Prof. Liang-Chih Yu
10:30-11:00	Coffee Break	
11:00-12:00	Oral Session 2: Machine Translation and Word Segmentation	Chair: Prof. Yuen-Hsien Tseng
12:00-13:00	Lunch	
13:00-14:30	Poster Session 2: Poster Papers	
14:30-15:00	Coffee Break	
15:00-16:00	Oral Session 3: Lexicon, Resources and NLP applications	Chair: Prof. June-Jei Kuo
16:00-16:20	Closing Ceremony and Best Paper Award	

Technical Program Details

Oral Session 1: Speech Recognition and Synthesis

Time: Thursday, September 8, 11:40-12:40

1. Empirical Comparisons of Various Discriminative Language Models for Speech Recognition
   Min-Hsuan Lai, Bang-Xuan Huang, Kuan-Yu Chen and Berlin Chen
2. Compensating the Speech Features via Discrete Cosine Transform for Robust Speech Recognition
   Hsin-Ju Hsieh, Wen-Hsiang Tu and Jeih-Weih Hung
3. 聯合語者、雜訊環境與說話內容因素分析之強健性語音辨認
   Sheng-Tang Wu, Wei-Te Fang and Yuan-Fu Liao
4. Evaluation of TTS Systems in Intelligibility and Comprehension Tasks
   Yu-Yun Chang

Oral Session 2: Machine Translation and Word Segmentation

Time: Friday, September 9, 11:00-12:00

1. 片語式機器翻譯中未知詞與落單字的問題探討
   蔣明撰, 黃仲淇, 顏合淨, 黃士庭, 張俊盛, 楊秉哲, 谷圳
2. 英文技術文獻中一般動詞與其受詞之中文翻譯的語境效用
   Yi-Hsuan Chuang, Jui-Ping Wang, Chia-Chi Tsai and Chao-Lin Liu
3. Unsupervised Overlapping Feature Selection for Conditional Random Fields Learning in Chinese Word Segmentation
   Ting-Hao Yang, Tian-Jian Jiang, Chan-Hung Kuo, Richard Tzong-han Tsai and Wen-Lian Hsu
4. 繁體中文文本中對於日文人名及異體字的處理策略
   林川傑, 詹嘉丞, 陳彥亨, 鮑建威

Oral Session 3: Lexicon, Resources and NLP applications

Time: Friday, September 9, 15:00-16:00

1. 動補結構的及物性及修飾對象
   You-Shan Chung and Keh-Jiann Chen
2. Predicting the Semantic Orientation of Terms in E-HowNet
   Cheng-Ru Li, Chi-Hsin Yu and Hsin-Hsi Chen
3. 聲符部件排序與形聲字發音規則探勘
   Chia-Hui Chang and Sean Lin
4. Frequency, Collocation, and Statistical Modeling of Lexical Items: A Case Study of Temporal Expressions in an Elderly Speaker Corpus
   Sheng-Fu Wang, Jing-Chen Yang, Yu-Yun Chang, Yu-Wen Liu and Shu-Kai Hsieh

Invited Speaker: Haizhou Li

Machine Transliteration - Translating the Untranslatables

Abstract

Machine transliteration is the process of automatically rewriting the script of a word from one language to another, while preserving pronunciation. The last decade has seen a tremendous progress and a growth of interests from theory to practice of machine transliteration. In this talk, I will present an overview of the fundamentals, algorithms and applications, in particular, transliteration between English and Chinese. I will also report the findings in the most recent transliteration evaluation campaigns - NEWS 2009 and NEWS 2010 Machine Transliteration Shared Tasks.

Biography

Dr. Haizhou Li is currently the Principal Scientist and Department Head of Human Language Technology at the Institute for Infocomm Research. Dr Li has worked on speech and language technology in academia and industry since1988. He taught in the University of Hong Kong (1988-1990), South China University of Technology (1990-1994), and Nanyang Technological University (2006-). He was a Visiting Professor at CRIN/INRIA in France (1994-1995), and at the University of New South Wales in Australia (2008). As a technologist, he was appointed as Research Manager in Apple-ISS Research Centre (1996-1998), Research Director in Lernout & Hauspie Asia Pacific (1999-2001), and Vice President in InfoTalk Corp. Ltd (2001-2003).

