Chu-Ren Huang

CURRICULUM VITAE
January 2008

Record

ADDRESS

Institute of Linguistics   tel. (886)2-2652-5019
Academia Sinica   fax. (886)2-2785-6622
Nankang, Taipei, Taiwan   E-mail: churen@gate.sinica.edu.tw

EDUCATION

1982-1986   Cornell University
Ph.D. in Linguistics (January 1987)

1976-1980   National Taiwan University
B.A. in English Language and Literature (June 1980)

CURRENT EMPLOYMENT

Research Fellow, Institute of Linguistics, Academia Sinica

Founding Director, Computational Linguistics and Chinese Language Processing PhD Program, Taiwan International Graduate Programs

CURRENT ACADEMIC AND PROFESSIONAL POSITIONS

General Assembly Member, International Permanent Committee on Linguistics (CIPL), 2007-

Board Member and Head of Division of Information Processing, National Languages Committee, Ministry of Education, Republic of China (2007-)

Standing Executive Board Member, Linguistic Society of Taiwan (2007-)

Steering Committee Member, Chinese Lexical Semantics Workshops (CLSW) (2007-)

Monitor Member, Association of Computational Linguistics and Chinese Language Processing (2005-)

AFNLP Liaison to ISO TC37 SC4 (2005-)

Permanent Member, International Committee on Computational Linguistics (ICCL) (September 2004-)

Steering Committee Member, Pacific Asia Conference on Language, Information, and Computation (PACLIC) (2003-)

Founding Council Member, Open Language Archives Community (OLAC)

(2003-)

Overseas Adjunct Research Professor, Institute of Computational Linguistics, Peking University (2003-)

Founding Co-chair, Asian Language Resources Committee (2001-)

Project Co-director, Chinese Knowledge Information Processing Group, Academia Sinica (1987-)

ACADEMIC POSITIONS HELD

2007 (April) Visiting Scholar, Department of Computing, Hong Kong Polytechnic University

2006 (March-April) Visiting Scholar, ICL-CNR, Pisa Italy

2004 (April) -2005 (July) Vice-Director, Institute of Linguistics, Academia Sinica

2003 (May-July) Visiting Professor, Institute of Chinese Linguistics, City University of Hong Kong

2000-2004 External Academic Advisor, BALIS Honor Program, Department of Chinese, Translation, and Linguistics, City University of Hong Kong

2002 (November-December) Visiting Scholar, Institute of Research on Cognitive Sciences, University of Pennsylvania

2001 (September-October) Visiting Scholar, Institute of Research on Cognitive Sciences, University of Pennsylvania

2001 (June-August) Adjunct Professor and Visiting Scholar, Dept. of Linguistics, UC-Santa Barbara

1999 (June-July) Visiting Scholar, C.R.L.A.O. EHESS/CNRS, France

1997 Adjunct Professor, Institute of Translation, National Taiwan Normal University

1993-1997 Research Fellow, Institute of History and Philology, Academia Sinica

1995-1996 Adjunct Professor, Institute of Linguistics, National Chung Cheng University

1993-1995 Adjunct Professor, Institute of Linguistics, National Tsing, Hua University

1987-1993 Associate Research Fellow, Institute of History and Philology, Academia Sinica

1987-1993 Adjunct Associate Professor, Institute of Linguistics, National Tsing Hua University

1992 (Jan.-June) Visiting Scholar, Department of Linguistics, University of California at San Diego

1991 (Sept.-Dec.) Visiting Scholar, the Center for the Study of Language and

Information, Stanford University

1987 (March-July)   Post-doctoral Associate Research Fellow, Institute of History and Philology, Academia Sinica

1986(Sept.)-1987(Feb.) Computational Linguist. Weidner Communications Corporation. Northbrook, Illinois

PROFESSIONAL POSITIONS HELD

2005-2007   President, Linguistic Society of Taiwan

2002-2005   Vice President, Linguistic Society of Taiwan

2001-2005   Chair, Preparatory Committee, Computational Linguistics and Chinese Language Processing PhD Program, Taiwan International Graduate Programs

2001-2003   Founding Advisory Committee Member, Open Language Archives Community (OLAC)

2001-2003   Founding Executive Board Member, ACL Special Interest Group on Chinese Language Processing (SigHAN)

1998-2002 Executive Secretary, Association of Computational Linguistics and Chinese Language Processing (a.k.a. ROCLING)

1998-2001 Founding Executive Board Member, Linguistic Society of Taiwan

1997-1998 Executive Committee Member, International Association of Chinese Linguistics

1994-1995   President, the Computational Linguistics Society of the R.O.C. (ROCLING)

1988-1993 Founding Executive Board Member, ROCLING

Vision and Research Accomplishments

My research over the past twenty years covers four major areas:

♦ Exploration of linguistic theories,

♦ Construction of explanatory accounts of language and cognition,

♦ Construction of Language Archives and Infrastructures for Linguistic Resources, and

♦ Creation of synergy between Human Language Technology and Linguistic Theory.

These four major areas are inter-connected and systemized by three important scientific perspectives. These three perspectives underlie my research vision:

1. Language is an organism that does not exist independent of species.

2. A functional account of language must be based on the structural representation of knowledge.

3. Words are atoms in both representation of and access to linguistic knowledge.

My research accomplishments in the above research areas are:

1. Exploration of linguistic theories

My early research concentrates on understanding and integrating different linguistic theories. Based my research expertise covering three grammatical theories, I have focused in recent years on developing a theory as well as a representational framework for lexical semantics. I attempt to approach the issue with a comprehensive account of the facts of a language, and develop a sense and meaning driven account of the data. The theory must account for semantics, morphology, syntax, as well as how they interact to generate actual linguistic behaviors. Important accomplishments in the past include:

1) Pioneering the application of GPSG (Generalized Phrase Structure Grammar) to the study of Mandarin Chinese and Southern-Min. I was involved in the development of the theory of GPSG. From 1984 to 1991, I published one book, 5 journal papers, and over 10 conference papers using this theory.

