2006

COLING • ACL

COLING • ACL 2006

21st International Conference on
Computational Linguistics
and
44th Annual Meeting of the
Association for Computational Linguistics

Proceedings of the Main Conference Poster Sessions

17-18 July 2006
Sydney, Australia

©2006 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

Table of Contents

Using Machine Learning Techniques to Build a Comma Checker for Basque
Iñaki Alegria, Bertol Arrieta, Arantza Diaz de Ilarraza, Eli Izagirre and Montse Maritxalar 1

A Rote Extractor with Edit Distance-Based Generalisation and Multi-Corpora Precision Calculation
Enrique Alfonseca, Pablo Castells, Manabu Okumura and Maria Ruiz-Casado 9

MT Evaluation: Human-Like vs. Human Acceptable
Enrique Amigó, Jesús Giménez, Julio Gonzalo and Lluís Màrquez 17

The Effect of Corpus Size in Combining Supervised and Unsupervised Training for Disambiguation
Michaela Atterer and Hinrich Schütze 25

A Phrase-Based Statistical Model for SMS Text Normalization
AiTi Aw, Min Zhang, Juan Xiao and Jian Su 33

Evaluating the Accuracy of an Unlexicalized Statistical Parser on the PARC DepBank
Ted Briscoe and John Carroll 41

N Semantic Classes are Harder than Two
Ben Carterette, Rosie Jones, Wiley Greiner and Cory Barr 49

Towards Conversational QA: Automatic Identification of Problematic Situations and User Intent
Joyce Y. Chai, Chen Zhang and Tyler Baldwin 57

A Pipeline Framework for Dependency Parsing
Ming-Wei Chang, Quang Do and Dan Roth 65

A Hybrid Convolution Tree Kernel for Semantic Role Labeling
Wanxiang Che, Min Zhang, Ting Liu and Sheng Li 73

A High-Accurate Chinese-English NE Backward Translation System Combining Both Lexical Information and Web Statistics
Conrad Chen and Hsin-Hsi Chen 81

Unsupervised Relation Disambiguation Using Spectral Clustering
Jinxiu Chen, Donghong Ji, Chew Lim Tan and Zhengyu Niu 89

An Empirical Study of Chinese Chunking
Wenliang Chen, Yujie Zhang and Hitoshi Isahara 97

Soft Syntactic Constraints for Word Alignment through Discriminative Training
Colin Cherry and Dekang Lin 105

An Account for Compound Prepositions in Farsi
Zahra Abolhassani Chime 113

Techniques to Incorporate the Benefits of a Hierarchy in a Modified Hidden Markov Model
Lin-Yi Chou 120

Analysis and Synthesis of the Distribution of Consonants over Languages: A Complex Network Approach
Monojit Choudhury, Animesh Mukherjee, Anupam Basu and Niloy Ganguly 128

Using Machine-Learning to Assign Function Labels to Parser Output for Spanish
Grzegorz Chrupała and Josef van Genabith 136

Constraint-Based Sentence Compression: An Integer Programming Approach
James Clarke and Mirella Lapata 144

Topic-Focused Multi-Document Summarization Using an Approximate Oracle Score
John M. Conroy, Judith D. Schlesinger and Dianne P. O'Leary 152

Using WordNet to Automatically Deduce Relations between Words in Noun-Noun Compounds
Fintan J. Costello, Tony Veale and Simon Dunne 160

Automatically Extracting Nominal Mentions of Events with a Bootstrapped Probabilistic Classifier
Cassandre Creswell, Matthew J. Beal, John Chen, Thomas L. Cornell, Lars Nilsson and Rohini K. Srihari 168

A Bio-Inspired Approach for Multi-Word Expression Extraction
Jianyong Duan, Ruzhan Lu, Weilin Wu, Yi Hu and Yan Tian 176

Towards A Modular Data Model For Multi-Layer Annotated Corpora
Richard Eckart 183

A Modified Joint Source-Channel Model for Transliteration
Asif Ekbal, Sudip Kumar Naskar and Sivaji Bandyopadhyay 191

Chinese-English Term Translation Mining Based on Semantic Prediction
Gaolin Fang, Hao Yu and Fumihito Nishino 199

Automatic Creation of Domain Templates
Elena Filatova, Vasileios Hatzivassiloglou and Kathleen McKeown 207

Using Lexical Dependency and Ontological Knowledge to Improve a Detailed Syntactic and Semantic Tagger of English
Andrew Finch, Ezra Black, Young-Sook Hwang and Eiichiro Sumita 215

