2nd International Joint Conference on Natural Language Processing
IJCNLP-05, Jeju Island, Korea, October 11-13, 2005

Program Guide

Table Of Content

Version 10: September 20, 2005 (draft)

(1) Master Schedule

First Day: 11 October 2005 (Tuesday)

09:00 – 09:30	**Opening Ceremony** **Welcome Speech:** Benjamin T'sou, City University of Hong Kong, China (President, AFNLP) **Chair:** Key-Sun Choi, KAIST, Korea		
09:30 - 10:30	**Invited Speech:** Situated Natural Language Understanding by Tanaka Hozumi, Chukyo University, Japan (Honorary Chair, IJCLP05) **Chair:** Jong-Hyeok Lee, POSTECH, Korea		
10:30 - 11:00	Coffee Break		
	Information Retrieval **Chair:** I*ryna Gurevych* *EML Research, Germany*	**Corpus-based Parsing** **Chair:** *Oi Yee Kwong* *City University of Hong Kong, China*	**Web Mining** **Chair:** *Virach Sornlertlamvanich* *Thai Computational Linguistics Lab., Thailand*
11:00 - 12:40	A New Method for Sentiment Classification in Text Retrieval *Yi Hu, Jianyong Duan, Xiaomin Chen, Binzhen Pei and Ruzhan Lu*	Corpus-based Analysis of Japanese Relative Clause Constructions *Takeshi Abekawa and Manabu Okumura*	Entropy as Indicator of Context Boundaries - An experiment using Web Search Engine *Kumiko Tanaka-Ishii*
	Topic Tracking based on Linguistic Features *Fukumoto Fumiyo and Yamaji Yusuke*	Parsing Biomedical Literature *Matthew Lease and Eugene Charniak*	Automatic Discovery of Attribute Words from Web Documents *Kosuke Tokunaga, Jun'ichi Kazama and Kentaro Torisawa*

	The Use of Monolingual Context Vectors for Missing Translations in Cross-Language Information Retrieval *Yan Qu, Gregory Grefenstette and David A. Evans*	Parsing the Penn Chinese Treebank with Semantic Knowledge *Deyi Xiong, Shuanglong Li, Qun Liu, Shouxun Lin and Yueliang Qian*	Aligning Needles in a Haystack: Paraphrase Acquisition Across the Web *Marius Pasca and Peter Dienes*
	Automatic Image Annotation and Retrieval using Maximum Entropy Model *Wei Li and Maosong Sun*	Using a Partially Annotated Corpus to Build a Dependency Parser for Japanese *Manabu Sassano*	Confirmed Knowledge Acquisition Using Mails Posted to a Mailing List *Yasuhiko Watanabe, Ryo Nishimura and Yoshihiro Okada*
12:40 - 14:00	Lunch		
	Rule-Based Parsing **Chair:** *Marshall R. Mayberry* Saarland University, Germany	**Disambiguation** **Chair:** *Di Jiang* *CASS, China*	**Text Mining** **Chair:** *Zhu Zhang,* *University of Arizona, USA*
14:00 - 15:40	Automatic Partial Parsing Rule Acquisition Using Decision Tree Induction *Myung-Seok Choi, Chul Su Lim, and Key-Sun Choi*	PP-Attachment Disambiguation Boosted by a Gigantic Volume of Unambiguous Examples *Daisuke Kawahara and Sadao Kurohashi*	Acquiring Synonyms from Monolingual Comparable Texts *Mitsuo Shimohata and Eiichiro Sumita*
	Chunking using Conditional Random Fields in Korean Texts *Yong-Hun Lee, Mi-Young Kim and Jong-Hyeok Lee*	Adapting a probabilistic disambiguation model of an HPSG parser to a new domain *Tadayoshi Hara, Yusuke Miyao, and Jun'ichi Tsujii*	A method of recognizing entity and relation *Xinghua Fan and Maosong Sun*

	High Efficiency Realization for a Wide-Coverage Unification Grammar *John Carroll and Stephan Oepen*	A Hybrid Approach to Single and Multiple PP attachment using WordNet *Akshar Bharati, Rohini U., Vishnu P., S. M. Bendre and Rajeev Sangal*	Inversion Transduction Grammar Constraints for Mining Parallel Sentences from Quasi-Comparable Corpora *Dekai Wu and Pascale Fung*
	Linguistically-motivated Grammar Extraction, Generalization and Adaptation *Yu-Ming Hsieh, Duen-Chi Yang and Keh-Jiann Chen*	Period Disambiguation with Maxent Model *Chunyu Kit and Xiaoyue Liu*	Automatic Term Extraction based on Perplexity of Compound Words *Minoru Yoshida and Hiroshi Nakagawa*
15:40 - 16:00	Break		
	Document Analysis **Chair:** *Tetsuya Sakai* *Toshiba Corp. R&D Center, Japan*	**Ontology and Thesaurus** **Chair:** *Ruvan Weerasinghe* *University of Colombo School of Computing, Sri Lanka*	**Semantic Analysis – I** **Chair:** *Gary Lee* *POSTECH, Korea*
16:00 - 17:40	Document Clustering with Grouping and Chaining Algorithms *Yllias Chali and Soufiane Noureddine*	Analogy as Functional Recategorization Abstraction with HowNet Semantics *Tony Veale*	Using the Structure of a Conceptual Network in Computing Semantic Relatedness *Iryna Gurevych*
	Using Multiple Discriminant Analysis Approach for Linear Text Segmentation *Jingbo Zhu, Na Ye, Xinzhi Chang, Wenliang Chen, Benjamin Tsou*	PLSI Utilization for Automatic Thesaurus Construction *Masato Hagiwara, Yasuhiro Ogawa and Katsuhiko Toyama*	Semantic Role Labelling of Prepositional Phrases *Patrick Ye and Timothy Baldwin*

