

CCL 2021

**The 20th Chinese National Conference on
Computational Linguistic**

**Proceedings of the 20th Chinese National Conference on
Computational Linguistics**

August 13 - August 15, 2021

Hohhot, China

©The 19th Chinese National Conference on Computational Linguistic

Order copies of this and other CCL proceedings from:

Chinese National Conference on Computational Linguistic (CCL)
Courtyard 4, South Fourth Street, Zhongguancun , Haidian District, Beijing
100190, China
Tel: + 010-62562916
Fax: + 010-62661046
cips@iscas.ac.cn

Introduction

Welcome to the proceedings of the twentieth China National Conference on Computational Linguistics (20th CCL). The conference and symposium were hosted and co-organized by Inner Mongolia University, China.

CCL is an annual conference (bi-annual before 2013) that started in 1991. It is the flagship conference of the Chinese Information Processing Society of China (CIPS), which is the largest NLP scholar and expert community in China. CCL is a premier nation-wide forum for disseminating new scholarly and technological work in computational linguistics, with a major emphasis on computer processing of the languages in China such as Mandarin, Tibetan, Mongolian, and Uyghur.

The Program Committee selected 108 papers (77 Chinese papers and 31 English papers) out of 303 submissions for publication. The acceptance rate is 35.64%. The 108 papers cover the following topics:

- Machine Translation and Multilingual Information Processing (7)
- Fundamental Theory and Methods of Computational Linguistics (8)
- Minority Language Information Processing (11)
- Social Computing and Sentiment Analysis (5)
- Text Generation and Summarization (4)
- Information Retrieval, Dialogue and Question Answering (12)
- Linguistics and Cognitive Science (15)
- Language Resource and Evaluation (8)
- Knowledge Graph and Information Extraction (13)
- NLP Applications (25)

The final program for the 20th CCL was the result of intense work by many dedicated colleagues. We want to thank, first of all, the authors who submitted their papers, contributing to the creation of the high-quality program. We are deeply indebted to all the Program Committee members for providing high-quality and insightful reviews under a tight schedule, and extremely grateful to the sponsors of the conference. Finally, we extend a special word of thanks to all the colleagues of the Organizing Committee and secretariat for their hard work in organizing the conference, and to ACL Anthology for their assistance in publishing the proceedings in due time.

We thank the Program and Organizing Committees for helping to make the conference successful, and we hope all the participants enjoyed the first online CCL conference.

July 2021

Sheng Li, Maosong Sun, Yang Liu
Hua Wu, Kang Liu, Wanxiang Che

Organizers

Conference Chairs

Sheng Li Harbin Institute of Technology
Maosong Sun Tsinghua University, China
Yang Liu Tsinghua University, China

Program Chairs

Hua Wu Baidu, Inc.
Kang Liu Institute of Automation, Chinese Academy of Sciences, China
Wanxiang Che Harbin Institute of Technology, China

Area Co-Chairs

Renfen Hu Beijing Normal University, China
Xiaofei Lu The Pennsylvania State University, USA
Zhenghua Li Soochow University, China
Pengfei Liu Carnegie Mellon University, USA
Jiafeng Guo Institute of Computing Technology, CAS, China
Qingyao Ai The university of Utah, USA
Rui Yan Renmin University of China, China
Jing Li The Hong Kong Polytechnic University
Wenliang Chen Soochow University, China
Bowe Zou Institute for Infocomm Research, Singapore
Shujian Huang Nanjing University, China
Hui Huang University of Macau, China
Yating Yang Xinjiang Physics and Chemistry Institute, CAS, China
Chenhui Chu Kyoto University, Japan
Junsheng Zhou Nanjing Normal University, China
Wei Lu Singapore University of Technology and Design, Singapore
Rui Xia Nanjing University of Science and Technology, China
Lin Gui The University of Warwick, England
Chenliang Li Wuhan University, China
Jiashu Zhao Wilfrid Laurier University, Canada

Local Arrangement Chairs

Guanglai Gao Inner Mongolia University, China
Hongxu Hou Inner Mongolia University, China

Evaluation Chairs

Hongfei Lin Dalian University of Technology, China
Wei Song Capital Normal University, China

Publications Chairs

Shizhu He Institute of Automation, CAS, China
Gaoqi Rao Beijing Language and Culture University, China

Workshop Chairs

Xianpei Han Institute of Software, CAS, China
Yanyan Lan Tsinghua University, China