The Benefit of Stochastic PP Attachment to a Rule-Based Parser
Kilian A. Foth and Wolfgang Menzel 223

Using Bilingual Comparable Corpora and Semi-supervised Clustering for Topic Tracking
Fumiyo Fukumoto and Yoshimi Suzuki 231

Robust Word Sense Translation by EM Learning of Frame Semantics
Pascale Fung and Benfeng Chen 239

Coreference Handling in XMG
Claire Gardent and Yannick Parmentier 247

Conceptual Coherence in the Generation of Referring Expressions
Albert Gatt and Kees van Deemter 255

Discriminative Reranking for Semantic Parsing
Ruifang Ge and Raymond J. Mooney 263

Multilingual Lexical Database Generation from Parallel Texts in 20 European Languages with Endogenous Resources
Emmanuel Giguet and Pierre-Sylvain Luquet 271

Factoring Synchronous Grammars by Sorting
Daniel Gildea, Giorgio Satta and Hao Zhang 279

Low-Cost Enrichment of Spanish WordNet with Automatically Translated Glosses: Combining General and Specialized Models
Jesús Giménez and Lluís Màrquez 287

Speeding Up Full Syntactic Parsing by Leveraging Partial Parsing Decisions
Elliot Glaysher and Dan Moldovan 295

Parsing Aligned Parallel Corpus by Projecting Syntactic Relations from Annotated Source Corpus
Shailly Goyal and Niladri Chatterjee 301

Reduced n-gram Models for English and Chinese Corpora
Le Q Ha, P Hanna, D W Stewart and F J Smith 309

Discriminative Classifiers for Deterministic Dependency Parsing
Johan Hall, Joakim Nivre and Jens Nilsson 316

Detection of Quotations and Inserted Clauses and Its Application to Dependency Structure Analysis in Spontaneous Japanese
Ryoji Hamabe, Kiyotaka Uchimoto, Tatsuya Kawahara and Hitoshi Isahara 324

Improving English Subcategorization Acquisition with Diathesis Alternations as Heuristic Information
Xiwu Han, Tiejun Zhao and Xingshang Fu 331

Local Constraints on Sentence Markers and Focus in Somali
Katherine Hargreaves and Allan Ramsay 337

A Collaborative Framework for Collecting Thai Unknown Words from the Web
Choochart Haruechaiyasak, Chatchawal Sangkeettrakarn, Pornpimon Palingoon, Sarawoot Kongyoung and Chaianun Damrongrat 345

Japanese Idiom Recognition: Drawing a Line between Literal and Idiomatic Meanings
Chikara Hashimoto, Satoshi Sato and Takehito Utsuro 353

Graph Branch Algorithm: An Optimum Tree Search Method for Scored Dependency Graph with Arc Co-Occurrence Constraints
Hideki Hirakawa 361

Exploring the Potential of Intractable Parsers
Mark Hopkins and Jonas Kuhn 369

Transformation-Based Interpretation of Implicit Parallel Structures: Reconstructing the Meaning of "vice versa" and Similar Linguistic Operators
Helmut Horacek and Magdalena Wolska 377

When Conset Meets Synset: A Preliminary Survey of an Ontological Lexical Resource Based on Chinese Characters
Shu-Kai Hsieh and Chu-Ren Huang 385

Spontaneous Speech Understanding for Robust Multi-Modal Human-Robot Communication
Sonja Hüwel and Britta Wrede 391

Efficient Sentence Retrieval Based on Syntactic Structure
Hiroshi Ichikawa, Keita Hakoda, Taiichi Hashimoto and Takenobu Tokunaga 399

Towards the Orwellian Nightmare: Separation of Business and Personal Emails
Sanaz Jabbari, Ben Allison, David Guthrie and Louise Guthrie 407

Exploiting Non-Local Features for Spoken Language Understanding
Minwoo Jeong and Gary Geunbae Lee 412

Analysis and Repair of Name Tagger Errors
Heng Ji and Ralph Grishman 420

Unsupervised Segmentation of Chinese Text by Use of Branching Entropy
Zhihui Jin and Kumiko Tanaka-Ishii 428

A FrameNet-Based Semantic Role Labeler for Swedish
Richard Johansson and Pierre Nugues 436

Obfuscating Document Stylometry to Preserve Author Anonymity
Gary Kacmarcik and Michael Gamon 444

Automatic Construction of Polarity-Tagged Corpus from HTML Documents
Nobuhiro Kaji and Masaru Kitsuregawa 452

Minority Vote: At-Least-N Voting Improves Recall for Extracting Relations
Nanda Kambhatla 460