<table>
<tr><td></td><td>Classifying Chinese Texts in Two Steps
Xinghua Fan, Maosong Sun, Key-Sun Choi and Qin Zhang</td><td>Analysis of an Iterative Algorithm for Term-Based Ontology Alignment
Shisanu Tongchim, Canasai Kruengkrai, Virach Sornlertlamvanich, Prapass Srichaivattana and Hitoshi Isahara</td><td>Global Path-based Refinement of Noisy Graphs Applied to Verb Semantics
Timothy Chklovski and Patrick Pantel</td></tr>
<tr><td></td><td>Assigning Polarity Scores to Reviews using Machine Learning Techniques
Daisuke Okanohara and Jun'ichi Tsujii</td><td>Finding Taxonomical Relation from an MRD for Thesaurus Extension
SeonHwa Choi and HyukRo Park</td><td>Exploiting Lexical Conceptual Structure for Paraphrase Generation
Atsushi Fujita, Kentaro Inui and Yuji Matsumoto</td></tr>
</table>

Second Day: 12 October 2005 (Wednesday)

09:00 - 09:50	**Keynote Speech I:** Software and NLP R&D Strategy in Korea by Seyoung Park, Kyungbuk National University Korea **Chair:** Jun'Ichi Tsujii, University of Tokyo, Japan		
09:50 - 09:55	Break		
	Text Classification **Chair:** *Ozlem Uzuner* MIT, USA	**Transliteration** **Chair:** *Daisuke Kawahara* *University of Tokyo, Japan*	**Machine Translation - I** **Chair:** *Jong C. Park* *KAIST, Korea*
09:55 - 10:45	A Preliminary Work on Classifying Time Granularities of Temporal Questions *Wei Li, Wenjie Li and Qin Lu*	A Rule Based Syllabification Algorithm for Sinhala *Ruvan Weerasinghe, Asanka Wasala and Kumudu Gamage*	Improving Statistical Word Alignment with Ensemble Methods *Hua Wu and,Haifeng Wang*
	Classification of Multiple-Sentence Questions *Akihiro Tamura, Hiroya Takamura and Manabu Okumura*	An Ensemble of Grapheme and Phoneme for Machine Transliteration *Jong-Hoon Oh and Key-Sun Choi*	Empirical Study of Utilizing Morph-Syntactic Information in SMT *Young-Sook Hwang, Taro Watanabe and Yutaka Sasaki*
10:45 - 11:00	Break		
	Question & Answering **Chair:** *Yasuhiko Watanabe* *Ryukoku University, Japan*	**Morphological Analysis** **Chair:** *Maosung Sun,* *Tsinghua University*	**Machine Translation - II** **Chair:** *Stephan Oepen* *Universitetet i Oslo & CSLI Stanford*
11:00 - 12:40	Instance-Based Generation for Interactive Restricted Domain Question Answering Systems *Matthias Denecke and Hajime Tsukada*	A Chunking Strategy towards Unknown Word Detection in Chinese Word Segmentation *GuoDong Zhou*	Phrase-based Statistical Machine Translation: A Level of Detail Approach *Hendra* Setiawan, *Haizhou Li, Min Zhang and Beng Chin Ooi*

	Answering Definition Questions Using Web Knowledge Bases *Zhushuo Zhang, Yaqian Zhou, Xuangjing Huang and Lide Wu*	A Knowledge-constrained Character Model for Chinese Morphology *Yao Meng, Hao Yu and Fumihito Nishino*	Why is Zero Marking Important in Korean Treebanks? *Sun-Hee Lee, Donna Byron, and Seok Bae Jang*
	Exploring Syntactic Relation Patterns for Question Answering *Dan Shen, Greet-Jan M. Kruijff and Dietrich Klakow*	Relative compositionality of multi-word expressions a study of verb-noun (V-N) collocations *Sriram Venkatapathy and Aravind K. Joshi*	A Phrase-based Context-dependent Joint Probability Model for Named Entity Translation *Min Zhang, Haizhou Li, Jian Su and Hendra Setiwan*
	Web-Based Unsupervised Learning for Query Formulation in Question Answering *Yi-Chia Wang, Jian-Cheng Wu, Tyne Liang and Jason S. Chang*	Automatic extraction of fixed multiword expressions *Campbell Hore, Masayuki Asahara and Yuji Matsumoto*	Machine Translation Based on Constraint-Based Synchronous Grammar *Fai Wong, Dong-Cheng Hu, Yu-Hang Mao, Ming-Chui Dong and Yi-Ping Li*
12:40 - 14:00	Lunch		
	Poster Session – I **Chair:** *Rejeev Sangal,* *IIT India*		**Demo Session – I** **Chair:** *Kam-Fai Wong* *Chinese University of Hong Kong, China*
14:00 - 15:40	To be provided by poster chair		To be provided by poster chair
15:40 - 16:00	Break		
	Poster Session – II **Chair:** *Kentaro Inui* *Nara Inst. of Science and Technology, Japan*		**Demo Session – II** **Chair:** *Kam-Fai Wong* *Chinese University of Hong Kong, China*
16:00 - 17:40	To be provided by poster chair		To be provided by poster chair

Third Day: 13 October 2005 (Thursday)

<table>
<tr><td>09:00 - 09:50</td><td colspan="3">Keynote Speech II: Progress in NLIP. What does the summarising task tell us?
by Karen Spärck Jones, University of Cambridge, UK
Chair: Kam-Fai Wong, The Chinese University of Hong Kong, China</td></tr>
<tr><td>09:50 - 09:55</td><td colspan="3">Break</td></tr>
<tr><td></td><td>Text Summarization
Chair: Chun-Yu Kit
City University of Hong Kong, China</td><td>Panel:
“PAN Localization Projects: Status and Challenges Faced Across Developing Asia”</td><td>Named Entity Recognition
Chair: Kentaro Torisawa
Japan Advanced Inst. of Sc & Tech, Japan</td></tr>
<tr><td rowspan="2">9:55 - 10:45</td><td>A Machine Learning Approach to Sentence Ordering for Multidocument Summarization and its Evaluation
Danushka Bollegala, Naoaki Okazaki and Mitsuru Ishizuka</td><td rowspan="2">Chair:
Key-Sun Choi, KAIST, Korea