Dr Li's research interests include automatic speech recognition, natural language processing and social robotics. He has published over 150 technical papers in international journals and conferences. He holds five international patents. Dr Li now serves as an Associate Editor of IEEE Transactions on Audio, Speech and Language Processing, ACM Transactions on Speech and Language Processing, and Springer International Journal of Social Robotics. He is an elected Board Member of the International Speech Communication Association (ISCA, 2009-2013), an Executive Board Member of the Asian Federation of Natural Language Processing (AFNLP, 2006-2010), and a Senior Member of IEEE since 2001. Dr Li was the Local Organizing Chair of SIGIR 2008 and ACL-IJCNLP 2009. He was appointed the General Chair of ACL 2012 and Interspeech 2014. He was the recipient of National Infocomm Award of Singapore in 2001. He was named one of the two Nokia Professors 2009 by Nokia Foundation in recognition of his contribution to speaker and language recognition technologies.

Invited Speaker: Lee-Feng Chien

Opportunities and Technology Challenges for Search Engines in the Mobile Internet

Abstract

The web started on the PC, within the recent years it started arriving for mobile devices. It will soon arrive for many other types of devices we haven't even thought of yet. This is going to open up some pretty amazing business opportunities and technology challenges for search engine development, and online marketing that can seek to promote businesses by increasing their visibility when users access the mobile Internet. So what I'd like to do is walk you through some of the macro trends that are converging right now to set us up for explosive growth in the mobile Internet over the next couple of years and then walk you through some of the technology challenges that await those who understand and invest in -- or at least start experimenting in -- this area.

Biography

Dr. Lee-Feng Chien is working with Google as GM of Google Taiwan and engineering site director of Taiwan/Hong Kong R&D center. He is known for his work on Chinese natural language processing, has researched Chinese analysis systems, language models, speech recognition systems, and search engineering technology for many years. He has served on program committees for major conferences and journal editorial boards in the related academic areas, and is the author of a hundred of technical papers. Prior to joining Google, he was research fellow and deputy director of the Institute of information Science, Academia Sinica, Taiwan, and also jointly appointed as a professor of the Information Management Department of National Taiwan University. He received his Ph.D. in CS from National Taiwan University in 1991.

Invited Speaker: Jingbo Zhu

Some Issues on Statistical Machine Translation Using Source and (or) Target Syntax

Abstract

Machine Translation (MT) is one of the oldest sub-fields in Natural Language Processing (NLP) and Artificial Intelligence (AI). During the last decade, syntax-based approaches have received growing interests in MT community, showing state-of-the-art performance for many language pairs such as Chinese-English. In this talk, I will present our recent work on syntax-based MT, and some approaches to performing translation using source and (or) target syntax, involving string-to-tree, tree-to-string and tree-to-tree SMT paradigms. Also, an empirical study is shown to compare the strengths and weaknesses among various syntax-based SMT approaches. Furthermore, several interesting issues are further addressed to investigate what the major problems in current (syntax-based) MT paradigm are. Finally, I will spend a little time to introduce a new open-source SMT toolkit (named NEUTrans) which was developed by the NLPLab of Northeastern University, and our current efforts on incorporating syntax-based SMT paradigms into this open SMT platform.

Biography

Dr. Jingbo Zhu is a full professor of Computer Science at the Northeastern University at Shenyang, China, and is in charge of research activities within the Natural Language Processing Laboratory (NEU-NLPlab, htttp://www.nlplab.com). He received his Ph.D. degree in computer software and theory from the Northeastern University in 1999. He was a visiting researcher at the City University of Hongkong (2004) and ISI, University of Southern California at Los Angeles (2006-2007), and was selected by the Program for New Century Excellent Talents in University, Ministry of Education (2005). His research interests include machine translation, syntactic parsing, sentiment analysis and text mining. He has published 100+ papers in many high-level journals and conferences including IEEE Transactions on Affective Computing, IEEE Transactions on Audio, Speech and Language Processing, ACM Transactions on Speech and Language Processing, ACM Transactions on Asian Language Information Processing, and ACL/EMNLP/Coling, etc.