2) Pioneering the application of LFG (Lexical Functional Grammar) theory in the study of Mandarin Chinese. I was involved in the development of the theory as well as studies to apply the theory to computational linguistic research.

From 1985 to 1997, I published one book, 7 journal papers, and over 30 conference papers adopting LFG.

3) Introducing the theory of CG (Construction Grammar). From 1995 to 1999, I published one journal paper and 3 conference papers using this theory.

4) Developed and proposed the Module-Attribute Representation of Verbal Semantics (MARVS).This theory integrates the empirical research methodology of corpus linguistics and adopted a verbal semantic representational model that is event-driven. The theory also establishes a framework to predict the behavior and distribution of syntactic forms based lexical meaning. From 1996 on, I have published 10 journal papers and over 20 conference papers in this theory.

2. Construction of explanatory accounts of language and cognition

An explanatory account of language and cognition is one of the most fascinating research issues in the scientific study of language. I have adopted both a mathematical model (Huang 1989) as well as a simulation of a psychological model based on large quantity of corpus data (Huang, Ahrens, and Chen 1998, etc.) in order to approach an explanatory account. My recent, and hopefully more significant, contribution is to try to free my research from the constraints of formalized theories and instead concentrate on pursuing explanations to two uniquely human facts in language cognition. These two facts are the inherent ambiguity in human languages and the use of linguistic metaphor.

The crucial issue involved in ambiguity in human language is the fact that all artificial languages and formal communication systems eliminate ambiguity by design, yet human languages contain ambiguity by nature. In other words, ambiguity is a unique design feature of human languages. Since ambiguity reduces the effective informational load, it cannot be motivated in an efficient driven theory such as the Communication Theory of Shannon. I began by studying the nature and complete behavior of ambiguity in a specific language. From 1996, I have studied the complete data involving categorical ambiguity in the five million-word Sinica Corpus. Based on the analysis of all the instances of categorical ambiguity, I proposed several generalizations as well as explanatory accounts; I also proposed two definitions of ambiguity, with one referring to the potential for a certain linguistic element to be ambiguous, whereas the other refers to the complexity of ambiguity when it does occur. My findings demonstrate that the degree of ambiguity, with regard to potential of an element to be ambiguous, is dependent on its frequency and category, but irrelevant to the number of categories in the classification system. On the other hand, the degree of ambiguity with regard to the complexity of ambiguity is dependent on the category of that element as well as the number of categories used in the

classification system. I proposed an explanatory account of the above generalizations integrating evolutionary perspectives in Huang (2004).

Metaphor, on the other hand, is the use of knowledge or conceptual structure based on more familiar or concrete domains to describe an element from a unfamiliar or more abstract domain, such as 'LIFE is a JOURNEY.' What metaphor offers is a way for human being to 'map' from the knowledge of a familiar domain to a new and unfamiliar domain. Adopting a mapping model will allow us to verify and explain how knowledge is structured in human cognition. We (including Kathleen Ahrens, and Siaw-fong Chung) pioneered the research method of extracting metaphors from corpora, and adopting the structure of an upper ontology to explain and account for the mapping. Our group has designed both on-line processing as well as ontology-based test to verify our account. I have already published 4 journal papers and over 10 conference articles in this area.

3. Construction of Language Archives and Infrastructures for Linguistic Resources

Language archives are both the foundation and last resort for sustaining cultural tradition as well as saving endangered languages. In addition, they also provide an essential infrastructure for empirical studies of languages. They also represent the obligatory resources utilized in language technology and knowledge engineering. I have directed or co-directed (with Keh-jiann Chen and other colleagues) the completion of the following language archives:

1) Academia Sinica Balanced Corpus of Modern Mandarin Chinese (Sinica Corpus): This is the first Chinese corpus in the world to have grammatical category tagging (1994); it is also the first corpus in the world to be fully accessible on the web (1996). It has become a respected and often-used tool for computational and theoretical studies in Chinese linguistics.

2) Sinica TreeBank (2001): This is the first syntactically annotated Chinese corpus as well as one of the first corpus marked with argument role information in the world.

3) Academia Sinica Bilingual Ontological WordNet (Sinica BOW 2003): This is the first complete English-Chinese bilingual wordnet as well as the first wordnet to be fully integrated with an ontology.

Additional accomplishments include the proposal, overseeing and implementation of the 'Language Archives' and 'Linguistic Anchoring of Digital Archives' project under the National Digital Archives Program. The pioneering design provided linguistic anchoring in addition to the conventional spatio-temporal coordinates and prepared a foundation for rich and effective

content manipulation. I also bridged our research with the current state of the art in the world for digital archives. I am involved in the Open Language Archives Community (OLAC), serving first as a consulting committee member before becoming one of the founding executive council members. I am also the founding co-chair of the Asian Language Resources Committee.

Last, but not least, I have also spearheaded the use of language archives in a web-based learning environment. Of the three completed language learning resources sites that I directed or co-directed, 'The Adventures in Wen-land' is worth noting. Since its inauguration at the end of 2002, it has received over quarter of a million visits, mostly from grade school teachers and students.

4. Creation of synergy between Human Language Technology and Linguistics Theory

I have played a central role in the development and promotion of Chinese computational linguistics since 1987. On the one hand, I strive to bring up the research in Taiwan to international standard, and raise the visibility as well as influence of Chinese language processing in the field at the same time. On the other hand, I aim to find an optimal synergy between rule-based linguistic accounts and stochastically oriented computational studies, promoting studies that are both theoretically significant but robust enough for computational applications. One of the most significant milestones achieved is my organization of COLING2002, the premier computational linguistics conference, in the world in Taipei. In this area, I have published two compiled books, edited 4 journal special issues, published 20 journal papers and over 40 conference papers.