Panelists:
- Sarmad Hussain, FAST National University of Computer and Emerging Sciences, Pakistan
- Ruvan Weerasinghe, Univ. of Colombo School of Computing, Sri Lanka
- Mumit Khan, BRAC University, Bangladesh</td><td>Two-phase Biomedical Named Entity Recognition Using A Hybrid Method
Seonho Kim, Juntae Yoon, Kyung-Mi Park, and Hae-Chang Rim</td></tr>
<tr><td>Significant Sentence Extraction by Euclidean Distance based on Singular Value Decomposition
Changbeom Lee, Hyukro Park and Cheolyoung Ock</td><td>Heuristic Methods for Reducing Errors of Geographic Named Entities Learned by Bootstrapping
Seungwoo Lee and Gary Geunbae Lee</td></tr>
<tr><td>10:45 - 11:00</td><td colspan="3">Break</td></tr>
</table>

	Linguistic Resources and Tools **Chair:** *Keh-Jian. Chen* *Academic Sinica, Taiwan*	**Discourse Analysis** **Chair:** *Wenjie Li* Polytechnic University, Hong Kong, China	**Relation Extraction** **Chair:** *Patrick Pantel* *University of Southern California, USA*
11:00 - 12:40	Building a Japanese-Chinese Dictionary Using Kanji/Hanzi Conversion *Chooi-Ling Goh, Masayuki Asahara and Yuji Matsumoto*	A Twin-Candidate Model of Coreference Resolution with Non-Anaphor Identification Capability *Xiaofeng Yang, Jian Su and Chew Lim Tan*	Relation Extraction Using Support Vector Machine *Gumwon Hong*
	Automatic Acquisition of Basic Katakana Lexicon from a Given Corpus *Toshiaki Nakazawa, Daisuke Kawahara and Sadao Kurohash*i	Improving Korean Speech Acts Analysis by Using Shrinkage and Discourse Stack *Kyungsun Kim, Youngjoong Ko, and Jungyun Seo*	Discovering Relations from a Large Raw Corpus Using Tree Similarity-based Clustering *Min Zhang, Jian Su, Danmei Wang and Guodong Zhou*
	CTEMP A Chinese Temporal Parser for Extracting and Normalizing Temporal Information *Mingli Wu, Wenjie Li, Qin Lu and Baoli Li*	Anaphora Resolution for Biomedical Literature by Exploiting Multiple Resources *Tyne Liang and Yu-Hsiang Lin*	Automatic Relation Extraction with Model Order Selection and Discriminative Label Identification *Jinxiu Chen, Donghong Ji, Chew Lim Tan and Zhengyu Niu*
	French-English terminology extraction from comparable corpora *Beatrice Daille and Emmanuel Morin*	Automatic Slide Generation Based on Discourse Structure Analysis *Tomohide Shibata and Sadao Kurohashi*	Mining Inter-entity Semantic Relations using Improved Transductive Learning *Zhu Zhang*
12:40 - 14:00	Lunch		
	NLP Applications **Chair:** *Elizabeth Kandall,* *Monash University, Australia*	**Tagging** **Chair:** *Jingbo Zhu,* *Northeastern University, China*	**Semantic Analysis – II** **Chair:** *Jian Su* *Institute of Infocomm Research, Singapore*

14:00 - 15:40	Detecting article errors based on the mass count distinction *Ryo Nagata, Takahiro Wakana, Fumito Masui, Atsuo Kawa, and Naoki Isu*	Automatically Inducing a Part-of-Speech Tagger by Projecting from Multiple Source Languages Across Aligned Corpora *Victoria Fossum and Steven Abney*	Semantic Role Tagging for Chinese at the Lexical Level *Oi Yee Kwong and Benjamin K. Tsou*
	Principles of Non-stationary Hidden Markov Model and its Applications on Sequence Labeling Task *JingHui Xiao, BingQuan Liu and XiaoLong Wang*	The Verbal Entries and their Description in a Grammatical Information-Dictionary of Contemporary Tibetan *Di Jiang, Congjun Long and Jichuan Zhang*	Word Sense Disambiguation by Relative Selection *Hee-Cheol Seo, Hae-Chang Rim and Myung-Gil Jang*
	Integrating Punctuation Rules and Naive Bayesian Model for Chinese Creation Title Recognition *Conrad Chen and Hsin-Hsi Chen*	Tense Tagging for Verbs in Cross-Lingual Context a Case Study *Yang Ye and Zhu Zhang*	Towards Robust High Performance Word Sense Disambiguation of English Verbs Using Rich Linguistic Features *Jinying Chen and Martha Palmer*
	A Connectionist Model of Anticipation in Visual Worlds *Marshall R. Mayberry, III Matthew W. Crocker Pia Knoeferle*	Regularization Techniques for Conditional Random Fields Parameterized versus Parameter-free *Andrew Smith and Miles Osborne*	Automatic Interpretation of Noun Compounds using WordNet Similarity *Su Nam Kim and Timothy Baldwin*
15:40 - 16:00	Break		
	Language Model **Chair:** *Hyeok-Cheol Kwon* *Busan National University, Korea*	**Spoken Language** **Chair:** Changbeom Lee *University o Ulsam, Korea*	**Terminology Mining** **Chair:** *Beatrice Daille* *Universite de Nantes, France*
16:00 - 16:50	An Empirical Study on Language Model Adaptation Using a Metric of Domain Similarity *Wei Yuan, Jianfeng Gao and Hisami Suzuki*	Lexical Choice via Topic Adaptation for Paraphrasing Written Language to Spoken Language *Nobuhiro Kaji and Sadao Kurohashi*	Web-based Terminology Translation Mining *Gaolin Fang, Hao Yu and Fumihito Nishino*

	A Comparative Study of Language Models for Book and Author Recognition *Ozlem Uzuner and Boris Katz*	A Case-Based Reasoning Approach for Speech Corpus Generation *Yandong Fan and Elizabeth Kendall*	Extracting Terminologically Relevant Collocations in the Translation of Chinese Monograph *Byeong-Kwu Kang and Bao-Bao Chang*
16:50 – 16:55	*Break*		
16:55 – 17:45	**Keynote Speech III:** Exploiting Models of Semantic Overlap for Applications by Bill Dolan, Microsoft Research, USA **Chair:** Benjamin Tsou, City University of Hong Kong, China		
17:45 - 18:15	**Closing Ceremony** **IJCNLP05 Report:** Kam-Fai Wong, The Chinese University of Hong Kong, China **Best Paper Award:** Key-Sun Choi, KAIST, Korea (IJCNLP General Chair) **Closing Speech:** Jun'Ichi Tsujii, University of Tokyo, Japan (Vice President, AFNLP)		