ROCLONG 2011 Abstracts

Oral Session 1: Speech Recognition and Synthesis
Time: Thursday, September 8, 11:40-12:40

1. **Empirical Comparisons of Various Discriminative Language Models for Speech Recognition**

   Min-Hsuan Lai, Bang-Xuan Huang, Kuan-Yu Chen and Berlin Chen

   傳統語言模型(Language Models)是藉由使用大量的文字語料訓練而成，以機率模型來描述自然語言的規律性。*N* 連(*N*-gram)語言模型是最常見的語言模型，被用來估測每一個詞出現在已知前*N*-1 個歷史詞之後的條件機率。此外，傳統語言模型大多是以最大化相似度為訓練目標；因此，當它被使用於語音辨識上時，對於降低語音辨識錯誤率常會有所侷限。近年來，有別於傳統語言模型的鑑別式語言模型(Discriminative Language Model)陸續地被提出；與傳統語言模型不同的是，鑑別式語言模型是以最小化語音辨識錯誤率做為訓練準則，期望所訓練出的語言模型可以幫助降低語音辨識的錯誤率。本論文探究基於不同訓練準則的鑑別式語言模型，分析各種鑑別式語言模型之基礎特性，並且比較它們被使用於大詞彙連續語音辨識(Large Vocabulary Continuous Speech Recognition,LV CSR)時之效能。同時，本論文亦提出將邊際(Margin)概念引入於鑑別式語言模型的訓練準則中。實驗結果顯示，相較於傳統*N* 連語言模型，使用鑑別式語言模型能對於大詞彙連續語音辨識有相當程度的幫助；而本論文所提出的基於邊際資訊之鑑別式語言模型亦能夠進一步地提升語音辨識的正確率。

2. **Compensating the Speech Features via Discrete Cosine Transform for Robust Speech Recognition**

   Hsin-Ju Hsieh, Wen-Hsiang Tu and Jeih-Weih Hung

   In this paper, we develop a series of algorithms to improve the noise robustness of speech features based on discrete cosine transform (DCT). The DCT-based modulation spectra of clean speech feature streams in the training set are employed to generate two sequences representing the reference magnitudes and magnitude weights, respectively. The two sequences are then used to update the magnitude spectrum of each feature stream in the training and testing sets. The resulting new feature streams have shown robustness against the noise distortion. The experiments conducted on the Aurora-2 digit string database reveal that the proposed DCT-based approaches can provide relative error reduction rates of over 25% as compared with the baseline system using MVN-processed MFCC features. Experimental results also show that these new algorithms are well additive to many noise robustness methods to produce even higher recognition accuracy rates.

3. **聯合語者、雜訊環境與說話內容因素分析之強健性語音辨認**
Sheng-Tang Wu, Wei-Te Fang and Yuan-Fu Liao

本論文主要研究於強健性語音辨認上，我們提出聯合語者、雜訊環境與語音內容因素分析(Joint Speaker and Noisy Environment and Speech Content Factor Analysis；JSEC)，主要是透過聯合因素分析，在特徵空間做即時語音辨認模型補償(online recognition model compensation)，使得調適出來的模型與測試環境能夠盡量匹配，進而提升辨識效果。此外，我們先將 JSEC 分解成語音和非語音二個模型做模型調適、估算影響因素，接著每個模型再利用階層式的概念，語音特性考慮之因素分成雜訊環境特徵空間、語者特徵空間、說話內容特徵空間與獨特因素空間分別估算，非語音特性考慮之因素則分成雜訊特徵空間和獨特因素空間分別估算，最後再把語音和非語音組合回辨認用的模型，用此方式來降低我們的參數量。我們使用 Aurora2 語料庫做實驗，在複合情境的訓練模式下，我們得到最佳的辨識錯誤率為 4.37%，比傳統強健性參數求取方法 MVA (Mean subtraction ，Variance normalization，and ARMA filtering)[1][2]的錯誤率 4.99%低了許多，也比我們先前提出的 JSE (Joint Speaker and Noisy Environment Factor Analysis)[11]方法的錯誤率相當甚至好一點。除了辨認率之外，我們提出的方法也能使得調適模型的參數量大幅下降，JSEC 參數量約為傳統 MVA 的 4 倍，也比 JSE 方法少了十分之一的參數量，因此為更有效率的調適方法。

4. **Evaluation of TTS Systems in Intelligibility and Comprehension Tasks**
Yu-Yun Chang

This paper aims at finding the relationships between intelligibility and comprehensibility in speech synthesizers, and tries to design an appropriate comprehension task for evaluating the speech synthesizers’ comprehensibility. It is predicted that speech synthesizer with higher intelligibility, will have greater performance in comprehension. Also, since the two most popular used speech synthesis methods are HMM-based and unit selection, this study tries to compare whether the HTS-2008 (HMM-based) or Multisyn (unit selection) speech synthesizer has better performance in application. Natural speech is applied in the experiment as a controlled group to the speech synthesizers. The results in the intelligibility test shows that natural speech is better than HTS-2008, and HTS-2008 is much better than Multisyn system. Whereas, in the comprehension task, all the three speech systems present not much differences in speech comprehending process. This is because that the two speech synthesizers have reached the threshold of enough intelligibility to provide high speech comprehension quality. Therefore, although with equal comprehensible speech quality between HTS-2008 and Multisyn systems, HTS-2008 speech synthesizer is more recommended and preferable due to its higher intelligibility.

