

计算语言学教育部重点实验室

2010 成果清单

一、论文

李月伦, 常宝宝. 基于最大间隔马尔可夫网模型的汉语分词方法, 中文信息学报, 2010 年第 1 期.

Baobao Chang, Dongxu Han, Enhancing domain portability of Chinese segmentation model using chi-square statistics and bootstrapping. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010). 2010.

Tiansang Du, Baobao Chang. Training MT Model Using Structural SVM. In Proceedings of the 2010 International Conference on Asian Language Processing (IALP 2010), 2010

Lei Wang, Baobao Chang, Janet Harkness: A Method of Automatic Translation of Words of Multiple Affixes in Scientific Literature. In Proceedings of the 2010 International Conference on Asian Language Processing (IALP 2010), 2010

Baobao Chang, Mairgup Mansur, Chinese word segmentation model using bootstrapping. In Proceedings of the 2010 CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP 2010). 2010.

Baobao Chang, Byeong-Kwu Kang, Shiwen Yu, Developing CAT tools for translating Chinese scientific monographs, *Journal of Translation Studies*, Volume 13, Number 1-2:165-179. 2010.

李月伦, 李湘, 常宝宝, 袁毓林. 一种基于认知情景框架的文本分类方法. 第五届全国青年计算语言学研讨会文集 (YWCL2010), 2010.

王雷, 常宝宝, 俞士汶, 汉语定语从句结构的一种自动翻译方法, 第五届全国青年计算语言学研讨会文集 (YWCL2010), 2010

董秀芳.汉语的句法演变与词汇化, 《中国语文》2009 年第 5 期: 399-409。又收入《北大中文学刊》: 630-645, 北京大学出版社, 2010 年 8 月。

董秀芳.汉语光杆名词指称特性的历时演变, 《语言研究》2010 年第 1 期: 11-20 页。

董秀芳.汉语中表示承诺的言语施为动词, 《汉语学习》2010 年第 2 期, 23-30 页。

董秀芳.来源于完整小句的话语标记“我告诉你”, 《语言科学》2010 年第 3 期: 279-286 页。

董秀芳.量与强调, 徐丹主编《量与复数的研究——中国境内语言的跨时空研究》, 商务印书馆, 2010年。

董秀芳.共性和类型: 词汇语义关系研究的方向——评《词汇语义学》, 《中国社会科学报》2010年12月21日第8版。

Li Decong, Li Sujian, Li Wenjie, Gu Congyun, Li Yun, Keyphrases Extraction Based on Topic Relevance and Term Association, Journal of Information and Computational Science,(SEWM 2010, Chengdu, May 2010.) 7:1 (2010) 293-299.
(EI)

李素建, 谭雨田, 邵艳秋, 李芸, 穗志方, 服务于关键词标引的韵律语法研究, 第十一届汉语词汇语义学研讨会 (CLSW2010), 2010年5月, 苏州。

Li Decong, Li Sujian, Li Wenjie, Wang Wei, A Semi-Supervised Key Phrase Extraction Approach: Learning from Title Phrases through a Document Semantic Network, ACL 2010, July 11-16. Uppsala, Sweden.

You Ouyang, Wenjie Li, Sujian Li, Qin Lu, Intertopic information mining for query-based summarization, Journal of the American Society for Information Science and Technology (JASIST), vol 61 issue 5, May 2010.

宋涛, 李素建, 基于流形排序的领域词抽取方法, 第五届全国青年计算语言学研讨会 (YWCL2010), 2010 年 10 月, 武汉.

Li Sujian, Song Tao, Wang Xun, TAC 2010 Update Summarization and AESOP of ICL, TAC 2010 workshop, November 15-16, 2010, Gaithersburg, Maryland USA.

吴轲, 李素建, 王厚峰, 基于 CRF 和词信息的缩略语预测模型, 认识计算学术交流会, 2010.12.23-24, 北京

Zhifang Sui, Yao Liu, Jun Zhao, Hong Zhang, The Development of an NLP-based Chinese Ontology Construction, The 2010 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology(WI-IAT 2010), 2010.

Zhifang Sui, Yongwei Hu, Hong Zhang, An Interactive Approach to Term Relation Extraction and Term Extraction, Journal of Computational Information Systems, 6(1), pp 229-235, 2010.

Shao Yanqiu, Sui Zhifang, Mao Ning, Chinese Semantic Role Labeling Based on Semantic Knowledge, the 6th International Conference on Natural Language Processing and Knowledge Engineering, NLPKE 2010, 2010.

Xin Wang, Weiwei Sun, Zhifang Sui, A Lightweight Chinese Semantic Dependency Parsing Model Based on Sentence Compression, The 6th International Conference on Natural Language Processing and Knowledge Engineering, NLPKE 2010, 2010.