(2) Invited and Keynote Speeches

(A) Invited Speech by Honorary Chair	
Title	Situated Natural Language Understanding
Speaker	Tanaka Hozumi, Chukyo University, Japan
Time	October 11, 2005, 0930h-1030h (Day 1)
Abstract	To make our robot more intelligent, it has to understand a conversation through natural language. The mile stone of such an NLU system was SHRDLU system developed at MIT more than 30 years ago. NLU research environment has changed drastically in the past two decade. Not only tremendous amount of computing power but also better technologies in speech recognition, natural language processing and computer graphics are now available. The advancement in such technologies enables to build a next-generation NLU system, where situated natural language understanding will be the most important research theme. In this talk we will explain why the situated natural language understanding is so important and difficult. Along with showing some examples, we will give you a few of solutions obtained thus far through our NLU research.
Bio	Hozumi Tanaka was born in Japan on October 02, 1941. He graduated from Department of control engineering, Faculty of Engineering, Tokyo Institute of Technology in 1964, and completed his Master's degree in 1966. He joined Electro-technical Laboratory of Ministry of Trade and Industry (MITI) in 1966, where he engaged in developing a TSS and natural language processing systems, the latter of which has been his current main research theme. He became a chief of machine inference section and received his Dr. of Eng. from Tokyo Institute of Technology in 1980. He joined the Fifth Generation Computer Project supported by MITI from 1983 to 1992. He moved from Electro-technical Laboratory to Tokyo Institute of Technology in 1983 and retired from the Institute in 2005. Since April 2001, he has conducted a 5-year project named "Natural Language Understanding and Action Control" supported by the Ministry of Education, Culture, Sports, Science and Technology and Japan Society for the Promotion of Sciences. At present, he is a professor of Department of Information Science at Chukyo University.

(B) Keynote Speech I	
Title	Software and NLP R&D Strategy in Korea
Speaker	Seyoung Park, Kyungbuk National University Korea
Time	October 12, 2005, 0900h-0955h (Day 2)
Abstract	I will begin the talk with an introduction to the R&D strategy for the Information and Communication Industry in Korea. This strategy has turned out to provide IT services successfully to competitors, commercialize them and lead the IT service industry market. The Ministry of Information and Communication initiated the IT839 Strategy that envisions enhancing competitiveness of the Information and Communication Industry through eight services, three infrastructures and nine new growth engines. I will then introduce the SW R&D policy in the context of the IT839 Strategy. Many important SW R&D issues such as open source SW, embedded SW including NLP will be dealt with. Especially, I will present NLP and semantic web technologies as a new SW infrastructure that enables the interoperability among ubiquitous IT services.
Bio	GENERAL INFORMATION Name: Se Young Park Birthday: March 21, 1957 Address: 357-5 BoRyung Villa, JangDae-Dong, YuSeong, DaeJeon, Korea EDUCATION Feb. 1980: Kyungpook National University, Department of Electronics, B.S Feb. 1982: KAIST, Department of Computer Science, M.S Feb. 1989: Paris 7 University, France, Ph.D EXPERIENCES Feb. 1982 – Apr. 2000: Director of Natural Language Department, Electronics & Telecommunications Research Institute (ETRI) Apr. 2000 – Aug. 2003: President, Searchcast Inc. Aug. 2003 – Aug. 2005: Project Manager of Digital Contents & S/W Solution, Institute of Information Technology Assessment (IITA) Project Manager of Digital Contents & S/W Solution, IT Policy Advisory Group, Ministry of Information and Communication, Republic of KOREA Apr. 2005 – Current: Professor, Department of Computer Engineering, Kyungpook National University

(C) Keynote Speech II	
Title	Progress in NLIP. What does the summarising task tell us?
Speaker	Karen Spärck Jones, University of Cambridge, UK
Time	October 13, 2005, 0900h-0955h (Day 3)
Abstract	Natural language information processing (NLIP) has made significant progress, in important ways, in the last twenty years. We have developed fairly comprehensive and robust tools like grammars and parsers, and have gained experience with applications including multilingual ones. We have been able not only to take advantage of the general advance in computing and communications technology but, more significantly, to exploit by-now vast text corpora to adapt our tools to actual patterns of language use. We have learnt, in particular, that many NLIP tasks can be sufficiently well done to be useful in many practical contexts by exploiting shallow text processing, ie by relying on surface indications of discourse meaning and communicative intent. We have also been learning how to do NLIP system evaluation. Summarising illustrates what we have learnt, where we are, and where we need to go, very well. The first experiments in automatic summarizing used very simple technology, a simple statistical sentence extraction technology that seemed too simple for useful summaries. Subsequent research focused on deeper text analysis that could sometimes work better could not readily be scaled up to large heterogenous data sources or to some user needs. More recent work on summarising has largely returned to the simpler, extractive approach, though it has also sought to refine or enrich this by, for example, incorporating parsing or by exploiting machine learning. Summarising has also been better contextualised, partly by being seen as encompassing a spectrum of types ranging from basic index descriptions for individual documents to multi-source syntheses of specific types of information, for example biographies. At the same time, summarising is increasingly, and rightly, seen as a task that is only one activity within a set that may all be useful for some larger purpose so that, for example, summarising may be related to search queries or to the need to encapsulate extended information-seeking interactions. But all of this richer view of summarising presents significant challenges for system evaluation. NLIP research has been transformed since 1990 by the major task evaluation programs that have been running, notably for information extraction and document retrieval and, later, question answering, that have served to establish whether plausible ideas actually work and to disseminate effective techniques. Summarising itself has been the focus of its own evaluation programmes for five years. This evaluation work, and the summarising evaluation work in particular, has been important both in promoting a better understanding of NLIP tasks and the impact of their

	application conditions. The summarising evaluations have, in particular, served to demonstrate both how crucial application contexts are for how tasks are handled, and how extremely challenging evaluation in itself is.
Bio	Karen Sparck Jones is emeritus Professor of Computers and Information at the Computer Laboratory, University of Cambridge. She has worked in automatic language and information processing research since the late fifties, and has many publications including nine books. She is a Fellow of the British Academy and of the American Association for Artificial Intelligence. She has received three awards for information retrieval research as well as, in 2004, the Association for Computational Linguistics' Lifetime Achievement Award. Her more recent research has been on information retrieval models and practice, on automatic summarising, and on system evaluation, where she is involved in international programmes.