Sponsorship Chairs

Zhongyu Wei Fudan University, China
Yang Feng Institute of Computing Technology, CAS, China

Publicity Chair

Xipeng Qiu Fudan University, China

Website Chair

Liang Yang Dalian University of Technology, China

System Demonstration Chairs

Jiajun Zhang Institute of Automation, CAS, China
Haofen Wang Tongji University, China

Student Counseling Chair

Zhiyuan Liu Tsinghua University, China

Student Seminar Chairs

Xin Zhao Renmin University of China, China
Libo Qin Harbin Institute of Technology, China

Finance Chair

Yuxing Wang

Tsinghua University, China

Table of Content

<i>融合零指代识别的篇章级机器翻译</i>	
汪浩, 李军辉, 贡正仙·····	1
<i>融合 XLM 词语表示的神经机器译文自动评价方法</i>	
胡纬, 李茂西, 裘白莲, 王明文·····	13
<i>利用语义关联增强的跨语言预训练模型的译文质量评估</i>	
叶恒, 贡正仙·····	23
<i>基于模型不确定性约束的半监督汉缅神经机器翻译</i>	
王琳钦, 余正涛, 毛存礼, 高盛祥, 满志博, 王振晗·····	35
<i>基于枢轴语言系统融合的词汇混淆网络神经机器翻译</i>	
赵小兵, 金波, 孙媛·····	46
<i>基于义原表示学习的词向量表示方法</i>	
于宁, 王江萍, 石宇, 刘建毅·····	57
<i>一种新的处理汉语动名超常搭配的方法</i>	
汪梦翔·····	66
<i>基于双编码器的医学文本中文分词</i>	
宗源, 常宝宝·····	76
<i>中文连动句语义关系识别研究</i>	
孙超, 曲维光, 魏庭新, 顾彦慧, 李斌, 周俊生·····	86
<i>中文词语离合现象识别研究</i>	
周露, 曲维光, 魏庭新, 周俊生, 李斌, 顾彦慧·····	97
<i>先秦词网构建及梵汉对比研究</i>	
卢雪晖, 徐会丹, 陈思瑜, 李斌·····	108
<i>基于 BERT 的意图分类与槽填充联合方法</i>	
覃俊, 马天宇, 刘晶, 帖军, 后琦·····	121
<i>基于依存语法的偷抢类动词研究</i>	
王珊, 刘晓骏·····	130
<i>基于中文信息与越南语句法指导的越南语事件检测</i>	
陈龙, 郭军军, 张亚飞, 高盛祥, 余正涛·····	143
<i>基于多层次预训练策略和多任务学习的端到端蒙汉语音翻译</i>	

王宁宁, 飞龙, 张晖·····	156
<i>基于层间知识蒸馏的神经机器翻译</i>	
金畅, 段仁翀, 肖妮妮, 段湘煜·····	166
<i>融合多层语义特征图的缅甸语图像文本识别方法</i>	
刘福浩, 毛存礼, 余正涛, 高盛祥, 王琳钦, 谢旭阳·····	176
<i>JCapsR: 一种联合胶囊神经网络的藏语知识图谱表示学习模型</i>	
孙媛, 梁家亚, 陈安东, 赵小兵·····	186
<i>基于阅读理解的汉越跨语言新闻事件要素抽取方法</i>	
朱恩昌, 余正涛, 高盛祥, 黄宇欣, 郭军军·····	196
<i>面向机器阅读理解的高质量藏语数据集构建</i>	
孙媛, 刘思思, 陈超凡, 旦正错, 赵小兵·····	208
<i>Ti-Reader: 基于注意力机制的藏文机器阅读理解端到端网络模型</i>	
孙媛, 陈超凡, 刘思思, 赵小兵·····	219
<i>藏文文本校对评测集构建</i>	
三毛措, 才智杰, 道吉扎西·····	229
<i>结合标签转移关系的多任务笑点识别方法</i>	
张童越, 张绍武, 徐博, 杨亮, 林鸿飞·····	238
<i>基于时间注意力胶囊网络的维吾尔语情感分类模型</i>	
罗涵天, 杨雅婷, 董瑞, 马博·····	248
<i>基于HowNet的无监督汉语动词隐喻识别方法</i>	
张明昊, 张冬瑜, 林鸿飞·····	258
<i>基于风格化嵌入的中文文本风格迁移</i>	
王晨光, 林鸿飞, 杨亮·····	269
<i>基于双星型自注意力网络的搜索结果多样化方法</i>	
秦绪博, 窦志成, 朱余韬, 文继荣·····	280
<i>基于迭代信息传递和滑动窗口注意力的问题生成模型研究</i>	
陈千, 高晓影, 王素格, 郭鑫·····	293
<i>面向中文口语理解的基于依赖引导的字特征槽填充模型</i>	
朱展标, 黄沛杰, 张业兴, 刘树东, 张华林, 黄均曜·····	304
<i>利用图像描述与知识图谱增强表示的视觉问答</i>	
王屹超, 朱慕华, 许晨, 张琰, 王会珍, 朱靖波·····	316
<i>面向对话文本的实体关系抽取</i>	