Other significant accomplishments include drafting the national standard for Chinese segmentation for information processing (CNS14366, 1999), the proposal of the ICG (Information-based Case Grammar) framework in conjunction with Keh-jiann Chen, and construction of a lexical knowledgebase based on ICG (CKIP Lexicon,1998). In the past two years, my study on lexical semantics has lead to a set of theoretically felicitous criteria for differentiating senses in Chinese as well as a set of operational criteria, which have been applied to the large scale Sinica Sense Management System.

Current Research Focus

1. A sense-based grammatical and knowledge presentation system, and construction of sense-based KnowlegeBase for Mandarin Chinese

Based on the three research perspectives outlined at the beginning of this proposal, what is essential for my research and for the scientific study of language in general is a corpus that contains comprehensive coverage of linguistic content (i.e. meaning and sense) as well as linguistics forms (i.e.

including morphological, phonological, and syntactic items) from one particular language. Such a corpus must be built upon a complete and robust theory or lexical semantic and representational framework, must have a comprehensive coverage of all lexical items, and also have a rich, yet felicitous, semantic markup. Works in progress include:

a) The lexical semantic representational theory of MARVS. It has already been developed and has been verified and modified based on on-going studies.

b) The criteria and operational principles for sense differentiation has been proposed. Studies are now being carried out to explore the possibility of automatic tagging of senses in Chinese.

c) Complete the design and implementation of a sense-management system for Sinica Corpus. We have completed analysis and account of over 2,000 senses in Mandarin Chinese and have designed and implemented these senses within the Sinica Sense Management system.

2. Comparative Studies of lexicon-driven conceptual structures and ontologies

Once we can automatically differentiate word senses, we will be able to carry out systematic comparative studies of conceptual structures and knowledge structures. The research topics will include:

1) A complete linguistic knowledge structure of Chinese based on word senses. This study largely follows the WordNet paradigm set up by George Miller, Christiane Fellbaum and colleagues at Princeton University for English. The basic idea is that the complete network of semantic relations extended and linked from the complete inventory of word senses in a language represent a conceptual shared by speakers of that language. We have established a theory to identify senses and semantic relations in Chinese in our on-going research and will build up the semantic network of Chinese WordNet based on our thorough and robust account of the word senses in Chinese.

2) A contrastive study of cross-lingual semantic relations. Based on the framework of WordNet, I have argued previously (Huang et al. 2003) that cross-lingual translation equivalents must also be described in terms of semantic relations. We are now constructing a bilingual WordNet with semantic relation tagging, which will enable us to pursue studies of comparative and contrastive studies of the different knowledge structures between two languages.

3) Comparative studies of the structural correspondences between conceptual structures and lexical knowledge. The study of conceptual structure has been a favorite topic in cognitive sciences. However, even though there have been a few theories of conceptual structures, none has become a commonly accepted standard. In computer science, ontology was proposed to provide a shared framework for knowledge exchanges. SUMO (Suggested Upper Merged Ontology), an upper ontology construction which

was sanctioned by IEEE, offers a promising shared representational scheme that can overcome both domain and language barriers. It can also serve as the reference point to compare two different knowledge systems. Our recent study has not only verified a Chinese version of SUMO but also verified the conceptual links from over 100,000 Chinese lemmas to SUMO nodes. This resource allows the conceptual location of each Chinese lexical term to be determined. In terms of comparative studies of knowledge systems, we also completed a pilot study based on 300 Tang Poem (Huang et al. 2004). In this study we propose a method for constructing a domain ontology based on domain lexicon. The constructed domain ontology facilitates a global comparison of knowledge structures and allows a holistic understanding of the knowledge structure of that domain.

Future Research Plans

Lexicon-driven Study of Ontology and Conceptual Structure

The success of a lexicon-driven research is dependent on the completeness and quality of the lexical knowledgebase. Hence my research will start with building such lexical knowledge-bases to allow diachronic, cross-lingual, and cross-domain study of knowledge systems.

1. Construction of a complete periodical database of historical Chinese lexicons. This will be the basic resource for diachronic comparative studies of Chinese texts.

2. Construction of ontology-based, cross-domain and historical knowledge systems. Based on the constructed ontologies, we will also study establish a framework and methodology of descriptive and comparative studies of knowledge systems based on ontology.

3. Application of the research results integrating lexical semantics and ontology to the Semantic Web. When the Semantic Web and its related technologies mature, it will popularize the juxtaposition of web technology and ontology. It is conceivable that web usage will entail dealing with ontology at that stage. Our basic research on lexical semantics and ontology can be applied to the following issues of Human Language Technology:

    (1) Multi-lingual and cross-lingual information processing

    (2) Meaning-driven knowledge acquisition

    (3) Interpretation and value-added reproduction of human knowledge

Current Research Grants

2005-2009   Lexicon Driven Studies of Ontology and Conceptual Structure (Academia Sinica Investigator Grant, US$555,000 in total, I was the only recipient in humanities and social science in the

inaugural version of this prestigious grant from Academia Sinica)

2005-2008   Towards A Standard Lexicon for Asian Languages for the Application of Semantic Web. (NEDO Foundation, Japan. International Collaborative Project. US.$120,000 in total for Taiwan portion of project.)