(D) Keynote Speech III	
Title	Exploiting Models of Semantic Overlap for Applications
Speaker	Bill Dolan, Microsoft Research, USA
Time	October 13, 2005, 1655h-1745h (Day 3)
Abstract	The last few years have seen increased interest in measuring the semantic overlap between text segments. Work on paraphrase recognition, for instance, attempts to identify when two sentences "mean the same thing" at some abstract level, despite superficial differences: e.g. *"On its way to an extended mission at Saturn, the Cassini probe on Friday makes its closest rendezvous with Saturn's dark moon Phoebe."* *"The Cassini spacecraft, which is en route to Saturn, is about to make a close pass of the ringed planet's mysterious moon Phoebe."* Work on inference or "semantic entailment", meanwhile, aims to identify when the meaning of one sentence can be wholly inferred from the meaning of another: e.g. *"Mary Mallon unknowingly caused several typhoid outbreaks, leading to many illnesses and deaths."* *"Mary Mallon was a carrier of typhoid."* Modeling semantic overlap poses major challenges, as it encompasses complex issues of lexical choice, syntactic alternation, and reference/discourse structure. The assumption driving work in this area is that reliable metrics for identifying when two pieces of text overlap in specific ways will play a crucial role in building applications that appear to "understand" natural language. Problems as diverse as question answering, multi-document summarization, proofing tools, and translation could all benefit from advances in this area. This talk will focus on (1) the role of semantic overlap metrics in real applications and (2) defining shared tasks/datasets that will promote application-oriented advances in modeling these phenomena.
Bio	Bill Dolan is a Senior Researcher and the manager of the Natural Language Processing Group at Microsoft Research in Redmond, Washington. He has worked on many aspects of semantic processing, including word sense disambiguation and MindNet, a large-scale lexical knowledge base built automatically from free text. His current interests include paraphrase recognition/generation and machine translation.

(3) Session Chair Information

Session Chair Index *(sorted by last name)*		
Chair	**Session**	**Time**
Keh-Jiann Chen	Linguistic Resources and Tools	October13, 2005, 1100h-1240h (Day 3)
Key-Sun Choi	Opening Ceremony	October 11, 2005, 0900h-0930h (Day 1)
Key-Sun Choi	Panel	October 13, 2005, 0955h-1045h (Day 3)
Key-Sun Choi	Closing Ceremony	October 13, 2005, 1745h-1815h (Day 3)
Beatrice Daille	Terminology Mining	October 13, 2005, 1600h-1650h (Day 3)
Iryna Gurevych	Information Retrieval	October 11, 2005, 1100h-1240h (Day 1)
Sarmad Hussain	Panel	October 13, 2005, 0955h-1045h (Day 3)
Kentaro Inui	Poster Session – II	October 12, 2005, 1600h-1740h (Day 2)
Di Jiang	Disambiguation	October 11, 2005, 1400h-1540h (Day 1)
Elizabeth Kandall	NLP Applications	October 13, 2005, 1400h-1540h (Day 3)
Daisuke Kawahara	Transliteration	October 12, 2005, 0955h-1045h (Day 2)
Mumit Khan	Panel	October 13, 2005, 0955h-1045h (Day 3)
Chun-Yu Kit	Text Summarization	October 13, 2005, 0955h-1045h (Day 3)
Hyeok-Cheol Kwon	Language Model	October 13, 2005, 1600h-1650h (Day 3)
Oi Yee Kwong	Corpus-based Parsing	October 11, 2005, 1100h-1240h (Day 1)
Gary Lee	Semantic Analysis – I	October 11, 2005, 1600h-1740h (Day 1)
Changbeom Lee	Spoken Language	October 13, 2005, 1600h-1650h (Day 3)
Jong-Hyeok Lee	Invited Speech by Honorary Chair	October 11, 2005, 0930h-1030h (Day 1)
Wenjie Li	Discourse Analysis	October13, 2005, 1100h-1240h (Day 3)
Marshall R. Mayberry	Rule-based Parsing	October 11, 2005, 1400h-1540h (Day 1)
Stephan Oepen	Machine Translation – II	October 12, 2005, 1100h-1240h (Day 2)
Patrick Pantel	Relation Extraction	October 11, 2005, 1600h-1740h (Day 1)

Jong C. Park	Machine Translation – I	October 12, 2005, 0955h-1045h (Day 2)
Rejeev Sangal	Poster Session – I	October 12, 2005, 1400h-1540h (Day 2)
Tetsuya Sakai	Document Analysis	October 11, 2005, 1600h-1740h (Day 1)
Jian Su	Semantic Analysis – II	October 13, 2005, 1400h-1540h (Day 3)
Maosong Sun	Morphological Analysis	October 12, 2005, 1100h-1240h (Day 2)
Kentaro Torisawa	Named Entity Recognition	October 13, 2005, 0955h-1045h (Day 3)
Benjamin Tsou	Welcome Speech, Opening Ceremony	October 11, 2005, 0900h-0930h (Day 1)
Benjamin Tsou	Keynote Speech III	October 13, 2005, 1655h-1745h (Day 3)
Jun'Ichi Tsujii	Keynote Speech I	October 12, 2005, 0900h-0950h (Day 2)
Jun'Ichi Tsujii	Closing Ceremony	October 13, 2005, 1745h-1815h (Day 3)
Ozlem Uzuner	Text Classification	October 12, 2005, 0955h-1045h (Day 2)
Virach Sornlertlamvanich	Web Mining	October 11, 2005, 1100h-1240h (Day 1)
Yasuhiko Watanabe	Question & Answering	October 12, 2005, 1100h-1240h (Day 2)
Ruvan Weerasinghe	Panel	October 13, 2005, 0955h-1045h (Day 3)
Ruvan Weerasinghe	Ontology and Thesaurus	October 13, 2005, 1100h-1240h (Day 3)
Kam-Fai Wong	Demo Session – I	October 12, 2005, 1400h-1540h (Day 2)
Kam-Fai Wong	Demo Session – II	October 12, 2005, 1600h-1740h (Day 2)
Kam-Fai Wong	Keynote Speech II	October 13, 2005, 0900h-0950h (Day 3)
Kam-Fai Wong	Closing Ceremony	October 13, 2005, 1745h-1815h (Day 3)
Zhu Zhang,	Text Mining	October 11, 2005, 1400h-1540h (Day 1)
Jingbo Zhu	Tagging	October 13, 2005, 1400h-1540h (Day 3)