陆亮, 孔芳·····	327
<i>结合边界预测和动态模板方法的槽填充模型</i>	
朱展标, 黄沛杰, 张业兴, 刘树东, 张华林, 黄均曜·····	339
<i>近十年来澳门的词汇增长</i>	
王珊, 陈钊, 张昊迪·····	350
<i>基于结构树库的状位动词语义分类及搭配库构建</i>	
邵田, 翟世权, 饶高琦, 荀恩东·····	361
<i>基于序列到序列的中文 AMR 解析</i>	
黄子怡, 李军辉, 贡正仙·····	374
<i>基于词信息嵌入的汉语构词结构识别研究</i>	
郑嫫, 殷雅琦, 王悦, 代达勋, 刘扬·····	386
<i>汉语语体特征的计量与分类研究</i>	
邰沁清, 饶高琦·····	398
<i>多模态表述视域下的小学数学课堂语言计量初探</i>	
郑泽芝, 赵骞·····	413
<i>替换类动词的句法语义分析</i>	
王珊, 吴乐·····	422
<i>基于自动识别的委婉语历时性发展变化与社会共变研究</i>	
张辰麟, 王明文, 谭亦鸣, 尹明, 张心怡·····	435
<i>基于篇章结构攻击的阅读理解任务探究</i>	
马树楷, 邹家杰, 丁鼎·····	446
<i>中美学者学术英语写作中词汇难度特征比较研究 ——以计算语言学领域论文为例</i>	
谢永慧, 刘洋, 杨尔弘, 杨麟儿·····	454
<i>基于词汇链强化表征的篇章修辞结构分析研究</i>	
王金锋, 孔芳·····	467
<i>不同类型噪声环境下言语理解的脑机制研究</i>	
耿立波, 薛紫炫, 杨亦鸣·····	477
<i>回避类动词的句法语义</i>	
王珊, 刘晓骏·····	490
<i>欺骗类动词的句法语义研究</i>	
王珊, 周洁·····	501
<i>基于结构检索的汉语介动搭配知识库构建</i>	

王诚文, 饶高琦, 荀恩东·····	513
<i>数据标注方法比较研究: 以依存句法树标注为例</i>	
周明月, 龚晨, 李正华, 张民·····	525
<i>字里行间的道德: 中文文本道德句识别研究</i>	
彭诗雅, 刘畅, 邓雅月, 于东·····	537
<i>古汉语词义标注语料库的构建及应用研究</i>	
舒蕾, 郭懿鸾, 王慧萍, 张学涛, 胡韧奋·····	549
<i>中文句子级性别无偏数据集构建及预训练语言模型的性别偏度评估</i>	
赵继舜, 杜冰洁, 朱述承, 刘鹏远·····	564
<i>基于多任务标签一致性机制的中文命名实体识别</i>	
吕书宁, 刘健, 徐金安, 陈钰枫, 张玉洁·····	576
<i>面向法律文本的实体关系联合抽取算法</i>	
宋文辉, 周翔, 杨萍, 孙媛媛, 杨亮, 林鸿飞·····	589
<i>融入篇章信息的文学作品命名实体识别</i>	
贾玉祥, 晁睿, 咎红英, 窦华溢, 曹帅, 徐硕·····	600
<i>基于人物特征增强的拟人句要素抽取方法研究</i>	
李婧, 王素格, 陈鑫, 王典·····	612
<i>糖尿病电子病历实体及关系标注语料库构建</i>	
叶娅娟, 胡斌, 张坤丽, 咎红英·····	622
<i>脑卒中疾病电子病历实体及实体关系标注语料库构建</i>	
常洪阳, 咎红英, 马玉团, 张坤丽·····	633
<i>中文关系抽取的句级语言学特征探究</i>	
邢百西, 赵继舜, 刘鹏远·····	643
<i>基于数据选择和局部伪标注的跨语义依存分析研究</i>	
毛达展, 喻快, 邵艳秋·····	655
<i>SaGE: 基于句法感知图卷积神经网络和 ELECTRA 的中文隐喻识别模型</i>	
张声龙, 刘颖, 马艳军·····	667
<i>基于预训练语言模型的繁体古文自动句读研究</i>	
唐雪梅, 苏祺, 王军, 陈雨航, 杨浩·····	678
<i>阅读分级相关研究综述</i>	
饶思敏, 郑娅, 李素建·····	689
<i>融合情感分析的隐式反问句识别模型</i>	