Oral Session 2: Machine Translation and Word Segmentation
Time: Friday, September 9, 11:00-12:00

1. **片語式機器翻譯中未知詞與落單字的問題探討**

   蔣明撰, 黃仲淇, 顏合淨, 黃士庭, 張俊盛, 楊秉哲, 谷圳

   近年來，機器翻譯技術蓬勃發展並越顯重要。然而，現存的機器翻譯系統對於（系統未收錄）未知詞多採直接輸出到目標翻譯的方式。此忽略的舉動可能造成未知詞附近的選字錯誤，或是其附近的翻譯字詞順序錯置，因而降低翻譯品質或降低閱讀者對翻譯文章的理解。經過我們的初步分析，大約有 25%的系統未知詞可用重述（paraphrase）的方式來作翻譯，另外的 25%可利用組合單字翻譯來翻譯。另外，現有的片語式（phrase-based）機器翻譯系統對於落單字（singleton）的翻譯效果也未加重視。所謂的落單字是指系統在翻譯此字時必須單獨翻譯：此字沒法與前面或是後面的字組合成連續字詞片語或是文法翻譯結構。本研究將建構於片語式機器翻譯處理技術，開發未知詞翻譯模組和落單字翻譯模組。實驗結果顯示即使在不假額外的雙語資料，我們的未知詞翻譯模組仍勝出片語式翻譯系統，尤其是在包含有未知詞的句子上。

2. **英文技術文獻中一般動詞與其受詞之中文翻譯的語境效用**

   Yi-Hsuan Chuang, Jui-Ping Wang, Chia-Chi Tsai and Chao-Lin Liu

   We investigate the potential contribution of a very specific feature to the quality of Chinese translations of English verbs. Researchers have studied the effects of the linguistic information about the verbs being translated, and many have reported how considering the objects of the verbs will facilitate the quality of translations. In this paper, we take an extreme assumption and examine the results: How will the availability of the Chinese translations of the objects help the translations of the verbs. We explored the issue with thousands of samples that we extracted from 2011 NTCIR PatentMT workshop and Scientific American. The results indicated that the extra information improved the quality of the translations, but not quite significantly. We plan to refine and extend our experiments to achieve more decisive conclusions.

3. **Unsupervised Overlapping Feature Selection for Conditional Random Fields Learning in Chinese Word Segmentation**

   Ting-Hao Yang, Tian-Jian Jiang, Chan-Hung Kuo, Richard Tzong-han Tsai and Wen-Lian Hsu

   This work represents several unsupervised feature selections based on frequent strings that help improve conditional random fields (CRF) model for Chinese word segmentation (CWS). These features include character-based N-gram (CNG), Accessor Variety based string (AVS), and Term Contributed Frequency (TCF) with a specific manner of boundary overlapping. For the experiment, the baseline is the *6-tag*, a state-of-the-art labeling scheme of CRF-based CWS; and the data set is acquired from SIGHAN CWS bakeoff 2005. The experiment results show that all of those features improve our system's F_1 measure (*F*) and Recall of Out-of-Vocabulary (*Roov*). In particular, the feature collections which contain AVS feature outperform other types of features in terms of *F*, whereas the feature collections containing TCB/TCF information has better *Roov*.

4. **繁體中文文本中對於日文人名及異體字的處理策略**

   林川傑, 詹嘉丞, 陳彥亨, 鮑建威

   本論文提出一個可於進行繁體中文文章斷詞時，處理非繁體中文詞彙的方法。包括以日文漢字或中文書寫的日文人名，或是以異體字書寫的同義詞等。處理人名時，我們提出了姓名組合機率模型。處理日文人名時，我們也提出一個異體字對應的方法，可將日文姓氏及名用字對應至繁體中文用字。這方法甚至可以處理同一句子中同時出現日文及繁簡中文書寫方式的情形。在加入各種特殊類別以及中日人名處理方法後，斷詞效能 F-measure 由 94.16%提昇至 96.06%。另外對 109 篇標有日文人名的中文新聞文章進行斷詞實驗，測試集裡 862 個日文人名被成功斷成詞的比例為 83.18%。論文中亦針對以異體字書寫的中文詞提出了一套可行的處理方式。