(4) Author Information

Regular Paper Author Index *(sorted by the last name of the first author)*		
First Author	**Paper**	**Session**
Takeshi Abekawa	Corpus-based Analysis of Japanese Relative Clause Constructions *Takeshi Abekawa and Manabu Okumura*	Corpus-based Parsing (day 1, am)
Akshar Bharati	A Hybrid Approach to Single and Multiple PP attachment using WordNet *Akshar Bharati, Rohini U., Vishnu P., S. M. Bendre and Rajeev Sangal*	Disambiguation (day 1, early pm)
Danushka Bollegala	A Machine Learning Approach to Sentence Ordering for Multidocument Summarization and its Evaluation *Danushka Bollegala, Naoaki Okazaki and Mitsuru Ishizuka*	Text Summarization (day 3, early am)
John Carroll	High Efficiency Realization for a Wide-Coverage Unification Grammar *John Carroll and Stephan Oepen*	Rule-based Parsing (day 1, early pm)
Yllias Chali	Document Clustering with Grouping and Chaining Algorithms *Yllias Chali and Soufiane Noureddine*	Document Analysis (day 1, late pm)
Conrad Chen	Integrating Punctuation Rules and Naive Bayesian Model for Chinese Creation Title Recognition *Conrad Chen and Hsin-Hsi Chen*	NLP Applications (day 3, early pm)
Jinxiu Chen	Automatic Relation Extraction with Model Order Selection and Discriminative Label Identification *Jinxiu Chen, Donghong Ji, Chew Lim Tan and Zhengyu Niu*	Relation Extraction (day 3, late am)
Jinying Chen	Towards Robust High Performance Word Sense Disambiguation of English Verbs Using Rich Linguistic Features *Jinying Chen and Martha Palmer*	Semantic Analysis – II (day 3, early pm)
Timothy Chklovski	Global Path-based Refinement of Noisy Graphs Applied to Verb Semantics *Timothy Chklovski and Patrick Pantel*	Semantic Analysis – I (day 1, late pm)
Myung-Seok Choi	Automatic Partial Parsing Rule Acquisition Using Decision Tree Induction *Myung-Seok Choi, Chul Su Lim, and Key-Sun Choi*	Rule-based Parsing (day 3, late am)

SeonHwa Choi	Finding Taxonomical Relation from an MRD for Thesaurus Extension *SeonHwa Choi and HyukRo Park*	Ontology and Thesaurus (day 1, late pm)
Beatrice Daille	French-English terminology extraction from comparable corpora *Beatrice Daille and Emmanuel Morin*	Linguistic Resources and Tools (day 3, late am)
Matthias Denecke	Instance-Based Generation for Interactive Restricted Domain Question Answering Systems *Matthias Denecke and Hajime Tsukada*	Question & Answering (day 2, late am)
Xinghua Fan	A method of recognizing entity and relation *Xinghua Fan and Maosong Sun*	Text Mining (day 1, early pm)
Xinghua Fan	Classifying Chinese Texts in Two Steps *Xinghua Fan, Maosong Sun, Key-Sun Choi and Qin Zhang*	Document Analysis (day 1, late pm)
Yandong Fan	A Case-Based Reasoning Approach for Speech Corpus Generation *Yandong Fan and Elizabeth Kendall*	Spoken Language (day 3, late pm)
Gaolin Fang	Web-based Terminology Translation Mining *Gaolin Fang, Hao Yu and Fumihito Nishino*	Terminology Mining (day 3, late pm)
Victoria Fossum	Automatically Inducing a Part-of-Speech Tagger by Projecting from Multiple Source Languages Across Aligned Corpora *Victoria Fossum and Steven Abney*	Tagging (day 3, early pm)
Atsushi Fujita	Exploiting Lexical Conceptual Structure for Paraphrase Generation *Atsushi Fujita, Kentaro Inui and Yuji Matsumoto*	Semantic Analysis – II (day 3, early pm)
Fukumoto Fumiyo	Topic Tracking based on Linguistic Features *Fukumoto Fumiyo and Yamaji Yusuke*	Information retrieval (day 1, am)
Chooi-Ling Goh	Building a Japanese-Chinese Dictionary Using Kanji/Hanzi Conversion *Chooi-Ling Goh, Masayuki Asahara and Yuji Matsumoto*	Linguistic Resources and Tools (day 3, late am)
Iryna Gurevych	Using the Structure of a Conceptual Network in Computing Semantic Relatedness *Iryna Gurevych*	Semantic Analysis – I (day 1, late pm)
Masato Hagiwara	PLSI Utilization for Automatic Thesaurus Construction *Masato Hagiwara, Yasuhiro Ogawa and Katsuhiko Toyama*	Ontology and Thesaurus (day 3, late am)