2006   Seeking Excellence in Linguistics: Taiwan Summer Institute in Linguistics, National Science Council (Co-PI, US $35,000)

2002-2006   Linguistic Anchoring of Digital Archives, National Digital Archives Program (PI, US$240,000 for 2004, est. US$1,200,000 in total)

2002-2006   Language Archives, National Digital Archives Program (Coordinator: 2000-2003, Co-PI: 2002-2006, US$220,000 for 2004, est. US$ 1,250,000 in total)

2002-2005   Sense and Sense-Ability: linguistic, computational, and neuro-psychological approaches to Mandarin Chinese senses, National Science Council Research Group Project (PI and coordinator of four projects, US$60,000 for 2004, est. US$150,000 in total)

2004   Seeking Excellence in Linguistics: Taiwan Summer Institute in Linguistics, National Science Council (PI, US $35,000)

2004-2005   Language E-Learning for Chinese: Center for Resources (co-PI, US $50,000)

2004-2007   Linguistic KnowledgeBase for Mandarin Chinese, National Languages Committee, Ministry of Education (PI, US$300,000 for three years, pending final approval)

Past Research Grants

2001   Language Archives-Digital Archives Pilot Project. Academia Sinica. NT$ 4,150,000.

2001   Summer Institute of Linguistics in Chinese Corpus Linguistics (in

collaboration with UC-Santa Barbara). Chiang Ching-kuo Foundation. NT$899,114.

2001-2002   Chinese-English Translation Equivalents Database. Global View Co. and Academia Sinica Collaborative Project. NT$1,896,000.

2000.08-2002.07 International Digital Library Project: Lexicon-based knowledge linking—Towards a wordnet infrastructure for multilingual digital library (in collaboration with NSF-Simmons College). National Science Council. NT$2,405,000.

1999.08-2002.07 Knowledge Representation: A corpus-based and lexicon-driven study on ambiguity. National Science Council. NT$2,507,000.

1999.1-2002.12   A Lexical KnowledgeBase for Mandarin Chinese. (One of the 5 projects in the Thematic Program of Academia Sinica Project Group 'Chinese Knowledge Infrastructure and Language Engineering'). NT$1,500,000/year (estimated, budget awarded in sum for the group.)

1998.12.-2001.10 Linguistic KnowledgeNet (I & II). National Science Council Digital Library Project. NT$3,820,000.
I.*SouWenJieZi*-A Lexico-Ideographic Knowledge Network for Sinology.
II. *Adventures in WenLand*-A linguistic KnowledgeNet for Children.

1997.08-1999.07 Lexical-Syntactic and Lexical Semantic Studies of Mandarin Chinese. National Science Council. NT$1,688,000.

1997-1998   Chinese Word Segmentation Standard and Standard Reference Lexicon for Chinese Information Processing. National Standardization Bureau. NT$900,000.

1996.08-1997.07 Corpus-based Lexical Semantic Study of Mandarin Chinese. National Science Council. NT$542,000.

1996   Readings in Chinese Natural Language Processing (In collaboration with City University of Hong Kong and POLA, UC-Berkeley). Chiang Ching-kuo Foundation. NT$600,000.

1995-1996　Preliminary Study of National Standard for Chinese Word Segmentation in Information Processing. National Standardization Bureau. NT$750,000.

1994-1996　Collection of Old Chinese Lexical Entries and Construction of an Electronic Lexicon (In collaboration with SOAS and Stanford University). Chiang Ching-kuo Foundation. NT$ 3,000,000.

1994.02.-1996.07 Balanced Corpus for Modern Mandarin Chinese – A first study in Chinese Corpus Linguistics. National Science Council. NT$1,668,000.

1992.08-1993.07 Analysis and Processing of Chinese Morphological Rules (Co-PI). National Science Council. NT$ 769,000.

1992.07.-2000.06 Knowledge-based Chinese Lexicon for Information Processing. (Co-PI). (One of the 5 projects in the Thematic Program of Academia Sinica Project Group: Collaboration on Chinese Information Processing). NT$1,200,000/year (estimated, budget awarded in sum for the group.)

1992.07.-2000.06 Mandarin Chinese Corpora. (One of the 5 projects in the Thematic Program of Academia Sinica Project Group: Collaboration on Chinese Information Processing.). NT$1,000,000/year (estimated, budget awarded in sum for the group.)

1991-1993　Linguistic Analysis of Computerized Classical and Modern Chinese Text-Corpora --A Pilot Study in Chinese Corpus Linguistic. Chiang Ching-kuo Foundation. NT$4,500,000.

1991.08-1992.07 Representation and Parsing Algorithm of Nominalized Verb Phrases (Co-PI). National Science Council. NT$732,000.

1990.08-1991.07 Chinese KnowlegeBase for Information Processing (Co-PI). National Science Council. NT$1,495,000.

1988-1992　A Syntactic Parsing System for Chinese Sentences. (Co-PI).

Academia Sinica and Institute of Research on Industrial Technology joint project.

1988-1991   A Syntactic Parsing Formalism for Chinese Sentences. (Co-PI). National Science Council, R.O.C.

1987-1989   A Knowledge-based Electronic Dictionary for Chinese. (Co-PI). Academia Sinica and Institute of Research on Industrial Technology joint project.

Web Resources Created

Over the past 10 years, world-wide web has become an essential source of research information. In addition, corpus-based research has also become a dominant paradigm in linguistics. I combined these two new directions and led efforts to provide web-based infrastructure of research resources for Chinese linguistics and language processing.

1. **Academia Sinica Balanced Corpus of Mandarin Chinese**
http://www.sinica.edu.tw/SinicaCorpus/

[The first web-accessible Chinese corpus in the world. On-line since November 1996. Over 150,000 hits since 2001 (v.3.0). Co-PI with Keh-Jiann Chen]

2. **Sinica BOW** (Academia Sinica Bilingual Ontological WordNet)
http://BOW.sinica.edu.tw

[The first wordnet to combine with ontology in the world. On-line since October 2003. About 30,000 hits since October 2003.]

3. **SouWenJieZi (Chinese Language KnowledgeNet)**

http://words.sinica.edu.tw/

[A Chinese knowledgenet combining linguistic, literary, and paleographic work. On-line since November 1999.   Over 1,500,000 combined hits for four site since 1999.]

4. **Adventures in WenLand**

http://www.sinica.edu.tw/wen/

[A Chinese knowledgenet for language learners. On-line since November 2002. Over 500,000 hits since 2002.]