Integration of Speech to Computer-Assisted Translation Using Finite-State Automata
Shahram Khadivi, Richard Zens and Hermann Ney 467

GF Parallel Resource Grammars and Russian
Janna Khegai 475

Automatic Identification of Pro and Con Reasons in Online Reviews
Soo-Min Kim and Eduard Hovy 483

Interpreting Semantic Relations in Noun Compounds via Verb Semantics
Su Nam Kim and Timothy Baldwin 491

Unsupervised Analysis for Decipherment Problems
Kevin Knight, Anish Nair, Nishit Rathod and Kenji Yamada 499

Mildly Non-Projective Dependency Structures
Marco Kuhlmann and Joakim Nivre 507

Parsing and Subcategorization Data
Jianguo Li and Chris Brew 515

The Role of Information Retrieval in Answering Complex Questions
Jimmy Lin 523

Examining the Content Load of Part of Speech Blocks for Information Retrieval
Christina Lioma and Iadh Ounis 531

Stochastic Iterative Alignment for Machine Translation Evaluation
Ding Liu and Daniel Gildea 539

Discriminating Image Senses by Clustering with Multimodal Features
Nicolas Loeff, Cecilia Ovesdotter Alm and David A. Forsyth 547

Modeling Adjectives in Computational Relational Lexica
Palmira Marrafa and Sara Mendes 555

Segmented and Unsegmented Dialogue-Act Annotation with Statistical Dialogue Models
Carlos D. Martínez Hinarejos, Ramón Granell and José Miguel Benedí 563

ARE: Instance Splitting Strategies for Dependency Relation-Based Information Extraction
Mstislav Maslennikov, Hai-Kiat Goh and Tat-Seng Chua 571

Integrating Pattern-Based and Distributional Similarity Methods for Lexical Entailment Acquisition
Shachar Mirkin, Ido Dagan and Maayan Geffet 579

Machine-Learning-Based Transformation of Passive Japanese Sentences into Active by Separating Training Data into Each Input Particle
Masaki Murata, Toshiyuki Kanamaru, Tamotsu Shirado and Hitoshi Isahara 587

Reinforcing English Countability Prediction with One Countability per Discourse Property
Ryo Nagata, Atsuo Kawai, Koichiro Morihiro and Naoki Isu 595

An Automatic Method for Summary Evaluation Using Multiple Evaluation Results by a Manual Method
Hidetsugu Nanba and Manabu Okumura 603

Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews
Vincent Ng, Sajib Dasgupta and S. M. Niaz Arifin 611

Semantic Parsing with Structured SVM Ensemble Classification Models
Le-Minh Nguyen, Akira Shimazu and Xuan-Hieu Phan 619

Whose Thumb Is It Anyway? Classifying Author Personality from Weblog Text
Jon Oberlander and Scott Nowson 627

Analysis of Selective Strategies to Build a Dependency-Analyzed Corpus
Kiyonori Ohtake 635

A Term Recognition Approach to Acronym Recognition
Naoaki Okazaki and Sophia Ananiadou 643

Combining Association Measures for Collocation Extraction
Pavel Pecina and Pavel Schlesinger 651

Using Machine Learning to Explore Human Multimodal Clarification Strategies
Verena Rieser and Oliver Lemon 659

URES : an Unsupervised Web Relation Extraction System
Benjamin Rosenfeld and Ronen Feldman 667

Argumentative Feedback: A Linguistically-Motivated Term Expansion for Information Retrieval
Patrick Ruch, Imad Tbahriti, Julien Gobeill and Alan R. Aronson 675

Simultaneous English-Japanese Spoken Language Translation Based on Incremental Dependency Parsing and Transfer
Koichiro Ryu, Shigeki Matsubara and Yasuyoshi Inagaki 683

A Best-First Probabilistic Shift-Reduce Parser
Kenji Sagae and Alon Lavie 691

Implementing a Characterization of Genre for Automatic Genre Identification of Web Pages
Marina Santini, Richard Power and Roger Evans 699

Translating HPSG-Style Outputs of a Robust Parser into Typed Dynamic Logic
Manabu Sato, Daisuke Bekki, Yusuke Miyao and Jun'ichi Tsujii 707

ATLAS – A New Text Alignment Architecture
Bettina Schrader 715

Continuous Space Language Models for Statistical Machine Translation
Holger Schwenk, Daniel Dechelotte and Jean-Luc Gauvain 723

On-Demand Information Extraction
Satoshi Sekine 731

Using Comparable Corpora to Solve Problems Difficult for Human Translators
Serge Sharoff, Bogdan Babych and Anthony Hartley 739