Tadayoshi Hara	Adapting a probabilistic disambiguation model of an HPSG parser to a new domain *Tadayoshi Hara, Yusuke Miyao, and Jun'ichi Tsujii*	Disambiguation (day 1, early pm)
Gumwon Hong	Relation Extraction Using Support Vector Machine *Gumwon Hong*	Relation Extraction (day 3, late am)
Campbell Hore	Automatic extraction of fixed multiword expressions *Campbell Hore, Masayuki Asahara and Yuji Matsumoto*	Morphological Analysis (day 2, late am)
Yu-Ming Hsieh	Linguistically-motivated Grammar Extraction, Generalization and Adaptation *Yu-Ming Hsieh, Duen-Chi Yang and Keh-Jiann Chen*	Rule-based Parsing (day 1, early pm)
Yi Hu	A New Method to Sentiment Classification in Text Retrieval *Yi Hu, Jianyong Duan, Xiaomin Chen, Binzhen Pei and Ruzhan Lu*	Information retrieval (day 1, am)
Young-Sook Hwang	Empirical Study of Utilizing Morph-Syntactic Information in SMT *Young-Sook Hwang, Taro Watanabe and Yutaka Sasaki*	Machine Translation – I (day 2, early am)
Di Jiang	The Verbal Entries and their Description in a Grammatical Information-Dictionary of Contemporary Tibetan *Di Jiang, Congjun Long and Jichuan Zhang*	Tagging (day 3, early pm)
Nobuhiro Kaji	Lexical Choice via Topic Adaptation for Paraphrasing Written Language to Spoken Language *Nobuhiro Kaji and Sadao Kurohashi*	Spoken Language (day 3, late pm)
Byeong-Kwu Kang	Extracting Terminologically Relevant Collocations in the Translation of Chinese Monograph *Byeong-Kwu Kang and Bao-Bao Chang*	Terminology Mining (day 3, late pm)
Daisuke Karwahara	PP-Attachment Disambiguation Boosted by a Gigantic Volume of Unambiguous Examples *Daisuke Kawahara and Sadao Kurohashi*	Disambiguation (day 1, early pm)
Kyungsun Kim	Improving Korean Speech Acts Analysis by Using Shrinkage and Discourse Stack *Kyungsun Kim, Youngjoong Ko, and Jungyun Seo*	Discourse Analysis (day 3, late am)
Seonho Kim	Two-phase Biomedical Named Entity Recognition Using A Hybrid Method *Seonho Kim, Juntae Yoon, Kyung-Mi Park, and Hae-Chang Rim*	Named Entity Recognition (day 3, early am)
Su Nam Kim	Automatic Interpretation of Noun Compounds using WordNet Similarity *Su Nam Kim and Timothy Baldwin*	Semantic Analysis – II (day 3, early pm)
Chunyu Kit	Period Disambiguation with Maxent Model *Chunyu Kit and Xiaoyue Liu*	Disambiguation (day 1, early pm)

Oi Yee Kwong	Semantic Role Tagging for Chinese at the Lexical Level *Oi Yee Kwong and Benjamin K. Tsou*	Semantic Analysis – II (day 3, early pm)
Matthew Lease	Parsing Biomedical Literature *Matthew Lease and Eugene Charniak*	Corpus-based Parsing (day 1, am)
Changbeom Lee	Significant Sentence Extraction by Euclidean Distance based on Singular Value Decomposition *Changbeom Lee, Hyukro Park and Cheolyoung Ock*	Text Summarization (day 3, early am)
Seungwoo Lee	Heuristic Methods for Reducing Errors of Geographic Named Entities Learned by Bootstrapping *Seungwoo Lee and Gary Geunbae Lee*	Named Entity Recognition (day 3, early am)
Sun-Hee Lee	Why is Zero Marking Important in Korean Treebanks? *Sun-Hee Lee, Donna Byron, and Seok Bae Jang*	Machine Translation – II (day 2, late am)
Yong-Hun Lee	Chunking using Conditional Random Fields in Korean Texts *Yong-Hun Lee, Mi-Young Kim and Jong-Hyeok Lee*	Rule-based Parsing (day 1, early pm)
Wei Li	A Preliminary Work on Classifying Time Granularities of Temporal Questions *Wei Li, Wenjie Li and Qin Lu*	Text Classification (day 2, early am)
Wei Li	Automatic Image Annotation and Retrieval using Maximum Entropy Model *Wei Li and Maosong Sun*	Information retrieval (day 1, am)
Tyne Liang	Anaphora Resolution for Biomedical Literature by Exploiting Multiple Resources *Tyne Liang and Yu-Hsiang Lin*	Discourse Analysis (day 3, late am)
Marshall R. Mayberry	A Connectionist Model of Anticipation in Visual Worlds *Marshall R. Mayberry, III Matthew W. Crocker Pia Knoeferle*	NLP Applications (day 3, early pm)
Yao Meng	A Knowledge-constrained Character Model for Chinese Morphology *Yao Meng, Hao Yu and Fumihito Nishino*	Morphological Analysis (day 2, late am)
Ryo Nagata	Detecting article errors based on the mass count distinction *Ryo Nagata, Takahiro Wakana, Fumito Masui, Atsuo Kawa, and Naoki Isu*	NLP Applications (day 3, early pm)
Toshiaki Nakazawa	Automatic Acquisition of Basic Katakana Lexicon from a Given Corpus *Toshiaki Nakazawa, Daisuke Kawahara and Sadao Kurohash*i	Linguistic Resources and Tools (day 3, late am)
Jong-Hoon Oh	An Ensemble of Grapheme and Phoneme for Machine Transliteration *Jong-Hoon Oh and Key-Sun Choi*	Transliteration (day 1, early am)