5. Chinese WordNet (Prototype)

http://cwn.ling.sinica.edu.tw/

[The first Chinese Wordnet. On-line since January 2006. An ongoing project. Currently containing over 6,500 lemmas and nearly 20,000 senses.]

6. **Chinese Word Sketch**

[A powerful corpus search platform especially designed for automatic extraction of grammatical knowledge. Online since January 2007. An collaborative effort with Adam Kilgarriff of LexicoComputing Co., UK.]

PROGRAM DIRECTOR

2006   Co-Director, 2006 Taiwan Summer Institute in Linguistics:
Meaning in Context
The institute offered four short series of intensive courses I provided the vision outline, and recruited the 4 faculty members from leading universities all over the world. There were more than 200 students and young faculty members who attended this institute.

2005-   Director, Computational Linguistics and Chinese Language Processing, Doctoral Program, Academia Sinica
During the organizational period of over three years, I laid out the vision for CLCLP to be a leading program in Chinese Language Processing, negotiated for collaboration with several universities, and set up the organizational guidelines. I also recruited both regular and consulting faculty members from leading departments in computational linguistics.

2004   Director, 2004 Taiwan Summer Institute in Linguistics:
Corpus and Computational Linguistics
The institute offered two-weeks of intensive courses, in addition to 7 invited speeches and two workshops. I provided the vision, drafted the course outline, and recruited the 8 faculty members from leading universities all over the world. There were more than 200 students and young faculty members who attended this institute.

2001   Director, Sub-Institute on Chinese Corpus Linguistics,
LSA Summer Institute in Linguistics, UC-Santa Barbara
This was one of the sub-institutes under the Pacific Rim theme. I laid out the vision and secured grants from four different sources to support both the faculty and students. Over 70 students took the five 6 or 3 week courses offered.

PROGRAMS TAUGHT

Computational Linguistics and Chinese Language Processing, Taiwan International Graduate Program (In collaboration with National Taiwan University, Institute of Linguistics, and National Tsing Hua University, Institute of Information Systems and Applications)

National Chung Cheng University, Institute of Linguistics

National Taiwan Normal University, Institute of Translation

National Tsing Hua University, Institute of Linguistics

University of California at Santa Barbara, Department Linguistics (LSA Summer Institute of Linguistics)

COURSES TAUGHT

Chinese Lexical Semantics, Computational Linguistics, Corpus Linguistics, Generalized Phrase Structure Grammar, Lexical-functional Grammar, Machine Translation, Linguistic Theory and Computational Linguistics

POST-DOCTORAL FELLOWSHIP SUPERVISION

2005-2006

Shu-Kai Hsieh (Phd, University of Tuebigen, Assistant Professor National Taiwan Normal University), Ontology of Chinese Writing System.

Laurent Prevot (Phd, Toulouse Research Institute of Computer Science. Research Scientist, University of Toulouse) Ontology and NLP.

2003   Simon Smith (Phd, Birmingham; Asst. Professor Mingchuan University), Chinese Wordnet

1999   Oliver Streiter (PhD, University of Saarland; Asst. Professor, National Kaohsiung University) Chinese Parsing

1998   Zhao-Ming Gao (PhD, University of Manchester Institute of Science and Technology; Asst. Professor, National Taiwan University), Collocation and Treebank

1995-1996   Mei-Chih Tsai (PhD, University of Paris VII; Ass. Professor and Language Center Director, National Chungshan University), Chinese Lexical Semantics and Near Synonyms

1995   Charles C. Li (PhD, Stanford University; Prof. Christ's College, Taiwan), Chinese Language Processing

1993-1994   Mei-chun Liu (PhD, University of Colorado; Professor and Chair of Foreign Languages Department, National Chiao-tung University), Interaction between Semantics and Pragmatics

1991-1993   Cheng-hui Liu (PhD, Ohio State University; Professor and Director of Writing Center, National Tsing Hua University), Chinese historical grammar

PHD THESIS DIRECTED: 1
The student currently holds a tenure-track position. His thesis received a honorable mention, best thesis of the year award from Taiwan's Computer Science Association in 2006.
[Currently supervising 4 doctoral students from Czech, Hong Kong, and

Malaysia]

MA THESIS DIRECTED: 20

Four of them now hold tenure-track positions. One third of them received or are getting their doctoral degrees. All except two are holding academic or teaching positions.

RESEARCH ASSISTANTS DIRECTED: over 150

Eight of them now hold tenure-track positions. At least 12 of them are now pursuing their doctoral degrees.

INTERNS AND VISITING STUDENTS SUPERVISED: over 20

These students are from:

Europe: University of Bielefeld, Germany, University of Bonn, Germany, University of Manchester Institute of Science and Technology (UMIST), UK**,** University of Paris 7, France**,** University of Saarland, Germany**,** Trinity College Dublin, Ireland

US: University of Buffalo**,** University of Colorado, Georgetown University**,** University of Illinois-Urbana-Champaign**,** University of Pennsylvania**,** Stanford University

Others: University of Montreal, CANADA**,** University of Melbourne

PHD THESIS COMMITTEES SERVED

Disciplines: Linguistics, Computer Science, Information Management, Psychology, Building and Planning

Universities: City University of Hong Kong, National Chung Cheng University, National Taiwan University, National Tsing Hua University, University of Hawaii, University of Hong Kong, University of Pennsylvania

ACADEMIA SINICA

2004-	Director, Computational Linguistics and Chinese Language Processing International Phd Program, TIGP
2003-	Digital Archives Council Member, Academia Sinica
2001-	Organizing Committee, Taiwan International Graduate Programs
2004-2005	Vice-Director, Institute of Linguistics
2004-2005	Academic Committee, Institute of Linguistics
2004-2005	Space Planning Committee, Institute of Linguistics
2003-2004	Library Committee, Institute of Linguistics
1999-2002	Information and Computer Committee, Institute of Linguistics
1997-2002	Director, Corpus and Textual Database Research Group, Institute of Linguistics
1998-1999	New Hiring Committee, Institute of Linguistics
1998-1999	Convener, Space Planning Taskforce, Institute of Linguistics
1995-1999	Textual Database Coordination Committee, Academia Sinica
1997-1998	Publication Committee, Institute of Linguistics
1997-1998	Chair, Information and Computer Committee, Institute of Linguistics
1997-1998	Bylaw Drafting Taskforce, Institute of Linguistics
1995-1997	Director, Corpus and Textual Database Research Group, Institute of History and Philology
1993-1994	New Dormitory Construction Committee, Academia Sinica
1989-1991	Convener, Library Committee, Institute of History and Philology
1987-1989	Library Committee, Institute of History and Philology

CONFERENCES CO-FOUNDED

1988 **ROCLING:** ROC Computational Linguistics Conference, the first computational linguistics conference in Taiwan. The conference is held annually.