Adding Syntax to Dynamic Programming for Aligning Comparable Texts for the Generation of Paraphrases
Siwei Shen, Dragomir R. Radev, Agam Patel and Güneş Erkan 747

Unsupervised Topic Identification by Integrating Linguistic and Visual Information Based on Hidden Markov Models
Tomohide Shibata and Sadao Kurohashi 755

Exact Decoding for Jointly Labeling and Chunking Sequences
Nobuyuki Shimizu and Andrew Haas 763

Compiling a Lexicon of Cooking Actions for Animation Generation
Kiyoaki Shirai and Hiroshi Ookawa 771

Morphological Richness Offsets Resource Demand – Experiences in Constructing a POS Tagger for Hindi
Smriti Singh, Kuhoo Gupta, Manish Shrivastava and Pushpak Bhattacharyya 779

Minimum Risk Annealing for Training Log-Linear Models
David A. Smith and Jason Eisner 787

Unsupervised Induction of Modern Standard Arabic Verb Classes Using Syntactic Frames and LSA
Neal Snider and Mona Diab 795

Discourse Generation Using Utility-Trained Coherence Models
Radu Soricut and Daniel Marcu 803

A Comparison of Alternative Parse Tree Paths for Labeling Semantic Roles
Reid Swanson and Andrew S. Gordon 811

A Logic-Based Semantic Approach to Recognizing Textual Entailment
Marta Tatu and Dan Moldovan 819

Infrastructure for Standardization of Asian Language Resources
Takenobu Tokunaga, Virach Sornlertlamvanich, Thatsanee Charoenporn, Nicoletta Calzolari, Monica Monachini, Claudia Soria, Chu-Ren Huang, YingJu Xia, Hao Yu, Laurent Prevot and Kiyoaki Shirai 827

Statistical Phrase-Based Models for Interactive Computer-Assisted Translation
Jesús Tomás and Francisco Casacuberta 835

Using Word Support Model to Improve Chinese Input System
Jia-Lin Tsai 842

Trimming CFG Parse Trees for Sentence Compression Using Machine Learning Approaches
Yuya Unno, Takashi Ninomiya, Yusuke Miyao and Jun'ichi Tsujii 850

Word Vectors and Two Kinds of Similarity
Akira Utsumi and Daisuke Suzuki 858

Finding Synonyms Using Automatic Word Alignment and Measures of Distributional Similarity
Lonneke van der Plas and Jörg Tiedemann 866

Word Alignment for Languages with Scarce Resources Using Bilingual Corpora of Other Language Pairs
Haifeng Wang, Hua Wu and Zhanyi Liu 874

Combining Statistical and Knowledge-Based Spoken Language Understanding in Conditional Models
Ye-Yi Wang, Alex Acero, Milind Mahajan and John Lee 882

Sinhala Grapheme-to-Phoneme Conversion and Rules for Schwa Epenthesis
Asanka Wasala, Ruvan Weerasinghe and Kumudu Gamage 890

From Prosodic Trees to Syntactic Trees
Andi Wu and Kirk Lowery 898

A Grammatical Approach to Understanding Textual Tables Using Two-Dimensional SCFGs
Dekai Wu and Ken Wing Kuen Lee 905

Boosting Statistical Word Alignment Using Labeled and Unlabeled Data
Hua Wu, Haifeng Wang and Zhanyi Liu 913

Aligning Features with Sense Distinction Dimensions
Nianwen Xue, Jinying Chen and Martha Palmer 921

Word Sense Disambiguation Using Lexical Cohesion in the Context
Dongqiang Yang and David M. W. Powers 929

Stochastic Discourse Modeling in Spoken Dialogue Systems Using Semantic Dependency Graphs
Jui-Feng Yeh, Chung-Hsien Wu and Mao-Zhu Yang 937

HAL-Based Cascaded Model for Variable-Length Semantic Pattern Induction from Psychiatry Web Resources
Liang-Chih Yu, Chung-Hsien Wu and Fong-Lin Jang 945

Inducing Word Alignments with Bilexical Synchronous Trees
Hao Zhang and Daniel Gildea 953

Subword-Based Tagging for Confidence-Dependent Chinese Word Segmentation
Ruiqiang Zhang, Genichiro Kikui and Eiichiro Sumita 961

BiTAM: Bilingual Topic AdMixture Models for Word Alignment
Bing Zhao and Eric P. Xing 969

An HMM-Based Approach to Automatic Phrasing for Mandarin Text-to-Speech Synthesis
Jing Zhu and Jian-Hua Li 977

Author Index 983