孙斌. 网络信息检索基础平台. 第六届自然语言处理青年学者研讨会 (YSSNLP 2009). 大连理工大学, 2010 年 9 月.

周彬, 孙斌. 中文网页内容提取方法. 第八届全国搜索引擎和网上信息挖掘学术研讨会 (SEWM2010), 2010 年 1 月, 成都.

B. Zhou, et al. Chinese Web Page Content Extraction based on Page Content Analysis. Journal of Computational Information Systems, Vol. 5, May 2010.

王婵娟, 孙斌, 张璐. 汉语词汇语义在网页关键词提取算法中的应用. 第十一届汉语词汇语义学研讨会 (CLSW2010), 东方与中文语言信息处理学会 (新加坡), 2010 年 5 月, 苏州大学.

张璐, 苏祺, 孙斌, 王婵娟. 基于 HittingTime 模型的跨语言图片检索. 第十一届汉语词汇语义学研讨会 (CLSW2010), 东方与中文语言信息处理学会 (新加坡), 2010 年 5 月, 苏州大学.

Weiwei Sun. 2010. Word-based and character-based word segmentation models: comparison and combination. COLING.

Weiwei Sun. 2010. Improving Chinese semantic role labeling with rich syntactic features. ACL short.

Weiwei Sun. 2010. Semantics-driven shallow parsing for Chinese semantic role labeling. ACL short.

Xu Sun, Jianfeng Gao, Daniel Micol, Chris Quirk. Learning Phrase-Based Spelling Error Models from Clickthrough Data. In ACL. 266-274. 2010.

Xiaojun Wan, Jianguo Xiao. Exploiting Neighborhood Knowledge for Single Document Summarization and Keyphrase Extraction. ACM Transactions on Information Systems(TOIS), Volume 28, Issue 2, Article 8, 34 pages.

Xiaojun Wan, Huiying Li, Jianguo Xiao. Cross-Language Document Summarization Based on Machine Translation Quality Prediction. In ACL2010, pages 917-926. (Full Paper, Acceptance Rate = 25%)

Xiaojun Wan, Huiying Li, Jianguo Xiao. EUSUM: Extracting Easy-to-Understand English Summaries for Non-Native Readers. In SIGIR2010, pages 491-498. (Full Paper, Acceptance Rate = 16.7%)

Xiaojun Wan, Jianwu Yang. A practical system for harvesting and monitoring hot topics on the web. In WWW2010, pages 1197-1198. (Poster Paper)

Xiaojun Wan. Towards a Unified Approach to Simultaneous Single-Document and Multi-Document Summarizations. In COLING2010, pages 1137-1145. (Oral Paper, Acceptance Rate = 19%)

Tengfei Ma, Xiaojun Wan. Opinion Target Extraction in Chinese News Comments. In COLING2010, pages 782-790. (Poster Paper)

Tengfei Ma, Xiaojun Wan. Multi-Document Summarization Using Minimum Distortion. In ICDM2010, pages 354-363. (Regular Paper, Acceptance Rate=9%)

Liang Zong, Xiaojun Wan, Lihong Zhao, Jianwu Yang, Yuqian Wu. Named Entity Resolution in Chinese News Comments on the Web. In APWeb2010, pages 307-313. (Full Paper)

Chenfeng Wang, Tengfei Ma, Liqiang Guo, Xiaojun Wan, Jianwu Yang. PKUTM Experiments in NTCIR-8 MOAT Task. In Proceedings of the 8th NTCIR Workshop Meeting (NTCIR-8), pages 228-233.

Houping Jia, Xiaojun Wan, Tengfei Ma, Xiaojun Wan, Jianguo Xiao. PKUTM Participation at TAC 2010 RTE and Summarization Tracks. In Proceedings of the 2010 Text Analysis Conference (TAC2010).

何径舟,王厚峰. 基于特征选择和最大熵模型的汉语词义消歧, 软件学报, 2010(6).

Yang Song, Zhengyan He, Chen Chen, Houfeng Wang, A Pipeline Approach to Chinese Personal Name Disambiguation. In Proc. of CIPS-SIGHAN Joint Conference on Chinese Language Processing.

Ge Xu, Xinfan Meng, Houfeng Wang. Build Chinese Emotion Lexicons Using A Graph-based Algorithm and Multiple Resources. In Processing of the 23rd International Conference of Computational Linguistics(COLING-2010)

王厚峰 计算语言学歧义消解研究, 《术语标准化与信息技术》, 2010(3)

Yang Song, Zhengyan He, Houfeng Wang, ICL KBP Approaches to Knowledge Base Population at TAC2010, In Proc. of TAC2010.