1992 **PACLIC**: Pacific Asia Conference on Language, Information, and Computation. This is a regional conference that pulls theoretical and computational linguists together. This was formed by merging two existing conference and held annually.

2000 **CLSW**: Chinese Lexical Semantics Workshops. This is the first conference dedicated to Chinese lexical semantics while embracing both computational and linguistics approaches.

2001   **Asian Language Resources Workshops.** The goal of this workshop is to raise awareness of linguistic diversity in Asia as well as to coordinate efforts to build infrastructures for language processing for Asian languages.

CONFERENCE ORGANIZATION: CHAIR/SECRETARY
2008
Co-Chair, Program Committee. Th The Sixth Asian Language Resources Workshop. In Conjunction with 3rd IJCNLP. Hydrabad, India..

Organizer, Workshop on Linguistic Studies of Ontology at 18th International Congress of Linguists. Seoul, Korea

2007

Conference Co-Chair. The Seventh Chinese Lexical Semantics Workshop. Hong Kong.

Track Chair, Program Committee Area: Ontologies and Semantics. The Second IJCNLP. Jeju, Korea.

Co-Chair, Program Committee. Ontolex 2007. Held In Conjunction with the International Conference on Semantic Web. Busan, Korea.

2006

Co-Chair, Program Committee. *OntoLex 2006: Interfacing Ontologies and Lexical Resources for Semantic Web Technologies* at LREC (International Conference on Language Resources and Evaluation).Genoa, Italy.

Co-Chair, Program Committee. *Joint COCOSDA-WRITE Workshop on Roadmap for Language Resources* at LREC (International Conference on Language Resources and Evaluation), Genoa, Italy.

Area Chair, Program Committee Area: Lexical Semantics and Ontologies. 2006 COLING/ACL Joint Conference. Sydney, Australia.

Conference Co-Chair, *20th Pacific Asia Conference on Language, Information and Computation. Wuhan, China*

2005

Co-Chair, Program Committee. The Sixth Chinese Lexical Semantics Workshop. Xiamen, China.

Track Chair, Program Committee Area: Ontologies and Semantics. The Second IJCNLP. Jeju, Korea.

Conference Chair, PACLIC19. Taipei, Taiwan.

Co-Chair, Program Committee. The Fifth Asian Language Resources Workshop. Jeju, Korea.

Co-Chair, Program Committee. The Fourth SigHan Workshop on Chinese Language Processing. Jeju, Korea.

Co-Chair, Program Committee. The Fourth OntoLex Workshop on Ontology and Lexical Resources. Jeju, Korea.

2004

Co-Chair, Program Committee. PACLIC18. Waseda University, Japan.

Co-Chair, Program Committee. The Fourth Asian Language Resources Workshop. Hainan, China.

Area Co-Chair, Program Committee Area: Lexical Semantics and Ontology. The First IJCNLP. Hainan, China.

2003

International Co-Chair, PACLIC17, Singapore.

Chair, Organizing Committee, Special Session on Upper Ontology and Natural Language Processing. The 2003 IEEE Conference on Natural Language Processing and Knowledge Engineering. Peking China. October 2003.

2002

Chair, Organizing Committee, The 19th International Conference on Computational Linguistics (COLING2002). Taipei. August 25-September 1

Area Chair, Program Committee Area on Asian Language Processing. COLING2002. Taipei. August 25-September 1

Co-Chair, PALCIC16, Jeju, Korea.

Chair, the Third Chinese Lexical Semantics Workshop, Taipei

2001

Co-Chair, the Second Chinese Lexical Semantics Workshop, Peking University. Beijing

Co-Chair, PACLIC15, Hong Kong

Co-Chair, Program Committee, The 2nd Language Resources in Asia Workshop, Tokyo, Japan. November

1988 - 2000

Co-Chair, PALCIC14, Tokyo, Japan, 2000

Co-Chair, PALCIC12, Singapore, 1998

Co-Chair, Program Committee. The International Conference on Research on
Computational Linguistics (ROCLING X), Taipei, 1997

Co-Chair, PALCIC11, Seoul, Korea, 1996

Chair, PACFOCOL, Taipei, 1995

Executive Secretary, IsCLL II, Taipei, 1991

Co-Chair, Program Committee, ROCLING II, Sun Moon Lake, Taiwan, 1989

Co-Chair, Program Committee, ROCLING I, Hsitou, Taiwan, 1988

CONFERENCE PROGRAM COMMITTEE MEMBER

ACL, CLSW, COLING, GWA, IACL, ICECE, IJCNLP, ISCLL, IWIC,
LREC, NLPKE, PACLIC, ROCLING, SIGHAN

INTERNATIONAL AND REGIONAL STANDARDS

OLAC: Open Language Archives Community

Set up standards and best practice recommendations involving all aspects of
language archives as one of OLAC's council member.

ISLE: International Standards for Language Engineering

Set up recommendation for standards for monolingual and multilingual
computational lexicons. Main responsibilities included coordinating
responses from Asian colleagues.