李翔, 刘承伟, 朱晓旭·····	703
<i>面向我国英语学习的英语文本可读性评测模型研究</i>	
王志娟, 曹晓丽·····	714
<i>基于小句复合体的中文机器阅读理解研究</i>	
王瑞琦, 罗智勇, 刘祥, 韩瑞昉, 李舒馨·····	723
<i>融合自编码器和对抗训练的中文新词发现方法</i>	
潘韦, 刘天元, 孙宇清, 龚斌, 张永满, 杨萍·····	736
<i>融合多粒度特征的低资源语言词性标记和依存分析联合模型</i>	
陆杉, 毛存礼, 余正涛, 高盛祥, 黄于欣, 王振晗·····	747
<i>融合外部知识的开放域复述模板获取方法</i>	
金波, 刘明童, 张玉洁, 徐金安, 陈钰枫·····	758
<i>基于堆叠式注意力网络的复杂话语领域分类方法</i>	
梁超杰, 黄沛杰, 丁健德, 朱建恺, 林丕源·····	769
<i>基于大规模语料库的《古籍汉字分级字表》研究</i>	
许长伟, 冯敏萱, 李斌, 袁义国·····	781
<i>一种基于IDLSTM+CRF的中文主地域抽取方法</i>	
童逸琦, 叶培根, 付彪, 陈毅东, 史晓东·····	792
<i>基于信息交互增强的时序关系联合识别</i>	
戴倩颖, 孔芳·····	803
<i>基于字词粒度噪声数据增强的中文语法纠错</i>	
汤泽成, 纪一心, 赵怡博, 李军辉·····	813
<i>基于多质心异质图学习的社交网络用户建模</i>	
宁上毅, 李冠颖, 陈琴, 黄增峰, 周葆华, 魏忠钰·····	825
<i>软件标识符的自然语言规范性研究</i>	
汶东震, 张帆, 张晓, 杨亮, 林原, 徐博, 林鸿飞·····	837
<i>基于改进Conformer的新闻领域端到端语音识别</i>	
张济民, 早克热·卡德尔, 申云飞, 艾山·吾买尔, 汪烈军·····	851
<i>基于BPE分词的中国古诗主题模型及主题可控的诗歌生成</i>	
张家瑞, 李文浩, 孙茂松·····	862
<i>Reducing Length Bias in Scoring Neural Machine Translation via a Causal Inference Method</i>	
Xuwen Shi, Heyan Huang, Ping Jian and Yi-Kun Tang·····	874
<i>Low-Resource Machine Translation based on Asynchronous Dynamic Programming</i>	