Daisuke Okanohara	Assigning Polarity Scores to Reviews using Machine Learning Techniques *Daisuke Okanohara and Jun'ichi Tsujii*	Document Analysis (day 1, late pm)
Yan Qu	The Use of Monolingual Context Vectors for Missing Translations in Cross-Language Information Retrieval *Yan Qu, Gregory Grefenstette and David A. Evans*	Information retrieval (day 1, am)
Marius Pasca	Aligning Needles in a Haystack: Paraphrase Acquisition Across the Web *Marius Pasca and Peter Dienes*	Web Mining (day 1, am)
Manabu Sassano	Using a Partially Annotated Corpus to Build a Dependency Parser for Japanese *Manabu Sassano*	Corpus-based Parsing (day 1, am)
Hee-Cheol Seo	Word Sense Disambiguation by Relative Selection *Hee-Cheol Seo, Hae-Chang Rim and Myung-Gil Jang*	Semantic Analysis – II (day 3, early pm)
Hendra Setiawan	Phrase-based Statistical Machine Translation: A Level of Detail Approach *Hendra Setiawan, Haizhou Li, Min Zhang and Beng Chin Ooi*	Machine Translation – II (day 2, late am)
Dan Shen	Exploring Syntactic Relation Patterns for Question Answering *Dan Shen, Greet-Jan M. Kruijff and Dietrich Klakow*	Question & Answering (day 2, late am)
Tomohide Shibata	Automatic Slide Generation Based on Discourse Structure Analysis *Tomohide Shibata and Sadao Kurohashi*	Discourse Analysis (day 3, later am)
Mitsuo Shimohata	Acquiring Synonyms from Monolingual Comparable Texts *Mitsuo Shimohata and Eiichiro Sumita*	Text Mining (day 1, early pm)
Andrew Smith	Regularization Techniques for Conditional Random Fields Parameterized versus Parameter-free *Andrew Smith and Miles Osborne*	Tagging (day 3, early pm)
Akihiro Tamura	Classification of Multiple-Sentence Questions *Akihiro Tamura, Hiroya Takamura and Manabu Okumura*	Text Classification (day 2, early am)
Kumiko Tanaka-Ishii	Entropy as Indicator of Context Boundaries - An experiment using Web Search Engine Kumiko Tanaka-Ishii	Web Mining (day 1, am)
Kosuke Tokunaga	Automatic Discovery of Attribute Words from Web Documents *Kosuke Tokunaga, Jun'ichi Kazama and Kentaro Torisawa*	Web Mining (day 1, am)
Shisanu Tongchi	Analysis of an Iterative Algorithm for Term-Based Ontology Alignment *Shisanu Tongchim, Canasai Kruengkrai, Virach Sornlertlamvanich, Prapass Srichaivattana and Hitoshi*	Ontology and Thesaurus (day 3, late am)

	Isahara	
Ozlem Uzuner	A Comparative Study of Language Models for Book and Author Recognition *Ozlem Uzuner and Boris Katz*	Language Model (day 3, late pm)
Tony Veale	Analogy as Functional Recategorization Abstraction with HowNet Semantics *Tony Veale*	Ontology and Thesaurus (day 3, late am)
Sriram Venkatapathy	Relative compositionality of multi-word expressions a study of verb-noun (V-N) collocations *Sriram Venkatapathy and Aravind K. Joshi*	Morphological Analysis (day 2, late am)
Yi-Chia Wang	Web-Based Unsupervised Learning for Query Formulation in Question Answering *Yi-Chia Wang, Jian-Cheng Wu, Tyne Liang and Jason S. Chang*	Question & Answering (day 2, late am)
Ruvan Weerasinghe	A Rule Based Syllabification Algorithm for Sinhala *Ruvan Weerasinghe, Asanka Wasala and Kumudu Gamage*	Transliteration (day 1, early am)
Fai Wong	Machine Translation Based on Constraint-Based Synchronous Grammar *Fai Wong, Dong-Cheng Hu, Yu-Hang Mao, Ming-Chui Dong and Yi-Ping Li*	Machine Translation – II (day 2, late am)
Dekai Wu	Inversion Transduction Grammar Constraints for Mining Parallel Sentences from Quasi-Comparable Corpora *Dekai Wu and Pascale Fung*	Text Mining (day 1, early pm)
Hua Wu	Improving Statistical Word Alignment with Ensemble Methods *Hua Wu and,Haifeng Wang*	Machine Translation – I (day 2, early am)
Mingl Wu	CTEMP A Chinese Temporal Parser for Extracting and Normalizing Temporal Information *MingliWu , Wenjie Li, Qin Lu and Baoli Li*	Linguistic Resources and Tools (day 3, late am)
Yasuhiko Watanabe	Confirmed Knowledge Acquisition Using Mails Posted to a Mailing List *Yasuhiko Watanabe, Ryo Nishimura and Yoshihiro Okada*	Web Mining (day 1, am)
JingHui Xiao	Principles of Non-stationary Hidden Markov Model and its Applications on Sequence Labeling Task *JingHui Xiao, BingQuan Liu and XiaoLong Wang*	NLP Applications (day 3, early pm)
Deyi Xiong	Parsing the Penn Chinese Treebank with Semantic Knowledge *Deyi Xiong, Shuanglong Li, Qun Liu, Shouxun Lin and Yueliang Qian*	Corpus-based Parsing (day 1, am)
Xiaofeng Yang	A Twin-Candidate Model of Coreference Resolution with Non-Anaphor Identification Capability *Xiaofeng Yang, Jian Su and Chew Lim Tan*	Discourse Analysis (day 3, late am)

Patrick Ye	Semantic Role Labelling of Prepositional Phrases *Patrick Ye and Timothy Baldwin*	Semantic Analysis – I (day 1, late pm)
Yang Ye	Tense Tagging for Verbs in Cross-Lingual Context a Case Study *Yang Ye and Zhu Zhang*	Tagging (day 3, early pm)
Minoru Yoshida	Automatic Term Extraction based on Perplexity of Compound Words *Minoru Yoshida and Hiroshi Nakagawa*	Text Mining (day 1, early pm)
Wei Yuan	An Empirical Study on Language Model Adaptation Using a Metric of Domain Similarity *Wei Yuan, Jianfeng Gao and Hisami Suzuki*	Language Model (day 3, late pm)
Min Zhang	A Phrase-based Context-dependent Joint Probability Model for Named Entity Translation *Min Zhang, Haizhou Li, Jian Su and Hendra Setiwan*	Machine Translation – II (day 2, late am)
Min Zhang	Discovering Relations from a Large Raw Corpus Using Tree Similarity-based Clustering *Min Zhang, Jian Su, Danmei Wang and Guodong Zhou*	Relation Extraction (day 3, late am)
Zhu Zhang	Mining Inter-entity Semantic Relations using Improved Transductive Learning *Zhu Zhang*	Relation Extraction (day 3, late am)
Zhushuo Zhang	Answering Definition Questions Using Web Knowledge Bases *Zhushuo Zhang, Yaqian Zhou, Xuangjing Huang and Lide Wu*	Question & Answering (day 2, late am)
GuoDong Zhou	A Chunking Strategy towards Unknown Word Detection in Chinese Word Segmentation *GuoDong Zhou*	Morphological Analysis (day 2, late am)
Jingbo Zhu	Using Multiple Discriminant Analysis Approach for Linear Text Segmentation *Jingbo Zhu, Na Ye, Xinzhi Chang, Wenliang Chen, Benjamin Tsou*	Document Analysis (Day 1, late pm)