ISO TC37/SC4/WG04: Language Resources Management-Lexicon

Invited participation in the drafting and testing of the set of proposed ISO
standards involving language processing, especially those involving
monolingual and multilingual lexicons.

CNS14366:

Directed the drafting of Taiwan's standard of segmentation for Chinese
information processing. It was approved as CNS in 1998.

EDITORIAL BOARDS

2007-   Linguistic Issues in Language Technology

2007-   Journal of Chinese Information Processing (中文信息學報)

2005-   Cambridge Studies in Natural Language Processing

2005-   Language Resources and Evaluation (formally Computer and the
Humanities, SCI, AHCI)

2005-2007　Journal of Computational Linguistics (SCI, SSCI, AHCI)

2003-　Language and Linguistics (Associate Editor, 2006-2007)

2002-　Taiwan Journal of Linguistics

1996-　Computational Linguistics and Chinese Language Processing

1992-　Journal of Chinese Linguistics (Associate Editor, AHCI)

1994 -1998　Communications of the Chinese and Oriental Languages Information Processing Society

1996　International Review of Chinese Linguistics

RESEARCH GRANT REVIEWER

University Grant Council, Hong Kong: Since 1994 (Both in Humanities and Engineering)

University Research Grants, City University of Hong Kong

Chiang Ching-Kuo Foundation for International Scholarly Exchanges

National Science Council, Taiwan

Ministry of Education, Taiwan

TENURE/PROMOTION REVIEW [not including universities in Taiwan]

Chinese University of Hong Kong

City University of Hong Kong

Cornell University

Stanford University

UC-Santa Barbara

University of Hawaii

National University of Singapore

JOURNAL ARTICLE REVIEWER

Journal of Computational Linguistics

Journal of Chinese Linguistics

Journal of Chinese Language and Computing

Computational Linguistics and Chinese Language Processing

Computer Processing of Chinese and Oriental Languages

Concentric

Journal of East Asian Linguistics

Journal of Information Science and Engineering

Bulletin of the Institute of History and Philology 史語所集刊

Language and Linguistics

Language Resources and Evaluation

Lingua

Linguistics

Natural Language and Linguistic Theory

Journal of Natural Language Engineering

Journal of Building and Planning 城鄉學報

Sinological Studies 漢學研究

Humantista Taiwanica 文史哲

RECENT INVITED TALKS AND TITLES [2002- present]

2007

Invited speaker, 4th National Natural Language Processing (NLP) Research Symposium, to be held at the De La Salle University-Manila, Philippines, June 14-17, I. From Synergy to Knowledge: Integrating multiple language resources. II. Towards a shared upper ontology of basic lexicon.

Keynote Speaker at National Conference on Linguistics at Chengkung University, Tainan, June 1-2, 2007. From Linguistic Insights to Theoretical Relevance: A personal retrospective on Chinese grammar

Invited Paper at International Conference on Endangered Austronesian Language Documentation at Providence University, Taichung, June 5-7. Towards a Common Conceptual Framework of Language Documentation

Invited Panel talk at the CRPP Conference: "Redesigning Pedagogy: Culture, Knowledge and Understanding", Panel on Corpus Research, NIE-Singapore Nanyang Technological University, May 24-27, From Synergy to Knowledge: Corpus as a natural format for integrating multiple educational resources

2006

Invited Speaker. The 4th Annual Meeting of Society of Chinese Teachers in Taiwan. October 28, Kaohsiung, Taiwan. 大數與求真：如何以十億字語料庫進行語言分析與研究.

Invited Speaker, ROCLING2006. Hsinchu, Taiwan. September 7-8. Towards Global Wordnet Grid: Infrastructure for language technology in the age of multilingualism 全球詞網網格倡議：多語社會中的語言科技基礎建設.

Keynote Speaker, Shenyang, China. August 23. 2006 Student Workshop on Computational Linguistics (全國學生計算語言學會議 Biennial National Conference in China). 語言學理論與分析在計算語言學中的應用.

2005

Keynote Speaker, The Fifth Asian Language Resources Workshop. October 14, From Writing Systems to Ontology: A New Perspective for Asian Language Resources. Jeju, Korea.

Invited Tutorial. The 19th ROCLING Conference. September 15. Bilingual Ontological Wordnet. Tinan, Taiwan.

Invited Speaker, 2005 Joint Conference on Computational Linguistics. August 27-29, Nanjing. Ontology and Natural Language Processing.

Keynote Speaker, 2005 National Conference on Linguistics, July 3-4. Linguistics in the Age of Semantic Web: Challenges and Opportunities.

Hsinchu, Taipei

Plenary Talk. National Science Council Research Accomplishment Seminar in the Area of Linguistics. May 28-29. Lexicon-driven Studies on Semantics.

Plenary Talk. The Sixth Chinese Lexical Semantics Workshop. May 14-15. A Cognitive Account of the Lexical Semantics of Sense Verbs.

Plenary Talk. 漢字與全球化國際學術研討會 (International Conference on Chinese Characters and Globalization). January 28-30. Taipei. 漢字知識表達的幾個層面：字、義、與詞義關係 (Knowledge Representation with Hanzi: The relationship among characters, words, and senses).

2004

Invited Speaker, PACLIC18. Waseda University, Japan. December 8-10. Text-based Construction and Comparison of Domain Ontology: A Study Based on Classical Poetry.

The Linguistics Society of Providence University. October 13. From Distributional Patterns to Semantic Fields: A Corpus-based MARVS Approach.

Invited Panelist, Masters on Linguistics and Literature Studies. A Special Workshop in celebration of the 10th Anniversary of the Foreign Language and Literature Department, National Chiao-tung University. October 2.

Invited talk at the 2004 Taiwan Summer Institute in Linguistics. Academia Sinica. July 14. From Lexical KnowledgeBase to Domain KnowledgeBase: The Design and Implementation of Domain Lexico-Taxonomy.