Xiaoning Jia, Hongxu Hou, Nier Wu, Haoran Li and Xin Chang·····	886
<i>Incorporating translation quality estimation into Chinese-Korean neural machine translation</i>	
Feiyu Li, Yahui Zhao, Feiyang Yang and Rongyi Cui·····	906
<i>Emotion Classification of COVID-19 Chinese Microblogs based on the Emotion Category Description</i>	
Xianwei Guo, Hua Lai, Yan Xiang, Zhengtao Yu and Yuxin Huang·····	916
<i>Multi-level Emotion Cause Analysis by Multi-head Attention Based Multi-task Learning</i>	
Xiangju Li, Shi Feng, Yifei Zhang and Daling Wang·····	928
<i>Using Query Expansion in Manifold Ranking for Query-Oriented Multi-Document Summarization</i>	
Quanye Jia, Rui Liu and Jianying Lin·····	940
<i>Jointly Learning Saliency and Redundancy by Adaptive Sentence Reranking for Extractive Summarization</i>	
Ximing Zhang and Ruifang Liu·····	952
<i>Incorporating Commonsense Knowledge into Abstractive Dialogue Summarization via Heterogeneous Graph Networks</i>	
Xiachong Feng, Xiaocheng Feng and Bing Qin·····	964
<i>Enhancing Question Generation with Commonsense Knowledge</i>	
Xin Jia, Hao Wang, Dawei Yin and Yunfang Wu·····	976
<i>Topic Knowledge Acquisition and Utilization for Machine Reading Comprehension in Social Media Domain</i>	
Zhixing Tian, Yuanzhe Zhang, Kang Liu and Jun Zhao·····	988
<i>Category-Based Strategy-Driven Question Generator for Visual Dialogue</i>	
Yanan Shi, Yanxin Tan, Fangxiang Feng, Chunping Zheng and Xiaojie Wang·····	1000
<i>From Learning-to-Match to Learning-to-Discriminate: Global Prototype Learning for Few-shot Relation Classification</i>	
Fangchao Liu, Xinyan Xiao, Lingyong Yan, Hongyu Lin, Xianpei Han, Dai Dai, Hua Wu and Le Sun·····	1012
<i>Multi-Strategy Knowledge Distillation Based Teacher-Student Framework for Machine Reading Comprehension</i>	
Xiaoyan Yu, Qingbin Liu, Shizhu He, Kang Liu, Shengping Liu, Jun Zhao and Yongbin Zhou·····	1024
<i>LRRA: A Transparent Neural-Symbolic Reasoning Framework for Real-World Visual Question Answering</i>	
Zhang Wan, Keming Chen, Yujie Zhang, Jinan Xu and Yufeng Chen·····	1037
<i>Meaningfulness and unit of Zipf's law: evidence from danmu comments</i>	

Yihan Zhou	1046
<i>Unifying Discourse Resources with Dependency Framework</i>	
Yi Cheng, Sujian Li and Yueyuan Li	1058
<i>A Chinese Machine Reading Comprehension Dataset Automatic Generated Based on Knowledge Graph</i>	
Hanyu Zhao, Sha Yuan, Jiahong Leng, Xiang Pan, Zhao Xue, Quanyue Ma and Yangxiao Liang	1066
<i>Morphological Analysis Corpus Construction of Uyghur</i>	
Gulinigeer Abudouwaili, Kahaerjiang Abiderexiti, Jiamila Wushouer, Yunfei Shen, Turenisha Maimaitimin and Tuergen Yibulayin	1076
<i>Improving Entity Linking by Encoding Type Information into Entity Embeddings</i>	
Tianran Li, Erguang Yang, Yujie Zhang, Yufeng Chen and Jinan Xu	1087
<i>A Unified Representation Learning Strategy for Open Relation Extraction with Ranked List Loss</i>	
Renze Lou, Fan Zhang, Xiaowei Zhou, Yutong Wang, Minghui Wu and Lin Sun	1096
<i>NS-Hunter: BERT-Cloze Based Semantic Denoising for Distantly Supervised Relation Classification</i>	
Tielin Shen, Daling Wang, Shi Feng and Yifei Zhang	1109
<i>A Trigger-Aware Multi-Task Learning for Chinese Event Entity Recognition</i>	
Yangxiao Xiang and Chenliang Li	1121
<i>Improving Low-Resource Named Entity Recognition via Label-Aware Data Augmentation and Curriculum Denoising</i>	
Wenjing Zhu, Jian Liu, Jinan Xu, Yufeng Chen and Yujie Zhang	1131
<i>Global entity alignment with Gated Latent Space Neighborhood Aggregation</i>	
Wei Chen, Xiaoying Chen and Shengwu Xiong	1143
<i>Few-Shot Charge Prediction with Multi-Grained Features and Mutual Information</i>	
Han Zhang, Yutao Zhu, Zhicheng Dou and Ji-Rong Wen	1154
<i>Sketchy Scene Captioning: Learning Multi-Level Semantic Information from Sparse Visual Scene Cues</i>	
Lian Zhou, Yangdong Chen and Yuejie Zhang	1167
<i>BDCN: Semantic Embedding Self-explanatory Breast Diagnostic Capsules Network</i>	
Dehua Chen, Keting Zhong and Jianrong He	1178
<i>GCN with External Knowledge for Clinical Event Detection</i>	
Dan Liu, Zhichang Zhang, Hui Peng and Ruirui Han	1190
<i>A Prompt-independent and Interpretable Automated Essay Scoring Method for Chinese Second Language Writing</i>	

Yupei Wang and Renfen Hu.....	1202
<i>A Robustly Optimized BERT Pre-training Approach with Post-training</i>	
Zhuang Liu, Wayne Lin, Ya Shi, Jun Zhao.....	1218