Oral Session 3: Lexicon, Resources and NLP applications
Time: Friday, September 9, 15:00-16:00

1. **動補結構的及物性及修飾對象**

   You-Shan Chung and Keh-Jiann Chen

   動補結構〈VR〉的分析一直是中文裡一個棘手的問題。其中，動補的及物性以及 V2〈第二個動詞，表達某動作的結果 R〉是修飾主語還是賓語，更是很多理論試圖解釋的現象。本篇論文透過 V1〈第一個動詞，表達造成某結果的動作〉和 V2 本身是及物或不及物動詞以及 V1 及 V2 和主語、賓語搭配的可能性，成功預測大多數動補句型的及物性以及 V2 是修飾主語還是賓語。除了預測正確性及覆蓋率高，我們的方法在處理多數句型時只需知道 V1 和 V2 本身是否是及物動詞以及 V1 和 V2 和主語賓語搭配的可能性，因此也較其他須先辨識 V2 是修飾有生命還是無生命物體、V1 和 V2 的域外和域內論元為何的分析方法更符合自動處理的需要。

2. **Predicting the Semantic Orientation of Terms in E-HowNet**
   Cheng-Ru Li, Chi-Hsin Yu and Hsin-Hsi Chen

   詞彙的意見極性是句子及文件層次意見分析的重要基礎，雖然目前已經存在一些人工標記的中文情緒字典，但如何自動標記詞彙的意見極性，仍是一個重要的工作。這篇論文的目的是為廣義知網的詞彙自動標記意見極性。我們運用監督式機器學習的方法，抽取不同來源的各種有用特徵並加以整合，來預測詞彙的意見極性。實驗結果顯示，廣義知網詞彙意見極性預測的準確率可到達 92.33%，這個結果跟人的標記準確率不相上下。

3. **聲符部件排序與形聲字發音規則探勘**
   Chia-Hui Chang and Sean Lin

   近年來台灣有相當多的新移民的加入，這些新移民在口語的學習上雖然有地利之變，但是在漢字的認識上則是相當弱勢。由於漢字乃是圖形文字，學習單一字的成本相對的高。如果可以讓漢字教一個字，可以學到十個字，對於漢字教學的成效應有相當的助益。本文從部件教學的概念出發，考慮聲符的發音強度、出現頻率、及筆劃數，做為聲符部件教學順序的準則。我們利用部件發音強度[8]，以線性加總、幾合乘積、及調和平均三種方法對部件排序。根據此部件排序學習，前五個部件便可延伸學習多達 140 個相似發音的漢字。進一步，我們應用中研院文獻處理實驗室所建立的「漢字構形資料庫」，以及標記所得之形聲字，拆解形聲字組成的部件，挖掘串連漢字之間關係的形音關聯規則。我們從 600 萬條發音規則中篩選與分群出 8 條高信賴度與兩組各約 10 條高支持度的規則，並藉由這些規則來輔助漢語發音的學習效率。

4. **Frequency, Collocation, and Statistical Modeling of Lexical Items: A Case Study of Temporal Expressions in an Elderly Speaker Corpus**
   Sheng-Fu Wang, Jing-Chen Yang, Yu-Yun Chang, Yu-Wen Liu and Shu-Kai Hsieh

   This study examines how different dimensions of corpus frequency data may affect the outcome of statistical modeling of lexical items. The corpus used in our analysis is an elderly speaker corpus in its early development, and the target words are temporal expressions, which might reveal how the speech produced by the elderly is organized. We conduct divisive hierarchical clustering based on two different dimensions of corpus data, namely raw frequency distribution and collocation-based vectors. Results show when different dimensions of data were used as the input, the target terms were indeed clustered in different ways. Analyses based on frequency distributions and collocational patterns are distinct from each other. Specifically, statistically-based collocational analysis produces more distinct clustering results that differentiate temporal terms more delicately than do the ones based on raw frequency.

國科會工程科技推展中心
Engineering and Technology Promotion Center
National Science Council

中央研究院 資訊科學研究所
Institute of Information Science, Academia Sinica

中華電信研究所

財團法人資訊工業策進會
INSTITUTE FOR INFORMATION INDUSTRY

工業技術研究院
Industrial Technology
Research Institute

BESTA
無敵科技

Cyberon

BEHAVIOR Behavior Design Corp.
致遠科技股份有限公司