Invited paper at the Workshop on the Possibilities of a Knowledgebase of Tang Civilization.The Institute for Research In Humanities of Kyoto University. February 20-21. Sinica BOW and 300 Tang Poems: An overview of a bilingual ontological wordnet and its application to a small ontology of Tang poetry.

2003

Invited Talk. The Third Language and Technology Conference. December 20. National Taiwan Normal University. The Language Web and Knowledge Infrastructure: Meaning-triggered Language Technology. [In Chinese.]

Invited Panel Talk. The Seventh World Chinese Language Conference. December 30. Knowledge Chinese-Looking ahead for the Information Processing for Chinese in the Age of Semantic Web.

Invited Panel talk: Synergy Between Language Resources and Knowledge Resources. The 2003 IEEE International Conference on Natural Language Processing and Knowledge Engineering (NLPKE2003), Special Session on Upper Ontology and Natural Language Processing. Beijing. Oct. 28. SINICA BOW: Integrating Bilingual WordNet and SUMO Ontology.

Institute of Computational Linguistics, Peking University. Oct. 21. Knowledge Representations: Ontology, WordNet, and Metaphor. [In Chinese].

Institute of Computational Linguistics, Peking University. October 14. Sense and Meaning Facets: Linguistic Units of Meaning in Chinese and Criteria for Their Identification [In Chinese].

Invited Speech at the Symposium on Library Resources for Medical Libraries. National Cheng Kung University. September 23. From Lexicon to Ontology: A brave new world for special domain knowledgebase in the time of the semantic web [In Chinese].

CTL/ICL/LEL, City University of Hong Kong. July 8. Issues in the Structure of Lexicons: A Multilingual Perspective

CTL/ICL/LISRC, City University of Hong Kong. June 30. Sense and Meaning Facet: Operational Criteria for sense distinction in Chinese.

CTL/ICL/LISRC, City University of Hong Kong. June 23. Word Knowledge, World Knowledge, and Ontology: Towards a Linguistic Infrastructure for Knowledge Representation and Knowledge Engineering.

Department of English. National Dong Hwa University. April 4. Multi-Resources Language Learning on the Web: From *Wen-Land* to *LiJingJi* and Beyond.

2002

Institute of Automation, Chinese Academy of Sciences. Nov. 8. Language Archives, Linguistic Anchoring, and OLAC.

Institute of Computational Linguistics, Peking University. Nov. 5. Computational Linguistics in Taiwan-a perspective from COLING2002 at Taipei.

Language Training and Testing Center. Taipei. September 25. In the Name of Semantics: Meaning, WordNet, and the Future of Webs.

Invited Keynote Speech. Symposium on Knowledge Management for Buddhism. September 13. Semantic Web, WordNet, and Ontology: On knowledge management on the future web. [In Chinese]

Invited Panelist. The First SigHAN Workshop. Panel on Chinese Language Processing 10 Years from Now. September 1. Chinese Language Processing in 10 Years.

Invited Panelist. COLING2002 Panel on Semantic Web and Natural Language Processing. August. 30. Semantic Web: A New Challenge for Language Technology.

School of Humanities. Hong Kong University of Science and Technology. March 6. Open Language Archives Community and the Digital archives/Language Archives.

CTL/LISRC, City University of Hong Kong. March 5. The Nature of Categorical Ambiguity and Its Implications for Language Processing: A Corpus-Based Study of Mandarin Chinese.

INVITED TALKS GIVEN ABROAD
-**Asia**:
Chinese Academy of Sciences (2007, 2002)
City University of Hong Kong (2003x3, 2002, 2000, 1998x2, 1993, 1990)
De LaSalle University, Manila (2007)
Hong Kong Polytechnic University (2007)
Hong Kong University of Science and Technology (2002)
Huazhong Normal University (2006)
Kyoto University (2004)
Peking University (2007, 2006, 2005, 2004, 2003, 2002, 2001)
Peking Normal University (2007)
Vietnam University of Science and Technology (2006)
Waseda University (2004)
Wuhan University (2006)

-**Europe:**
CRLAO, EHESS, France (1988, 1989, 1999)
ICL-CNR, Italy (2006)
Xerox European Research Centre (1999)
University of Bologna (2006)

-**USA:**
Cornell University (1990)

Microsoft Research Center (1998)
Ohio State University (1990, 1989)
Stanford University (1991, 1990)
UCSD (2003)
University of Pennsylvania (2001x2, 2000)

Awards

2005   Academia Sinica Investigatorship (中央研究院深耕研究獎助)
This is the most prestigious research award given by Academia Sinica with guaranteed five year support. I won this inaugural award as the only recipient in either humanities or social sciences.

2004   Elected, Permanent Member, International Committee on Computational Linguistics (One of the Highest Honors in Computational Linguistics. Currently there are only 19 ICCL permanent members. I am the first and only Chinese member, as well as the fifth (and only non-Japanese member) elected from Asia.)

1988-2006 NSC Research Award

2005   Honorable Mention (Doctoral Thesis), Best Thesis Award, ROC Computer Science Association in 2006. (Chou Ya-Min, 'An Ontology of Chinese Characters' Award Given jointly to the student and thesis director) 佳作（博士論文）中華民國資訊學會 2005 年碩博士最佳論文獎 （周亞民）

2002   Honorable Mention, Hai Hua Award (Chinese Language Teaching Software)九十一年海華獎優良華語文教學軟體　網路組佳作（文國尋寶記）

2001   Finalists, 10 Best Quality Website for Children [文國尋寶記]

2001   Finalists, 10 Best Quality Website for Children [搜文解字]

1998   Little Sun Award (小太陽獎), Best Children/Young Adult book of the year in the category of reference books [國語日報量詞典]

1994   NSC Research Excellence Award (優良研究獎)

1993   NSC Research Excellence Award (優良研究獎)

1992   NSC Research Excellence Award (優良研究獎)

1988   Best Paper, International Computer Symposium