黎锦熙先生的汉语方言分区方案述评, 《民俗典籍文字研究》第 7 辑, 商务印书馆, 2010 年, 北京。

评罗杰瑞教授的汉语方言分区方案, 余蔼芹、柯蔚南主编《罗杰瑞先生七秩晋三寿庆论文集》279-293, 香港中文大学中国文化研究所吴多泰中国语文研究中心, 2010 年, 香港。

韶山如意方言同音字汇, 《开篇》第 29 卷, 日本好文出版, 2010 年, 东京。

陈振宇、袁毓林。《汉语亲属关系的语义表示和自动推理》, 《中国语文》2010 年, 第 1 期, 第 44-56 页。中国人民大学复印报刊资料《语言文字学》2010 年第 5 期, 第 17-27 页。全文转载。

袁毓林。《汉语和英语在语法范畴的实现关系上的平行性——也谈汉语里的名词/动词与指称/陈述、主语与话题、句子与话段》，《汉藏语学报》2010年，第5期，第139-168页。

袁毓林、陈振宇、张秀松、李湘、周强、高嵩。《从认知假设到计算分析和程序实现——一种认知语言学研究的计算范式与技术路线》，《当代语言学》2010年第2期，第97-114页。中国人民大学复印报刊资料《语言文字学》2010年第2期，第页。全文转载。

袁毓林。《中国语言学在信息科学的话语权——建设语言资源，争取有话可说》，《中国语言资源动态》2010年第1期，第10-14页，商务印书馆内部刊物。

袁毓林、王明华。《文本蕴涵的推理模型与识别模型》，《中文信息学报》2010年，第2期，第3-13页。

陈振宇、袁毓林、张秀松、周强。《一种基于大知识库的亲属关系自动推理模型》，《中文信息学报》2010年，第3期，第117-123页。

袁毓林。《一种由问题推动的语法研究路线——以副词“都”的语义功能的探索为例》，《苏州教育学院学报》2010年，第2期，第2-11页。

袁毓林。《零形式和零成分的确立条件》，《当代语言学》2010年第3期，第220-231页。中国人民大学复印报刊资料《语言文字学》2010年第11期，第20-29页。全文转载。

袁毓林。《汉语不能承受的翻译之轻——从去范畴化角度看汉语动词和名词的关系》，《语言学论丛》第41辑，第15-61页；商务印书馆，2010年。中国人民大学复印报刊资料《语言文字学》2011年第3期，第页。全文转载。

袁毓林。The Family Resemblances in Word-classes (《词类范畴的家族相似性》)，刊 *Linguistics in China*, No 1, pp. 259-271 (《语言学文选》第一辑)，世界图书出版公司(北京)，2010年。袁毓林。《中国语言学在信息科学的话语权——建设语言资源，争取有话可说》，《中国语言资源动态》2010年第1期(总第7期)，教育部语言文字信息管理司指导，商务印书馆主办，内部刊物。

Lei Wang, Baobao Chang, A Method of Automatic Translation of Words of Multiple Affixes in Scientific Literature, 亚洲语言处理国际会议 (IALP2010) 论文集 (EI 检索号 20110613644393), IEEE Computer Society, 2010年12月

Wang Lei, Chang Baobao, Research on the Human-aided Auto-assessment System for Online Translation Tests of College English, 语言学习与新技术—北京大学第二届英语教学国际研讨会论文集, 北京大学出版社, 2010年10月

二、专著

《汉语配价语法研究》，商务印书馆，2010年，第1-578页，袁毓林。

《汉语词类的认知研究和模糊划分》，上海教育出版社，2010年，第1-443页，袁毓林。

三、科研项目

类型	批准部门	编号	负责人	时间	经费
国家自然科学基金重大研究计划/培育项目	国家自然科学基金委	90920011	李素建	2010.1 - 2012.12	50万
国家自然科学基金	基金委	60975054	常宝宝	2010.1- 2012.12	30万

北京市科技计划	北京市科委	Z101101005010003 三维 廉政风险预警及综合防 控技术研发	赵东岩	2010.5-2011.5	90 万
国家社会科学基金青年项 目	国家社会科学基金委	10CYY023	李素建	2010.7 - 2012.12	10 万
国家自然科学基金（面上）	国家自然科学基金委	60973053	王厚峰	2010-2012	30 万
面上项目	国家自然科学基金研究基 金会	30970889	周晓林	2010 年 1 月- 2012 年 12 月	35 万
科技部 973 项目	科技部	2010CB833904	周晓林	2010 年 1 月- 2014 年 12 月	453 万